



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

**Modelos predictivos de eventos extremos en espacio
tiempo**

José del Carmen Jiménez Hernández

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

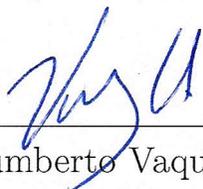
MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2016

La presente tesis titulada: **Modelos predictivos de eventos extremos en espacio tiempo**, realizada por el alumno: **José del Carmen Jiménez Hernández**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

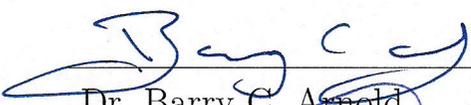
DOCTOR EN CIENCIAS

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA**

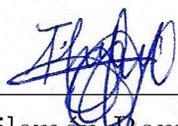
CONSEJO PARTICULAR

CONSEJERO 
Dr. Humberto Vaquera Huerta

ASESOR 
Dr. Paulino Pérez Rodríguez

ASESOR 
Dr. Barry C. Arnold

ASESOR 
Dr. José A. Villaseñor Alva

ASESOR 
Dr. Filemon Ramírez Pérez

Montecillo, Texcoco, Estado de México, septiembre de 2016.

Modelos predictivos de eventos extremos en espacio tiempo

José del Carmen Jiménez Hernández, MC

Colegio de Postgraduados, 2016

La presente investigación se enfoca en el análisis espacial y espacio–temporal de eventos extremos, en particular valores extremos del monóxido de carbono en la Zona Metropolitana de la Ciudad de México. En el análisis espacial, se aplican los modelos de Smith y Schlather, se proponen tendencias espaciales a los parámetros de la distribución de valores extremos lo que permite producir mapas predictivos y de nivel de retorno, también se propone un modelo jerárquico bayesiano para el análisis de estos datos, aquí también se pueden producir mapas predictivos y de nivel de retorno. Finalmente se presenta la propuesta sobre la extensión del modelo de Schlather al espacio–tiempo, se obtiene de manera analítica la función de distribución acumulada bivariada y con base en esta se propone usar la metodología de verosimilitud compuesta para la estimación de los parámetros del modelo.

Palabras clave: Distribución de valores extremos, procesos max-stable, verosimilitud compuesta, estadística bayesiana, procesos espacio temporales.

Predictive models of extreme events in space-time

José del Carmen Jiménez Hernández, MC

Colegio de Postgraduados, 2016

This research focuses on spatial analysis and space-time of extreme events, particularly extreme values of carbon monoxide in the metropolitan area of Mexico City. In spatial analysis, models of Smith and Schlather were applied, we proposed spatial trends to the parameters of the distribution of extreme values which can produce predictive and return level maps, a Bayesian hierarchical model for the analysis also proposed these data, here also can produce predictive and return level maps. Finally, the proposal on the extension of the model Schlather space – time is presented, here obtained analytically the bivariate cumulative distribution function and based on this methodology intends to use the composite likelihood to estimate parameters of the model.

Key words: Extreme value distribution, max–stable processes, composite likelihood, bayesian statistics, space–time processes

AGRADECIMIENTOS

Al Colegio de Postgraduados Campus Montecillo, por brindarme la oportunidad de realizar mis estudios de doctorado.

Al Consejo Nacional de Ciencia y Tecnología por el apoyo económico recibido para la realización de mis estudios.

Al programa Para el Mejoramiento del Profesorado (PROMEP) por el apoyo brindado durante mi estancia en el Colegio.

A mi consejero el Dr. Humberto Vaquera Huerta por su invaluable apoyo durante el tiempo que estuve estudiando.

A todos los miembros de mi consejo particular, Dr. José A. Villaseñor Alva, Dr. Paulino Pérez Rodríguez, Dr. Barry C. Arnold y Dr. Filemón Ramírez Pérez, así como al Dr. Gilberto Rendón Sánchez por fungir como sinodal, por sus comentarios y observaciones durante el desarrollo de la investigación.

DEDICATORIA

A mi esposa, Marisol López Cerino.

A mi hija, Julieta Jiménez López.

A mis padres, José del Carmen Jiménez Arcia (†) y Concesa Hernández Sanchez.

Contenido

Contenido	VII
Lista de tablas	IX
Lista de figuras	X
Introducción	XI
1. Teoría clásica de valores extremos	1
1.1. Teoría de valores extremos univariada	1
1.1.1. Dominio de atracción del máximo	5
1.1.2. Extensión a series estacionarias	6
1.1.3. Inferencia	9
1.2. Teoría de valores extremos multivariada	10
1.2.1. Distribuciones multivariadas y max-stabilidad	11
1.2.2. La medida del exponente	12
1.2.3. Representación espectral	12
1.2.4. Función de dependencia de Pickands	13
1.2.5. Modelos paramétricos	13
1.2.6. Inferencia	15
1.2.7. Medida de dependencia extrema	16
2. Procesos max-stable para extremos espaciales	17
2.1. Fundamentos de procesos aleatorios	18
2.1.1. Funciones de covarianza	19
2.2. Procesos max-stable	20
2.3. Modelos paramétricos estacionarios	23
2.3.1. El modelo de Smith	23
2.3.2. El modelo de Schlather	24
2.4. Medida de dependencia extremal	26

Contenido

2.5. Inferencia en procesos max-stable	29
2.5.1. Función de verosimilitud compuesta	29
2.5.2. Verosimilitud a pares para extremos espaciales	30
2.5.3. Comparación de modelos	31
2.5.4. Estimación de la varianza asintótica	32
2.6. Aplicación	32
2.6.1. Datos	32
2.6.2. Metodología	33
2.6.3. Resultados	33
3. Modelos jerárquicos bayesianos para extremos espaciales	40
3.1. Formulación del modelo	41
3.2. Metodología	42
3.3. Resultados	43
4. El modelo de Schlather espacio-temporal	48
4.1. Generalidades	49
4.2. El modelo espacio-temporal	51
4.3. Inferencia para el modelo propuesto	59
5. Conclusiones	60
Referencias	61
Anexo	65
Anexo A: Códigos en R	66

Lista de tablas

1.1. Clasificación de varias distribuciones conocidas de acuerdo al dominio de atracción del máximo, con sus correspondientes constantes de normalización $a_n > 0$ y $b_n \in \mathbb{R}$	7
2.1. Algunas funciones de correlación.	28
2.2. Estaciones metereológicas bajo estudio.	34
2.3. Parámetros de dependencia espacial estimados para el modelo de Smith, su error estándar esta entre paréntesis.	34
2.4. Criterio de información de verosimilitud compuesta.	34
2.5. Parámetros estimados para el modelo de Schalther. La cantidad que esta entre paréntesis indica su error estándar.	35
2.6. Resumen del modelo de Schlather ajustado a los datos para el año 2012, los errores estándar están en paréntesis, CLIC* es el criterio de información de verosimilitud compuesta y l_p es la función de log-verosimilitud compuesta maximizada.	36
2.7. Parámetros estimados y sus errores estándar del modelo de Schlather ajustado a los datos para el año 2012 suponiendo tendencias espaciales en los parámetros de la distribución GEV.	36
3.1. Hiperparámetros para el proceso latente usados en nuestra aplicación. Las distribuciones a priori para α . y λ . son gamma inversa y gamma, respectivamente.	43
3.2. Parámetros estimados para la distribución a posterior y del proceso latente. La estimación corresponde a la media a posterior y su correspondiente intervalo creíble del 95 %, el cual se muestra en paréntesis.	43

Lista de figuras

1.1.	Distribución de M_n (izquierda) y distribución de M_n^* (derecha), para $a_n = 3$ y $b_n = 4$. Aquí, $n = 1$ (negro), $n = 2$ (rojo), $n = 4$ (verde), $n = 16$ (azul) y $n = 32$ (azul cielo), y F es la función de distribución normal estándar.	2
1.2.	Función de densidad Gumbel estándar (verde), Fréchet unitaria (azul) y Weibull (roja).	4
2.1.	Realizaciones del modelo de Smith en \mathbb{R}^2 para diferentes matrices de covarianzas $\Sigma = (\Sigma_{ij})$, $i, j = 1, 2$. (a) $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0$; (b) $\Sigma_{11} = 2$, $\Sigma_{22} = 3$, $\Sigma_{12} = \Sigma_{21} = -0.5$; (c) $\Sigma_{11} = 5$, $\Sigma_{22} = 2$, $\Sigma_{12} = \Sigma_{21} = 0.5$; (d) $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0.5$	25
2.2.	Realizaciones del modelo de Schlather con marginales Fréchet unitarias en \mathbb{R}^2 usando la función de correlación exponencial $\rho(\mathbf{h}) = \exp\left[-\left(\frac{\mathbf{h}}{c_2}\right)^\nu\right]$, para diferentes valores de los parámetros. (a) $c_2 = 2$, $\nu = 0.5$; (b) $c_2 = 3$, $\nu = 1$; (c) $c_2 = 3$, $\nu = 1.5$; (d) $c_2 = 4$, $\nu = 1.9$	27
2.3.	Región de estudio.	33
2.4.	Coeficiente extremal no paramétrico. La línea continua estima la tendencia de la dependencia espacial.	35
2.5.	Gráfica de símbolos para los datos del año 2012.	37
2.6.	Validación del modelo para el modelo max-stable ajustado.	37
2.7.	Predicciones puntuales del proceso de Schalter ajustado. (a) parámetro de localidad; (b) de escala y (c) el nivel de retorno a dos años.	38
2.8.	Predicción del nivel de retorno. (a) 2 años; (b) 4 años y (c) 5 años.	39
3.1.	Distribuciones a posterior para α y λ para cada parámetro de la distribución GEV. EL primer renglón corresponde a μ , el de en mitad a σ y el de abajo a ξ	44
3.2.	Distribuciones a priori y a posteriori para α y λ para cada parámetro de la distribución GEV. EL primer renglón corresponde a μ , el de en mitad a σ y el de abajo a ξ	45
3.3.	Mapa para el parámetro de localización de la distribución GEV.	45
3.4.	Mapa para el parámetro de escala de la distribución GEV.	46

Lista de figuras

3.5. Mapa para el parámetro de forma de la distribución GEV.	46
3.6. Mapa del nivel de retorno estimado de un año para los datos de CO mediante el proceso latente.	47
4.1. Relaciones entre los diferentes tipos de funciones de covarianza espacio- temporales.	51

Introducción

La modelación estadística de eventos raros ha presentado, en los últimos años, un gran número de aplicaciones en muchas áreas de la ciencia, pero dentro de estos, los eventos extremos ambientales son de suma importancia por el impacto negativo que puedan ocasionar en la salud humana, por ejemplo, en la contaminación ambiental. Aquí, la teoría de valores extremos univariada y multivariada, ha sido muy bien aceptada para la modelación estadística de éstos ya que su sustento matemático (de Haan y Pereira, 2006, Resnick, 2007, 2013) y su fundamento estadístico (Beirlant *et al.*, 2006, Coles *et al.*, 2001) esta bien cimentado. Muchos de los problemas que se presentan en esta teoría, son de naturaleza espacial o espacio-temporal, esto ha dado lugar a la teoría de valores extremos espacial y espacio-temporal, en esta tesis nos enfocamos en ambas.

La teoría clásica de extremos espaciales, se basa, en uno de sus enfoques, en los procesos max-stable, estos procesos son una generalización infinito dimensional de la teoría de valores extremos multivariada (Davison *et al.*, 2012, de Haan, 1984, Huser y Davison, 2014, Smith, 1990b). Otro enfoque para modelar extremos espaciales se basa en modelos jerárquicos bayesianos con una estructura espacial latente, en ésta se puede describir la variación espacial de los parámetros en la distribución marginal.

Dentro del primer enfoque, Smith (1990b) propuso el primer modelo y lo llamo perfil de tormentas, usando una función determinista, más tarde, Schlather (2002) propuso modelar los extremos mediante un proceso estocástico lo cual le permite ser más flexible al analizar la dependencia espacial. Otros modelo se ha propuestos, pero son extensiones de estos.

La teoría de valores extremos espacio-temporal, presenta poco avance, y es Davis *et al.* (2013a) quienes proponen extender el modelo de Smith (Smith, 1990b) al espacio-tiempo, obtienen de manera analítica la función de distribución conjunta para dos ubicaciones y también muestran que la función de covarianza en el espacio-tiempo del proceso se puede interpretar en términos de la función de dependencia del proceso max-stable espacio-temporal, esta extensión aun no presenta una aplicación a datos reales, el proceso de inferencia para esta extensión la muestran Davis *et al.* (2013b). Huser y Davison (2014), presentan una extensión del modelo de Schlather (Schlather, 2002) al espacio-tiempo considerando el conjunto aleatorio que Schlather propone, ellos lo aplican a datos de lluvias en Suiza.

El objetivo general de esta tesis es generar metodología estadística para la modelación de eventos extremos con dependencia en el espacio-tiempo, específicamente, proponer un modelo que permita predecir eventos extremos espacio temporales, para ello nos

centramos en el modelo de Schlather y proponemos una generalización de este.

La estructura de la tesis es la siguiente, en el [Capítulo 1](#) se presentan los fundamentos de la teoría de valores extremos en el caso univariado y multivariado, se pretende que sean conceptos que ayuden al desarrollo de misma.

En el [Capítulo 2](#), se presenta la teoría referente a procesos max-stable para extremos espaciales, se presenta la aplicación de ésta a datos de contaminación por monóxido de carbono (CO) en la Zona Metropolitana del Valle de México, se proponen tendencias espaciales a los parámetros de distribución de valores extremos generalizada y se obtienen mapas predictivos para estos parámetros.

En el siguiente Capítulo, se propone analizar los mismos datos de CO desde el enfoque bayesiano, aquí se propone un modelo jerárquico con una estructura espacial latente.

En el [Capítulo 4](#) proponemos la generalización del modelo de Schlather al espacio-tiempo, se obtiene la función de distribución acumulada bivariada y se propone usar verosimilitud compuesta para la estimación de los parámetros del modelo.

Finalmente en el [Capítulo 5](#) se presentan las conclusiones de la tesis.

Teoría clásica de valores extremos

En este capítulo se hace una revisión de los conceptos básicos de la teoría de valores extremos en el caso univariado y multivariado, se pretende proporcionar antecedentes sólidos para el resto de la tesis, la mayor parte del contenido esta bien establecido en la literatura de valores extremos y se puede encontrar, por ejemplo, en [Resnick \(2007, 2013\)](#), de [Haan y Pereira \(2006\)](#), [Beirlant *et al.* \(2006\)](#) y [Coles *et al.* \(2001\)](#).

1.1. Teoría de valores extremos univariada

Iniciamos con Y, Y_1, Y_2, \dots variables aleatorias independientes que se distribuyen de acuerdo a una distribución común F con soporte Supp_F , esto es, el conjunto de puntos $y \in \mathbb{R}$ con densidad estrictamente positiva (o masa de probabilidad en el caso discreto) con respecto a F . Se esta interesado en fluctuaciones del máximo de n variables aleatorias, el cual se denota por $M_n = \max(Y_1, \dots, Y_n)$. La función de distribución acumulada de M_n es

$$P(M_n \leq y) = P(Y_1 \leq y, \dots, Y_n \leq y) = \prod_{i=1}^n P(Y_i \leq y) = F^n(y),$$

la cual converge a una función de distribución no degenerada poniendo masa uno al punto final $y_F = \sup\{y : F(y) < 1\}$ de la distribución de F . De hecho, para todo $y < y_F$, se tiene, $F^n(y) \rightarrow 0$ cuando $n \rightarrow \infty$.

Al igual que el caso de variables aleatorias independientes, que bajo ciertas condiciones de normalización se puede obtener la distribución de la suma –muchos de los cuales son atraídos a la distribución normal por el teorema del límite central–, el comportamiento estocástico de M_n también se puede estabilizar después de ciertas transformaciones y bajo ciertas condiciones.

La [Figura 1.1](#) muestra como evoluciona la distribución de máximo de n variables aleatorias normales estándar conforme n aumenta, sin (izquierda) y con normalización (derecha). La figura izquierda muestra que la distribución del máximo M_n , se mueve hacia la derecha y que la probabilidad se encuentra centrada más o menos en un punto. La figura de la derecha muestra la distribución del máximo normalizado $M_n^* = (M_n - b_n)/a_n$, para elecciones adecuadas de sucesiones $a_n > 0$ y b_n . Aquí la

1.1. Teoría de valores extremos univariada

distribución de M_n^* converge a una distribución Gumbel estándar. Similarmente al Teorema del límite central para suma de variables aleatorias, la esperanza de esta normalización para máximos es caracterizar todas las posibles distribuciones límites no degeneradas por una simple familia paramétrica, y así tener asintóticamente modelos para máximos. Para establecer este resultado, se necesita introducir la noción de max-stable.

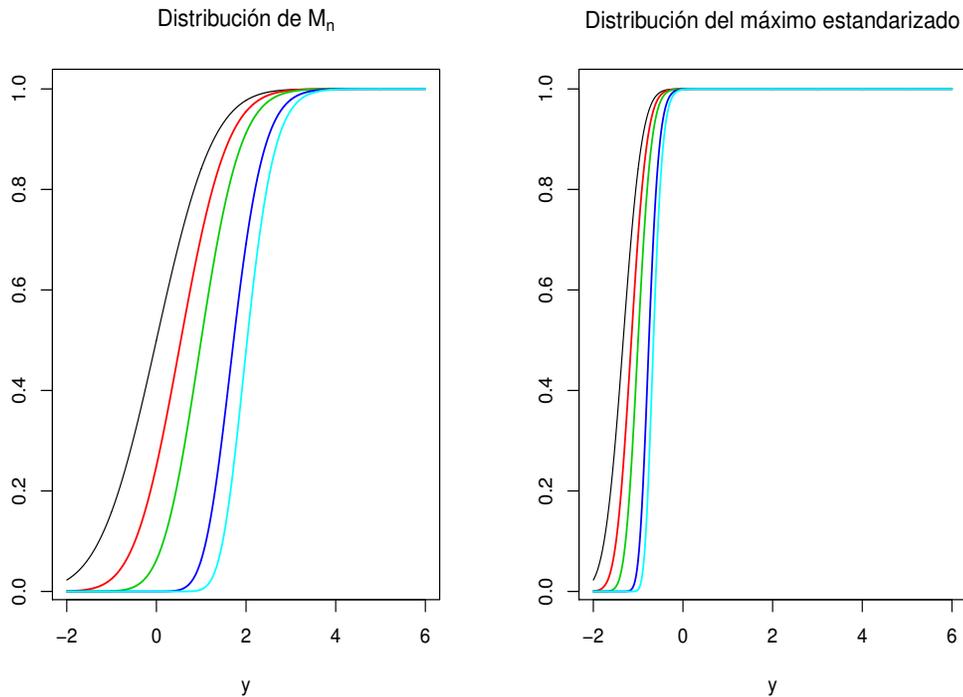


Figura 1.1: Distribución de M_n (izquierda) y distribución de M_n^* (derecha), para $a_n = 3$ y $b_n = 4$. Aquí, $n = 1$ (negro), $n = 2$ (rojo), $n = 4$ (verde), $n = 16$ (azul) y $n = 32$ (azul cielo), y F es la función de distribución normal estándar.

Definición 1.1. (Distribución y variables aleatorias del mismo tipo). Dos distribuciones G_1 y G_2 se dicen que son del mismo tipo, si existen constantes de normalización $a > 0$ y $b \in \mathbb{R}$ tal que

$$G_1(ay + b) = G_2(y), \quad \text{para todo } y \in \mathbb{R}.$$

Similarmente, dos variables aleatorias Y_1 y Y_2 son del mismo tipo, si sus distribuciones son del mismo tipo.

Definición 1.2. (Max-stable). Una distribución G es max-stable si para cualquier $k \in \mathbb{N}$ la distribución de G^k es del mismo tipo que la de G . Una variable aleatoria Z se dice que es max-stable si su distribución es max-stable.

1.1. Teoría de valores extremos univariada

La condición max-stable se satisface para distribuciones para las cuales la operación de tomar el máximo de muestras conduce a la misma distribución, aparte de cambios en la escala y localización.

Considere ahora una sucesión de variables aleatorias max-stable independientes e idénticamente distribuidas (i.i.d.) $Z_1, Z_2, \dots \sim G$. Por definición la distribución del máximo es del mismo tipo que la de G , esto es, para cada $k \in \mathbb{N}$ se pueden encontrar constantes reales $a_k > 0$ y b_k tal que $G^k(a_k y + b_k) = G(y)$ para todo $y \in \mathbb{N}$. Aquí la variable aleatoria $Z_n^* = (\max\{Z_1, \dots, Z_n\} - b_n)/a_n$ y $Z \sim G$ tienen la misma distribución. En particular, Z_n^* converge en distribución a Z cuando $n \rightarrow \infty$ ($Z_n^* \xrightarrow{D} Z$), lo cual implica que todas las distribuciones max-stable son límites de máximos normalizados de variables aleatorias i.i.d. El resultado interesante reside en el hecho de que lo inverso también es cierto: las distribuciones max-stable son las únicas posibles distribuciones no degeneradas de máximos normalizados. Ver el siguiente Teorema.

Teorema 1.1. (*de Haan, 1970, Embrechts et al., 2013*). *La clase de todas las posibles distribuciones límites no degeneradas (correctamente normalizadas) para máximos de variables aleatorias i.i.d. coincide con la clase de distribuciones max-stable.*

La prueba se basa en el hecho de que el máximo de un bloque de longitud pq puede escribirse como el máximo de p máximos de pequeños bloques, cada uno de longitud q , escribiendo $M_{i;j} = \max\{Y_i, \dots, Y_j\}$, se tiene

$$M_{1;pq} = \max\{M_{1;pq}, \dots, M_{(p-1)q+1;pq}\}.$$

Aquí la distribución límite G satisface la ecuación $G\{(y - b_{pq})/a_{pq}\} = G^p\{(y - b_q)/a_q\}$, y así es max-stable.

El siguiente Teorema, mostrado primero por [Fisher y Tippett \(1928\)](#), es la base de la teoría clásica de valores extremos. Afirma que si el máximo de variables aleatorias se puede desplazar y escalar de tal manera que converja en distribución a una distribución límite no degenerada, entonces ésta última tiene que ser una de tres tipos especiales. Ver [Teorema 1.2](#) y [Figura 1.2](#). Específicamente, la clase de distribuciones max-stables está completamente descrita sólo por esta familia paramétrica.

Teorema 1.2. (*Teorema de tipos de extremos; Fisher y Tippett (1928), Gnedenko (1943), Resnick (2013)*). *Sea $\{Y_i\}_{i \geq 1}$ una sucesión de variables aleatorias i.i.d. y sea $M_n = \max(Y_1, \dots, Y_n)$ si existen sucesiones de constantes $a_n > 0$ y b_n tal que*

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

donde $G(y)$ es una función de distribución no degenerada, entonces $G(y)$ debe ser del mismo tipo como una de las siguientes distribuciones:

Tipo I (Gumel):

$$\Lambda(y) = \exp\{-\exp(-y)\}, \quad y \in \mathbb{R},$$

1.1. Teoría de valores extremos univariada

Tipo II (Fréchet):

$$\Phi_\alpha(y) = \begin{cases} \exp\{-y^{-\alpha}\} & \text{si } y > 0; \\ 0 & \text{si } y \geq 0, \end{cases}$$

Tipo III (Weibull)

$$\Phi_\alpha(y) = \begin{cases} \exp\{-(-y^\alpha)\} & \text{si } y < 0; \\ 1 & \text{si } y \geq 0, \end{cases}$$

para algún $\alpha > 0$.

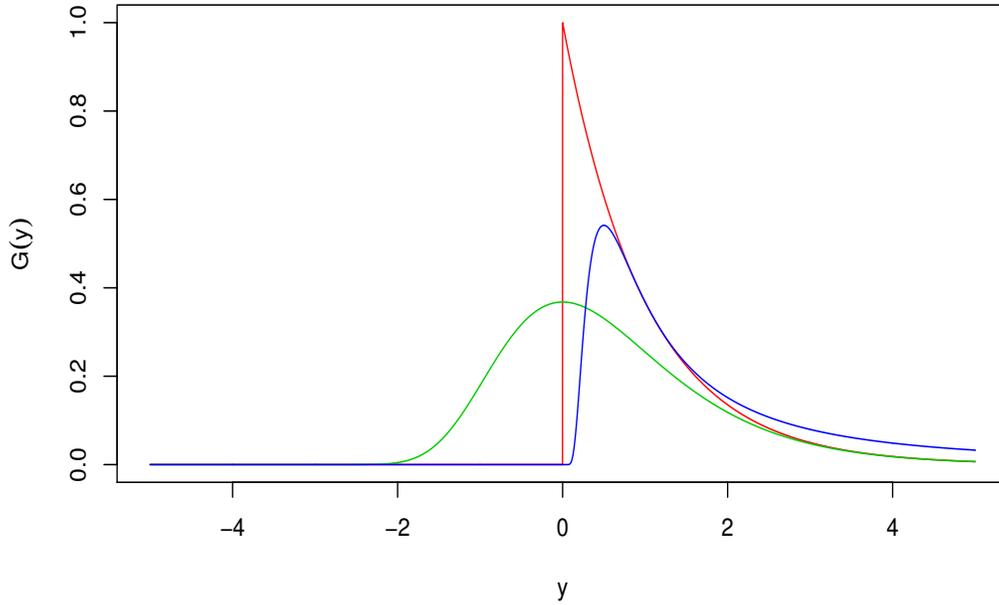


Figura 1.2: Función de densidad Gumbel estándar (verde), Fréchet unitaria (azul) y Weibull (roja).

Los tres tipos se pueden resumir en una sola familia paramétrica, llamada la distribución de valores extremos generalizada, $GEV(\mu, \sigma, \xi)$:

$$G(y) = \begin{cases} \exp\left[-\left\{1 + \xi \left(\frac{y - \mu}{\sigma}\right)_+^{-1/\xi}\right\}\right], & \xi \neq 0; \\ \exp\left[-\exp\left\{-\left(\frac{y - \mu}{\sigma}\right)_+\right\}\right], & \xi = 0, \end{cases}$$

con parámetro de localización $\mu \in \mathbb{R}$, parámetro de escala $\sigma > 0$ y parámetro de forma $\xi \in \mathbb{R}$, y donde $t_+ = \max(t, 0)$. La distribución tiene soporte $\text{Supp}_G = \{y \in \mathbb{R} : 1 + \xi(y - \mu/\sigma) > 0\}$. El principal parámetro es ξ , ya que determina el tipo de distribución

1.1. Teoría de valores extremos univariada

que se tiene y de esta forma si el soporte de G esta acotado por arriba, por abajo o no lo está. Si $\xi < 0$ la distribución tiene una cota superior en $y = \mu - \sigma/\xi$ y se tiene la distribución Weibull con $\alpha = -\xi^{-1}$; cuando $\xi = 0$, $G(y)$ corresponde a la distribución Gumbel cuyo soporte no esta acotado y cuya cola superior decae exponencialmente; y cuando $\xi > 0$ la distribución tiene una cota inferior en $y = \mu - \sigma/\xi$ y se tiene la distribución Fréchet con $\alpha = \xi^{-1}$. Además, para la distribución Fréchet el parámetro ξ controla como cae la cola superior y de aquí la posible gravedad de eventos extremos futuros. El r -ésimo momento de la distribución GEV es finito si $r\xi < 1$, por lo tanto cuando $\xi = 1$ (el caso de la distribución Fréchet unitaria), la media no esta definida.

Esta distribución es especialmente conveniente para inferencia; aunque en la práctica ξ es difícil de estimar, uno puede dejar que los datos decidan su valor sin tener que elegir de antemano una distribución particular. A menudo son de particular interés cuantiles altos, ya que dan una descripción cuantitativa de la severidad de un evento extremo que podría ocurrir en periodos de tiempo futuro. El nivel que se espera sea excedido una vez, en promedio, en un periodo de tiempo específico se denomina *nivel de retorno* y esta asociado con el *periodo de retorno*. Ver [Definición 1.4](#).

Definición 1.3. (Función cuantil). La inversa generalizada, o función cuantil, de una función de distribución F es

$$F^{-1}(p) = \inf\{y \in \mathbb{R} : F(y) \geq p\}, \quad 0 < p < 1.$$

Definición 1.4. (Niveles de retorno y periodos de retorno). Suponga que los máximos anuales de alguna variable aleatoria son modelados con la distribución $G \sim \text{GEV}(\mu, \sigma, \xi)$. el nivel y_p el cual es excedido con probabilidad p , esto es uno cada $1/p$ años, en promedio, se llama el nivel de retorno, así

$$y_p = G^{-1}(1 - p) = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}], & \text{para } \xi \neq 0 \\ \mu - \sigma \log \{-\log(1 - p)\}, & \text{para } \xi = 0. \end{cases}$$

El periodo de retorno es $1/p$. La definición se puede extender a otros bloques de máximos (por ejemplo bloques diarios, mensuales, etc.)

1.1.1. Dominio de atracción del máximo

El [Teorema 1.2](#) identifica las tres posibles distribuciones límites para máximos normalizados de variables aleatorias i.i.d. Sin embargo, si existen sucesiones de constantes $a_n > 0$ y $b_n \in \mathbb{R}$ tal que $(M_n - b_n)/a_n$ converge a una distribución no degenerada G , entonces G se determina de manera única por una transformación afín, esto es, si existen sucesiones de constantes $a'_n > 0$ y $b'_n \in \mathbb{R}$ tal que $(M_n - b'_n)/a'_n \xrightarrow{D} G'$, entonces G' y G deben ser del mismo tipo. Por lo tanto se puede definir el dominio de atracción del máximo (MDA) como la clase de distribuciones cuyos máximos son atraídos a una distribución en particular cuando $n \rightarrow \infty$.

Definición 1.5. (Dominio de atracción del máximo). La variable aleatoria Y pertenece al dominio de atracción del máximo de la distribución de valores extremos G si existen constantes $a_n > 0$ y $b_n \in \mathbb{R}$ tal que $(M_n - b_n)/a_n \xrightarrow{D} G$. Se escribe $Y \in \text{MDA}(G)$.

1.1. Teoría de valores extremos univariada

La caracterización del MDA ha sido bastante estudiada en la literatura, ver por ejemplo, [von Mises \(1954\)](#), [Resnick \(2013\)](#), [Embrechts *et al.* \(2013\)](#), [Beirlant *et al.* \(2006\)](#) y [de Haan y Pereira \(2006\)](#), donde los autores dan una discusión de condiciones necesarias y suficientes para diferentes dominios de atracción.

La noción de distribuciones equivalentes en la cola es primordial para describir completamente el MDA de cada una de las tres distribuciones de tipos de extremos.

Definición 1.6. (Equivalencia en la cola de distribuciones). Las funciones de distribuciones F_1 y F_2 se dice que son equivalentes en su cola si tienen el mismo punto final en la derecha $y_{F_1} = y_{F_2} = y_F$, y si

$$\lim_{y \rightarrow y_F} \frac{1 - F_1(y)}{1 - F_2(y)} = c,$$

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

Cada MDA es cerrado bajo equivalencia en la cola de su distribución, esto es, si F_1 y F_2 son equivalentes en su cola, $F_1 \in \text{MDA}(G)$ si y sólo si $F_2 \in \text{MDA}(G)$, de aquí se puede decir que el MDA esta compuesto por distribuciones cuyas colas derechas caen en la misma proporción que sus puntos derechos finales.

Teorema 1.3. (*Caracterización del MDA*). Sea F una función de distribución con punto final superior y_F , y sea \sim que denota equivalencia asintótica. Las siguientes afirmaciones son ciertas:

- $F \in \text{MDA}(\Phi_\alpha)$ si y sólo si $1 - F(y) \sim Ky^{-\alpha}$, cuando $y \rightarrow y_F = \infty$.
- $F \in \text{MDA}(\Psi_\alpha)$ si y sólo si $1 - F(y) \sim K(y_F - y)^\alpha$, cuando $y \rightarrow y_F < \infty$ para alguna $\alpha > 0$ y $K \in \mathbb{R}$, una constante que no dependen de y .
- $F \in \text{MDA}(\Delta)$ si y sólo si existe alguna $z < y_F \leq \infty$ tal que $F(y)$ tiene la representación $1 - F(y) = c(y) \exp \left[- \int_z^y \{1/a(t)\} dt \right]$, $z < y < y_F$, donde c es una función medible que satisface $c(y) \rightarrow c > 0$ cuando $y \rightarrow y_F$, y $a(y)$ es una función positiva absolutamente continua (con respecto a la medida de Lebesgue) con densidad $a'(y)$ que cumple $\lim_{y \rightarrow y_F} a'(y) = 0$.

Para distribuciones que son suficientemente suaves en su punto final derecho, [von Mises \(1954\)](#) establece condiciones suficientes para la convergencia de $(M_n - b_n)/a_n \rightarrow Z \sim \text{GEV}(0, 1, \xi)$, cuando $n \rightarrow \infty$, proporcionando herramientas útiles para determinar el tipo de distribución límite y la elección de las constantes de normalización $a_n > 0$ y $b_n \in \mathbb{R}$. Estas condiciones se pueden emplear para clasificar alguna distribución conocida en los diferentes MDA. Ver [Tabla 1.1](#) y [Embrechts *et al.* \(2013\)](#) para más detalles.

1.1.2. Extensión a series estacionarias

Un supuesto fuerte hecho en el teorema de tipos de extremos es que la sucesión $\{Y_i\}_{i \geq 1}$ son independientes e idénticamente distribuidas, sin embargo, los datos frecuentemente

1.1. Teoría de valores extremos univariada

MDA(Δ)			
Distribución	$F(y)$	a_n	b_n
$N(0, 1)$	$\Phi(y), y \in \mathbb{R}$	$(2 \log n)^{-1/2}$	$(2 \log n)^{1/2} - (2 \log n)^{-1/2}$ $\times (\log 4\pi + \log \log n)/2$
Weibull(α)	$1 - \exp(-y^\alpha), y > 0, \alpha > 0$	$(\log n)^{1/\alpha}$	$\frac{1}{a}(\log n)^{1/\alpha-1}$
Exp(λ)	$1 - \exp(-\lambda y), y > 0, \lambda > 0$	λ^{-1}	$\lambda^{-1} \log n$
MDA(Ψ_α)			
Distribución	$F(y)$	a_n	b_n
Uniforme	$(y - a)/(b - a), y \in [a, b]$	$(b - a)/n$	$b - (b - a)/n$
Beta	$\int_0^y \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} s^{a-1} (1-s)^{b-1} ds,$ $y \in (0, 1), a, b > 0$	1	$\left(n \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b+1)}\right)^{-1/b}$
MDA(Φ_α)			
Distribución	$F(y)$	a_n	b_n
Pareto	$1 - (a/y)^\alpha, y \geq a, \alpha > 0$	$a\alpha^{-1}n^{1/\alpha}$	$an^{1/\alpha}$
Cauchy	$1/2 + \pi^{-1} \arctan(y), y \in \mathbb{R}$	n/π	0

Tabla 1.1: Clasificación de varias distribuciones conocidas de acuerdo al dominio de atracción del máximo, con sus correspondientes constantes de normalización $a_n > 0$ y $b_n \in \mathbb{R}$.

no cumplen con este supuesto en dos aspectos; primero, las observaciones pueden no ser independientes. Por ejemplo, cuando se consideran datos ambientales puede existir dependencia temporal y las condiciones severas del clima pueden persistir en varias observaciones consecutivas. Segundo, pueden no tener la misma distribución, en este caso, por ejemplo, puede existir un comportamiento estacional o los datos pueden tener una tendencia global.

En esta sección, sólo estudiaremos la primera de ellas, la extensión de independencia y estacionariedad. La no estacionariedad, normalmente se trabaja modelando directamente los parámetros marginales, por ejemplo, usando regresión lineal [Katz *et al.* \(2002\)](#), modelos semi-paramétricos basados en splines ([Chavez-Demoulin y Davison, 2005, 2012](#)), modelos jerárquicos bayesianos ([Cooley *et al.*, 2007, Sang y Gelfand, 2009, 2010](#)).

Definición 1.7. (Series estacionarias). Sea $\{Y_i\}_{i \geq 1}$ una serie de tiempo, $T \subset \mathbb{N}$ algún conjunto finito y sea \mathbf{Y}_T que denota la colección de las Y_i 's tal que $i \in T$. Se dice que la serie de tiempo $\{Y_i\}_{i \geq 1}$ es estrictamente estacionaria si la distribución conjunta del vector \mathbf{Y}_T y \mathbf{Y}_{h+T} es la misma para cualquier longitud de $h \in \mathbb{N}$. El proceso se llama débilmente estacionario si para cualquier i_1, i_2 , $E(Y_{i_1}) = E(Y_{i_1+h})$ y $\text{cov}(Y_{i_1}, Y_{i_2}) = \text{cov}(Y_{i_1+h}, Y_{i_2+h})$.

Estacionariedad estricta significa que la traslación no afecta las propiedades probabilísticas del proceso. En otras palabras, corresponde a series cuyas variables pueden ser mutuamente dependientes, pero cuyas propiedades estocásticas son homogéneas a través del tiempo. Por otro lado, estacionariedad débil, sólo asume homogeneidad temporal de los primeros dos momentos.

En la práctica, la dependencia puede tomar diferentes formas y sólo se puede tener

1.1. Teoría de valores extremos univariada

a distancias cortas o distancias muy largas. De hecho, los procesos de memoria larga pueden afectar la convergencia de máximos normalizados a la distribución GEV. El contra ejemplo surge para sucesiones perfectamente dependientes: Si $Y_1, \dots, Y_n \sim F$ con $Y_i = Y_j$ casi seguramente, entonces $\max(Y_1, \dots, Y_n) \stackrel{D}{=} Y_1 \sim F$, por lo que la distribución del máximo puede tomar esencialmente cualquier forma. Como consecuencia, note que se puede obtener un resultado análogo al Teorema de tipos de extremos bajo estacionariedad estricta y dependencia corta. Leadbetter (1983) formalizó esta idea de dependencia corta con la llamada condición $D(u_n)$, que se define a continuación.

Definición 1.8. (Condición $D(u_n)$). Sean $\mathcal{A}, \mathcal{B} \subset \{1, \dots, n\}$ que denotan subconjuntos de índices $i_1 < \dots < i_p$ y $j_1 < \dots < j_q$ respectivamente, tal que $j_1 - i_p > l$, y sea $Y_{\mathcal{A}} \leq u$ que denota el evento $\cap_{i \in \mathcal{A}} \{Y_i \leq u\}$, similarmente para $Y_{\mathcal{B}} \leq u$. Entonces se satisface la condición $D(u_n)$ si

$$|P(Y_{\mathcal{A}} \leq u_n \cap Y_{\mathcal{B}} \leq u_n) - P(Y_{\mathcal{A}} \leq u_n)P(Y_{\mathcal{B}} \leq u_n)| \leq \alpha(n, l_n),$$

donde $\alpha(n, l_n) \rightarrow 0$ para alguna sucesión l_n tal que $l_n/n \rightarrow 0$ cuando $n \rightarrow \infty$.

Esta condición asegura que los eventos raros suficientemente separados se pueden considerar que sean cercanos a independencia, de modo que su distribución conjunta no tiene efecto sobre las probabilidades límites de extremos. Para sucesiones gaussianas con auto correlación ρ_n con retraso n , por ejemplo la condición $D(u_n)$ se satisface tan rápido como $\rho_n \log n \rightarrow 0$ cuando $n \rightarrow \infty$ (Beirlant *et al.*, 2006). De hecho, bajo esta condición y estacionariedad, se cumple que para cualquier sucesión de umbrales u_n , $P(M_n \leq u_n) = \{P(M_{\lfloor n/k_n \rfloor} \leq u_n)\}^{k_n} + o(1)$ para alguna sucesión creciente $k_n \rightarrow \infty$ (ver Beirlant *et al.* (2006)). Esto es, para series de tiempo estacionarias que satisfacen la condición $D(u_n)$, M_n se puede considerar como el máximo de bloques cercanos a independencia. Esta es la relación principal que se presenta en el siguiente Teorema.

Teorema 1.4. Sea $\{Y_i\}_{i \geq 1}$ una serie de tiempo estrictamente estacionaria. Suponga que $(M_n - b_n)/a_n \xrightarrow{D} Z \sim G$ para alguna distribución no degenerada G y adecuadas sucesiones $a_n > 0$ y $b_n \in \mathbb{R}$. Si $D(u_n)$ se cumple para $u_n = a_n y + b_n$ para cualquier $y \in \mathbb{R}$, entonces G es una distribución de valores extremos.

Aquí, la familia paramétrica de distribuciones derivada como el límite para máximos normalizados de series estacionarias con dependencia corta es la misma que en el caso de independencia. Sea $\{Y_i^*\}_{i \geq 1}$ que denota el equivalente a un proceso estacionario $\{Y_i\}_{i \geq 1}$, con máximo M_n^* . Bajo las condiciones del Teorema 1.4 tanto $(M_n - b_n)/a_n$ como $(M_n^* - b_n)/a_n$ convergen a una distribución GEV, pero los parámetros pueden diferir. El Teorema 1.5 y la Proposición 1.1 muestran como hacer la liga entre estas dos probabilidades límites.

Teorema 1.5. Bajo condiciones adecuadas, $(M_n^* - b_n)/a_n \xrightarrow{D} Z^* \sim G^*$, cuando $n \rightarrow \infty$ donde G^* es una función de distribución no degenerada y $a_n > 0$ y $b_n \in \mathbb{R}$ son sucesiones de normalización, si y sólo si, $(M_n - b_n)/a_n \xrightarrow{D} Z \sim G$, y $G = G^{*\theta}$, para alguna constante $0 < \theta \leq 1$.

1.1. Teoría de valores extremos univariada

Definición 1.9. (Índice extremo). Al parámetro θ en el [Teorema 1.5](#) se le llama índice extremo. Si θ decrece, la dependencia de series a niveles altos es fuerte, mientras que independencia extrema se alcanza cuando $\theta = 1$.

Los parámetros de la distribución límite derivados en el [Teorema 1.5](#) son ligados como sigue.

Proposición 1.1. (Parámetros de la distribución GEV bajo dependencia corta). Usando la misma notación que el [Teorema 1.5](#) y tomando a $G \sim \text{GEV}(\mu, \sigma, \xi)$ y $G^* \sim \text{GEV}(\mu^*, \sigma^*, \xi^*)$ se tiene que $\mu = \mu^* - \sigma^*(1 - \theta^{\xi^*})/\xi^*$, $\sigma = \sigma^*\theta^{\xi^*}$ y $\xi = \xi^*$, donde θ es el índice extremo.

De acuerdo a la [Proposición 1.1](#), el parámetro de forma de la distribución límite no se ve afectado por la dependencia temporal. Sin embargo, la dependencia fuerte se tiene en los parámetros de localización y de dispersión. En otras palabras, M_n es estocásticamente más pequeño que M_n^* , lo cual significa que la dependencia reduce el tamaño de los extremos para una serie de longitud dada. Consecuentemente, el nivel de retorno crece conforme la dependencia es más fuerte.

1.1.3. Inferencia

Motivados por el Teorema de tipos de extremos y el [Teorema 1.4](#), el uso de la distribución GEV se justifica asintóticamente para modelar la distribución de máximos por bloques. Supóngase que un proceso estacionario a tiempo continuo $\{Y_t\}_{t \in T}$ se ha muestreado a $n = MN$ puntos regulares en el tiempo y que la serie resultante cae en el dominio de atracción de alguna distribución GEV y sean y_1, \dots, y_n que denotan estas observaciones.

La aproximación clásica para inferencia es formar N bloques de M observaciones y ajustar la distribución $\text{GEV}(\mu, \sigma, \xi)$ a los bloques de máximos $m_1 = \max(y_1, \dots, y_M), \dots, m_N = \max(y_{M(N-1)+1}, \dots, y_{MN})$ por máxima verosimilitud. Si la longitud de los bloques M se elige suficientemente larga de tal manera que los máximos m_1, \dots, m_N se puedan considerar mutuamente independientes, la función de log-verosimilitud es de la forma,

$$l(\psi) = -N \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^N \log \left\{1 + \xi \left(\frac{m_i - \mu}{\sigma}\right)\right\}_+ - \sum_{i=1}^N \left\{1 + \xi \left(\frac{m_i - \mu}{\sigma}\right)\right\}_+^{-1/\xi}, \quad (1.1)$$

donde $\psi = (\mu, \sigma, \xi)$ es el vector de parámetros. En la práctica esta expresión se puede maximizar numéricamente con rutinas de optimización disponibles en algún software estadístico (por ejemplo la función `optim` en `R`, ([R Development Core Team, 2015](#))).

La elección de la longitud del bloque M es crucial debido a que corresponde a un equilibrio entre sesgo y varianza. Si M crece, el sesgo puede ser pequeño, pero como el tamaño de la muestra para ajustar disminuye al mismo tiempo, la varianza de los parámetros estimados puede ser muy grande. En la práctica, usualmente se elige a M de forma pragmática, de tal forma que los bloques correspondan a periodos naturales tal como semanas meses o años.

1.2. Teoría de valores extremos multivariada

En situaciones regulares, los estimadores de máxima verosimilitud son asintóticamente normales, consistentes y tienen varianza asintótica igual al recíproco de la matriz de información de Fisher (Davison, 2003, p.118). Las condiciones de regularidad para estimadores de máxima verosimilitud en la teoría de valores extremos son discutidas en Smith (1985), él ha establecido que el comportamiento límite de los estimadores de máxima verosimilitud depende del valor del parámetro de forma:

- cuando $\xi > -1/2$, los estimadores de máxima verosimilitud pertenecen a la teoría estándar.
- cuando $-1 \leq \xi \leq -1/2$, los estimadores son la solución de la ecuación escor, pero no tienen la distribución límite usual.
- cuando $\xi \leq -1$, los estimadores no son solución de la ecuación escor.

La evaluación de la incertidumbre de los parámetros estimados se puede basar en normalidad asintótica, suponiendo que los estimadores de máxima verosimilitud han alcanzado su distribución límite y que la inversa de la matriz de información de Fisher es un estimador adecuado para su varianza. Pero en la práctica, generalmente se prefieren los métodos basados en verosimilitud perfil (Coles *et al.*, 2001, p. 57-61), ya que reflejan mejor la asimetría que uno espera ver para la verosimilitud con respecto a ξ .

La estimación por máxima verosimilitud no es la única posibilidad para inferencia. Mientras que las técnicas basadas en el método de los momentos usualmente es deficiente para extremos debido a que éstos pueden no existir, los momentos ponderados han demostrado ser útiles debido a sus buenas propiedades para tamaños de muestra pequeños (Hosking *et al.*, 1985, Katz *et al.*, 2002), aunque es relativamente difícil extenderlos a datos más complejos. Las técnicas basadas en métodos bayesianos también se pueden aplicar y son muy eficientes para problemas de dimensiones altas donde los estimadores de máxima verosimilitud no se pueden calcular, por ejemplo en problemas que involucran el cálculo de integrales de dimensiones altas, tal como modelos jerárquicos (Blanchet y Davison, 2012, Cooley *et al.*, 2007, Sang y Gelfand, 2009). La aplicabilidad de los métodos bayesianos en ésta área ha sido posible gracias al desarrollo del algoritmo cadenas de Markov Monte Carlo (MCMC).

1.2. Teoría de valores extremos multivariada

En la práctica, la modelación conjunta de extremos resulta de interés por diversas razones, una de ellas es que en ocasiones se está interesado en tener una descripción cuantitativa de la estructura y el grado de dependencia extrema entre dos o más series de observaciones. Otra razón, es que el uso de modelos multivariados permite tratar a las observaciones de una manera más general ya que la interpretación de los resultados es mucho más fácil en este contexto. Y finalmente, si la dependencia extrema se modela adecuadamente, la modelación conjunta permite usar la serie para estimar mejor los parámetros marginales. En estadística espacial, esto se refiere a tener un equilibrio entre el espacio y el tiempo.

1.2. Teoría de valores extremos multivariada

Aunque parece ser atractiva, la modelación conjunta de extremos es difícil por varios aspectos. En primer lugar, no hay una forma obvia de ordenar observaciones multivariadas, de tal forma que la definición de observación multivariada extrema no es tan clara como en el caso univariado. En segundo lugar, la clase de distribuciones de valores extremos multivaridos es no paramétrica, es decir, no se pueden caracterizar por un número finito de parámetros. Además, aunque la teoría de valores extremos multivariada esta bien desarrollada, los modelos paramétricos para dimensiones mayores a dos siguen faltando.

En adelante, todos los vectores se denotarán por letras negritas, y se asume que son de dimensión D , esto es, $\mathbf{y} = (y_1, y_2, \dots, y_D)$, a menos que se especifique lo contrario, todas las operaciones se hacen componente a componente, por ejemplo, $\mathbf{a} \leq \mathbf{y}$ significa que $a_i \leq y_i$, para todo $i = 1, \dots, D$.

1.2.1. Distribuciones multivariadas y max-stabilidad

Sea $\mathbf{Y} = (Y_1, \dots, Y_D)$ un vector D -dimensional con distribuciones marginales F_1, \dots, F_D y distribución conjunta F , y sea $\{\mathbf{Y}_i\}_{i \geq 1}$ una sucesión i.i.d de vectores aleatorios distribuidos como \mathbf{Y} . Denote por $Y_{i,j}$ la componente j -ésima del vector \mathbf{Y}_i , y sea $M_{n,j} = \max(Y_{1,j}, \dots, Y_{n,j})$, $j = 1, \dots, D$, y sea $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,D})$ el vector de máximos.

Al igual que el caso univariado, consideramos una renormalización para \mathbf{M}_n con el fin de obtener un límite trivial cuando el tamaño de muestra tiende a infinito. Específicamente, suponga que existen sucesiones $\mathbf{a}_n = (a_{n,1}, \dots, a_{n,D}) \in \mathbb{R}_+^D$ y $\mathbf{b}_n = (b_{n,1}, \dots, b_{n,D}) \in \mathbb{R}^D$ tal que el vector de máximos renormalizados $\mathbf{M}_n^* = (\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n$ converge a una variable aleatoria con función de distribución conjunta G y marginales no degeneradas G_1, \dots, G_D . Si tales sucesiones se pueden encontrar, el límite de la distribución se llama *distribución de valores extremos multivariada*. [Balkema y Resnick \(1977\)](#) mostraron que si la convergencia ocurre, tal límite de la distribución G tiene que ser max-infinitamente divisible.

Definición 1.10. (Max-infinitamente divisible). Una distribución G es max-infinitamente divisible si para cualquier $k \in \mathbb{N}$, $G^{1/k}$ es una función de distribución. Una variable aleatoria Z es max-infinitamente divisible si su distribución es max-infinitamente divisible.

Una distribución max-infinitamente divisible produce una colección de distribuciones $\{F_k\}_{k \in \mathbb{N}}$ tal que $F_k^k = G$, para todo $k \in \mathbb{N}$; en otras palabras, G es la función de distribución del máximo de k variables aleatorias independientes que se distribuyen de acuerdo a F_k . Note que todas las distribuciones unidimensionales son max-infinitamente divisibles.

Debido a la teoría de valores extremos univariada y el teorema de tipos de extremos, uno sabe que si las marginales G_j son no degeneradas, tiene que ser distribuciones GEV, esto es, $M_{n,j}^* = (M_{n,j} - b_{n,j})/a_{n,j} \rightarrow Z_j \sim \text{GEV}(\mu_j, \sigma_j, \xi_j)$, cuando $n \rightarrow \infty$ para cualquier $j = 1, \dots, D$. O equivalentemente, los límites marginales son max-stable. Consecuentemente, como el límite de la distribución conjunta G es max-infinitamente divisible con marginales

1.2. Teoría de valores extremos multivariada

max-stable, tiene que ser en sí max-stable (de Haan y Ferreira, 2006), esto es, para cada $k \in \mathbb{N}$ y $\mathbf{z} = (z_1, \dots, z_D) \in \mathbb{R}_+^D$ pueden existir constantes $\mathbf{a}_k \in \mathbb{R}_+^D$ y $\mathbf{b}_k \in \mathbb{R}^D$ tal que

$$G^k(\mathbf{a}_k \mathbf{z} + \mathbf{b}_k) = G(\mathbf{z}). \quad (1.2)$$

Por definición, una distribución max-stable es max-infinitamente divisible; lo inverso no necesariamente es cierto.

En la práctica, es común para el estudio de extremos multivariados proceder en dos fases: las distribuciones marginales se estiman inicialmente usando la metodología univariada, y luego usando la transformada de la integral se convierten los datos a un escala común, con el fin de manejar la estructura de dependencia usando la teoría de valores extremos multivariada. Por razones de elegancia matemática, la transformación de los datos se hace frecuentemente a una distribución Fréchet unitaria, involucrando el mapeo $t_j(\cdot) = -1/\log\{G(\cdot)\}$, $j = 1, \dots, D$. Más precisamente, si $\mathbf{Z} = (Z_1, \dots, Z_D)$ tienen distribución conjunta G entonces las variables transformadas $t_1(Z_1), \dots, t_D(Z_D)$ son Fréchet unitarias con la misma estructura de dependencia, de hecho se cumple que, $\tilde{G}_j(z) = P\{t_j(Z_j) \leq z\} = \exp(-1/z)$, para $z > 0$, y $G(z_1, \dots, z_D) = \tilde{G}\{t_1(z_1), \dots, t_D(z_D)\}$.

1.2.2. La medida del exponente

Balkema y Resnick (1977) mostraron que las distribuciones max-infinitamente divisibles producen una medida μ sobre $(-\infty, \infty)^D$ tal que para todo $\mathbf{z} \in \mathbb{R}^D$,

$$G(\mathbf{z}) = \exp\{-\mu(A_{\mathbf{z}})\} = \exp\{-V(\mathbf{z})\},$$

donde $A_{\mathbf{z}} = (-\infty, \mathbf{z}]^c$, con B^c denota el complemento del conjunto B . La *medida del exponente* μ contiene toda la información acerca de la dependencia entre las variables Z_1, \dots, Z_D . En algunas ocasiones, a la función $V = -\log G$ también se le llama medida del exponente. Por simplicidad, pero sin pérdida de generalidad, limitaremos la discusión al caso de la distribución Fréchet unitaria, donde $\mathbf{a}_k = k^{-1}$ y $\mathbf{b}_k = 0$ en (1.2). Por lo tanto, se escribe $G_j \equiv \tilde{G}_j$, $\mu \equiv \tilde{\mu}$, $A_{\mathbf{z}} \equiv \tilde{A}_{\mathbf{z}}$ y $V \equiv \tilde{V}$, omitiendo la tilde por simplicidad. Como las marginales de G se supone son Fréchet unitarias, se puede verificar que la medida de exponente satisface la restricción $V(z_1, \infty, \dots, \infty) = 1/z_1$, y similarmente para las otras marginales. Además, debido a la propiedad max-stable (1.2), la función V es homogénea de orden -1 , esto es, $V(t\mathbf{z}) = t^{-1}V(\mathbf{z})$ para todo $\mathbf{z} > 0$ y $t > 0$.

1.2.3. Representación espectral

El siguiente Teorema proporciona una representación espectral de la medida del exponente, caracterizando así todas las posibles distribuciones límites para máximos bajo una renormalización afín.

Teorema 1.6. (*Caracterización de las distribuciones de valores extremos*). Si el vector normalizado $\mathbf{M}_n^* \xrightarrow{D} \mathbf{Z} \sim G$, donde G es una función de distribución no degenerada, entonces G tiene la forma,

$$G(\mathbf{z}) = \exp\{-V(\mathbf{z})\}, \quad \mathbf{z} > 0,$$

1.2. Teoría de valores extremos multivariada

donde

$$V(\mathbf{z}) = D \int_{S_D} \max(\mathbf{w}/\mathbf{z}) dH(\mathbf{w}), \quad (1.3)$$

y dH es una medida sobre el simplex $(D-1)$ -dimensional $S_D = \{\mathbf{w} \in \mathbb{R}_+^D : \sum_{j=1}^D w_j = 1\}$, satisfaciendo la restricción $\int_{S_D} w_j dH(\mathbf{w}) = 1/D$, $j = 1, \dots, D$.

Cuando $D = 2$, dH es una distribución sobre el intervalo $[0, 1]$, sujeto a la restricción $\int_0^1 w dH(w) = 1/2$. A la medida dH se le conoce frecuentemente como medida espectral debido a su interpretación en términos del pseudo-radio $r = z_1 + \dots + z_D$ y del pseudo-ángulo $w_1 = z_1/r, \dots, w_D = z_D/r$ (Beirlant *et al.*, 2006, p. 258). Contrario al caso univariado, donde una familia de distribuciones paramétrica cubre todos los posibles límites, la expresión (1.3) implica que las distribuciones de valores extremos multivariadas no se pueden describir por un número finito de parámetros.

1.2.4. Función de dependencia de Pickands

En dimensión $D = 2$, una representación alternativa de la ecuación (1.3) (Pickands, 1981) conduce a la llamada función de dependencia de Pickands, denotada por $A(w)$. Resulta que se puede escribir la medida del exponente como,

$$V(z_1, z_2) = (z_1^{-1} + z_2^{-1}) A\left(\frac{z_1}{z_1 + z_2}\right),$$

donde la función $A(w)$ satisface $A(w) = 2 \int_0^1 \max\{(1-w)q, w(1-q)\} dH(q)$. Aquí $A(w)$ es una función definida sobre el intervalo $[0, 1]$, y es tal que: i) $A(0) = A(1) = 1$, ii) $A(w)$ es convexa y iii) $A(w)$ esta contenida en una región triangular definida por $\max(w, 1-w) \leq A(w) \leq 1$ para todo $w \in [0, 1]$. La función $A(w)$ esta acotada por dos casos, completa dependencia cuando $A(w) = \max(w, 1-w)$ e independencia asintótica cuando $A(w) = 1$, también se puede interpretar como una medida de la dependencia entre Z_1 y Z_2 en la dirección w , donde $w = z_1/(z_1 + z_2)$ es el pseudo-ángulo entre z_1 y z_2 . La relación entre la función de dependencia de Pickands y la medida espectral dH se explica a detalle en Beirlant *et al.* (2006, p. 268-270).

1.2.5. Modelos paramétricos

La especificación de un modelo paramétrico para la medida del exponente $V(\cdot)$ en (1.3), o equivalentemente para la medida espectral dH , equivale a restringir la dependencia a tener una estructura particular, por lo que es esencial construir modelos de dependencia flexibles que sean fácil de interpretar. Por un lado, la independencia asintótica surge si $V(\mathbf{z}) = z_1^{-1} + \dots + z_D^{-1}$, o equivalentemente si $dH(\mathbf{e}_j) = D^{-1}$ para cada vértice \mathbf{e}_j del simplex $(D-1)$ -dimensional S_D . Por otro lado, la dependencia completa, se atrae cuando $V(\mathbf{z}) = \max(\mathbf{z}^{-1})$, o $dH(\mathbf{D}^{-1}) = 1$, donde $\mathbf{D}^{-1} = (D^{-1}, \dots, D^{-1})$. Muchos modelos razonables caen dentro de estos dos casos, sin embargo, en dimensiones grandes, la construcción de modelos flexibles para extremos es bastante difícil, actualmente sólo

1.2. Teoría de valores extremos multivariada

algunos han sido propuestos (por ejemplo, ver, [Tawn \(1990\)](#), [Coles y Tawn \(1991\)](#), [Cooley et al. \(2010\)](#), [Ballani y Schlather \(2011\)](#), [Segers \(2012\)](#)), pero por lo general sufren por la falta de flexibilidad para valores de D grandes y traen problemas computacionales e inferenciales.

El modelo paramétrico más famoso para el caso $D = 2$, es el modelo logístico, debido a [Gumbel \(1961\)](#):

$$V(z_1, z_2) = (z_1^{-1/\alpha} + z_2^{-1/\alpha})^\alpha, \quad z_1, z_2 > 0,$$

para algún parámetro de dependencia $\alpha \in (0, 1]$. El caso $\alpha \rightarrow 1$ corresponde a independencia, mientras que el caso $\alpha \rightarrow 0$ corresponde a una independencia completa. Este modelo se puede extender fácilmente a dimensiones altas, pero debido a su simetría, frecuentemente tiene poca flexibilidad.

Un modelo más sofisticado que es capaz de capturar asimetría en la estructura de dependencia es el modelo logístico asimétrico propuesto por [Tawn \(1988\)](#), extendido a dimensión $D \geq 3$ ([Coles y Tawn, 1991](#), [Stephenson, 2009](#)):

$$V(z_1, z_2) = (1 - \theta)/z_1 + (1 - \phi)/z_2 + \{(\theta/z_1)^{1/\alpha} + (\phi/z_2)^{1/\alpha}\}^\alpha, \quad z_1, z_2 > 0,$$

para algún parámetro de dependencia $\alpha \in (0, 1]$ y parámetros de asimetría $\theta, \phi \in (0, 1]$. Cuando $\theta = \phi = 1$, este modelo se reduce al modelo logístico. Un buen ajuste de este modelo de entre los distintos modelos de valores extremos bivariados se dio en un estudio realizado por [Ferrez et al. \(2011\)](#), en el análisis de temperaturas extremas bajo cubierta forestal comparada con un campo abierto.

Otra extensión del modelo logístico para el cual las variables no son intercambiables es el modelo bilogístico ([Smith, 1990a](#)):

$$V(z_1, z_2) = z_1^{-1}q^{1-\alpha} + z_2^{-1}(1-q)^{1-\beta}, \quad z_1, z_2 > 0,$$

aquí $q = q(z_1, z_2, \alpha, \beta)$ es la raíz de la ecuación $(1 - \alpha)z_1^{-1}(1 - q)^\beta - (1 - \beta)z_2^{-1}q^\alpha = 0$, con $\alpha, \beta > 0$. El valor de $|\alpha - \beta|$ determina el grado de asimetría en la estructura de dependencia. En particular, cuando $\alpha = \beta$, el modelo bilogístico se reduce al modelo logístico. La completa dependencia se alcanza en el límite cuando $\alpha = \beta$ se aproximan a cero.

[Coles y Tawn \(1991\)](#) propusieron otro modelo asimétrico, el modelo Dirichlet, para el cual,

$$V(z_1, z_2) = \{1 - B_{\alpha+1;\beta}(t)\}/z_1 + B_{\alpha;\beta+1}(t)/z_2, \quad z_1, z_2 > 0,$$

donde $\alpha, \beta > 0$ y $B_{a;b}(t)$ es la función de distribución acumulada beta con parámetros a, b evaluada en $t = \alpha z_1 / (\alpha z_1 + \beta z_2)$, éste modelo es simétrico cuando $\alpha = \beta$.

Entre los modelos simétricos, el modelo de Hüsler-Reiss ([Hüsler y Reiss, 1989](#)) basado en la distribución normal estándar ha ganado mucha atención. Este modelo es el único límite posible de máximos reescalados de variables aleatorias gaussianas, dando soporte para muchas aplicaciones. [Hüsler y Reiss \(1989\)](#) probaron que si $\{\mathbf{Y}_i\}_{i \geq 1}$ es una sucesión de variables aleatorias bivariadas normales independientes con media cero, varianza unitaria y correlación ρ_n que satisface $4(1 - \rho_n) \log n \rightarrow a^2$, cuando $n \rightarrow \infty$, entonces $P(\mathbf{M}_n^* \leq$

1.2. Teoría de valores extremos multivariada

$\log z) \rightarrow G(\mathbf{z}) = \exp\{-V(\mathbf{z})\}$ donde $\mathbf{M}_n^* = (\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n$ es el vector de máximos renormalizados, $\mathbf{a}_n = (2 \log n)^{1/2}$, $\mathbf{b}_n = \{2 \log n - \log \log n - \log(4\pi)\}^{1/2}$ y

$$V(z_1, z_2) = \frac{1}{z_1} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{z_1}{z_2} \right) \right\} + \frac{1}{z_2} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{z_2}{z_1} \right) \right\}, \quad (1.4)$$

donde $\Phi(\cdot)$ denota la función de distribución acumulada normal estándar, el parámetro $a > 0$ es un parámetro de dependencia. La independencia ocurre cuando $a \rightarrow \infty$, y dependencia cuando $a \rightarrow 0$. La expresión (1.4) se pueden extender para el caso general en dimensión D (Huser y Davison, 2013, Hüsler y Reiss, 1989).

1.2.6. Inferencia

Suponga que los datos se pueden descomponer en N bloques de M observaciones independientes de dimensión D , esto es $n = MN$, con $M, N \in \mathbb{N}$, se pueden formar N bloques de máximos $\mathbf{z}_i = (z_{1;i}, \dots, z_{D;i})$, $i = 1, \dots, N$ y ajustar un modelo de valores extremos a estos o uno no paramétrico. En el caso paramétrico el primer paso es elegir un modelo para $G = \exp(-V)$ de las distribuciones de valores extremos. La subfamilia de modelos que se elijan, debe estar bien balanceada entre flexibilidad y simplicidad con el fin de capturar una amplia gama de estructuras de dependencia. En la práctica, la función G depende de un número finito de parámetros desconocidos, incluidos en ψ . La función de verosimilitud es

$$L(\psi) = \prod_{i=1}^N g(\mathbf{z}_i),$$

donde $g(\mathbf{z}) = \partial^D G(\mathbf{z}) / (\partial z_1 \cdots \partial z_D)$, y el parámetro estimado $\hat{\psi}$ se obtiene maximizando $L(\psi)$ con respecto a ψ . Si el modelo elegido es correcto, la varianza de $\hat{\psi}$ se puede calcular por la inversa de la matriz de información de Fisher, esto es

$$\widehat{\text{var}}(\hat{\psi}) = \left\{ - \sum_{i=1}^N \partial^2 \log g(\mathbf{z}_i) / (\partial \psi \partial \psi^T) \right\}_{\psi=\hat{\psi}}^{-1}$$

en dimensión $D = 2$ la densidad conjunta se puede expresar como $g(z_1, z_2) = (V_1 V_2 - V_{12}) \exp(-V)$ donde por simplicidad se escribe $V_1 = \partial V(z_1, z_2) / \partial z_1$ y así para V_2 y V_{12} . Sin embargo, la función de verosimilitud es intratable conforme D aumenta, de hecho el número de términos involucrados en la densidad $g(\mathbf{z})$ es igual al número de particiones del conjunto $\{1, \dots, D\}$, por lo que se necesitan métodos alternativos para realizar inferencia en dimensiones altas. Una posibilidad es considerar sustituir la verosimilitud completa, como por ejemplo por verosimilitud compuesta, en el caso de extremos espaciales, se usa la verosimilitud a pares, bajo ciertas condiciones los estimadores obtenidos por esta vía son asintóticamente normales y fuertemente consistentes (Varin, 2008). Además, la inferencia por verosimilitud a pares es robusta contra falta de especificación de interacciones de orden superior, aunque con una pérdida en la eficiencia comparada con máxima verosimilitud. Los detalles de máxima verosimilitud compuesta en el enfoque espacial se dará en el siguiente capítulo.

1.2.7. Medida de dependencia extrema

En la literatura se han desarrollado diferentes medidas de dependencia extrema, con sus propios métodos de estimación, una revisión a detalle de éstas se puede encontrar en [Davison *et al.* \(2013\)](#). El coeficiente extremal es más adecuado para variables asintóticamente dependientes, esto es, en el dominio de atracción del máximo de alguna distribución max-stable no trivial, mientras que el coeficiente de dependencia de la cola de la distribución es mas conveniente para variables asintóticamente independientes. En lo siguiente, asumimos que $\mathbf{Y} = (Y_1, \dots, Y_D)$ tiene distribuciones marginales Fréchet unitarias.

Coeficiente extremal θ_D . Suponga que \mathbf{Y} esta en el dominio de atracción del máximo de alguna distribución de valores extremos G con medida del exponente V y marginales Fréchet unitarias, esto es, $G(\mathbf{y}) = \exp\{-V(\mathbf{y})\}$, $\mathbf{y} > 0$. La estructura de dependencia extrema está en la medida del exponente, como V es homogéneo de orden -1 la dependencia extrema entre las variables Y_1, \dots, Y_D se resume en el coeficiente extremal

$$\theta_D = V(1, \dots, 1) \in [1, D]. \quad (1.5)$$

Y es fácil ver que

$$P(\mathbf{Y} \leq y) = \{\exp(-1/y)\}^{\theta_D},$$

θ_D se puede interpretar como el número efectivo de variables independientes. Para una dependencia perfecta de los datos, $\theta_D = 1$, y para una independencia asintótica de éstos, $\theta_D = D$. En dimensión $D = 2$, se tiene $\theta_2 = V(1, 1) = 2A(1/2)$, donde A es la función de dependencia de Pickands. En realidad, este coeficiente mide la dependencia extrema “sobre la diagonal $y_1 = \dots = y_D = y$ ”, en este sentido, no proporciona una descripción exhaustiva de la dependencia extrema, especialmente para D grande, pero puede ser útil para análisis exploratorios o como una herramienta para la validación de modelos, comparando estimadores empíricos y ajustados ([Naveau *et al.*, 2009](#), [Schlather y Tawn, 2003](#)).

El estimador de Schlather-Taw del coeficiente extremal ([Schlather y Tawn, 2003](#)) se basa en la distribución asintótica de $Z = \max(Y_1, \dots, Y_D)$. Sea z_1, \dots, z_n que denotan el valor observado del estadístico Z , y sea N_u el número de z_i 's que exceden un umbral u . Como $P(Z \leq z) \approx \exp(-\theta_D/z)$, $z > u$, un estimador de máxima verosimilitud censurado $\hat{\theta}_D$ se puede obtener de forma cerrada como

$$\hat{\theta}_D = \left[\frac{1}{N_u} \sum_{i=1}^n \left\{ \frac{1}{z_i} I(z_i > u) + \frac{1}{u} I(z_i \leq u) \right\} \right]^{-1},$$

donde $I(\cdot)$ es la función indicadora. Otro estimador para el coeficiente extremal basado en el concepto de madogramas, analogo al variograma, fue propuesto por [Cooley *et al.* \(2006\)](#) y [Naveau *et al.* \(2009\)](#).

Procesos max-stable para extremos espaciales

En los últimos años la modelación estadística de eventos extremos ha tenido un gran número de aplicaciones en muchas áreas, por ejemplo, en hidrología, medio ambiente, por mencionar algunas, en las cuales la teoría de valores extremos se ha desarrollado ampliamente, [Resnick \(2007, 2013\)](#) y de [de Haan y Pereira \(2006\)](#) presentan el sustento matemático mientras que [Beirlant *et al.* \(2006\)](#), [Coles *et al.* \(2001\)](#) presentan su fundamento estadístico. Dentro de esta, la modelación con extremos espaciales presenta, recientemente, un gran interés, esto ha dado lugar a dos formas diferentes de modelación. La primera se basa en procesos max-stable, los cuales son una generalización infinito dimensional de la teoría de valores extremos multivariada ([Davison *et al.*, 2012](#), [Huser y Davison, 2014](#), [Smith, 1990b](#)). La segunda usa modelos jerárquicos bayesianos con una estructura espacial latente, en ésta se puede describir la variación espacial de los parámetros en la distribución marginal. En este capítulo abordaremos la primera de ellas en donde una de las primeras aplicaciones de modelación se encuentra en [Coles y Tawn \(1996\)](#), quienes modelan datos de lluvias extremas usando estos procesos. Otra aplicación a ráfagas de viento se presenta en [Coles y Tawn \(1994\)](#) quienes usan estos procesos para modelar la dependencia angular para la velocidad del viento. Más recientemente, una aplicación a datos de lluvias se puede encontrar en [Padoan *et al.* \(2010\)](#) quienes modelan precipitaciones pluviales y proponen un procedimiento práctico de estimación basado en verosimilitud a pares. [Westra y Sisson \(2011\)](#) aplican procesos de extremos espaciales para investigar si las tendencias espaciales de precipitación diarias son significativas en el tiempo y simulan campos aleatorios espaciales comparando observaciones en múltiples localizaciones. [Blanchet y Davison \(2011\)](#) ajustan modelos espaciales a datos de nevadas extremas. [Davison y Gholamrezaee \(2011\)](#) modelan máximos de temperatura anual, un fenómeno a gran escala en comparación con la precipitación. La segunda forma de modelación se aborda en el [Capítulo 3](#).

2.1. Fundamentos de procesos aleatorios

En esta sección se resumen algunos conceptos de procesos aleatorios espaciales, para más detalles ver, por ejemplo, Cressie (2015), Stein (2012), Banerjee *et al.* (2014), Diggle y Ribeiro (2007) y Cressie y Wikle (2015). A lo largo de este capítulo se usará la notación $\mathbf{s} \in \mathcal{S}$ para denotar una localización genérica en algún espacio $\mathcal{S} \subset \mathbb{R}^d$, en tal caso los puntos $\mathbf{s} \in \mathcal{S}$ serán llamados “estaciones”, “localizaciones” ó “ubicaciones”.

Definición 2.1. (Proceso aleatorio). Un proceso aleatorio es una colección de variables aleatorias $\{Y(\mathbf{s})\}_{\mathbf{s} \in \mathcal{S}}$ definidas sobre un espacio de probabilidad (Ω, \mathcal{F}, P) e indexadas por un parámetro $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d$.

Definición 2.2. (Procesos aleatorios idénticamente distribuidos). Dos procesos $Y_1(\mathbf{s})$ y $Y_2(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, no necesariamente definidos sobre el mismo espacio de probabilidad, se dice que tienen la misma distribución finito dimensional o que son idénticamente distribuidos si los vectores $\{Y_1(\mathbf{s}_1), \dots, Y_1(\mathbf{s}_k)\}$ y $\{Y_2(\mathbf{s}_1), \dots, Y_2(\mathbf{s}_k)\}$ tienen la misma distribución para cualquier k y cualquier elección de $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathcal{S}$.

Un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, está descrito por sus dos primeros momentos, esto es, $m(\mathbf{s}) = E\{Y(\mathbf{s})\}$, $\mathbf{s} \in \mathcal{S}$ y la función de covarianza $C(\mathbf{s}_i, \mathbf{s}_j) = E[\{Y(\mathbf{s}_i) - m(\mathbf{s}_i)\}\{Y(\mathbf{s}_j) - m(\mathbf{s}_j)\}]$, $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$, la varianza está definida como $v(\mathbf{s}) = C(\mathbf{s}, \mathbf{s})$ y la función de correlación es $\rho(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_i, \mathbf{s}_j)\{v(\mathbf{s}_i)v(\mathbf{s}_j)\}^{-1/2}$, $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$. Una categoría importante de procesos aleatorios es la clase de procesos gaussianos, los cuales están totalmente determinados por $m(\mathbf{s})$ y $C(\mathbf{s}_i, \mathbf{s}_j)$, $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$.

Definición 2.3. (Proceso gaussiano). Un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, se llama gaussiano, si todas sus distribuciones finito dimensionales son normales multivariadas y si para cualquier $k \in \mathbb{N}$ y para cualquier $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathcal{S}$ la distribución conjunta de $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_k)\}$ es normal. Un proceso gaussiano con media $m(\mathbf{s})$ y función de covarianza $C(\mathbf{s}_i, \mathbf{s}_j)$ se denota por $\text{GP}(m, C)$.

En extremos espaciales, los procesos gaussianos juega un papel importante debido a muchas propiedades que poseen, además se les considera la base de la geoestadística clásica.

Los fenómenos naturales frecuentemente se pueden modelar por procesos espaciales homogéneos, en la práctica se pueden suponer diferentes tipos de homogeneidad, en particular se han considerado diferentes tipos de estacionariedad. A continuación se definen tres tipos de ésta.

Definición 2.4. (Estacionariedad estricta). Se dice que un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, es estrictamente estacionario si la distribución de los vectores $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_k)\}$ y $\{Y(\mathbf{s}_1 + \mathbf{h}), \dots, Y(\mathbf{s}_k + \mathbf{h})\}$ es la misma para cualquier elección de $k \in \mathbb{N}$ y cualesquiera $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathcal{S}$ y cualquier $\mathbf{h} \in \mathbb{R}^d$ que cumpla que $\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_k + \mathbf{h} \in \mathcal{S}$.

Definición 2.5. (Estacionariedad débil). Un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, es débilmente estacionario, si su media es constante, esto es $m(\mathbf{s}) = m$, $\mathbf{s} \in \mathcal{S}$, y si su función de

2.1. Fundamentos de procesos aleatorios

covarianza sólo depende de la separación del vector, esto es, $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{h})$, donde $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ y $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$, en particular la varianza toma un valor constante $C(0)$ sobre \mathcal{S} y la función de correlación es $\rho(\mathbf{h}) = C(\mathbf{h})/C(0)$.

Definición 2.6. (Estacionariedad intrínseca y variograma). Se dice que un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, es intrínsecamente estacionario si sus incrementos son débilmente estacionarios, en particular, existe una función $\gamma(\mathbf{h})$ que se llama semivariograma, para la cual $2\gamma(\mathbf{h}) = \text{Var}\{Y(\mathbf{s}) - Y(\mathbf{s} + \mathbf{h})\}$ para cualquier $\mathbf{s} \in \mathcal{S}$ tal que $\mathbf{s} + \mathbf{h} \in \mathcal{S}$. La función $2\gamma(\mathbf{h})$ se llama variograma.

Los tres tipos de estacionariedad no son equivalentes, de la definición se puede ver que estacionariedad estricta implica estacionariedad débil, que a su vez implica estacionariedad intrínseca, pero la relación inversa generalmente no se cumple. Sin embargo, para procesos gaussianos, cuyas densidades finito dimensionales solo dependen de sus dos primeros momentos, estacionariedad estricta y débil son equivalentes.

La noción de isotropía esta estrechamente relacionada a la estacionariedad en el sentido de que ambas son propiedades de invarianza geométricas de procesos aleatorios, mientras que la estacionariedad es propiedad de cambio, la isotropía lo es de rotación.

Definición 2.7. (Campo aleatorio isotrópico). Se dice que un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d$ es isotrópico si para cualesquiera $k \in \mathbb{N}$, $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathcal{S}$ y cualquier rotación, esto es, si la matriz ortogonal $O \in \text{Isom}(\mathbb{R}^d)$ tal que $O\mathbf{s}_1, \dots, O\mathbf{s}_k \in \mathcal{S}$, la distribución conjunta de los vectores $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_k)\}$ y $\{Y(O\mathbf{s}_1), \dots, Y(O\mathbf{s}_k)\}$ es la misma.

Si el proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, es débilmente estacionario e isotrópico, su función de covarianza sólo depende de la longitud de separación del vector, no de su orientación, esto es, $C(\mathbf{s}_i, \mathbf{s}_j) = C(\|\mathbf{h}\|)$ donde $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ y $\|\cdot\|$ es alguna norma en \mathcal{S} (usualmente la norma euclidiana). Similarmente para cualquier otra estacionariedad, se puede hablar de isotropía espacial o temporal si se considera un proceso espacio-tiempo.

Por otro lado, un proceso aleatorio ergódico $Y(\mathbf{s})$ definido sobre un conjunto \mathcal{S} tiene la propiedad de que eventos distantes son cercanos a independencia. Para aplicaciones puramente espaciales, con $\mathcal{S} \subset \mathbb{R}^d$, el término “eventos distantes” se puede interpretar como la distancia euclidiana.

2.1.1. Funciones de covarianza

Suponga que se tiene un proceso aleatorio $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$ indexado por algún subespacio continuo $\mathcal{S} \subset \mathbb{R}^d$, y sean $C(\mathbf{h})$, $\rho(\mathbf{h})$ y $\gamma(\mathbf{h})$ su función de covarianza, su función de correlación y su semivariograma, entonces, $2\gamma(\mathbf{h}) = \text{Var}\{Y(\mathbf{s}) - Y(\mathbf{s} + \mathbf{h})\} = 2\{C(0) - C(\mathbf{h})\}$. Así, $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) = C(0)\{1 - \rho(\mathbf{h})\}$, lo que implica que se pueda recuperar $\gamma(\mathbf{h})$ a partir de $C(\mathbf{h})$. Si el proceso espacial es una mezcla, se tiene, $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h}) = \lim_{|u| \rightarrow \infty} \gamma(u) - \gamma(\mathbf{h})$, si el límite existe el proceso es débilmente estacionario con función de covarianza $C(\mathbf{h})$.

2.2. Procesos max-stable

Ambas funciones, de covarianza y de correlación, deben ser definidas positivas, esto es,

$$\sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \quad \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j \rho(\mathbf{s}_i - \mathbf{s}_j) \geq 0,$$

para cualquier entero positivo k y cualquier conjunto de ubicaciones $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k \in \mathcal{S}$ y cualquier conjunto de números $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$. Similarmente el variograma cumple la condición de que es una función condicionalmente definida negativa, es decir,

$$\sum_{i=1}^k \sum_{j=1}^k \beta_i \beta_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0,$$

para cualquier entero positivo k y cualquier conjunto de ubicaciones $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k \in \mathcal{S}$ y cualquier conjunto de números $\beta_1, \beta_2, \dots, \beta_k \in \mathbb{R}$, que satisfacen $\sum_{i=1}^k \beta_i = 0$. Debido a estas condiciones en ocasiones es difícil tener funciones de covarianza y variograma válidas. Sin embargo, el Teorema de Bochner's (Bochner, 1955) proporciona una caracterización de funciones no negativas las cuales se han usado ampliamente para construir funciones de covarianza.

2.2. Procesos max-stable

Los procesos max-stable son una extensión espacial de las distribuciones max-stable que satisface (1.2). Una definición formal es como sigue.

Definición 2.8. (Proceso max-stable). Sea $Z(\mathbf{s}), \mathbf{s} \in \mathcal{S}$ un proceso aleatorio indexado por un subespacio compacto $\mathcal{S} \subset \mathbb{R}^d$. Se dice que el proceso $Z(\mathbf{s})$ es max-stable si para cada $k = 1, 2, \dots$, existen funciones continuas $a_k(\mathbf{s}) > 0$ y $b(\mathbf{s})$ tal que para cualquier función $z(\mathbf{s})$,

$$P\{Z(\mathbf{s}) \leq a_k(\mathbf{s})z(\mathbf{s}) + b(\mathbf{s})\}^k = P\{Z(\mathbf{s}) \leq z(\mathbf{s})\}, \quad (2.1)$$

o equivalentemente $Z(\mathbf{s})$ y el máximo de k copias independientes de $\{Z(\mathbf{s}) - b(\mathbf{s})\}/a_k(\mathbf{s})$ tienen la misma distribución.

Como analogía al caso de dimensiones finitas, los procesos max-stable surgen como la única clase de límites de máximos reescalados de procesos espaciales. Específicamente, considere un proceso estocástico de variables aleatorias i.i.d $\{Y(\mathbf{s}), \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d\}, i = 1, 2, \dots$, con trayectorias muestrales continuas sobre un conjunto compacto \mathcal{S} con una norma infinita, y suponga que existen sucesiones de funciones continuas $a_k(\mathbf{s}) > 0$ y $b(\mathbf{s})$ tal que cuando $n \rightarrow \infty$ el proceso de máximos reescalados

$$Z^n(\mathbf{s}) = \frac{\max\{Y_1(\mathbf{s}), \dots, Y_n(\mathbf{s})\} - b_n(\mathbf{s})}{a_n(\mathbf{s})}, \quad (2.2)$$

converge en distribución a un proceso aleatorio $Z(\mathbf{s}), \mathbf{s} \in \mathcal{S}$ cuyas marginales univariadas son no degeneradas. Entonces se puede mostrar que la clase de procesos

2.2. Procesos max-stable

límites coincide con la clase de procesos max-stable (de Haan y Pereira, 2006). Para cada $D = 1, 2, \dots$, y cualquier colección finita de sitios $\mathcal{D} = \{\mathbf{s}_1, \dots, \mathbf{s}_D\} \subset \mathcal{S}$, las correspondientes variables $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_D)$ tienen una distribución de valores extremos multivariada. En particular, la distribución marginal de $Z(\mathbf{s})$ es $\text{GEV}(\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}))$. Considerando la transformación

$$\tilde{Z}(\mathbf{s}) = \left[1 + \frac{\xi(\mathbf{s})}{\sigma(\mathbf{s})} \{Z(\mathbf{s}) - \mu(\mathbf{s})\} \right]_+^{-1/\xi(\mathbf{s})}, \quad (2.3)$$

se tiene que para cada sitio $\mathbf{s} \in \mathcal{S}$, $P(\tilde{Z}(\mathbf{s}) \leq z) = \exp(-1/z)$, $z > 0$, es decir, la distribución marginal de $\tilde{Z}(\mathbf{s})$ es Fréchet unitaria; tal proceso se le conoce como *max-stable*.

Definición 2.9. (Proceso max-stable simple). Un proceso aleatorio $Z(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$ es max-stable simple si es *max-stable* con distribuciones marginales Fréchet unitarias, o equivalentemente, si el proceso $Z(\mathbf{s})$ y $k^{-1} \max\{Z_1(\mathbf{s}), \dots, Z_k(\mathbf{s})\}$ tienen la misma distribución finito-dimensional para cualquier $k \in \mathbb{N}$, donde $Z_1(\mathbf{s}), \dots, Z_k(\mathbf{s})$ denotan copias i.i.d de $Z(\mathbf{s})$.

Como en el caso multivariado, la transformación (2.3) es capaz de modelar extremos espaciales en dos pasos: primero se transforman las marginales usando distribuciones GEV y se puede ajustar un modelo para dependencia espacial, aquí sin pérdida de generalidad, se puede restringir la discusión a un proceso aleatorio max-stable simple.

Cuando un proceso max-stable $Z(\mathbf{s})$ es simple, las constantes de renormalización en (2.2) son $a_n(\mathbf{s}) \equiv n$, $b_n(\mathbf{s}) \equiv 0$ y la distribución conjunta de $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_D)$ se puede escribir como,

$$P\{Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_D) \leq z_D\} = \exp\{-V_{\mathcal{D}}(z_1, \dots, z_D)\}, \quad z_1, \dots, z_D > 0, \quad (2.4)$$

donde $V_{\mathcal{D}}(\cdot)$ es la medida del exponente derivada en (1.3), la cual resume la estructura de dependencia extremal. La medida del exponente es homogénea de orden -1 , es decir, $V_{\mathcal{D}}(t\mathbf{z}) = t^{-1}V_{\mathcal{D}}(\mathbf{z})$ para cualquier $\mathbf{z} = (z_1, \dots, z_D) \subset \mathbb{R}_+^d$ y satisface $V_{\mathcal{D}}(\infty, \dots, z, \dots, \infty) = 1/z$ para cualquier permutación de los D argumentos. Al igual que en el caso multivariado en (1.5), el coeficiente extremal definido por $\mathcal{D} = (\mathbf{s}_1, \dots, \mathbf{s}_D)$ como

$$\theta_D(\mathbf{s}_1, \dots, \mathbf{s}_D) = V_{\mathcal{D}}(1, \dots, 1) \in [1, D] \quad (2.5)$$

se puede ver como aquel que resume la dependencia extremal, en el caso estacionario, se puede reescribir como

$$\theta_D(\mathbf{s}_1, \dots, \mathbf{s}_D) = \theta_D(\mathbf{h}), \quad (2.6)$$

donde $\mathbf{h} \in \mathbb{R}^{D-1}$ es un vector $(D-1)$ -dimensional. Cuando $\theta_D(\mathbf{h}) = 1$ las variables $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_D)$ son perfectamente dependientes y si $\theta_D(\mathbf{h}) = D$ son independientes. El siguiente teorema proporciona un representación usual de un proceso max-stable simple.

2.2. Procesos max-stable

Teorema 2.1. (de Haan y Ferreira, 2006). Sea $Z(\mathbf{s})$ un proceso max-stable simple con trayectorias muestrales continuas definido sobre un subespacio compacto $\mathcal{S} \subset \mathbb{R}^d$. Entonces existe un proceso estocástico continuo positivo i.i.d. W, W_1, W_2, \dots , con $E\{W(\mathbf{s})\} = 1$, para todo $\mathbf{s} \in \mathcal{S}$ y $E\{\sup_{\mathbf{s} \in \mathcal{S}} W(\mathbf{s})\} < \infty$, tal que

$$Z(\mathbf{s}) = \sup_{i \geq 1} W_i(\mathbf{s})/P_i, \quad (2.7)$$

donde las P_i 's son los puntos de un proceso de Poisson de intensidad unitaria sobre \mathbb{R}_+ . E inversamente cada proceso con esta representación es max-stable.

Un proceso construido de (2.7) también es un proceso max-stable cuando $W(\mathbf{s})$ es estacionario pero no necesariamente un campo aleatorio continuo, y cuando la condición $E\{\sup_{\mathbf{s} \in \mathcal{S}} W(\mathbf{s})\} < \infty$ se relaja por $E[\max\{0, W(\mathbf{s})\}] < \infty$ (ver Schlather (2002)). Una interpretación de la expresión (2.7) es pensar que $Z(\mathbf{s})$ sea máximos puntuales de tormentas aleatorias $W_i(\mathbf{s})$ con intensidad P_i^{-1} . Además, las propiedades de $W_i(\mathbf{s})$ normalmente las hereda el proceso max-stable resultante $Z(\mathbf{s})$; por ejemplo, si $W(\mathbf{s})$ es estacionario sobre \mathcal{S} , entonces $Z(\mathbf{s})$ también es estacionario.

Sea $Q(w_1, \dots, w_D)$ que denota la distribución del vector aleatorio $\{W(\mathbf{s}_1), \dots, W(\mathbf{s}_D)\}$, se sigue de (2.4) que la medida del exponente del proceso max-stable $Z(\mathbf{s})$ definida através de (2.7) es,

$$\begin{aligned} V_{\mathcal{D}}(z_1, \dots, z_D) &= -\log[P\{Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_D) \leq z_D\}] \\ &= -\log\left[P\left\{\sup_{i \geq 1} W_i(\mathbf{s}_1)/P_i \leq z_1, \dots, \sup_{i \geq 1} W_i(\mathbf{s}_D)/P_i \leq z_D\right\}\right] \\ &= -\log[P\{P_i \geq W_i(\mathbf{s}_1)/z_1, \dots, P_i \geq W_i(\mathbf{s}_D)/z_D, i \geq 1\}] \\ &= -\log\left[P\left\{P_i \geq \max_{j=1, \dots, D} W_i(\mathbf{s}_j)/z_j, i \geq 1\right\}\right] \\ &= \int \int_{p < \max_{j=1, \dots, D} \mathbf{s}_j/z_j} dp dQ(w_1, \dots, w_D) \\ &= E\left[\max_{j=1, \dots, D} \left\{\frac{W(\mathbf{s}_j)}{z_j}\right\}\right], \end{aligned}$$

donde la penúltima igualdad se justifica por el hecho de que los puntos P_i provienen de un proceso de Poisson con intensidad unitaria. Más general, si $\mathcal{D} \subset \mathcal{S}$, se tiene,

$$P\{Z(\mathbf{s}) \leq z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\} = \exp\left\{-E\left[\sup_{\mathbf{s} \in \mathcal{D}} \left(\frac{W(\mathbf{s}_j)}{z_j}\right)\right]\right\}, \quad (2.8)$$

donde $z(\mathbf{s})$ es cualquier función definida sobre \mathcal{D} . La expresión (2.8) se puede usar para calcular $V_{\mathcal{D}}(z_1, \dots, z_D)$ para ciertas elecciones de $W(\mathbf{s})$, pero sólo se puede obtener explícitamente para $D = 2$ (ver Huser y Davison (2013), Wadsworth y Tawn (2014)), por lo que la inferencia con la función de verosimilitud completa es intratable de manera general. Una alternativa diferente para la representación espectral de procesos aleatorios max-stable simple se formula en el siguiente teorema.

2.3. Modelos paramétricos estacionarios

Teorema 2.2. (*de Haan, 1984, de Haan y Ferreira, 2006*). Sea $\{P_i, U_i\}_{i \geq 1}$ realizaciones de un proceso de Poisson definido en $[0, 1] \times R_+$ con medida de intensidad $dp \times d\lambda$, donde λ denota la medida de Lebesgue. Si $Z(\mathbf{s})$ es un a proceso max-stable simple con trayectorias continuas, entonces existe una familia de funciones $f(\mathbf{s}, u)$, $\mathbf{s} \in \mathcal{S}$, $u \in [0, 1]$ con las siguientes propiedades:

- para cada $u \in [0, 1]$, se tiene una función continua no negativa $f(\mathbf{s}, u) : \mathcal{S} \rightarrow [0, \infty)$,
- para cada $\mathbf{s} \in \mathcal{S}$, $\int_0^1 f(\mathbf{s}, u) du = 1$,
- para $\mathcal{D} \subset \mathcal{S}$, $\int_0^1 \sup_{\mathbf{s} \in \mathcal{D}} f(\mathbf{s}, u) du < \infty$, y $f(\mathbf{s}, u)$ es tal que

$$Z(\mathbf{s}) = \sup_{i \geq 1} f(\mathbf{s}, u) / P_i. \quad (2.9)$$

E inversamente, cada proceso que tenga la forma del lado derecho de (2.9) y con las condiciones descritas, es un proceso max-stable simple con trayectorias continuas en \mathcal{S} .

La función $f(\mathbf{s}, \cdot) : [0, 1] \rightarrow [0, \infty)$, la cual no esta definida de manera única, se llama *función espectral* o *perfil de tormenta* del proceso max-stable simple, y en tal caso,

$$V_{\mathcal{D}}(z_1, \dots, z_D) = \mathbb{E} \left[\max_{j=1, \dots, D} \left\{ \frac{f(\mathbf{s}_j, U)}{z_j} \right\} \right], \quad (2.10)$$

donde $U \sim Unif(0, 1)$. La representación (2.9) sigue siendo válida si las funciones espectrales se definen sobre un conjunto $\mathcal{C} \subset \mathbb{R}^p$, en vez de sólo el intervalo $[0, 1]$.

2.3. Modelos paramétricos estacionarios

Las representaciones de procesos max-stable dadas en (2.7) y (2.9) implican que existe una infinidad de tales procesos, pero en la práctica el reto es construir modelos parsimoniosos y flexibles que sean capaces de capturar un amplio rango de dependencias extremas. La parsimonia es importante ya que a menudo los datos extremos son escasos, pero la flexibilidad también lo es porque un mal ajuste podría llevar a una mala estimación. Actualmente se han propuesto varios modelos paramétricos para procesos max-stable; por ejemplo, [Smith \(1990b\)](#) propuso un modelo max-stable con función espectral determinista, [Schlather \(2002\)](#) propuso uno basado en procesos gaussianos, el proceso llamado Brow-Resnick ([Brown y Resnick, 1977](#), [Kabluchko y Schlather, 2010](#)) se construye de un proceso log-normal. Otros modelos incluyen el proceso t -extremal llamado proceso max-max-stable ([Davison et al., 2012](#), [Opitz, 2013](#), [Ribatet y Sedki, 2013](#)).

2.3.1. El modelo de Smith

[Smith \(1990b\)](#) propone tomar una función espectral del tipo $f(\mathbf{s}, \mathbf{u}) = g(\mathbf{s} - \mathbf{u})$ en (2.9) donde $g(\cdot)$ es una función de densidad de probabilidad definida sobre \mathbb{R}^d , por ejemplo si,

$$f(\mathbf{s}, \mathbf{u}) = \phi_d(\mathbf{s} - \mathbf{u}; \Sigma), \quad \mathbf{s}, \mathbf{u} \in \mathcal{S}, \quad (2.11)$$

2.3. Modelos paramétricos estacionarios

donde $\phi_d(\cdot; \Sigma)$ denota la densidad gaussiana d -variada con media cero y matriz de covarianza Σ , la medida del exponente para este proceso max-stable estacionario se puede calcular para $D = 2$ usando (2.10), y es,

$$V_{\mathcal{D}}(z_1, z_2) = \frac{1}{z_1} \Phi \left(\frac{a(\mathbf{h})}{2} - \frac{1}{a(\mathbf{h})} \log \frac{z_1}{z_2} \right) + \frac{1}{z_2} \Phi \left(\frac{a(\mathbf{h})}{2} - \frac{1}{a(\mathbf{h})} \log \frac{z_2}{z_1} \right), \quad (2.12)$$

donde Φ es la función de distribución normal estándar, $a^2(\mathbf{h}) = \mathbf{h}\Sigma^{-1}\mathbf{h}$ y $\mathcal{D} = \{\mathbf{s}, \mathbf{s} + \mathbf{h}\} \subset \mathcal{S}$. En este caso el coeficiente extremal bivariado es $\theta_2(\mathbf{h}) = 2\Phi(a(\mathbf{h})/2)$.

La matriz de covarianza Σ controla el rango de dependencia y el grado de anisotropía del campo aleatorio realizado. Si $\Sigma = \text{diag}(\lambda^2, \dots, \lambda^2)$, los contornos son esféricos (esto es, el proceso es isotrópico) y λ determina la extensión espacial, en general los contornos son elípticos. La Figura 2.1 muestra simulaciones de este proceso para diferentes valores de los parámetros en \mathbb{R}^2 . Debido a que el modelo tiene formas deterministas, no es muy flexible y es muy suave para ser usado en aplicaciones ambientales, a pesar de ello se considera estándar ya que fue el primero que se ajustó a datos reales (Padoan *et al.*, 2010).

2.3.2. El modelo de Schlather

Un modelo más flexible y más conveniente para modelar fenómenos ambientales, fue propuesto por Schlather (2002) y consiste en tomar un proceso estocástico de la forma,

$$W(\mathbf{s}) = \sqrt{2\pi} \max\{\varepsilon(\mathbf{s}), 0\}, \quad \mathbf{s} \in \mathcal{S}, \quad (2.13)$$

donde $\varepsilon(\mathbf{s}) \sim \text{GP}\{0, \rho(\mathbf{h})\}$, $W(\mathbf{s})$ es la parte positiva de un campo aleatorio gaussiano con media cero, varianza unitaria y función de correlación ρ , y se reescala para que cumpla que $E\{W(\mathbf{s})\} = 1$. El proceso estacionario max-stable que resulta de (2.7) tiene medida del exponente bivariada dada por,

$$V_{\mathcal{D}}(z_1, z_2) = \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - 2(\rho(\mathbf{h}) + 1) \frac{z_1 z_2}{(z_1 + z_2)^2}} \right), \quad (2.14)$$

donde $\mathcal{D} = \{\mathbf{s}, \mathbf{s} + \mathbf{h}\} \subset \mathcal{S}$. Al igual que en los procesos gaussianos, la suavidad del proceso max-stable construido a partir de (2.13) se puede controlar por la elección de la función de correlación $\rho(\mathbf{h})$ (ver realizaciones de este proceso en la Figura 2.2), el coeficiente extremal bivariado para este modelo es $\theta_2(\mathbf{h}) = 1 + \sqrt{\{1 - \rho(\mathbf{h})\}/2}$, aquí cuando $\|\mathbf{h}\| \rightarrow \infty$, $\theta_2(\mathbf{h})$ esta acotado por 1.707, lo que significa que la independencia completa no puede ser capturada por este modelo. Para resolver esto, Schlather (2002) propone introducir un conjunto aleatorio que garantice que ubicaciones que son suficientemente distantes no puedan ser cubiertos por la misma función aleatoria $W_i(\mathbf{s})$ en (2.7) y así produzca independencia exacta entre máximos de tales sitios. Específicamente,

$$W(\mathbf{s}) = \sqrt{2\pi} \frac{|\mathcal{S}|}{E(|\mathcal{A}|)} \max\{\varepsilon(\mathbf{s}), 0\} I_{\mathcal{A}}(\mathbf{s} - X), \quad \mathbf{s} \in \mathcal{S}, \quad (2.15)$$

2.3. Modelos paramétricos estacionarios

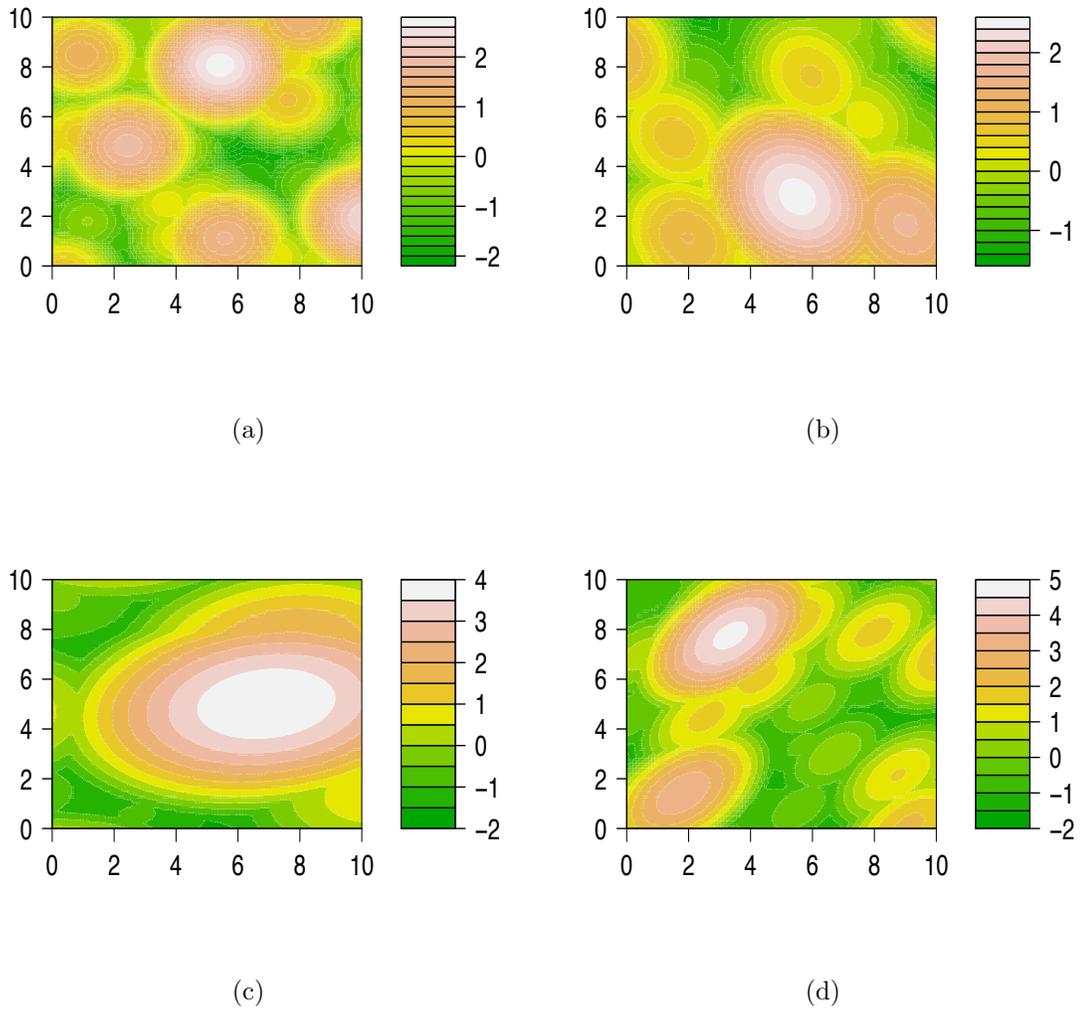


Figura 2.1: Realizaciones del modelo de Smith en \mathbb{R}^2 para diferentes matrices de covarianzas $\Sigma = (\Sigma_{ij})$, $i, j = 1, 2$. (a) $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0$; (b) $\Sigma_{11} = 2$, $\Sigma_{22} = 3$, $\Sigma_{12} = \Sigma_{21} = -0.5$; (c) $\Sigma_{11} = 5$, $\Sigma_{22} = 2$, $\Sigma_{12} = \Sigma_{21} = 0.5$; (d) $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0.5$.

2.4. Medida de dependencia extremal

donde $\mathcal{A} \subset \mathcal{S}$ es un conjunto compacto, $\varepsilon(\mathbf{s})$ se define como antes, $I_{\mathcal{A}}(\cdot)$ es la función indicadora sobre \mathcal{A} , X es un punto de un proceso de Poisson de intensidad unitaria sobre \mathcal{S} y $|\cdot|$ denota el volumen del conjunto. El proceso max-stable resultante de (2.7) es estacionario y construido de cada conjunto aleatorio con un proceso gaussiano. Aquí, la dependencia de corto alcance esta determinada por la función de correlación $\rho(\mathbf{h})$, mientras que la dependencia de largo alcance esta regulada por la geometría del conjunto aleatorio \mathcal{A} . La medida del exponente se puede escribir como,

$$V_D(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 - \frac{\delta(\mathbf{h})}{2} \left[1 - \sqrt{1 - 2(\rho(\mathbf{h}) + 1) \frac{z_1 z_2}{(z_1 + z_2)^2}} \right] \right), \quad (2.16)$$

donde $\delta(\mathbf{h}) = E\{|\mathcal{A} \cap (\mathbf{h} + \mathcal{A})|\} / E(|\mathcal{A}|)$ pertenece al intervalo unitario, este modelo es una extensión del modelo de Schlather sin conjunto aleatorio (2.14), $\delta(\mathbf{h}) \equiv 1$. El coeficiente extremal se puede expresar como $\theta_2(\mathbf{h}) = 2 - \delta(\mathbf{h}) \left[1 - \sqrt{\{1 - \rho(\mathbf{h})\}/2} \right]$, donde el conjunto \mathcal{A} se elige de tal forma que sea compacto, se puede elegir a \mathcal{A} de tal forma que $\delta(\mathbf{h}) \rightarrow 0$ y así $\theta_2(\mathbf{h}) \rightarrow 2$ cuando $\|\mathbf{h}\| \rightarrow \infty$ para cualquier función de correlación $\rho(\mathbf{h})$, de aquí que este modelo forme una mezcla para la elección adecuada de \mathcal{A} .

Un inconveniente de este modelo es que las realizaciones no son continuas. Sin embargo, se puede adaptar para modelar procesos con efectos muy locales y con cambios bruscos o fenómenos que reflejen independencia completa después de algún retraso. El modelo de Schlather con conjunto aleatorio se ajusto por primera vez a datos de temperatura máxima anual en Suiza por Davison y Gholamrezaee (2011) y una generalización al espacio-tiempo usando (2.15) la presenta Huser y Davison (2014). Usualmente $\rho(\mathbf{h})$ se elige de una forma cerrada de una familia paramétrica válida (Ver Tabla 2.1) en este caso c_2 y ν son el rango y el parámetro de suavizamiento de la función de correlación, $\Gamma(\cdot)$ es la función gamma, $J_\nu(\cdot)$ y $K_\nu(\cdot)$ son las funciones de Bessel y de Bessel modificadas de tercer tipo de orden ν y d es la dimensión del campo aleatorio. En este caso es posible añadir un efecto umbral c_1 y un efecto pepita α a estas funciones de correlaciones, de la siguiente manera,

$$\rho_*(\mathbf{h}) = \begin{cases} \nu + c_1, & \mathbf{h} = 0; \\ c_1 \rho(\mathbf{h}), & \mathbf{h} > 0, \end{cases}$$

aquí, c_1 y α satisfacen $\alpha = 1 - c_1$, Schlather (2002).

2.4. Medida de dependencia extremal

Actualmente, se han propuesto varias medidas de dependencia extremal para procesos estacionarios, muchas de ellas son una extensión del caso multivariado. El coeficiente extremal $\theta_D(\mathbf{h})$, introducido en (2.6), es adecuado para procesos asintóticamente dependientes cuyos máximos renormalizados convergen a un proceso max-stable. Un caso especial es considerar coeficientes extremales a pares esto es

$$\Pr [Z(\mathbf{s}_1) \leq z, Z(\mathbf{s}_2) \leq z] = \exp \left\{ -\frac{\theta(\mathbf{s}_1 - \mathbf{s}_2)}{z} \right\}. \quad (2.17)$$

2.4. Medida de dependencia extremal

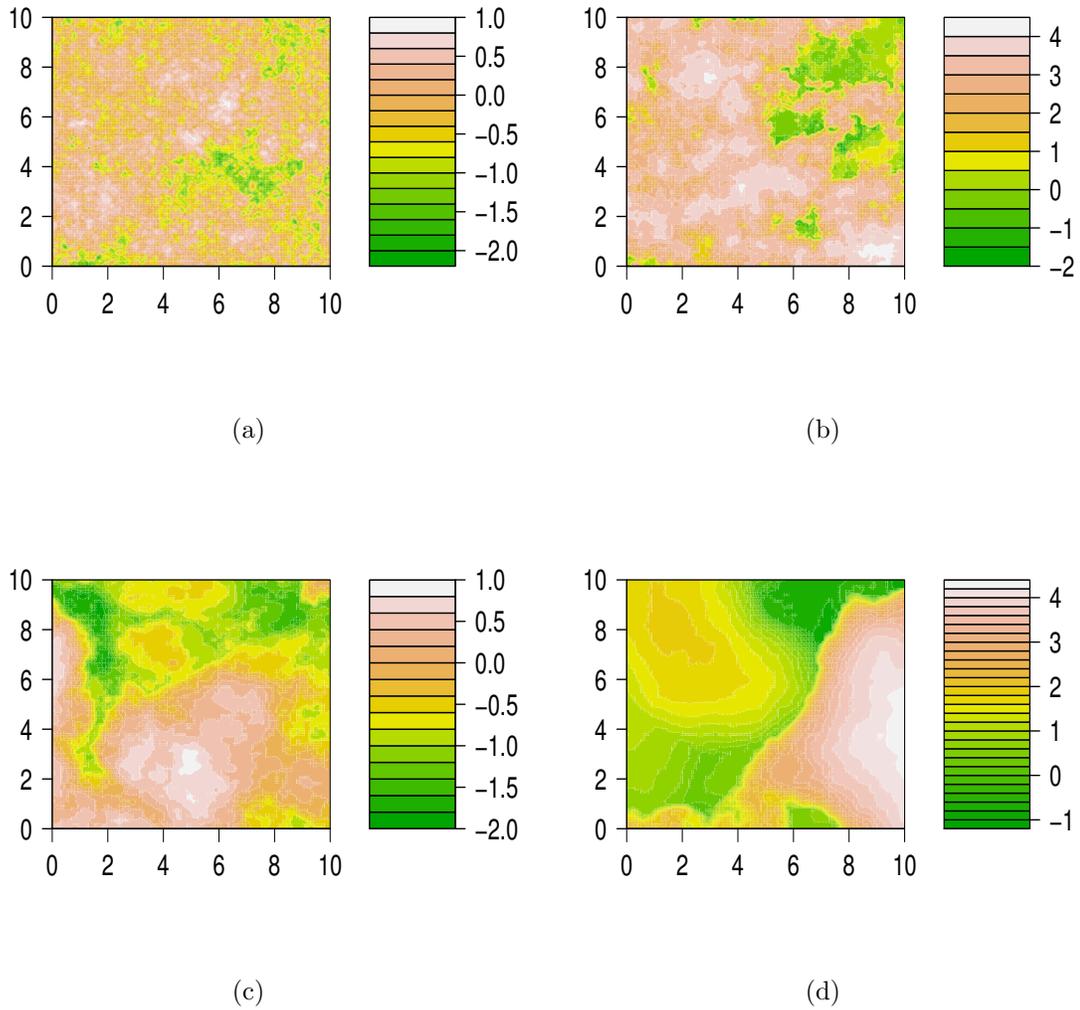


Figura 2.2: Realizaciones del modelo de Schlather con marginales Fréchet unitarias en \mathbb{R}^2 usando la función de correlación exponencial $\rho(\mathbf{h}) = \exp \left[- \left(\frac{\mathbf{h}}{c_2} \right)^\nu \right]$, para diferentes valores de los parámetros. (a) $c_2 = 2, \nu = 0.5$; (b) $c_2 = 3, \nu = 1$; (c) $c_2 = 3, \nu = 1.5$; (d) $c_2 = 4, \nu = 1.9$.

2.4. Medida de dependencia extremal

Tabla 2.1: Algunas funciones de correlación.

Familia	Función de correlación	Rango de los parámetros
Whittle–Matérn	$\rho(\mathbf{h}) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\mathbf{h}}{c_2}\right)^\nu K_\nu\left(\frac{\mathbf{h}}{c_2}\right),$	$c_2 > 0, \quad \nu > 0$
Cauchy	$\rho(\mathbf{h}) = \left[1 + \left(\frac{\mathbf{h}}{c_2}\right)^2\right]^{-\nu},$	$c_2 > 0, \quad \nu > 0$
Powered Exponential	$\rho(\mathbf{h}) = \exp\left[-\left(\frac{\mathbf{h}}{c_2}\right)^\nu\right],$	$c_2 > 0, \quad 0 < \nu \leq 2$
Bessel	$\rho(\mathbf{h}) = \left(\frac{2c_2}{\mathbf{h}}\right)^\nu \Gamma(\nu + 1) J_\nu\left(\frac{\mathbf{h}}{c_2}\right),$	$c_2 > 0, \quad \nu \geq \frac{d-2}{2}$

La función $\theta(\cdot)$ para los modelos descritos anteriormente se puede directamente de la función de distribución bivariada tomando $z_1 = z_2 = z$, es decir,

$$\theta(\mathbf{s}_1 - \mathbf{s}_2) = 2\Phi\left(\frac{\sqrt{(\mathbf{s}_1 - \mathbf{s}_2)^T \Sigma^{-1} (\mathbf{s}_1 - \mathbf{s}_2)}}{2}\right),$$

para el modelo de Smith, mientras que para el modelo de Schlather es,

$$\theta(\mathbf{s}_1 - \mathbf{s}_2) = 1 + \sqrt{\frac{1 - \rho(\mathbf{s}_1 - \mathbf{s}_2)}{2}}.$$

Schlather y Tawn (2003) prueban las siguientes propiedades para $\theta(\cdot)$:

- $2 - \theta(\cdot)$ es una función positiva semidefinida;
- $\theta(\cdot)$ no es diferenciable en cero, a menos que $\theta = 1$;
- si $d \geq 2$ y Z es isotrópico, $\theta(\cdot)$ tiene al menos un salto en cero y es continua donde quiera.

Estas propiedades tienen las siguientes consecuencias, la estructura de dependencia espacial de un proceso max-stable estacionario se puede caracterizar por la función de correlación y que una función de correlación válida no siempre conduce a una función extremal válida.

Un estimador propuesto por Smith (1990b) para el coeficiente extremal a pares es como sigue, como $Z(\mathbf{s})$ tiene distribuciones Fréchet unitaria para todo $\mathbf{s} \in \mathbb{R}^d$, entonces $1/Z(\mathbf{s})$ tiene distribución exponencial unitaria y, de acuerdo con la ecuación (2.17), $1/\max\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\}$ tiene distribución exponencial con parámetro $\theta(\mathbf{s}_1 - \mathbf{s}_2)$. Por la ley de los grandes números $\sum_{i=1}^n 1/Z_i(\mathbf{s}_1) = \sum_{i=1}^n 1/Z_i(\mathbf{s}_2) \approx n^1$, esto sugiere que un estimador para el coeficiente extremal para las ubicaciones \mathbf{s}_1 y \mathbf{s}_2 es,

$$\hat{\theta}(\mathbf{s}_1 - \mathbf{s}_2) = \frac{n}{\sum_{i=1}^n \min\{Z_i(\mathbf{s}_1)^{-1}, Z_i(\mathbf{s}_2)^{-1}\}}. \quad (2.18)$$

¹De hecho esta relación es exacta si las marginales fueron transformadas a Fréchet unitarias usando los estimadores de máxima verosimilitud.

2.5. Inferencia en procesos max-stable

En ocasiones no es posible evaluar u obtener analíticamente la función de verosimilitud completa, como lo es el caso de los procesos max-stable, la verosimilitud compuesta puede ser conveniente para realizar inferencia desde el enfoque frecuentista o bayesiano, por ejemplo [Varin et al. \(2011\)](#) presenta una revisión general de verosimilitud compuesta. En el contexto de extremos espaciales, usualmente se puede obtener la función de densidad bivariada (ver [sección 2.3](#)), así que el proceso de inferencia se puede basar en la verosimilitud a pares ([Blanchet y Davison, 2011](#), [Davison et al., 2012](#), [Padoan et al., 2010](#)). Estas pseudo verosimilitudes son robustas contra falta de especificación de las distribuciones marginales de orden superior y tienen buenas propiedades teóricas.

2.5.1. Función de verosimilitud compuesta

Considere un vector aleatorio \mathbf{Y} N -dimensional con función de densidad $g(\mathbf{y}; \boldsymbol{\psi})$, para algún vector de parámetros g -dimensional $\boldsymbol{\psi} \in \Psi$. Si se asume un conjunto de eventos marginales o condicionales $\{\mathcal{J}_1, \dots, \mathcal{J}_K\}$ con densidad asociada $g(\mathbf{y} \in \mathcal{J}_k; \boldsymbol{\psi})$, entonces la función de verosimilitud compuesta para n réplicas independientes de \mathbf{Y} se define como,

$$L_C(\boldsymbol{\psi}) = \prod_{i=1}^n \prod_{k=1}^K g(\mathbf{y}_i \in \mathcal{J}_k; \boldsymbol{\psi})^{w_k}, \quad (2.19)$$

donde $w_k > 0$, $k = 1, \dots, K$ son pesos adecuados ([Padoan et al., 2010](#), [Varin et al., 2011](#)). La función de log-verosimilitud compuesta es,

$$l_C(\boldsymbol{\psi}) = \sum_{i=1}^n \sum_{k=1}^K w_k \log g(\mathbf{y}_i \in \mathcal{J}_k; \boldsymbol{\psi}). \quad (2.20)$$

Si todos los pesos son iguales se pueden omitir, pero elecciones adecuadas pueden mejorar la eficiencia estadística. El estimador de máxima verosimilitud compuesta $\hat{\boldsymbol{\psi}}_C$, se obtiene maximizando (2.19) o equivalentemente (2.20), esto es,

$$\hat{\boldsymbol{\psi}}_C = \arg \max_{\boldsymbol{\psi} \in \Psi} L_C(\boldsymbol{\psi}) = \arg \max_{\boldsymbol{\psi} \in \Psi} l_C(\boldsymbol{\psi}). \quad (2.21)$$

La función score compuesta $U_C(\boldsymbol{\psi})$ es el vector de derivadas parciales de primer orden de $l_C(\boldsymbol{\psi})$ con respecto a las componentes de $\boldsymbol{\psi}$, mientras que la matriz derivadas parciales de segundo orden proporciona la matriz Hessiana $H_C(\boldsymbol{\psi})$, esto es,

$$U_C(\boldsymbol{\psi}) = \nabla_{\boldsymbol{\psi}} l_C(\boldsymbol{\psi}), \quad H_C(\boldsymbol{\psi}) = -\nabla_{\boldsymbol{\psi}} U_C(\boldsymbol{\psi}) \quad (2.22)$$

La ecuación score $U_C(\boldsymbol{\psi}) = 0$ se puede ver como una ecuación de estimación insesgada para $\hat{\boldsymbol{\psi}}_C$. Además se puede definir las matrices de variabilidad $K_C(\boldsymbol{\psi})$ y sensibilidad $J_C(\boldsymbol{\psi})$ como,

$$K_C(\boldsymbol{\psi}) = \text{var}\{U_C(\boldsymbol{\psi})\}, \quad J_C(\boldsymbol{\psi}) = \text{E}\{H_C(\boldsymbol{\psi})\}. \quad (2.23)$$

2.5. Inferencia en procesos max-stable

También se definen las matrices $k_C(\boldsymbol{\psi})$ y $j_C(\boldsymbol{\psi})$ en (2.23) normalizado por el número de réplicas,

$$k_C(\boldsymbol{\psi}) = \frac{K_C(\boldsymbol{\psi})}{n}, \quad j_C(\boldsymbol{\psi}) = \frac{J_C(\boldsymbol{\psi})}{n}. \quad (2.24)$$

Muchos resultados que involucran falta de especificación en la función de verosimilitud muestran que, bajo ciertas condiciones de regularidad y cuando $\boldsymbol{\psi}$ es identificable, el estimador de máxima verosimilitud compuesta, $\hat{\boldsymbol{\psi}}_C$, es fuertemente consistente y asintóticamente normal, (Cox y Reid, 2004, Davison, 2003, Padoan *et al.*, 2010, Varin *et al.*, 2011) esto es,

$$\sqrt{n}(\hat{\boldsymbol{\psi}}_C - \boldsymbol{\psi}_0) \xrightarrow{D} N_q\{0, v_C(\boldsymbol{\psi}_0)\},$$

donde \xrightarrow{D} significa convergencia en distribución, y donde la varianza asintótica tiene la forma $v_C(\boldsymbol{\psi}_0) = j_C(\boldsymbol{\psi}_0)^{-1}k_C(\boldsymbol{\psi}_0)j_C(\boldsymbol{\psi}_0)^{-1}$ y $\boldsymbol{\psi}_0$ es el verdadero valor del parámetro. De aquí la la distribución asintótica de $\hat{\boldsymbol{\psi}}_C$ se puede aproximar por,

$$\hat{\boldsymbol{\psi}}_C \sim N_q\{\boldsymbol{\psi}_0, V_C(\hat{\boldsymbol{\psi}}_C)\},$$

cuando $n \rightarrow \infty$, donde

$$V_C(\boldsymbol{\psi}) = nv_C(\boldsymbol{\psi}) = J_C(\boldsymbol{\psi})^{-1}K_C(\boldsymbol{\psi})J_C(\boldsymbol{\psi})^{-1}.$$

La inversa de la matriz de varianza asintótica $I_C(\boldsymbol{\psi}) = V_C(\boldsymbol{\psi})^{-1}$, se llama la matriz de información de Godambe y es la equivalente a la matriz de información de Fisher, para verosimilitud compuesta. Cuando la verosimilitud compuesta es exactamente la verosimilitud completa, la identidad de Bartlett resulta $I_C(\boldsymbol{\psi}) = K_C(\boldsymbol{\psi}) = J_C(\boldsymbol{\psi})$, lo que significa que la matriz de información esperada de la verosimilitud compuesta es igual a la matriz de información de Fisher.

2.5.2. Verosimilitud a pares para extremos espaciales

Dentro del enfoque espacial, se tienen $\{y_{m,k}\}$ observaciones, que son los máximos de s muestras sobre $m = 1, \dots, M$ bloques y para $k = 1, \dots, K$ ubicaciones en una región continua. Por ejemplo, para $s = 7$ implica que $\{y_{m,k}\}$ describe el máximo semanal. Se supone que $\{\mathbf{y}_{m,k}\}_{m=1}^M$ son M realizaciones independientes e idénticamente distribuidas de un proceso Gaussiano con marginales GEV denotadas por Y_1, \dots, Y_K . A pesar de no tener la distribución multivariada del proceso max-stable, la disponibilidad de la distribución bivariada implica la definición de la función de log-verosimilitud compuesta por pares como la suma de log-verosimilitudes que corresponden a cada contribución bivariada. Específicamente de (2.20) para cada m de las M réplicas independientes, definiendo los eventos $\mathcal{J}_k = \{y_{m,i}, y_{m,j}\}$ como el conjunto de subvectores bivariados de \mathbf{y} que se toman sobre todas las $K(K-1)/2$ pares de ubicaciones distintas i y j , la función de log-verosimilitud compuesta a pares o simplemente función de verosimilitud a pares, esta dada por,

$$l_p(\boldsymbol{\psi}) = \sum_{m=1}^M \sum_{i=1}^{K-1} \sum_{j=i+1}^K w_{ij} \log g(y_{m,i}, y_{m,j}; \boldsymbol{\psi}), \quad (2.25)$$

2.5. Inferencia en procesos max-stable

donde w_{ij} son los pesos y se toman tal que $w_{ij} > 0$ y $\sum_{ij} w_{ij} = 1$, con esto, se induce a la función de log-verosimilitud a pares construida para todos los pares de observaciones. A partir de aquí el proceso de estimación es muy similar al de verosimilitud compuesta, es decir, el estimador de máxima verosimilitud a pares $\hat{\psi}_p$ se obtiene maximizando (2.25) esto es,

$$\hat{\psi}_p = \arg \max_{\psi \in \Psi} l_p(\psi). \quad (2.26)$$

Las expresiones para las ecuaciones (2.22)–(2.24) se obtienen reemplazando $l_C(\psi)$ por $l_p(\psi)$. También por las propiedades de la función de verosimilitud compuesta, la distribución asintótica de $\hat{\psi}_p$ se puede aproximar por,

$$\hat{\psi}_p \sim N_q\{\psi_0, V_p(\hat{\psi}_p)\},$$

cuando $n \rightarrow \infty$, donde

$$V_p(\psi) = nv_p(\psi) = J_p(\psi)^{-1} K_p(\psi) J_p(\psi)^{-1}.$$

2.5.3. Comparación de modelos

Para comparar el desempeño de modelos anidados, se puede ejecutar la prueba de razón de verosimilitud compuesta. Suponga que el vector de parámetros ψ se particiona en $(\psi^1, \psi^2) \in \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}$, con $q_1 + q_2 = q$ y que se desea probar si la hipótesis nula $\psi^1 = \psi^1_*$ es cierta. En este sentido el parámetro $\psi^1 \in \mathbb{R}^{q_1}$ es el parámetro de interés mientras que $\psi^2 \in \mathbb{R}^{q_2}$ actúa como parámetro de ruido. Sea $\hat{\psi}_C = (\hat{\psi}^1, \hat{\psi}^2)$ que denota el estimador de máxima verosimilitud compuesta sin restricciones, y sea $\hat{\psi}_{C;*} = (\hat{\psi}^1_*, \hat{\psi}^2_{C;*})$ que denota el estimador de máxima verosimilitud bajo la hipótesis nula, es decir, $\hat{\psi}^2_{C;*}$ es el estimador de máxima verosimilitud compuesta de ψ^2 , cuando ψ^1 se mantiene fijo al valor ψ^1_* . la prueba de razón de verosimilitud compuesta se puede basar en el estadístico

$$W(\psi^1_*) = 2 \left[l_C(\hat{\psi}_C) - l_C(\hat{\psi}_{C;*}) \right].$$

Bajo la hipótesis nula se puede mostrar que la distribución de $W(\psi^1_*)$ se puede aproximar por $\sum_{j=1}^{q_1} c_j Z_j$ para n grande, donde las Z_j 's son variables aleatorias χ^2_1 independientes, y las c_j 's son los eigenvalores de la matriz $\left\{ \tilde{J}_C^1(\hat{\psi}_C) \right\}^{-1} I_C^1(\hat{\psi}_C)$, donde $\tilde{J}_C^1(\hat{\psi}_C)$ y $I_C^1(\hat{\psi}_C)$ son submatrices de tamaño $q_1 \times q_1$ de $J_C(\psi)^{-1}$ e $I_C(\psi)$, respectivamente, con elementos correspondientes a ψ^1 . Para comparar modelos no anidados, el criterio de información de la verosimilitud compuesta (CLIC), el análogo al criterio de información de Akaike (AIC), puede ser útil (Davison *et al.*, 2013, Padoan *et al.*, 2010). Se define como,

$$\text{CLIC} = -2 \left[l_C(\hat{\psi}_C) - \text{tr} \left\{ K_C(\hat{\psi}_C) J_C^{-1}(\hat{\psi}_C) \right\} \right], \quad (2.27)$$

donde $\text{tr}(\cdot)$ denota al operador traza, y la selección del modelo se basa en minimizar (2.27). Davison y Gholamrezaee (2011) sugieren usar una variante, el CLIC*, el cual se escala para que sea comparado con el AIC para datos independientes. Por ejemplo cuando se usa verosimilitud a pares, $\text{CLIC}^* = \text{CLIC}/(N - 1)$.

2.6. Aplicación

2.5.4. Estimación de la varianza asintótica

La varianza asintótica de los estimadores de máxima verosimilitud compuesta se pueden aproximar por

$$V_C(\boldsymbol{\psi}) \approx V_C(\hat{\boldsymbol{\psi}}_C) = J_C(\hat{\boldsymbol{\psi}}_C)^{-1} K_C(\hat{\boldsymbol{\psi}}_C) J_C(\hat{\boldsymbol{\psi}}_C)^{-1}.$$

La matriz $J_C(\hat{\boldsymbol{\psi}}_C)$, se puede aproximar por diferencias finitas, usando la matriz hessiana que se obtiene por la rutina de optimización. Más precisamente, $J_C(\hat{\boldsymbol{\psi}}_C)$ se puede estimar por la matriz $H_C(\hat{\boldsymbol{\psi}}_C)$ donde se calcula usando aproximaciones numéricas. La estimación de la matriz $K_C(\boldsymbol{\psi})$ es un poco más difícil de estimar, como $K_C(\boldsymbol{\psi}) = \text{var}\{U_C(\boldsymbol{\psi})\} = E\{U_C(\boldsymbol{\psi})U_C(\boldsymbol{\psi})^T\}$ una aproximación de $K_C(\hat{\boldsymbol{\psi}}_C)$ podría ser por $U_C(\hat{\boldsymbol{\psi}}_C)U_C(\hat{\boldsymbol{\psi}}_C)^T$, pero como por definición $U_C(\hat{\boldsymbol{\psi}}_C) = 0$, la matriz estimada contendrá puros ceros, para corregir esto, se han propuesto varias aproximaciones, asumiendo independencia en las observaciones, una primera aproximación (Varin *et al.*, 2011) es calcular la contribución de n scores, multiplicada por n , esto es,

$$K_C(\boldsymbol{\psi}) \approx n \widehat{\text{var}}_{i=1, \dots, n} \left\{ \sum_{k=1}^K w_k \nabla_{\boldsymbol{\psi}} \log g(y_i; \boldsymbol{\psi}) \right\}, \quad (2.28)$$

y evaluar $\hat{\boldsymbol{\psi}}_C$ en (2.28). Un inconveniente de esta expresión es que involucra todas las derivadas parciales del logaritmo de la densidad, cálculos que en la práctica pueden ser muy tediosos. También (2.28) puede tener imprecisiones cuando n no es suficientemente grande comparado con la dimensión de $\boldsymbol{\psi}$. Por lo que una ligera mala estimación de $K_C(\boldsymbol{\psi})$ puede tener repercusiones sobre $V_C(\boldsymbol{\psi})$ y sobre la selección del modelo usando el CLIC, de aquí que se deban tener estimaciones exactas. Si las fuentes computacionales lo permiten, se pueden usar otras estimaciones para $V_C(\boldsymbol{\psi})$, como por ejemplo métodos basados en submuestreo bootstrap (Thibaud *et al.*, 2013, Varin *et al.*, 2011).

2.6. Aplicación

2.6.1. Datos

En esta sección se presenta una aplicación espacial de la teoría presentada en este Capítulo, usando datos que consiste en mediciones diarias de los niveles de monóxido de carbono registrados de manera horaria en la Ciudad de México y su zona metropolitana. Los datos fueron obtenidos del Sistema Metereológico del Medio Ambiente de la Ciudad de México los cuales consisten en mediciones diarias de los niveles de éste contaminante registrados en 9 estaciones meteorológicas, desde el año de 2008 hasta 2013, disponibles en <http://www.aire.df.gob.mx>. En la Tabla 2.2 se presenta la lista de estaciones, su localización geográfica y su altitud, y en la Figura 2.3 se muestra la zona de estudio.

2.6. Aplicación

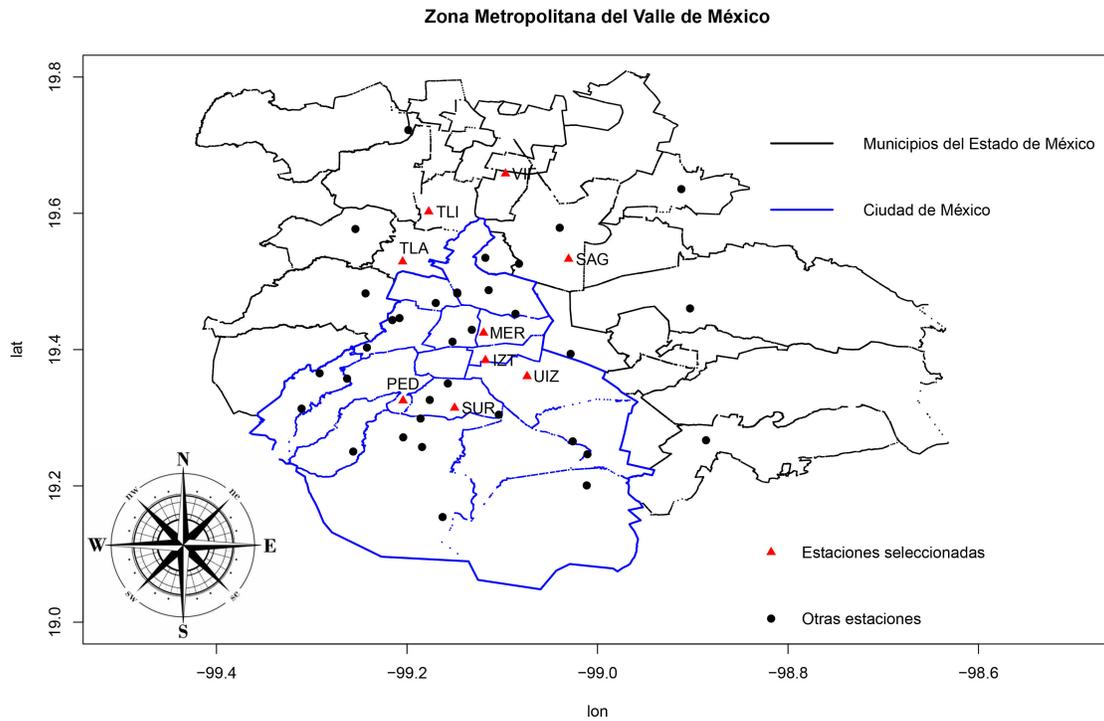


Figura 2.3: Región de estudio.

2.6.2. Metodología

Se tomó el máximo en cada bloque formado por observaciones consecutivas de tres días, para garantizar la independencia de éstos [Coles et al. \(2001\)](#), es decir, en cada año se tiene 121 datos. Se realiza un análisis exploratorio para cada estación en cada año y se ajusta la distribución GEV, en éste análisis no se encontraron violaciones a los supuestos distribucionales, después se transforman los datos para que tengan distribución Fréchet unitaria en cada ubicación de la siguiente forma, si $Y \sim \text{GEV}(\mu, \sigma, \xi)$ entonces la variable aleatoria definida como $Z = (1 + \xi \frac{Y - \mu}{\sigma})^{1/\xi}$ sigue una distribución Fréchet unitaria (ver [\(2.3\)](#)). Posteriormente se ajustaron los modelos max-stable de Smith y de Schalter en cada año, obteniendo las estimaciones de los parámetros en cada modelo. Para realizar los mapas predictivos se propone una tendencia espacial lineal en los parámetros de la distribución GEV y éstos se realizan usando los datos del año 2012. El análisis de los datos se llevó a cabo en el software estadístico R en particular en el paquete [SpatialExtremes](#) ([R Development Core Team, 2015](#), [Ribatet, 2015](#)).

2.6.3. Resultados

Para el modelo de Smith, los parámetros de dependencia espacial estimados se muestran en la [Tabla 2.3](#). El modelo de Schalter, se ajustó usando la función de correlación Whittle - Matérn (WM) y exponencial (Exp). Los resultados muestran que el modelo

2.6. Aplicación

Tabla 2.2: Estaciones metereológicas bajo estudio.

Clave	Estación	Longitud	Latitud	Altitud
IZT	Iztacalco	-99.11764	19.38441	2238
MER	Merced	-99.11959	19.42461	2245
PED	Pedregal	-99.20414	19.32515	2326
SAG	San Agustin	-99.03032	19.53297	2241
SUR	Santa Ursula	-99.14999	19.31448	2279
TLA	Tlalnepantla	-99.20460	19.52908	2311
TLI	Tultitlan	-99.17717	19.60254	2313
UIZ	UAM Iztapalapa	-99.07388	19.36079	2221
VIF	Villa de las Flores	-99.09659	19.65822	2242

Tabla 2.3: Parámetros de dependencia espacial estimados para el modelo de Smith, su error estándar esta entre paréntesis.

Año	$\hat{\sigma}_{11}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$
2008	0.0003051(9.96e-5)	-0.0015313(4.49e-4)	0.0131039(2.24e-3)
2009	0.002302(0.0001546)	-0.002188(0.0002323)	0.005717(0.0003743)
2010	0.002967(0.0006931)	-0.010289(0.0023580)	0.036004(0.0080169)
2011	0.0009467(4.89e-5)	0.0010138(7.32e-5)	0.0104137(3.07e-4)
2012	0.001396(0.0003106)	0.007702(0.0017259)	0.044424(0.0096874)
2013	0.001195(0.0001059)	-0.004073(0.0003724)	0.022085(0.0016569)

con el menor CLIC* fue cuando se utilizó la función de correlación exponencial, en la [Tabla 2.4](#) se muestran estos resultados. Los parámetros de dependencia espacial para el

Tabla 2.4: Criterio de información de verosimilitud compuesta.

	2008	2009	2010	2011	2012	2013
CLIC* WM	37049.25	36668.09	36914.81	37256.32	36933.39	36812.45
CLIC* Exp	37048.90	36667.56	36911.75	37233.72	36932.95	36803.19

modelo de Schalter usando esta función de correlación se muestran en la [Tabla 2.5](#).

Para analizar la dependencia espacial se graficaron los coeficientes extremales no paramétricos en cada año, para ello se usó el estimador propuesto por [Smith \(1990b\)](#) y dado en la ecuación (2.18), en la [Figura 2.4](#) se presentan éstos coeficientes para los años 2008, 2009 2012 y 2013, note que conforme avanza el tiempo ésta dependencia tiende a disminuir.

La [Tabla 2.6](#) resume el ajuste del modelo de Schlather para el año 2012 con las cuatro funciones de correlación, ambas tienen un CLIC* muy similar pero el ajuste con la función exponencial proporciona un CLIC* más pequeño, similarmente el valor de la función de log-verosimilitud es muy parecido en estos casos.

2.6. Aplicación

Tabla 2.5: Parámetros estimados para el modelo de Schalter. La cantidad que esta entre paréntesis indica su error estándar.

Año	$\hat{\alpha}$	\hat{c}_2	$\hat{\nu}$
2008	0.7723(0.1033)	0.3141(0.1283)	1.9999(3.26)
2009	0.7000(0.1178)	0.2315(0.0698)	1.9998(1.84)
2010	0.5562(0.3162)	0.0950(0.0398)	1.9887(2.43)
2011	0.1384(0.1708)	0.0342(0.2616)	1.9902(8.82)
2012	0.7818(0.2594)	0.1113(0.0732)	1.9997(4.34)
2013	0.9999(0.1274)	0.1066(0.0492)	1.9814(1.48)

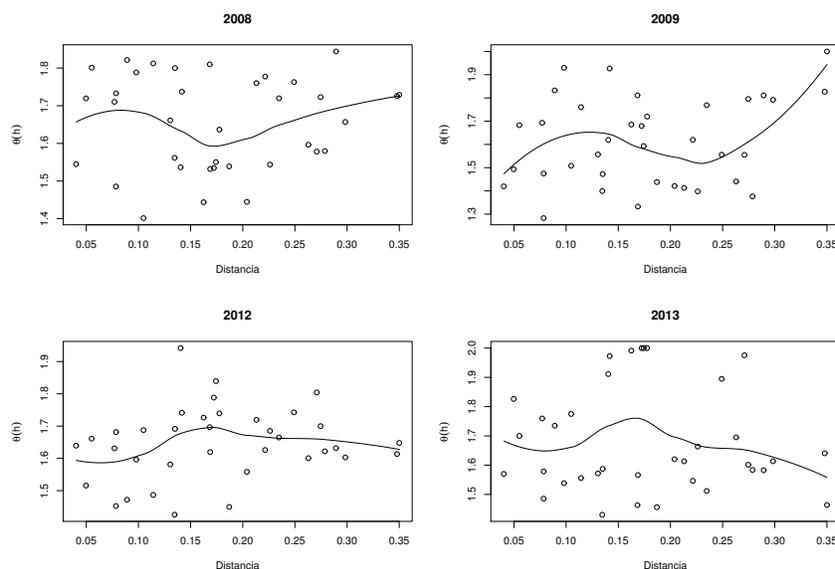


Figura 2.4: Coeficiente extremo no paramétrico. La línea continua estima la tendencia de la dependencia espacial.

Para realizar pronósticos es conveniente definir tendencias en los parámetros de la distribución GEV, ya que es probable que los parámetros presenten variación espacial de manera marginal, en este caso para cada $\mathbf{s} \in \mathcal{S}$, se supone que $Y(\mathbf{s}) \sim \text{GEV}(\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}))$ y de esta manera se puede definir, por ejemplo, $\mu(\mathbf{s}) = h(\mathbf{s}; \beta_\mu)$, $\mathbf{s} \in \mathcal{S}$, para alguna función paramétrica $h(\cdot; \beta_\mu)$. Para detectar una tendencia adecuada se realizó un análisis exploratorio a los datos, la [Figura 2.5](#) muestra una gráfica de símbolos para los datos del año 2012, los círculos están centrados en la ubicación de cada estación y cuyo radio es proporcional al valor de salida en esa posición con el valor medio del área, ésta sugiere las siguientes superficies

2.6. Aplicación

Tabla 2.6: Resumen del modelo de Schlather ajustado a los datos para el año 2012, los errores estándar están en paréntesis, CLIC* es el criterio de información de verosimilitud compuesta y l_p es la función de log-verosimilitud compuesta maximizada.

Correlación	α	c_2	ν	CLIC*	l_p
Whitmat	0.7545(0.1704)	0.0235(0.2351)	7.3450(140.30)	35897.60	-17943.46
Cauchy	0.7578(0.1518)	0.6800(24.1971)	30.1698(2105.10)	35897.68	-17943.44
Exponencial	0.7617(0.2202)	0.1252(0.0637)	1.9994(3.58)	35897.09	-17943.43
Bessel	0.7574(0.1335)	0.0051(0.7280)	150.3081(4347.12)	35897.42	-17943.43

para nuestros parámetros (Davison *et al.*, 2012),

$$\begin{aligned}
 \mu(\mathbf{s}) &= \beta_{0,\mu} + \beta_{1,\mu}\text{lon}(\mathbf{s}) + \beta_{2,\mu}\text{lat}(\mathbf{s}), \\
 \sigma(\mathbf{s}) &= \beta_{0,\sigma} + \beta_{1,\sigma}\text{lon}(\mathbf{s}) + \beta_{2,\sigma}\text{lat}(\mathbf{s}), \\
 \xi(\mathbf{s}) &= \beta_{0,\xi},
 \end{aligned}
 \tag{2.29}$$

donde $\text{lat}(\mathbf{s})$ y $\text{lon}(\mathbf{s})$ son la latitud y la longitud de las estaciones. Se pueden aplicar otras formas de elegir estas estructuras marginales, como por ejemplo ajustando varios modelos plausibles y eligiendo aquel que tenga el menor criterio de información de verosimilitud compuesta. En este caso el supuesto de tener distribuciones marginales de Fréchet en cada ubicación se puede omitir y ajustar el modelo max-stable a los datos originales (Davison *et al.*, 2012).

De esta forma, se ajusto el modelo de Schlather a los datos del año 2012 con la función de correlación exponencial y con las tendencias en los parámetros de la distribución GEV dadas en la ecuación (2.29). La Tabla 2.7 muestra los parámetros estimados con sus respectivos errores estándar de igual forma la Figura 2.6 presenta la validación del modelo, en la diagonal se grafican los niveles de retorno y debajo de ella las gráficas de probabilidad en la escala de Gumbel entre los máximos observados en cada bloque y los obtenidos mediante simulaciones del modelo ajustado, la gráfica superior compara el coeficiente extremal ajustado con el semiparamétrico, las dos gráficas restantes son las ubicaciones de las estaciones y la gráfica de probabilidad de máximos por bloques donde el tamaño de los bloques es cuatro, con esto se valida el buen ajuste del modelo. Posteriormente se realizaron mapas predictivos puntuales para los parámetros de localidad, de forma y para el nivel de retorno a dos años, esto lo muestra la Figura 2.7.

Tabla 2.7: Parámetros estimados y sus errores estándar del modelo de Schlather ajustado a los datos para el año 2012 suponiendo tendencias espaciales en los parámetros de la distribución GEV.

	α	c_2	$\beta_{0,\mu}$	$\beta_{1,\mu}$	$\beta_{2,\mu}$	$\beta_{0,\sigma}$	$\beta_{1,\sigma}$	$\beta_{2,\sigma}$	$\beta_{0,\xi}$
θ	0.00176	0.02575	132.50835	1.06559	-1.28067	68.12825	0.53656	-0.72515	-0.04232
$\text{se}(\hat{\theta})$	10.73064	0.05335	33.38795	0.32510	0.15991	29.18856	0.29139	0.14882	0.03082

2.6. Aplicación

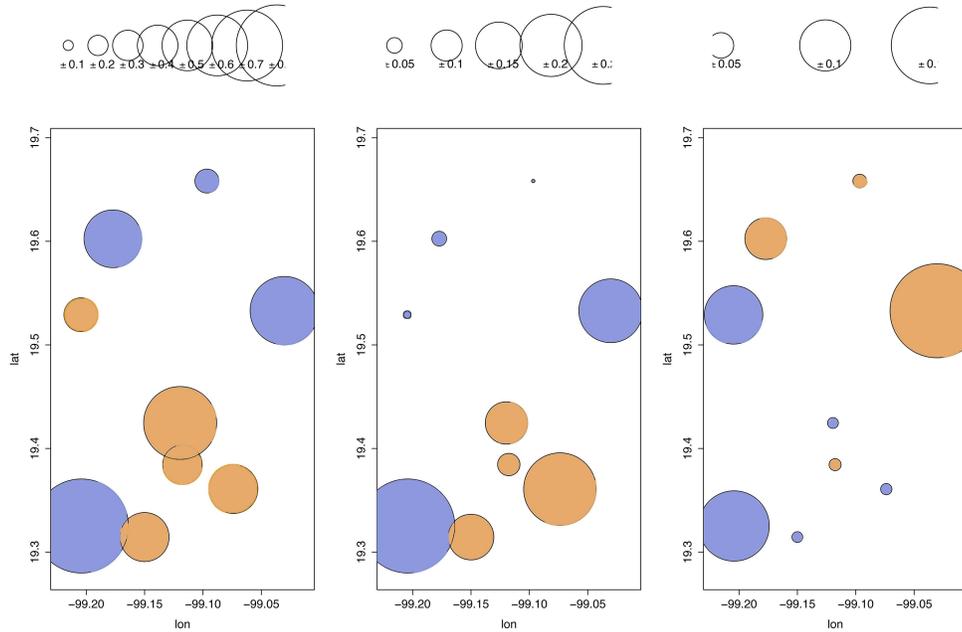


Figura 2.5: Gráfica de símbolos para los datos del año 2012.

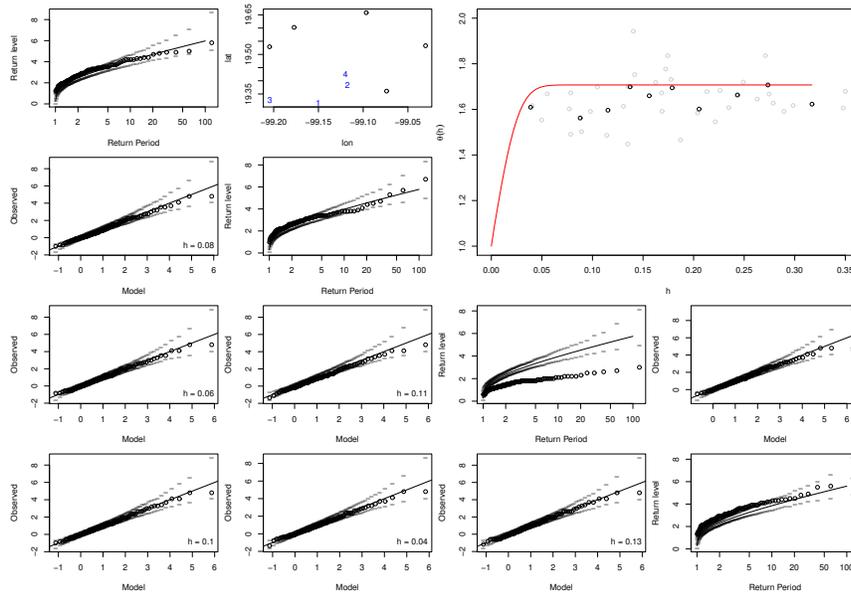


Figura 2.6: Validación del modelo para el modelo max-stable ajustado.

Finalmente se realizaron mapas predictivos usando las tendencias espaciales estimadas de los parámetros de la distribución GEV, se usaron los niveles de retorno dos, cuatro y cinco años. La Figura 2.8 muestra éstos mapas, observe que en la región sur del valle de México se pronostica un aumento en los niveles del contaminante, mientras que la región norte este permanece bajo. Cabe mencionar que en este caso no se toma en cuenta la dependencia espacial.

2.6. Aplicación

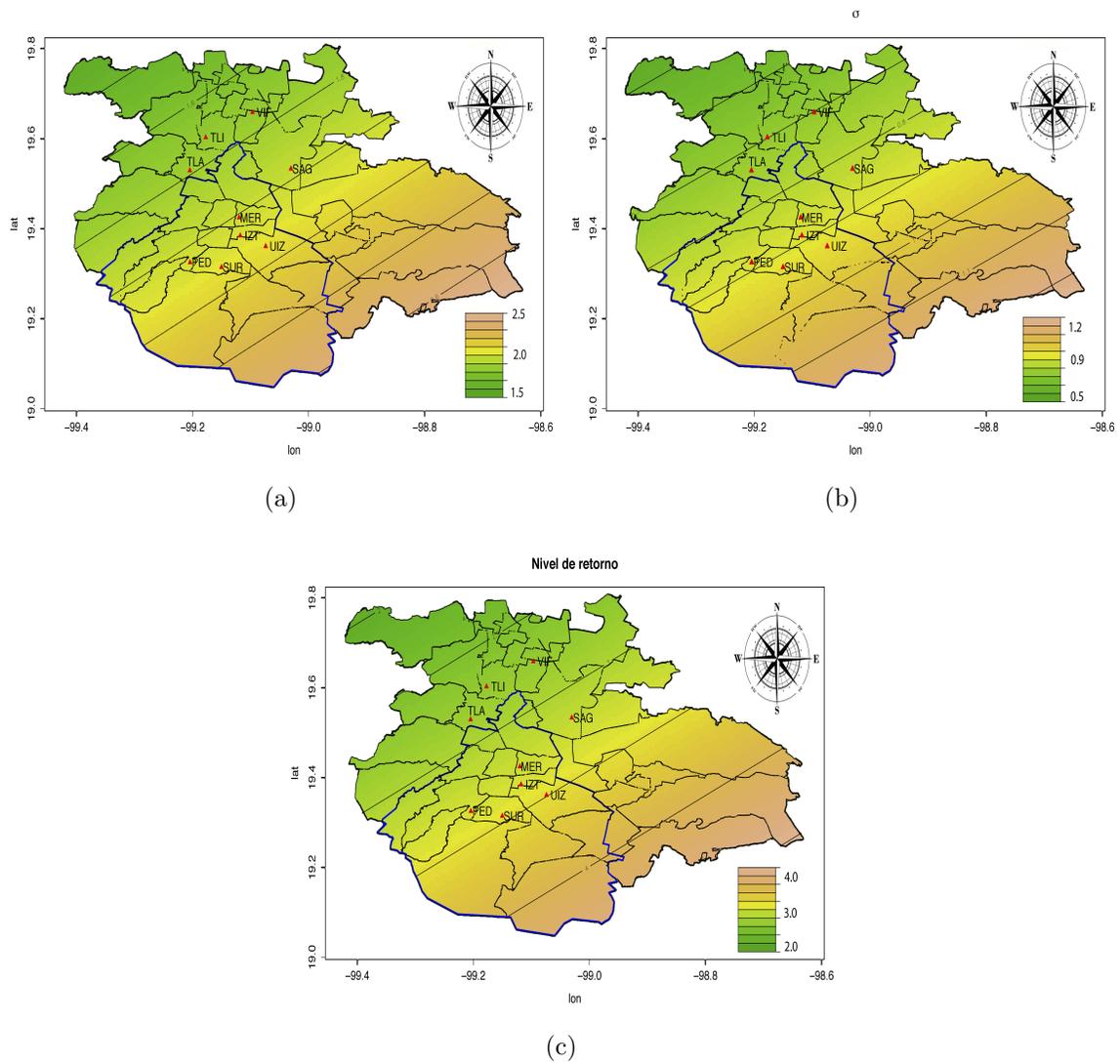


Figura 2.7: Predicciones puntuales del proceso de Schalter ajustado. (a) parámetro de localidad; (b) de escala y (c) el nivel de retorno a dos años.

2.6. Aplicación

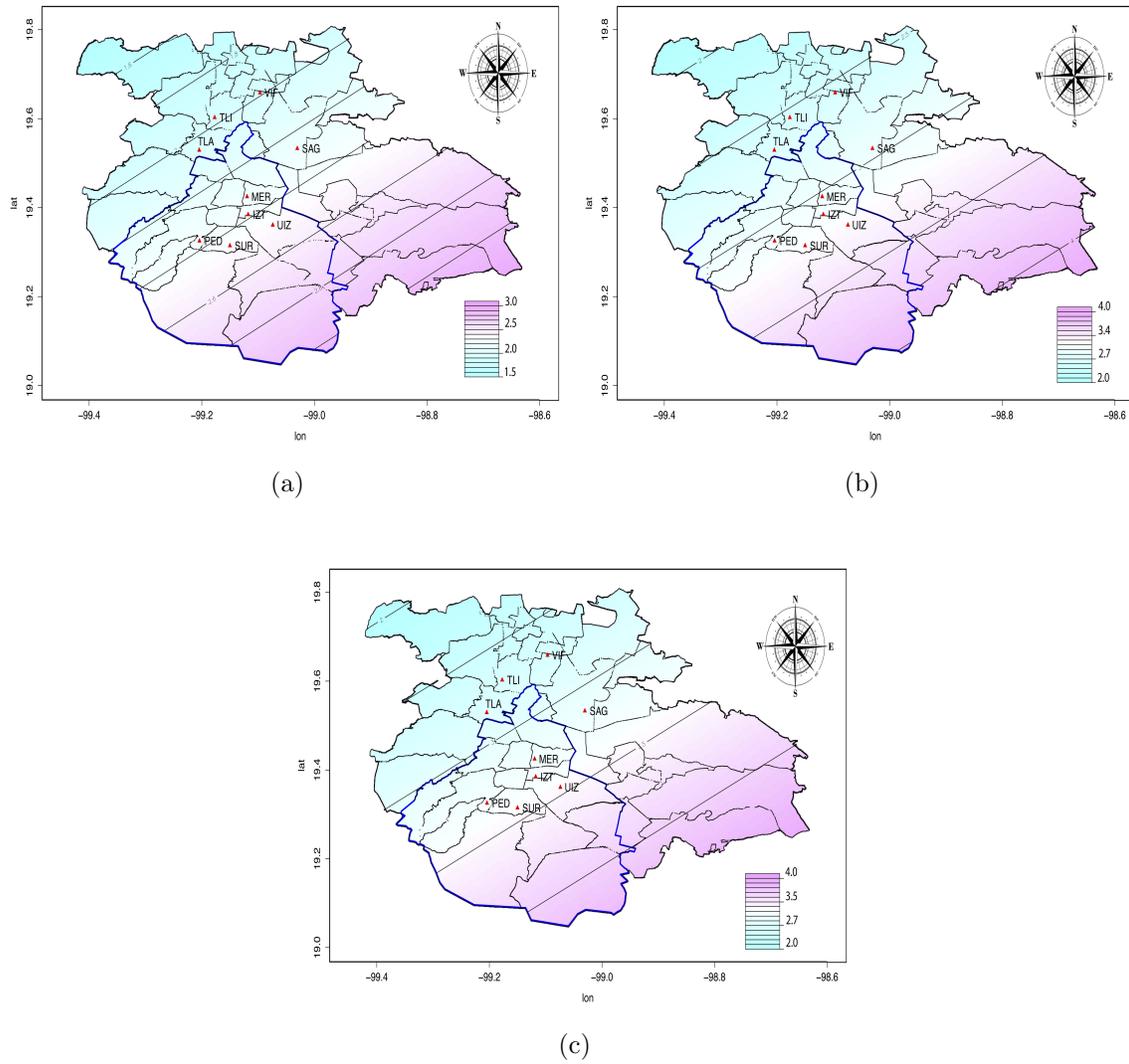


Figura 2.8: Predicción del nivel de retorno. (a) 2 años; (b) 4 años y (c) 5 años.

Modelos jerárquicos bayesianos para extremos espaciales

Muchos de los problemas en la teoría de valores extremos son de naturaleza espacial, las investigaciones aplicadas que se ha desarrollado actualmente, como se mencionó en el Capítulo anterior, han dando lugar a dos formas diferentes de modelación. En este Capítulo se aborda la segunda forma de modelación, en donde una de las primeras aplicaciones sobre estos procesos fue presentada por [Coles y Casson \(1998\)](#) y [Casson y Coles \(1999\)](#) quienes usan un modelo jerárquico bayesiano para modelar datos de velocidad del viento de huracanes en la costa del Golfo de los Estados Unidos. Aplicaciones recientes de este enfoque se pueden encontrar en, [Cooley *et al.* \(2007\)](#) quienes construyen un modelo jerárquico espacial para construir mapas de nivel de retorno de precipitación extrema para una región de Colorado, Estados Unidos, ellos modelan las excedencias sobre un umbral u común en todas las ubicaciones usando la distribución Pareto generalizada, usan una función de covarianza exponencial isotrópica y estacionaria para inducir la dependencia espacial en el proceso latente para los parámetros de intensidad y de escala, mientras que para producir el mapa de nivel de retorno modelan tanto la excedencia como su razón de ocurrencia y construyen para cada uno un modelo jerárquico. En este sentido, [Turkman *et al.* \(2010\)](#) usando una caminata aleatoria para caracterizar las propiedades temporales y suavizando la dependencia espacial, construyen un modelo espacio-temporal el cual lo aplican a velocidades del viento en Portugal. [Gaetan y Grigoletto \(2007\)](#) proponen un modelo espacial para analizar valores máximos anuales de lluvias sobre la región Triveneto en Italia, y usando dependencia espacial no estacionaria y una tendencia espacial en los parámetros de la distribución de valores extremos generalizada proponen un modelo jerárquico a dos niveles. [Daniel Cooley \(2010\)](#) desarrolla un modelo jerárquico bayesiano para caracterizar la precipitación extrema simulando un proceso para un modelo de clima regional sobre su dominio espacial. [Sang y Gelfand \(2009\)](#) ajustan un modelo jerárquico espacio-tiempo a precipitación extrema asumiendo independencia condicional en los datos, dados los parámetros correlacionados espacialmente y proporcionan pronósticos espacio-temporales evaluando la incertidumbre con el método Monte Carlo. [Nakajima *et al.* \(2012\)](#) y [Kunihama *et al.* \(2012\)](#) proponen un enfoque espacio estado para modelar la dependencia temporal en un proceso de valores extremos y proponen un algoritmo eficiente para implementar el método Monte Carlo de cadenas de Markov

3.1. Formulación del modelo

(MCMC) y así poder realizar inferencia bayesiana. Apputhurai y Stephenson (2013) usan un proceso latente espacio-temporal para modelar eventos de precipitación extrema en el Sur-Este de Australia, en donde la parte temporal es capturada modelando tendencias en los parámetros de localización y escala mientras que la parte espacial se captura usando un campo aleatorio gaussiano anisotrópico. Ghosh y Mallick (2011) consideran un enfoque espacio-temporal incorporando la dependencia espacial en la verosimilitud mientras la dependencia temporal se mantiene dentro de la estructura latente. Yan y Moradkhani (2015) proponen un modelo jerárquico bayesiano para el análisis de inundaciones frecuentes, su modelo es una alternativa al análisis tradicional ya que captura la dependencia espacial en su estructura interna. Wang y So (2016) proponen un modelo jerárquico bayesiano que aborda el problema de duraciones múltiples en extremos espaciales, el efecto de la duración de éstos extremos lo caracterizan por la combinación de la duración de estos extremos y la fusión de la duración en una estructura de un proceso espacial latente como una de las covariables.

3.1. Formulación del modelo

En este enfoque, proponemos un modelo que incluya variación espacial en los parámetros de la distribución de valores extremos, esta se puede describir mediante un proceso espacial latente (Casson y Coles, 1999, Davison *et al.*, 2012). Específicamente, se supone que las observaciones $\{Y(\mathbf{s})\}$ son condicionalmente independientes sobre un proceso espacial latente $\{S(\mathbf{s})\}$, $\mathbf{s} \in \mathcal{S}$ y que los parámetros de la distribución GEV $\{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}$ varían suavemente para cada $\mathbf{s} \in \mathcal{S}$, de acuerdo a éste proceso espacial, además se supone que $\{S(\mathbf{s})\}$ es mutuamente independiente con cada parámetro de la distribución GEV, aunque este supuesto se puede relajar (Davison *et al.*, 2012). Por ejemplo, se modela a $\mu(\mathbf{s})$ como:

$$\mu(\mathbf{s}) = f_{\mu}(\mathbf{s}; \boldsymbol{\beta}_{\mu}) + S_{\mu}(\mathbf{s}; \alpha_{\mu}, \lambda_{\mu}),$$

donde f_{μ} es una función determinística que depende de los parámetros de regresión $\boldsymbol{\beta}_{\mu}$ y S_{μ} es un proceso gaussiano estacionario con media cero y función de covarianza $\gamma_{\mu}(\mathbf{h}) = \alpha_{\mu} \exp(-\|\mathbf{h}\|/\lambda_{\mu})$, donde α_{μ} y λ_{μ} son el sill y el rango, respectivamente, ambos desconocidos, se consideran expresiones similares para los parámetros $\sigma(\mathbf{s})$ y $\xi(\mathbf{s})$, es decir,

$$\sigma(\mathbf{s}) = f_{\sigma}(\mathbf{s}; \boldsymbol{\beta}_{\sigma}) + S_{\sigma}(\mathbf{s}; \alpha_{\sigma}, \lambda_{\sigma}),$$

y

$$\xi(\mathbf{s}) = f_{\xi}(\mathbf{s}; \boldsymbol{\beta}_{\xi}) + S_{\xi}(\mathbf{s}; \alpha_{\xi}, \lambda_{\xi}).$$

De esta forma, si se condiciona sobre los tres procesos gaussianos en las ubicaciones $(\mathbf{s}_1, \dots, \mathbf{s}_D)$, se supone que los máximos son independientes con

$$Y_i(\mathbf{s}_d) \mid \{\mu(\mathbf{s}_d), \sigma(\mathbf{s}_d), \xi(\mathbf{s}_d)\} \sim \text{GEV}(\{\mu(\mathbf{s}_d), \sigma(\mathbf{s}_d), \xi(\mathbf{s}_d)\}),$$

$i = 1, \dots, n$ y $d = 1, \dots, D$.

3.2. Metodología

Desde el enfoque bayesiano se necesita definir distribuciones a priori para los parámetros $\alpha_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu$ de $\mu(\mathbf{s})$. Siguiendo a Davison *et al.* (2012), tomamos una distribución gama inversa para α_μ con parámetros $\kappa_{\alpha_\mu}^*$ y $\theta_{\alpha_\mu}^*$, es decir,

$$\alpha_\mu \sim GI(\kappa_{\alpha_\mu}^*, \theta_{\alpha_\mu}^*).$$

Una distribución normal multivariada para $\boldsymbol{\beta}_\mu$ con parámetros $\mu_{\alpha_\mu}^*$ y $\Sigma_{\alpha_\mu}^*$, es decir,

$$\boldsymbol{\beta}_\mu \sim N_p(\mu_{\alpha_\mu}^*, \Sigma_{\alpha_\mu}^*),$$

en esta caso tanto la distribución gamma inversa como la distribución normal multivariada son a priori conjugadas, y una distribución gamma para λ_μ con parámetros $\kappa_{\lambda_\mu}^*$ y $\theta_{\lambda_\mu}^*$, es decir,

$$\lambda_\mu \sim G(\kappa_{\lambda_\mu}^*, \theta_{\lambda_\mu}^*).$$

Para calcular las distribuciones a posterior usando el algoritmo de cadenas de Markov Monte Carlo (MCMC), se necesitan las distribuciones condicionales (full conditional), en esta caso toman la forma,

$$\begin{aligned} \pi(\mu(\mathbf{s}) \mid \text{resto}) &\propto \pi(\mu(\mathbf{s}) \mid \alpha_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu) \pi(\mathbf{y} \mid \mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})), \\ \pi(\alpha_\mu \mid \text{resto}) &\propto \pi(\alpha_\mu \mid \kappa_{\alpha_\mu}^*, \theta_{\alpha_\mu}^*) \pi(\mu(\mathbf{s}) \mid \alpha_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu), \\ \pi(\lambda_\mu \mid \text{resto}) &\propto \pi(\lambda_\mu \mid \kappa_{\lambda_\mu}^*, \theta_{\lambda_\mu}^*) \pi(\mu(\mathbf{s}) \mid \alpha_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu), \\ \pi(\boldsymbol{\beta}_\mu \mid \text{resto}) &\propto \pi(\boldsymbol{\beta}_\mu \mid \mu_{\alpha_\mu}^*, \Sigma_{\alpha_\mu}^*) \pi(\mu(\mathbf{s}) \mid \alpha_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu), \end{aligned}$$

donde $\kappa^*, \theta^*, \mu^*$ y Σ^* son los parámetros de las distribuciones a priori.

Las distribuciones a priori para los parámetros $\alpha_\sigma, \lambda_\sigma, \boldsymbol{\beta}_\sigma$ y $\alpha_\xi, \lambda_\xi, \boldsymbol{\beta}_\xi$ de $\sigma(\mathbf{s})$ y $\xi(\mathbf{s})$, respectivamente, son similares, por lo que las distribuciones condicionales también tienen expresiones similares.

Para obtener las muestras de la distribución conjunta a posteriori usamos una eficiente implementación del algoritmo Gibb sampler incluido en el paquete `SpatialExtremes` de R (R Development Core Team, 2015, Ribatet, 2015), éste algoritmo se puede encontrar a detalle en Davison *et al.* (2012).

3.2. Metodología

Los datos a analizar usando el modelo descrito en la sección anterior son los mismos que se usaron en la aplicación del Capítulo 2 sección 2.6, solo que para esta aplicación añadimos el año 2014 y 2015. Para esto, se propone la siguiente metodología:

- Usar la aproximación de máximos por bloques, cada bloque está formado por observaciones consecutivas de tres días, para garantizar la independencia de éstos, Coles *et al.* (2001).
- Ajustar la distribución de valores extremos generalizada, usando parámetros marginales descritos por las ecuaciones (2.29).

3.3. Resultados

Tabla 3.1: Hiperparámetros para el proceso latente usados en nuestra aplicación. Las distribuciones a priori para α . y λ . son gamma inversa y gamma, respectivamente.

	α .		λ .	
	κ^*	θ^*	κ^*	θ^*
$\mu(\mathbf{s})$	1	8	4	2
$\sigma(\mathbf{s})$	2	3	3	1
$\xi(\mathbf{s})$	2	1	5	1

- Proponer valores para los hiperparámetros de las distribuciones a priori, para los parámetros de regresión se propone una distribución normal multivariada con media cero y varianza grande, mientras que para los parámetros de la función de covarianza para el proceso latente $S(\mathbf{s})$, α . y λ . se proponen distribuciones a priori informativas (Banerjee *et al.*, 2014), gama inversa y gamma, respectivamente, los valores de los hiperparámetros para éstas distribuciones se muestran en la [Tabla 3.1](#). Éstos se eligieron después de un análisis exploratorio acerca del comportamiento de las distribuciones a posteriori.
- Realizar los mapas predictivos para cada parámetro de la distribución de valores extremos generalizada y para el nivel de retorno a un año, con las tendencias estimadas.

3.3. Resultados

Los parámetros de la distribución a posterior conjunta se resumen en la [Tabla 3.2](#), éstos se obtuvieron después de 100,000 iteraciones de la cadena de Markov con un adelgazamiento de 10 y un periodo de calentamiento de 5000 iteraciones.

Tabla 3.2: Parámetros estimados para la distribución a posterior y del proceso latente. La estimación corresponde a la media a posterior y su correspondiente intervalo creíble del 95 %, el cual se muestra en paréntesis.

	β_0	β_1	β_2	α	λ
$\mu(\mathbf{s})$	2.07(1.31,13.01)	-0.0526(-1.36,-1.25)	-0.2871(-6.26,5.63)	6.44(1.389,81)	3.66(0.7198,10.48)
$\sigma(\mathbf{s})$	3.12(9.23,14.74)	0.0388(-0.7719,0.8598)	0.0707(-1.77,1.88)	0.9604(0.3774,2.27)	2.80(0.8232,6.20)
$\xi(\mathbf{s})$	-0.0566(-0.9578,0.8256)	-	-	0.2105(0.0918,0.4560)	7.26(2.29,15.67)

Nótese que no existe variación (cambio de signo) en los parámetros β_1 y β_2 para el parámetro de localización $\mu(\mathbf{s})$, esto se ve razonable ya que no existe un cambio drástico en la altitud en la región de estudio, lo mismo sucede con el parámetro de escala σ . Debido a que el parámetro de forma $\xi < 0$, se tiene una distribución Weibull para nuestros datos, aunque podría corresponder una distribución Gumbel ya que es muy cercano a cero y el intervalo creíble contiene a la parte positiva.

3.3. Resultados

En estadística bayesiana es común llevar a cabo un análisis acerca de la sensibilidad de los resultados con respecto a la elección de las distribuciones a priori. La [Figura 3.1](#) muestra las distribuciones a posterior para α y λ para cada parámetro de la distribución GEV tomando diferentes distribuciones a priori, aunque estas son muy diferentes en todos los casos los resultados mostrados en la [Tabla 3.2](#) y los mapas que se muestran en las [Figuras 3.3–3.6](#), fueron similares.

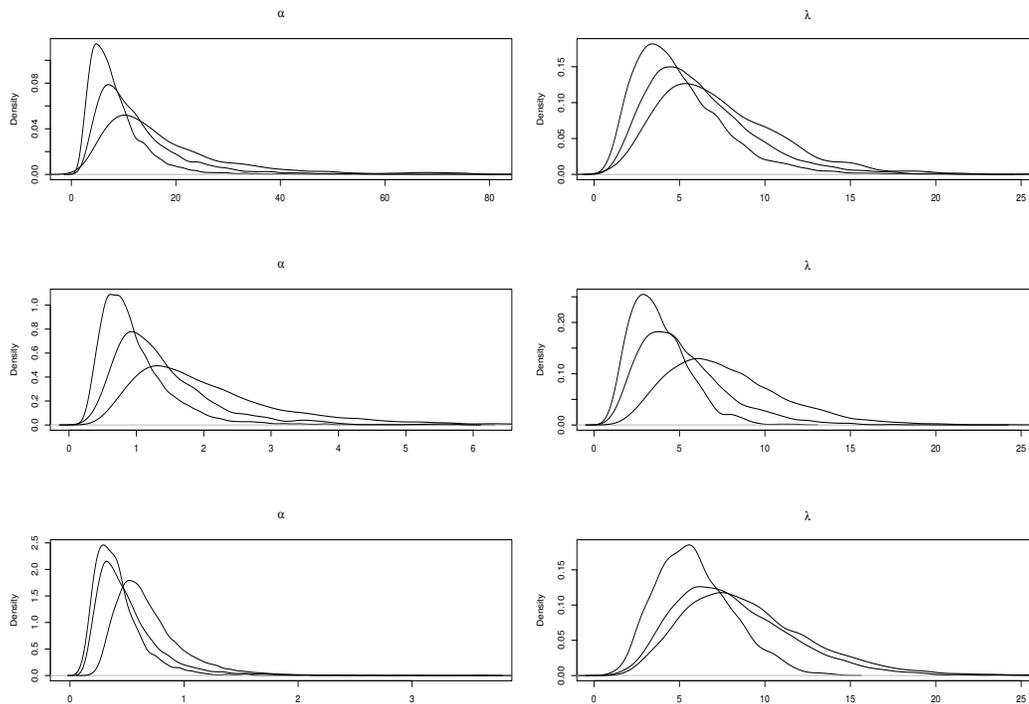


Figura 3.1: Distribuciones a posterior para α y λ para cada parámetro de la distribución GEV. EL primer renglón corresponde a μ , el de en mitad a σ y el de abajo a ξ .

Se encontró que no fue posible tener un aprendizaje simultáneo en los parámetros α y λ ya que la distribución a posterior para cada uno de ellos no difieren mucho de la a priori, esto se presento en cada parámetro de la distribución GEV, lo anterior se puede observar en la [Figura 3.2](#).

Las [Figuras 3.3](#), [3.4](#) y [3.5](#) muestran los mapas para cada parámetro de la distribución GEV y la [Figura 3.6](#) muestra el mapa de nivel de retorno a un año, estos mapas se realizan generando una simulación condicional de tres procesos gaussianos independientes para cada estado de la cadena de Markov dando el valor actual de μ , σ y ξ y después se usa esa realización para calcular los niveles de retorno en las ubicaciones ([Davison et al., 2012](#)). En estos mapas se puede observar el comportamiento espacial de los parámetros en forma marginal lo cual permite captar la variación local en los niveles de retorno, lo cual no es posible solamente con las tendencias determinísticas.

3.3. Resultados

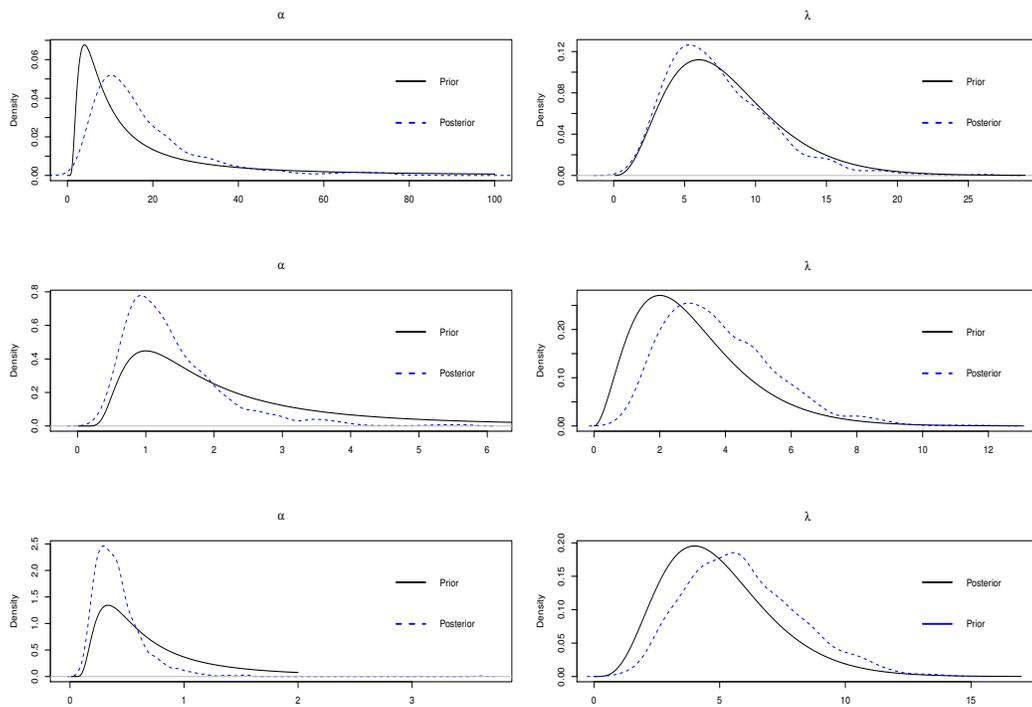


Figura 3.2: Distribuciones a priori y a posteriori para α y λ para cada parámetro de la distribución GEV. EL primer renglón corresponde a μ , el de en mitad a σ y el de abajo a ξ .

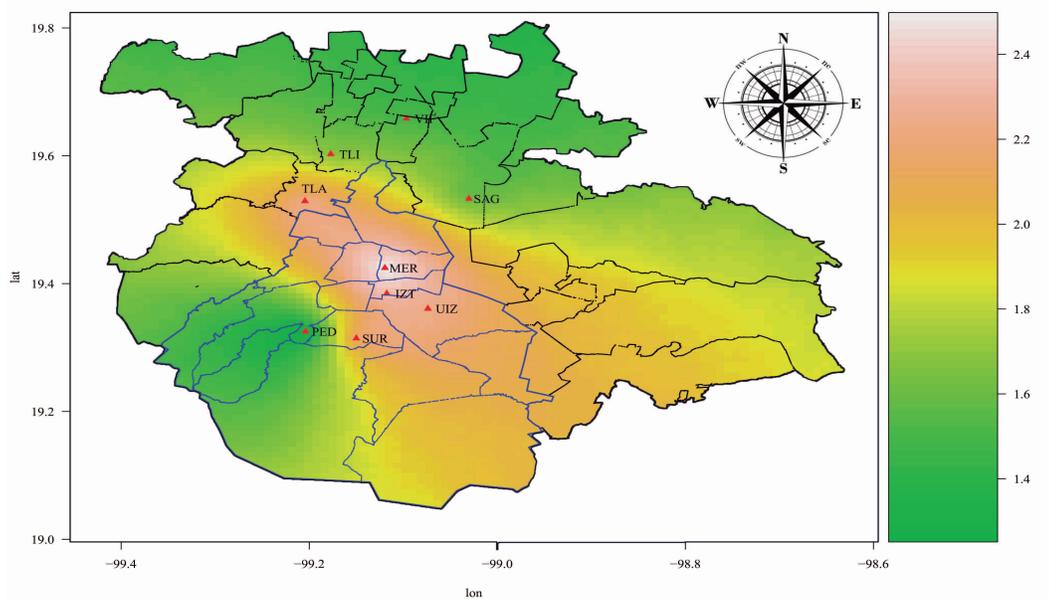


Figura 3.3: Mapa para el parámetro de localización de la distribución GEV.

3.3. Resultados

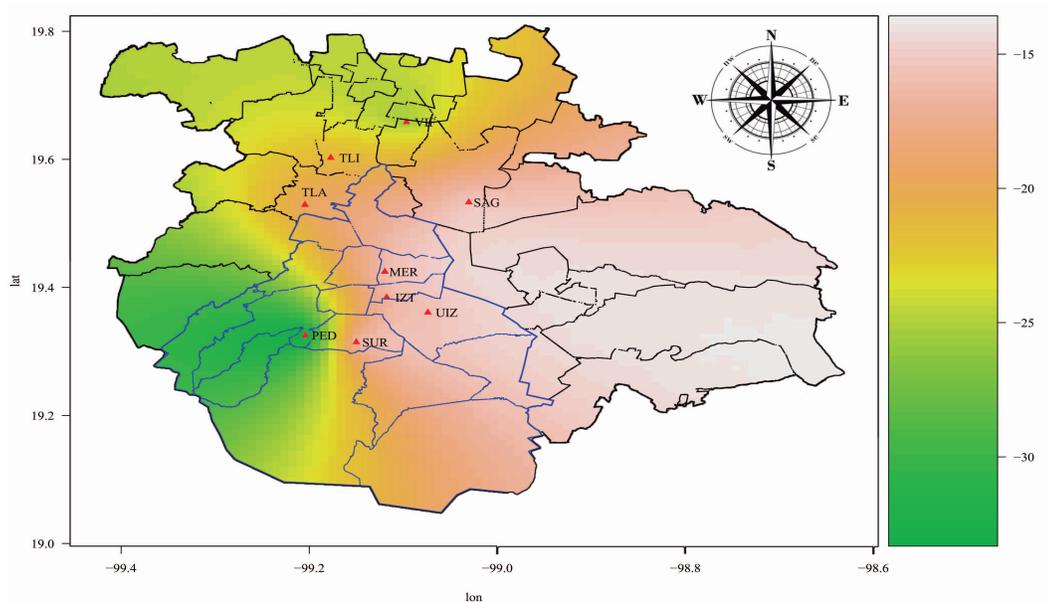


Figura 3.4: Mapa para el parámetro de escala de la distribución GEV.

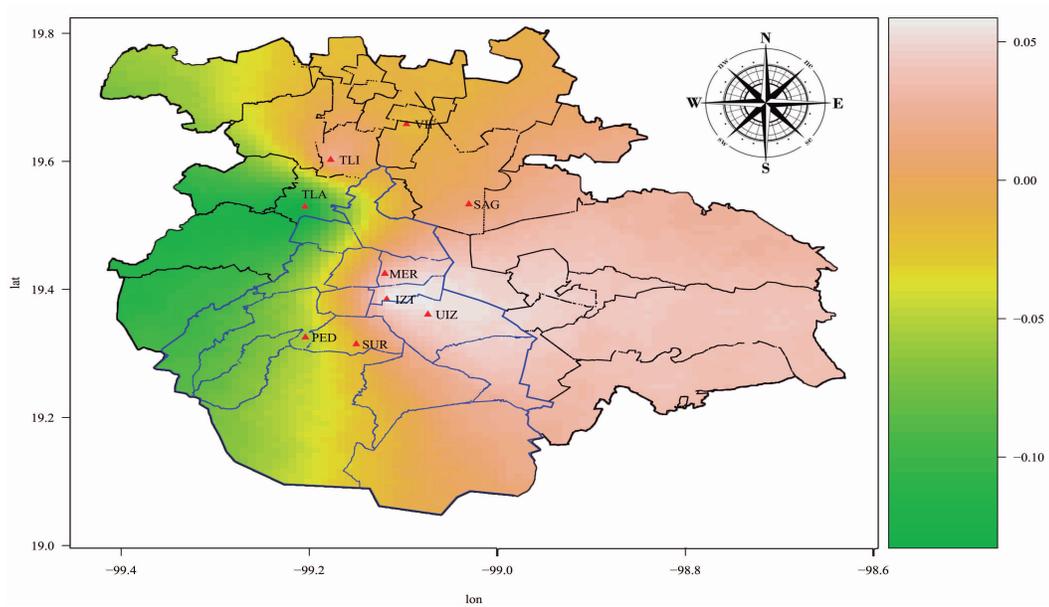


Figura 3.5: Mapa para el parámetro de forma de la distribución GEV.

3.3. Resultados

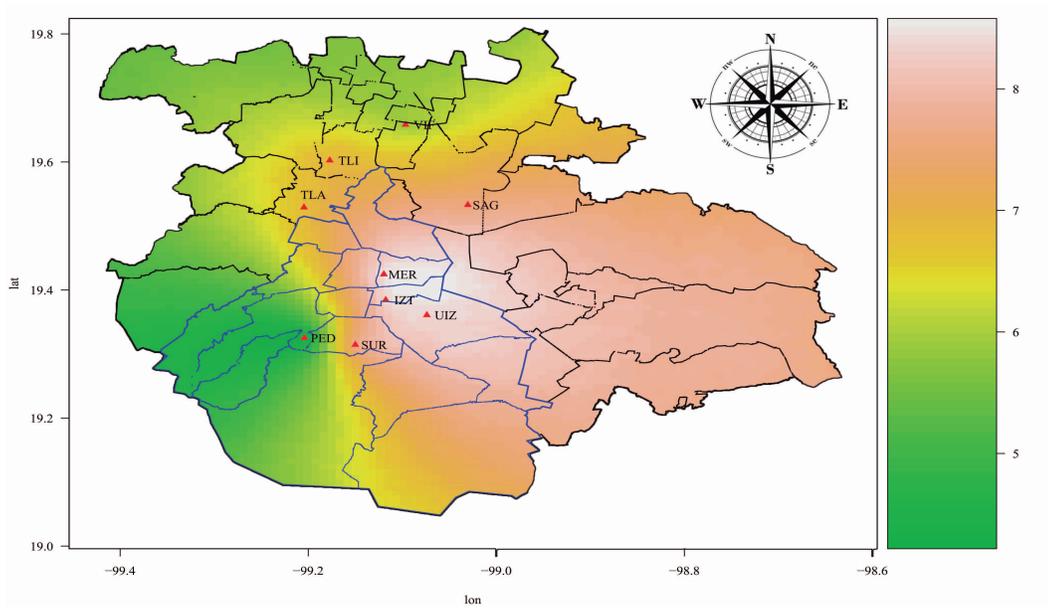


Figura 3.6: Mapa del nivel de retorno estimado de un año para los datos de CO mediante el proceso latente.

El modelo de Schlather espacio–temporal

En muchas de las aplicaciones que se presentan en la teoría de valores extremos se analizan procesos naturales que se desarrollan en el espacio, pero en algunos casos es de interés darle un seguimiento temporal. Por ejemplo, en la aplicación presentada en el [Capítulo 2](#) y [3](#), se estudió la contaminación por monóxido de carbono en la ZMVM y se disponen de estaciones que se distribuyen espacialmente por la ciudad y en donde se toman las mediciones de éste contaminante, y en las cuales sería de interés darles un seguimiento temporal a cada una de ellas. Suponemos que, en vez de considerar solamente la distribución espacial en ciertos instantes de tiempo (procesos puramente espaciales) o considerar solamente su evolución temporal sobre ciertas localizaciones (procesos puramente temporales), si se considera la modelación del proceso de manera conjunta en el espacio–tiempo, se pueden obtener algunos beneficios. En años recientes, esta modelación ha llamado la atención de algunos investigadores y ha propuesto algunas generalizaciones al espacio–tiempo de modelos que solo se han estudiado desde el enfoque espacial. Los primeros en considerar esto fueron [Davis *et al.* \(2013a\)](#), quienes extienden el modelo de Smith ([Smith, 1990b](#)) al espacio–tiempo, presentan de manera explícita la función de distribución conjunta para dos ubicaciones y también muestran que la función de covarianza en el espacio–tiempo del proceso gaussiano se puede interpretar en términos de la función de dependencia del proceso max-stable espacio–temporal, esta extensión aun no presenta una aplicación a datos reales, el proceso de inferencia para esta extensión la muestran [Davis *et al.* \(2013b\)](#). Una segunda generalización, se presenta en [Huser y Davison \(2014\)](#), ellos extienden el modelo de Schlather ([Schlather, 2002](#)) al espacio–tiempo considerando el conjunto aleatorio que Schlather propone y dado en la [ecuación \(2.15\)](#), ellos lo aplican a datos de lluvias en Suiza. Nuestro interés se centra en el modelo de Schlather presentado en la [ecuación \(2.13\)](#) y nuestra propuesta se basa en extender este modelo al espacio–tiempo, nuestro principal resultado es que se obtiene de manera explícita la función de distribución acumulada bivariada del proceso espacio–temporal que se propone, a partir de aquí también se propone realizar inferencia usando la función de verosimilitud compuesta.

4.1. Generalidades

Las definiciones que se presentan en este apartado, se refieren básicamente a la función de covarianza de un proceso espacio–temporal, y son generalizaciones del caso espacial dados en el [Capítulo 2, sección 2.1](#). Para más detalles, ver por ejemplo, [Cressie \(2015\)](#), [Stein \(2012\)](#), [Banerjee *et al.* \(2014\)](#), [Diggle y Ribeiro \(2007\)](#), [Gneiting *et al.* \(2006\)](#) y [Cressie y Wikle \(2015\)](#). Al igual que antes, se usará la notación $\mathbf{s} \in \mathcal{S}$ para denotar una localización genérica en algún espacio $\mathcal{S} \subset \mathbb{R}^d$ y $t \in \mathbb{R}, \mathbb{Z}$ o \mathbb{N} , por lo que se considera un proceso estocástico espacio–temporal dado por $\{Y(\mathbf{s}, t), \mathbf{s} \in \mathcal{S}, t \in \mathbb{R}\}$ el cual se observa en d coordenadas espacio–tiempo, $(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_d, t_d)$.

Definición 4.1. El proceso espacio–temporal $Y(\mathbf{s}, t)$ tiene función de covarianza espacialmente estacionaria si, para cualquier par de localizaciones, (\mathbf{s}_i, t_i) y $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$, la covarianza $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ depende únicamente de la distancia entre las localizaciones espaciales $\mathbf{s}_i - \mathbf{s}_j$ y de los tiempos $t_i - t_j$, $i, j = 1, 2, \dots, d$.

Definición 4.2. El proceso espacio–temporal $Y(\mathbf{s}, t)$ tiene función de covarianza temporalmente estacionaria si, para cualquier par de localizaciones, (\mathbf{s}_i, t_i) y $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$, la covarianza $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ depende únicamente de los tiempos $t_i - t_j$ y de localizaciones espaciales $\mathbf{s}_i - \mathbf{s}_j$, $i, j = 1, 2, \dots, d$.

Definición 4.3. Si el proceso espacio–temporal $Y(\mathbf{s}, t)$ tiene función de covarianza estacionaria tanto espacial como temporal, entonces se dice que tiene función de covarianza estacionaria. En ese caso, la función de covarianza se puede expresar como

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C(\mathbf{h}, u),$$

siendo $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ y $u = t_i - t_j$ las distancias espacial y temporal, respectivamente.

Observe que si un proceso tiene función de covarianza estacionaria, entonces su varianza no depende de su localización, ya que

$$V\{Y(\mathbf{s}, t)\} = C(\mathbf{0}, 0) = \sigma^2, \quad \forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}.$$

Definición 4.4. Se dice que un proceso espacio–temporal $Y(\mathbf{s}, t)$ es estacionario de segundo orden (o estacionario en sentido débil, o simplemente estacionario) si tiene media constante y función de covarianza estacionaria.

Definición 4.5. Se dice que un proceso espacio–temporal $Y(\mathbf{s}, t)$ es un proceso gaussiano si para cualquier conjunto de localizaciones espacio–temporales $(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_d, t_d)$ el vector aleatorio $(Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_d, t_d))'$ tiene una distribución normal multivariada.

Definición 4.6. Un proceso espacio–temporal $Y(\mathbf{s}, t)$ tiene función de covarianza separable si existe una función de covarianza puramente espacial $C_s(\cdot)$ y una función de covarianza puramente temporal $C_t(\cdot)$ tales que

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C_s(\mathbf{s}_i, \mathbf{s}_j)C_t(t_i, t_j), \quad i, j = 1, \dots, d,$$

para cualquier par de localizaciones (\mathbf{s}_i, t_i) y $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$. De lo contrario se dice que la función de covarianza no es separable.

4.1. Generalidades

Definición 4.7. Un proceso espacio-temporal $Y(\mathbf{s}, t)$ tiene función de covarianza completamente simétrica si

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C((\mathbf{s}_j, t_j), (\mathbf{s}_i, t_i)), \quad i, j = 1, \dots, d,$$

para cualquier par de localizaciones (\mathbf{s}_i, t_i) y $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$.

Definición 4.8. Un proceso espacio-temporal $Y(\mathbf{s}, t)$ tiene función de covarianza con soporte compacto si, para cualquier par de localizaciones (\mathbf{s}_i, t_i) y $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$ la covarianza $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ tiende a cero cuando la distancia espacial o temporal es suficientemente grande.

Definición 4.9. Un proceso espacio-temporal $Y(\mathbf{s}, t)$ tiene función de covarianza espacialmente isotrópica si

$$C(\mathbf{h}, u) = C(\|\mathbf{h}\|, u), \quad \forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}.$$

Definición 4.10. Un proceso espacio-temporal $Y(\mathbf{s}, t)$ tiene función de covarianza temporalmente isotrópica si

$$C(\mathbf{h}, u) = C(\mathbf{h}, |u|), \quad \forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}.$$

Note que si la función de covarianza de un proceso estacionario es espacial y temporalmente isotrópica, entonces es completamente simétrica. Cabe mencionar que una condición necesaria y suficiente para que una función $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ de valores reales definida sobre $\mathbb{R}^d \times \mathbb{R}$ sea una función de covarianza es que sea simétrica y definida positiva, es decir,

$$\sum_{i=1}^d \sum_{j=1}^d \varphi_i \varphi_j C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \leq 0,$$

para cualesquiera $(\mathbf{s}_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$ y $\varphi_i \in \mathbb{R}$, $i = 1, \dots, d$.

La [Figura 4.1](#) muestra la relación entre los diferentes tipos de funciones de covarianza para procesos espacio-temporales ([Gneiting et al., 2006](#)).

Otra función que se utiliza en la modelación de procesos espacio-temporales es la función de correlación espacio-temporal, que se define a continuación.

Definición 4.11. Sea $Y(\mathbf{s}, t)$ un proceso espacio-temporal estacionario de segundo orden con varianza $\sigma^2 = C(\mathbf{0}, 0) > 0$, la función de correlación espacio-temporal del proceso se define como

$$\rho(\mathbf{h}, u) = \frac{C(\mathbf{h}, u)}{C(\mathbf{0}, 0)}.$$

Observe que si $\rho(\mathbf{h}, u)$ es una función de correlación sobre $\mathbb{R}^d \times \mathbb{R}$, entonces sus marginales $\rho(\mathbf{0}, u)$ y $\rho(\mathbf{h}, 0)$, serán, respectivamente, funciones de correlación espacial sobre \mathbb{R}^d y temporal sobre \mathbb{R} .

4.2. El modelo espacio–temporal

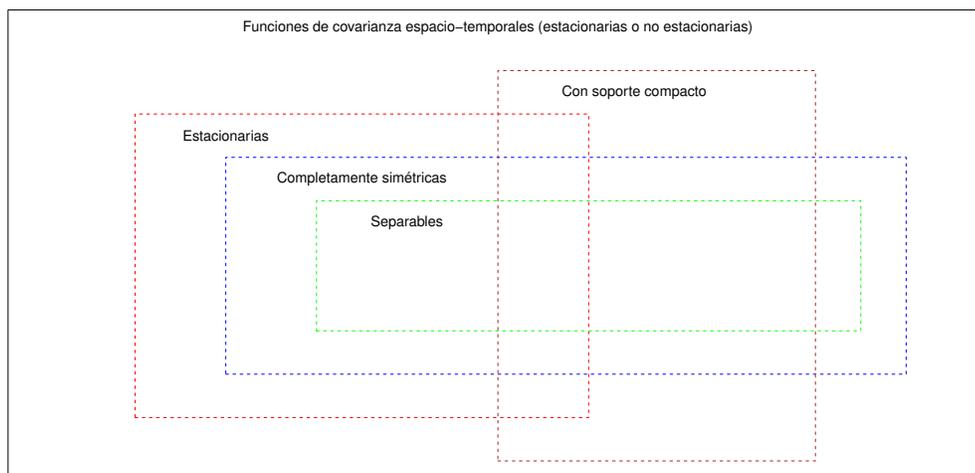


Figura 4.1: Relaciones entre los diferentes tipos de funciones de covarianza espacio–temporales.

4.2. El modelo espacio–temporal

En esta sección se presentan los resultados sobre la extensión del modelo de Schlather al espacio–tiempo. En el primero de ellos se muestra a continuación.

Teorema 4.1. Sea $\{Y(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$ un proceso aleatorio espacio–temporal estacionario con $E\{Y(\mathbf{s}, t)\} = \mu < \infty$ y sea $\rho(\mathbf{h}, u) = \rho_{hu}$ alguna función de correlación para $Y(\mathbf{s}, t)$, además sea Π un proceso de Poisson sobre $(0, \infty)$ con intensidad $\mu^{-1}\xi^{-2}d\xi$, entonces el proceso

$$Z(\mathbf{s}, t) = \max_{i \in \Pi} \xi_i Y_i(\mathbf{s}, t) = \max \xi_i \max \{0, Y_i(\mathbf{s}, t)\},$$

es un proceso max-stable con distribución marginal Fréchet unitaria, donde $\{Y_i(\mathbf{s}, t)\}$ son copias i.i.d de $\{Y(\mathbf{s}, t)\}$ para todo $i \in (0, \infty)$.

Demostración. Por demostrar que $Z(\mathbf{s}, t)$:

- a) es max-stable y que
- b) tiene distribución marginal Fréchet unitaria, es decir,

$$P\{Z(\mathbf{s}, t) \leq z\} = \exp\left(-\frac{1}{z}\right).$$

- a) El proceso $Z(\mathbf{s}, t)$ es max-stable debido a la infinita divisibilidad del proceso de Poisson (de Haan, 1984), en otras palabras $Z(\mathbf{s}, t)$ es max-stable debido a la superposición de n

4.2. El modelo espacio-temporal

procesos de Poisson idénticos e independientes.

b) Considere el conjunto

$$E = \{(\xi, y(\mathbf{s}, t)) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R} : \xi \max\{0, y(\mathbf{s}, t)\} > z\},$$

entonces para \mathbf{s} fija y $z > 0$,

$$\begin{aligned} P\{Z(\mathbf{s}, t) \leq z\} &= P(\text{no existan puntos en } E) \\ &= \exp\left\{-\int_{\mathbb{R}^d} \int_{\frac{z}{\max\{0, y(\mathbf{s}, t)\}}}^{\infty} \xi^{-2} \mu^{-1} d\xi \nu(dy(\mathbf{s}, t))\right\} \\ &= \exp\left\{-\mu^{-1} \int_{\mathbb{R}^d} z^{-1} \max\{0, y(\mathbf{s}, t)\} \nu(dy(\mathbf{s}, t))\right\}, \end{aligned}$$

pero como $E\{Y(\mathbf{s}, t)\} = E[\max\{0, Y_i(\mathbf{s}, t)\}] = \mu$,

$$P\{Z(\mathbf{s}, t) \leq z\} = \exp(-z^{-1}) = \exp\left(-\frac{1}{z}\right),$$

la cual es la FDA de una distribución Fréchet unitaria. ■

El segundo resultado muestra cual es la forma explícita de la función de distribución acumulada para el proceso max-stable dado en el teorema anterior, y es el siguiente.

Teorema 4.2. *La función de distribución acumulada bivariada de $Z(\mathbf{s}, t)$ definido en el Teorema anterior esta dada por,*

$$\begin{aligned} F(z_1, z_2) &= P\{Z(\mathbf{s}_1, t_1) \leq z_1, Z(\mathbf{s}_2, t_2) \leq z_2\} \\ &= \exp\left\{-\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2}\right) \left(1 + \sqrt{1 - \frac{2z_1 z_2 (1 + \rho_{hu})}{(z_1 + z_2)^2}}\right)\right\}, \quad z_1, z_2 > 0, \end{aligned}$$

cuando $Y_i(\mathbf{s}, t) = \sqrt{2\pi} \max\{0, \varepsilon_i(\mathbf{s}, t)\}$ y $\varepsilon_i(\mathbf{s}, t)$ son copias independientes de un proceso aleatorio gaussiano espacio-temporal con función de correlación $\rho(\mathbf{h}, u) = \rho_{hu}$.

Demostración. La función de distribución acumulada de $Z(\mathbf{s}, t)$ se calcula usando la representación espectral de [de Haan \(1984\)](#) para procesos max-stable, dada por,

$$P\{Z(\mathbf{s}_1, t_1) \leq z_1, Z(\mathbf{s}_2, t_2) \leq z_2\} = \exp\{-V(z_1, z_2)\}, \quad z_1, z_2 > 0,$$

donde $V(z_1, z_2) = E\left\{\max_{i=1,2} \frac{Y_i(\mathbf{s}, t)}{z_i}\right\}$ y $Y_i(\mathbf{s}, t) = \sqrt{2\pi} \max_{i=1,2} \{0, \varepsilon_i(\mathbf{s}, t)\}$. Así,

$$\begin{aligned} V(z_1, z_2) &= \sqrt{2\pi} \int \max\left\{0, \frac{\varepsilon_1(\mathbf{s}_1, t_1)}{z_1}, \frac{\varepsilon_2(\mathbf{s}_2, t_2)}{z_2}\right\} dP \\ &= \sqrt{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \max\left\{0, \frac{v}{z_1}, \frac{u}{z_2}\right\} \frac{\exp\left\{-\frac{v^2 - 2uv\rho_{hu} + u^2}{2(1 - \rho_{ht}^2)}\right\}}{2\pi(1 - \rho_{hu}^2)^{1/2}} dv du. \end{aligned}$$

4.2. El modelo espacio-temporal

Para simplificar, defina $f(u, v) = \frac{v^2 - 2uv\rho_{hu} + u^2}{2(1 - \rho_{hu}^2)}$ y note que $\frac{\sqrt{2\pi}}{2\pi(1 - \rho_{hu}^2)^{1/2}} = \frac{1}{\sqrt{2\pi(1 - \rho_{hu}^2)}} = \frac{1}{k}$, donde

$$k = \sqrt{2\pi(1 - \rho_{hu}^2)}, \quad (4.1)$$

resultando,

$$V(z_1, z_2) = \frac{1}{k} \int_0^\infty \int_0^\infty \max \left\{ 0, \frac{v}{z_1}, \frac{u}{z_2} \right\} \exp \{-f(u, v)\} dv du.$$

Aquí se tienen tres casos al tomar el máximo, cuando es i) 0, ii) $\frac{v}{z_1}$ y iii) $\frac{u}{z_2}$.

i) De esta forma, cuando el máximo es 0, $V(z_1, z_2)$ resulta,

$$\begin{aligned} A &= \frac{1}{k} \int_0^\infty \int_0^\infty \frac{v}{z_1} \exp \{-f(u, v)\} dv du + \frac{1}{k} \int_0^\infty \int_0^\infty \frac{u}{z_2} \exp \{-f(u, v)\} dv du \\ &= \frac{1}{kz_1} \int_0^\infty \int_0^\infty v \exp \{-f(u, v)\} dv du + \frac{1}{kz_2} \int_0^\infty \int_0^\infty u \exp \{-f(u, v)\} dv du. \end{aligned}$$

Definamos

$$I_1 = \int_0^\infty \int_0^\infty v \exp \{-f(u, v)\} dv du$$

e

$$I_{12} = \int_0^\infty \int_0^\infty u \exp \{-f(u, v)\} dv du.$$

Observe que el resultado de I_1 debe ser el mismo que el de I_{12} , por lo que se puede escribir lo siguiente,

$$\begin{aligned} I_1 + I_{12} &= \int_0^\infty \int_0^\infty (v + u) \exp \{-f(u, v)\} dv du \\ 2I_{11} &= \int_0^\infty \int_0^\infty (v + u) \exp \{-f(u, v)\} dv du \\ \therefore I_1 &= \frac{1}{2} \int_0^\infty \int_0^\infty (v + u) \exp \{-f(u, v)\} dv du. \end{aligned}$$

Así,

$$\begin{aligned} A &= \frac{1}{k} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \frac{1}{2} \int_0^\infty \int_0^\infty (v + u) \exp \{-f(u, v)\} dv du \\ &= \frac{1}{k} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) I_1. \end{aligned} \quad (4.2)$$

ii) Si el máximo es $\frac{v}{z_1}$, se debe cumplir que $\frac{v}{z_1} > \frac{u}{z_2} \Rightarrow v > u \frac{z_1}{z_2}$, por lo que $V(z_1, z_2)$ es,

$$\begin{aligned} B &= \frac{1}{k} \int_0^\infty \int_0^\infty \frac{v}{z_1} \exp \{-f(u, v)\} dv du \\ &= \frac{1}{kz_1} \int_0^\infty \int_0^\infty v \exp \{-f(u, v)\} dv du \\ &= \frac{1}{kz_1} \int_0^\infty \int_0^\infty 1_{\{v > u \frac{z_1}{z_2}\}} \exp \{-f(u, v)\} dv du. \end{aligned}$$

4.2. El modelo espacio-temporal

Si se define

$$I_2 = \int_0^\infty \int_0^\infty 1_{\{v > u \frac{z_1}{z_2}\}} \exp\{-f(u, v)\} dv du,$$

entonces,

$$B = \frac{1}{kz_1} I_2. \quad (4.3)$$

iii) Análogamente a ii), si el máximo es $\frac{u}{z_2} \Rightarrow u > v \frac{z_2}{z_1}$, y $V(z_1, z_2)$ resulta,

$$C = \frac{1}{kz_2} I_3, \quad (4.4)$$

con

$$I_3 = \int_0^\infty \int_0^\infty 1_{\{u > v \frac{z_2}{z_1}\}} \exp\{-f(u, v)\} dv du.$$

De esta forma con (4.2), (4.3) y (4.4),

$$V(z_1, z_2) = \frac{1}{k} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) I_1 + \frac{1}{kz_1} I_2 + \frac{1}{kz_2} I_3. \quad (4.5)$$

Ahora se procede a resolver cada una de las integrales anteriores. Para I_1 usemos la transformación

$$u = x + y$$

$$v = x - y.$$

El jacobiano es 2, $u + v = 2x$ y sustituyendo en $f(u, v)$,

$$f(x, y) = \frac{2[x^2(1 - \rho_{hu}) + y^2(1 + \rho_{hu})]}{2(1 - \rho_{hu}^2)} = k_1 x^2 + k_2 y^2$$

siendo

$$k_1 = \frac{1 - \rho_{hu}}{1 - \rho_{hu}^2} \quad \text{y} \quad k_2 = \frac{1 + \rho_{hu}}{1 - \rho_{hu}^2},$$

de esta forma,

$$\begin{aligned} I_1 &= \frac{1}{2} \int_0^\infty \int_{|y|}^\infty 2x \exp(-k_1 x^2 - k_2 y^2) 2dx dy \\ &= \int_0^\infty \left(\int_{|y|}^\infty 2x \exp(-k_1 x^2) dx \right) \exp(-k_2 y^2) dy, \end{aligned}$$

pero

$$\int_{|y|}^\infty 2x \exp(-k_1 x^2) dx = \frac{1}{k_1} \exp(-k_1 y^2),$$

así,

$$I_1 = \frac{1}{k_1} \int_0^\infty \exp(-k_1 y^2) \exp(-k_2 y^2) dy = \frac{1}{k_1} \int_0^\infty \exp\{-(k_1 + k_2) y^2\} dy.$$

4.2. El modelo espacio-temporal

Recuerde que $\int_{-\infty}^{\infty} \exp(-r^2) dr = \sqrt{\pi}$,

$$\therefore I_1 = (1 + \rho_{hu}) \frac{\sqrt{1 - \rho_{hu}^2}}{\sqrt{2}} \sqrt{\pi} = \frac{(1 + \rho_{hu})\sqrt{2\pi(1 - \rho_{hu}^2)}}{2} = \frac{(1 + \rho_{hu})k}{2}, \quad (4.6)$$

donde k esta dado en (4.1).

Para resolver I_2 e I_3 defina

$$g(z) = \int_0^{\infty} \int_0^{\infty} 1_{\{u > vz\}} \exp\{-f(u, v)\} dvdu,$$

con esto se observa que

$$I_2 = g(z_1 z_2^{-1}) \quad \text{e} \quad I_3 = g(z_2 z_1^{-1}),$$

lo anterior quiere decir que sólo basta calcular $g(z)$. Observe que

$$\begin{aligned} g(z) &= \int_0^{\infty} \int_{uz}^{\infty} v \exp\{-f(u, v)\} dvdu \\ &= \int_0^{\infty} \int_{uz}^{\infty} (v - u\rho_{hu}) \exp\{-f(u, v)\} dvdu - \rho_{hu} \int_0^{\infty} \int_0^{uz} u \exp\{-f(u, v)\} dvdu + \\ &\quad \rho_{hu} \int_0^{\infty} \int_0^{\infty} u \exp\{-f(u, v)\} dvdu. \end{aligned}$$

También defina,

$$h(z) = \int_0^{\infty} \int_{uz}^{\infty} (v - u\rho_{hu}) \exp\{-f(u, v)\} dvdu,$$

y observe que

$$\int_0^{\infty} \int_0^{uz} u \exp\{-f(u, v)\} dvdu = g(z^{-1})$$

y

$$\int_0^{\infty} \int_{-\infty}^{\infty} u \exp\{-f(u, v)\} dvdu = I_1,$$

así,

$$g(z) = h(z) - \rho_{hu}g(z^{-1}) + \rho_{hu}I_1.$$

Sustituyendo $g(z^{-1})$ en $g(z)$ y simplificando se obtiene,

$$g(z) = \frac{h(z) - \rho_{hu}h(z^{-1}) + \rho_{hu}(1 - \rho_{hu})I_1}{1 - \rho_{hu}^2}. \quad (4.7)$$

Ahora se resuelve $h(z)$ por sustitución simple, para ello sea $t = f(u, v)$, entonces

$$\frac{dt}{dv} = \frac{2v - 2u\rho_{hu}}{2(1 - \rho_{hu}^2)} = \frac{v - \rho_{hu}u}{(1 - \rho_{hu}^2)},$$

$$\therefore (1 - \rho_{hu}^2)dt = (v - u\rho_{hu})dv,$$

4.2. El modelo espacio-temporal

note que si $v = uz$ entonces $t = au^2$, con

$$a = \frac{z^2 - 2z\rho_{hu} + 1}{2(1 - \rho_{hu}^2)},$$

así,

$$\begin{aligned} h(z) &= \int_0^\infty \int_{uz}^\infty (v - u\rho_{hu}) \exp\{-f(u, v)\} dv du \\ &= \int_0^\infty \int_{au^2}^\infty (1 - \rho_{hu}^2) \exp(-t) dt du \\ &= (1 - \rho_{hu}^2) \int_0^\infty \exp(-au^2) du \\ &= (1 - \rho_{hu}^2) \frac{\sqrt{\pi}}{2\sqrt{a}}, \text{ sustityendo } a \\ &= \frac{(1 - \rho_{hu}^2) \sqrt{2\pi(1 - \rho_{hu}^2)}}{2\sqrt{z^2 - 2z\rho_{hu} + 1}} \\ &= \frac{(1 - \rho_{hu}^2)k}{2\sqrt{z^2 - 2z\rho_{hu} + 1}}. \end{aligned} \tag{4.8}$$

Si se sustituye z^{-1} en (4.8) el radical del denominador se puede escribir como

$$\sqrt{z^2 - 2z\rho_{hu} + 1} = \frac{1}{z} \sqrt{z^2 - 2z\rho_{hu} + 1}.$$

Regresando a (4.7), sustituyendo (4.8) y (4.6)

$$\begin{aligned} g(z) &= \frac{1}{1 - \rho_{hu}^2} \left[\frac{(1 - \rho_{hu}^2)k}{2\sqrt{z^2 - 2z\rho_{hu} + 1}} - \frac{z\rho_{hu}(1 - \rho_{hu}^2)k}{\sqrt{z^2 - 2z\rho_{hu} + 1}} + \frac{\rho_{hu}(1 - \rho_{hu})(1 + \rho_{hu})k}{2} \right] \\ &= \frac{k}{2\sqrt{z^2 - 2z\rho_{hu} + 1}} - \frac{z\rho_{hu}k}{2\sqrt{z^2 - 2z\rho_{hu} + 1}} + \frac{\rho_{hu}k}{2} \\ &= \frac{k}{2} \left(\frac{1}{\sqrt{z^2 - 2z\rho_{hu} + 1}} - \frac{z\rho_{hu}}{\sqrt{z^2 - 2z\rho_{hu} + 1}} + \rho_{hu} \right). \end{aligned}$$

Si $z = z_1/z_2$,

$$\sqrt{z^2 - 2z\rho_{hu} + 1} = \frac{1}{z_2} \sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}.$$

Así,

$$\begin{aligned} g\left(\frac{z_1}{z_2}\right) &= \frac{k}{2} \left(\frac{z_2}{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} - \frac{z_1z_2\rho_{hu}}{z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \rho_{hu} \right) \\ &= \frac{k}{2} \left[\frac{(z_2 - z_1\rho_{hu})}{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \rho_{hu} \right]. \end{aligned}$$

4.2. El modelo espacio-temporal

Similarmente, si $z = z_2/z_1$ se tiene,

$$g\left(\frac{z_2}{z_1}\right) = \frac{k}{2} \left[\frac{(z_1 - z_2\rho_{hu})}{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \rho_{hu} \right].$$

Lo anterior ya que el segundo y tercer sumando de $V(z_1, z_2)$ son,

$$\begin{aligned} \frac{1}{kz_1}I_2 + \frac{1}{kz_2}I_3 &= \frac{1}{kz_1}g\left(\frac{z_1}{z_2}\right) + \frac{1}{kz_2}g\left(\frac{z_2}{z_1}\right) \\ &= \frac{1}{kz_1} \frac{k}{2} \left[\frac{(z_2 - z_1\rho_{hu})}{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \rho_{hu} \right] + \\ &\quad \frac{1}{kz_2} \frac{k}{2} \left[\frac{(z_1 - z_2\rho_{hu})}{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \rho_{hu} \right] \\ &= \frac{1}{2} \left[\frac{(z_2 - z_1\rho_{hu})}{z_1\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \frac{\rho_{hu}}{z_1} \right] + \\ &\quad \frac{1}{2} \left[\frac{(z_1 - z_2\rho_{hu})}{z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \frac{\rho_{hu}}{z_2} \right] \\ 2 \left(\frac{1}{kz_1}I_2 + \frac{1}{kz_2}I_3 \right) &= \frac{(z_2 - z_1\rho_{hu}) + \rho_{hu}\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}}{z_1\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} + \\ &\quad \frac{(z_1 - z_2\rho_{hu}) + \rho_{hu}\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}}{z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} \\ &= \frac{z_1^2 + z_2^2 - 2z_1z_2\rho_{hu} + \rho_{hu}(z_1 + z_2)\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}}{z_1z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} \\ &= \frac{(z_1^2 + z_2^2) - 2z_1z_2(\rho_{hu} + 1) + \rho_{hu}(z_1 + z_2)\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}}{z_1z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} \\ &= \frac{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})} \left[\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})} + \rho_{hu}(z_1 + z_2) \right]}{z_1z_2\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}} \\ &= \frac{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})} + \rho_{hu}(z_1 + z_2)}{z_1z_2} \\ &= \frac{\sqrt{(z_1 + z_2)^2 - 2z_1z_2(1 + \rho_{hu})}}{z_1z_2} + \rho_{hu} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \\ &= \frac{(z_1 + z_2)\sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} + \rho_{hu}(z_1 + z_2)}{z_1z_2} \\ &= \frac{(z_1 + z_2)}{z_1z_2} \left[\sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} + \rho_{hu} \right] \end{aligned}$$

4.2. El modelo espacio-temporal

$$\therefore \frac{1}{kz_1}I_2 + \frac{1}{kz_2}I_3 = \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[\sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} + \rho_{hu} \right]. \quad (4.9)$$

Finalmente sustituyendo en 4.5,

$$\begin{aligned} V(z_1, z_2) &= \frac{1}{k} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) I_1 + \frac{1}{kz_1}I_2 + \frac{1}{kz_2}I_3 \\ &= \frac{1}{k} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \frac{(1 - \rho_{hu})k}{2} + \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[\sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} + \rho_{hu} \right] \\ &= \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[1 - \rho_{hu} \sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} + \rho_{hu} \right] \\ &= \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[1 + \sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} \right]. \end{aligned}$$

$$\begin{aligned} \therefore F(z_1, z_2) &= P \{ Z(\mathbf{s}_1, t_1) \leq z_1, Z(\mathbf{s}_2, t_2) \leq z_2 \} \\ &= \exp \left\{ -\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[1 + \sqrt{1 - \frac{2z_1z_2(1 + \rho_{hu})}{(z_1 + z_2)^2}} \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left[1 + \sqrt{1 - \frac{2z_1z_2(1 + \rho(\mathbf{h}, u))}{(z_1 + z_2)^2}} \right] \right\}, \quad z_1, z_2 > 0. \end{aligned}$$

■

Note que $F(z_1, z_2)$ es una FDA bien definida ya que si $z_1, z_2 \rightarrow \infty$, entonces $F(z_1, z_2) \rightarrow 1$ y si $z_1, z_2 \rightarrow 0$, entonces $F(z_1, z_2) \rightarrow 0$, así $0 \leq F(z_1, z_2) \leq 1$. Además, si $z_1 \rightarrow \infty$ y $z_2 = z$, $F(z) = \exp(-1/z)$, $z > 0$ que es la distribución Fréchet unitaria, esto coincide con el inciso b) del Teorema 4.1.

Ahora bien, si se define

$$a(z_1, z_2) = -\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right)$$

y

$$b(z_1, z_2) = 1 + \sqrt{1 - \frac{2z_1z_2(1 + \rho(\mathbf{h}, u))}{(z_1 + z_2)^2}},$$

entonces

$$F(z_1, z_2) = \exp \{ a(z_1, z_2) b(z_1, z_2) \}, \quad z_1, z_2 > 0,$$

así la función de densidad bivariada toma la forma,

$$\begin{aligned} f(z_1, z_2) &= \exp \{ a(z_1, z_2) b(z_1, z_2) \} \times \\ &\quad \left[\frac{\partial^2}{\partial z_1 \partial z_2} \{ a(z_1, z_2) b(z_1, z_2) \} + \frac{\partial}{\partial z_1} \{ a(z_1, z_2) b(z_1, z_2) \} + \frac{\partial}{\partial z_2} \{ a(z_1, z_2) b(z_1, z_2) \} \right]. \end{aligned}$$

4.3. Inferencia para el modelo propuesto

Suponga que el proceso espacio–temporal de interés $Z(\mathbf{s}, t) \in \mathcal{S} \times \mathbb{N}$ ha sido muestreado en S estaciones en T tiempos, resultando $N = ST$ datos en el espacio–tiempo. El estudio de los extremos de $Z(\mathbf{s}, t)$ se realiza suponiendo que el proceso max-stable es un modelo razonable para máximos por bloques, sin embargo, la inferencia clásica no es posible ya que la función de verosimilitud completa es intratable para esos modelos, en contraste se puede usar la verosimilitud a pares para realizar dicha inferencia, para ello, suponga que el proceso $Z(\mathbf{s}, t)$ se puede dividir en n bloques de máximos $M_i(\mathbf{s}), i = 1, \dots, n$. Sea $m_{\mathbf{s};i}$ el i -ésimo bloque registrado en la \mathbf{s} -ésima estación. Suponiendo que el proceso max-stable con función de densidad bivariada $f(z_1, z_2; \boldsymbol{\psi})$, donde $\boldsymbol{\psi}$ es el vector de parámetros a estimar, proporciona un ajuste razonable para el proceso espacio–temporal de máximos, se puede hacer inferencia basado en la log-verosimilitud a pares, dada por,

$$\ell_p(\boldsymbol{\psi}) = \sum_{i=1}^n \sum_{s_1=1}^{S-1} \sum_{s_2=s_1+1}^S \log f(m_{s_1;i}, m_{s_2;i}; \boldsymbol{\psi}). \quad (4.10)$$

Esta metodología ha sido usada satisfactoriamente en el enfoque espacial por [Padoan *et al.* \(2010\)](#) en el estudio de precipitaciones extremas en USA, [Davison y Gholamrezaee \(2011\)](#) usan esta aproximación para ajustar un proceso max-stable espacial para máximos de temperatura anual en Suiza.

De aquí,

$$\hat{\boldsymbol{\psi}}_p = \arg \max_{\boldsymbol{\psi} \in \Psi} \ell_p(\boldsymbol{\psi}).$$

Los estimadores de máxima verosimilitud compuesta, heredan propiedades de los estimadores de máxima verosimilitud clásica (los que se obtienen usando la verosimilitud completa). Bajo condiciones de regularidad, suponemos que son consistentes y asintóticamente normales, así,

$$\hat{\boldsymbol{\psi}}_p \sim N_q\{\boldsymbol{\psi}_0, V_p(\hat{\boldsymbol{\psi}}_p)\},$$

cuando $n \rightarrow \infty$, donde $V_p(\hat{\boldsymbol{\psi}}_p)$ es la varianza estimada.

Capítulo 5

Conclusiones

El trabajo de investigación que se presenta en esta tesis se puede dividir en dos partes. La primera de ellas, en el estudio de procesos max-stable y de modelos jerárquicos bayesianos para eventos extremos espaciales, los cuales se aplicaron a datos de contaminación por monóxido de carbono en la Zona Metropolitana de la Ciudad de México, y la segunda en el desarrollo de un nuevo modelo para eventos extremos en el espacio-tiempo.

En la aplicación de los procesos max-stable, se ajustaron los modelos de Smith ([Smith, 1990b](#)) y de Schlather ([Schlather, 2002](#)) a los datos del año 2008 a 2013 y con base en el criterio de información de Takeuchi se elige al modelo de Schlather como aquel que mejor se ajusta a los datos, posteriormente se proponen tendencias espaciales a los parámetros de la distribución de valores extremos generalizada y ajustando el modelo de Schlather al año 2012 se obtuvieron mapas predictivos para este contaminante, notando que en la región sur de la ciudad se espera tengan niveles más altos de contaminación. Mientras que en el enfoque bayesiano, se propone un modelo jerárquico espacial agregando tendencias espaciales a los parámetros de la distribución de valores extremos, en este caso se generaron mapas predictivos para cada parámetro en los que se puede observar el comportamiento espacial de cada uno de ellos en forma marginal, lo que no es posible solamente con tendencias deterministas.

La principal aportación de esta tesis, se centra en la generalización del modelo de Schlather al espacio-tiempo, aquí se propone introducir un proceso espacio-temporal y se demuestra que el proceso construido es max-stable, posteriormente se encuentra la función de distribución acumulada bivariada para el modelo propuesto, con ésta se propone usar la metodología de verosimilitud a pares para realizar inferencia para el modelo.

Referencias

- Apputhurai, P. y Stephenson, A. G. (2013). Spatiotemporal hierarchical modelling of extreme precipitation in Western Australia using anisotropic Gaussian random fields. *Environmental and ecological statistics*, 20, 4, 667–677.
- Balkema, A. A. y Resnick, S. I. (1977). Max-Infinite Divisibility. *Journal of Applied Probability*, 14, 2, 309–319. ISSN 00219002.
- Ballani, F. y Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika*, 98, 3, 633–645.
- Banerjee, S., Carlin, B. P. y Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press, second edition edición.
- Beirlant, J., Goegebeur, Y., Segers, J. y Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- Blanchet, J. y Davison, A. C. (2011). Spatial modeling of extreme snow depth. *The Annals of Applied Statistics*, 5, 3, 1699–1725. ISSN 19326157.
- Blanchet, J. y Davison, A. C. (2012). Statistical modelling of ground temperature in mountain permafrost. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. ISSN 1364-5021.
- Bochner, S. (1955). Harmonic analysis and the theory of probability. University of California Press.
- Brown, B. M. y Resnick, S. I. (1977). Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14, 4, 732–739. ISSN 00219002.
- Casson, E. y Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1, 4, 449–468.
- Chavez-Demoulin, V. y Davison, A. C. (2005). Generalized Additive Modelling of Sample Extremes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 54, 1, 207–222. ISSN 00359254, 14679876.
- Chavez-Demoulin, V. y Davison, A. C. (2012). Modelling Time Series Extremes. *REVSTAT*, 10, 1, 109–133.
- Coles, S., Bawa, J., Trenner, L. y Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, tomo 208. Springer.

Referencias

- Coles, S. y Casson, E. (1998). Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20, 3, 283–296.
- Coles, S. G. y Tawn, J. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 2, 377–392. ISSN 00359246.
- Coles, S. G. y Tawn, J. A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, 1, 1–48. ISSN 00359254, 14679876.
- Coles, S. G. y Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 329–347.
- Cooley, D., Davis, R. A. y Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101, 9, 2103 – 2117. ISSN 0047-259X.
- Cooley, D., Naveau, P. y Poncet, P. (2006). *Dependence in Probability and Statistics*. Springer New York, New York, NY. ISBN 978-0-387-36062-1.
- Cooley, D., Nychka, D. y Naveau, P. (2007). Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association*, 102, 479, 824–840. ISSN 01621459.
- Cox, D. R. y Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91, 3, 729–737.
- Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N. y Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.
- Daniel Cooley, S. R. S. (2010). Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 3, 381–402. ISSN 10857117.
- Davis, R. A., Kluppelberg, C. y Steinkohl, C. (2013a). Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society*, 42, 3, 399 – 414. ISSN 1226-3192.
- Davis, R. A., Kluppelberg, C. y Steinkohl, C. (2013b). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 5, 791–819. ISSN 1467-9868.
- Davison, A. C. (2003). *Statistical models*, tomo 11. Cambridge University Press.
- Davison, A. C. y Gholamrezaee, M. M. (2011). Geostatistics of extremes. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 468, 2138, 581–608. ISSN 1364-5021.
- Davison, A. C., Huser, R. y Thibaud, E. (2013). Geostatistics of Dependent and Asymptotically Independent Extremes. *Mathematical Geosciences*, 45, 5, 511–529. ISSN 1874-8953.

Referencias

- Davison, A. C., Padoan, S. A. y Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, 27, 2, 161–186. ISSN 08834237.
- de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Mathematisch Centrum.
- de Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12, 4, 1194–1204. ISSN 00911798.
- de Haan, L. y Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer Science & Business Media. ISBN 9780387239460.
- de Haan, T. L. y Pereira, T. (2006). Spatial Extremes: Models for the Stationary Case. *The Annals of Statistics*, 34, 1, 146–168. ISSN 00905364.
- Diggle, P. y Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics.
- Embrechts, P., Klüppelberg, C. y Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, tomo 33. Springer Science & Business Media.
- Ferrez, J., Davison, A. y Rebetez, M. (2011). Extreme temperature analysis under forest cover compared to an open field. *Agricultural and Forest Meteorology*, 151, 7, 992 – 1001. ISSN 0168-1923.
- Fisher, R. A. y Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. En *Mathematical Proceedings of the Cambridge Philosophical Society*, tomo 24, 180–190. Cambridge Univ Press.
- Gaetan, C. y Grigoletto, M. (2007). A hierarchical model for the analysis of spatial rainfall extremes. *Journal of agricultural, biological, and environmental statistics*, 12, 4, 434–449.
- Ghosh, S. y Mallick, B. K. (2011). A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22, 2, 192–204.
- Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44, 3, 423–453. ISSN 0003486X.
- Gneiting, T., Genton, M. G. y Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability*, 107, 151.
- Gumbel, E. J. (1961). Bivariate Logistic Distributions. *Journal of the American Statistical Association*, 56, 294, 335–349.
- Hosking, J. R. M., Wallis, J. R. y Wood, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, 27, 3, 251–261.
- Huser, R. y Davison, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*.
- Huser, R. y Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 2, 439–461. ISSN 1467-9868.

Referencias

- Hüsler, R. y Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics Probability Letters*, 7, 4, 283 – 286. ISSN 0167-7152.
- Kabluchko, Z. y Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Processes and their Applications*, 120, 3, 281 – 295. ISSN 0304-4149.
- Katz, R. W., Parlange, M. B. y Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25, 8, 1287–1304.
- Kunihama, T., Omori, Y. y Zhang, Z. (2012). Efficient estimation and particle filter for max-stable processes. *Journal of Time Series Analysis*, 33, 1, 61–80.
- Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, 65, 2, 291–306.
- Nakajima, J., Kunihama, T., Omori, Y. y Frühwirth-Schnatter, S. (2012). Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics & Data Analysis*, 56, 11, 3241–3259.
- Naveau, P., Guillou, A., Cooley, D. y Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96, 1, 1–17. ISSN 00063444.
- Opitz, T. (2013). Extremal processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122, 409 – 413. ISSN 0047-259X.
- Padoan, S. A., Ribatet, M. y Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, 105, 489, 263–277. ISSN 01621459.
- Pickands, J. (1981). Multivariate extreme value distributions. En *Proceedings 43rd Session International Statistical Institute*, tomo 2, 859–878.
- R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.
- Resnick, S. I. (2013). *Extreme values, regular variation and point processes*. New York: Springer.
- Ribatet, M. (2015). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-2.
- Ribatet, M. y Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 154, 1, 138–150.
- Sang, H. y Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16, 3, 407–426. ISSN 1573-3009.
- Sang, H. y Gelfand, A. E. (2010). Continuous Spatial Process Models for Spatial Extreme Values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 1, 49–65. ISSN 10857117.

Referencias

- Schlather, M. (2002). Models for Stationary Max-Stable Random Fields. *Extremes*, 5, 1, 33–44. ISSN 1572-915X.
- Schlather, M. y Tawn, M. (2003). A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference. *Biometrika*, 90, 1, 139–156. ISSN 00063444.
- Segers, J. (2012). Max-stable models for multivariate extremes. *RevStat Statistical Journal*, 10, 1, 61–82.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72, 1, 67–90.
- Smith, R. L. (1990a). Extreme value theory. *Handbook of applicable mathematics*, 7, 437–471.
- Smith, R. L. (1990b). Max-stable processes and spatial extremes. *Unpublished manuscript*.
- Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.
- Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extremes events. *Australian & New Zealand Journal of Statistics*, 51, 1, 77–88. ISSN 1467-842X.
- Tawn, J. A. (1988). Bivariate Extreme Value Theory: Models and Estimation. *Biometrika*, 75, 3, 397–415. ISSN 00063444.
- Tawn, J. A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika*, 77, 2, 245–253. ISSN 00063444.
- Thibaud, E., Mutzner, R. y Davison, A. (2013). Threshold modeling of extreme spatial rainfall. *Water resources research*, 49, 8, 4633–4644.
- Turkman, K. F., Turkman, M. A. y Pereira, J. (2010). Asymptotic models and inference for extremes of spatio-temporal data. *Extremes*, 13, 4, 375–397.
- Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, 92, 1, 1–28. ISSN 1863-818X.
- Varin, C., Reid, N. y Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21, 5–42.
- von Mises, R. (1954). La distribution de la plus grande de n valeurs, in selected papers, volume II, vol. 44 of American Mathematical Society. *Providence, RI*.
- Wadsworth, J. L. y Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101, 1, 1–15.
- Wang, Y. y So, M. K. (2016). A Bayesian hierarchical model for spatial extremes with multiple durations. *Computational Statistics & Data Analysis*, 95, 39–56.
- Westra, S. y Sisson, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406, 1-2, 119 – 128. ISSN 0022-1694.
- Yan, H. y Moradkhani, H. (2015). A regional Bayesian hierarchical model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, 29, 3, 1019–1036.

Anexo A: Códigos en R

En esta anexo se muestran algunos de los programas que se usaron para el análisis de los datos de monóxido de carbono presentado en la [sección 2.6](#) del [Capítulo ??](#) y [sección 3.3](#) del [Capítulo ??](#). En el siguiente se extraen los datos y se selecciona el contaminante de interés para los años de 2008 a 2013.

```
1 datos <- read.csv("/Users/jcjim/Documents/Doctorado/Datos/rama.csv",header=
  TRUE)
2 estaciones <- read.csv("/Users/jcjim/Documents/Doctorado/Datos/estacion.csv",
  header=TRUE)
3
4 a2008 <- rama[27164065:28666128,]
5 a2009 <- rama[28666129:30172848,]
6 a2010 <- rama[30172849:31670808,]
7 a2011 <- rama[31670809:33168768,]
8 a2012 <- rama[33168769:34943136,]
9 a2013 <- rama[34943137:36756456,]
10
11 library(extRemes)
12 #Selecciona el contaminante CO
13 a2005.CO <- subset(a2005,CONTAMINANTE=="CO")
14 a2006.CO <- subset(a2006,CONTAMINANTE=="CO")
15 a2007.CO <- subset(a2007,CONTAMINANTE=="CO")
16 a2008.CO <- subset(a2008,CONTAMINANTE=="CO")
17 a2009.CO <- subset(a2009,CONTAMINANTE=="CO")
18 a2010.CO <- subset(a2010,CONTAMINANTE=="CO")
19 a2011.CO <- subset(a2011,CONTAMINANTE=="CO")
20 a2012.CO <- subset(a2012,CONTAMINANTE=="CO")
21 a2013.CO <- subset(a2013,CONTAMINANTE=="CO")
```

En el siguiente se seleccionan las estaciones y se toman los bloques de máximos cada tres días, por ejemplo se muestra esto para el año 2013.

```
1 IZT2013.CO <- subset(a2013.CO, ESTACION == "IZT")
2 MER2013.CO <- subset(a2013.CO, ESTACION == "MER")
3 PED2013.CO <- subset(a2013.CO, ESTACION == "PED")
4 SAG2013.CO <- subset(a2013.CO, ESTACION == "SAG")
```

Referencias

```
5 SUR2013.CO <- subset(a2013.CO, ESTACION == "SUR")
6 TLA2013.CO <- subset(a2013.CO, ESTACION == "TLA")
7 TLI2013.CO <- subset(a2013.CO, ESTACION == "TLI")
8 UIZ2013.CO <- subset(a2013.CO, ESTACION == "UIZ")
9 VIF2013.CO <- subset(a2013.CO, ESTACION == "VIF")
10
11 IZT2013.CO <- subset(IZT2013.CO,CONCENTRACION != -99.9)
12 MER2013.CO <- subset(MER2013.CO,CONCENTRACION != -99.9)
13 PED2013.CO <- subset(PED2013.CO,CONCENTRACION != -99.9)
14 SAG2013.CO <- subset(SAG2013.CO,CONCENTRACION != -99.9)
15 SUR2013.CO <- subset(SUR2013.CO,CONCENTRACION != -99.9)
16 TLA2013.CO <- subset(TLA2013.CO,CONCENTRACION != -99.9)
17 TLI2013.CO <- subset(TLI2013.CO,CONCENTRACION != -99.9)
18 UIZ2013.CO <- subset(UIZ2013.CO,CONCENTRACION != -99.9)
19 VIF2013.CO <- subset(VIF2013.CO,CONCENTRACION != -99.9)
20
21 d1 <- dim(IZT2013.CO)[1]
22 d2 <- dim(MER2013.CO)[1]
23 d3 <- dim(PED2013.CO)[1]
24 d4 <- dim(SAG2013.CO)[1]
25 d5 <- dim(SUR2013.CO)[1]
26 d6 <- dim(TLA2013.CO)[1]
27 d7 <- dim(TLI2013.CO)[1]
28 d8 <- dim(UIZ2013.CO)[1]
29 d9 <- dim(VIF2013.CO)[1]
30
31 #Se toman los máximos por bloques semanales
32
33 IZT2013.CO.M <- blockmaxxer(IZT2013.CO,which="CONCENTRACION",blen=d1/K,span=K)
34 MER2013.CO.M <- blockmaxxer(MER2013.CO,which="CONCENTRACION",blen=d2/K,span=K)
35 PED2013.CO.M <- blockmaxxer(PED2013.CO,which="CONCENTRACION",blen=d3/K,span=K)
36 SAG2013.CO.M <- blockmaxxer(SAG2013.CO,which="CONCENTRACION",blen=d4/K,span=K)
37 SUR2013.CO.M <- blockmaxxer(SUR2013.CO,which="CONCENTRACION",blen=d5/K,span=K)
38 TLA2013.CO.M <- blockmaxxer(TLA2013.CO,which="CONCENTRACION",blen=d6/K,span=K)
39 TLI2013.CO.M <- blockmaxxer(TLI2013.CO,which="CONCENTRACION",blen=d7/K,span=K)
40 UIZ2013.CO.M <- blockmaxxer(UIZ2013.CO,which="CONCENTRACION",blen=d8/K,span=K)
41 VIF2013.CO.M <- blockmaxxer(VIF2013.CO,which="CONCENTRACION",blen=d9/K,span=K)
```

A continuación se agrupan los máximos para cada año y para todas las estaciones.

```
1 datos2007 <- data.frame(IZT=IZT2007.CO.M[,5],MER=MER2007.CO.M[5],PED=PED2007.
   CO.M[,5],SAG=SAG2007.CO.M[,5],SUR=SUR2007.CO.M[,5],TLA=TLA2007.CO.M[,5],TLI
   =TLI2007.CO.M[,5],UIZ=UIZ2007.CO.M[,5],VIF=VIF2007.CO.M[,5])
2
3 datos2008 <- data.frame(IZT=IZT2008.CO.M[,5],MER=MER2008.CO.M[5],PED=PED2008.
   CO.M[,5],SAG=SAG2008.CO.M[,5],SUR=SUR2008.CO.M[,5],TLA=TLA2008.CO.M[,5],TLI
   =TLI2008.CO.M[,5],UIZ=UIZ2008.CO.M[,5],VIF=VIF2008.CO.M[,5])
```

4

Referencias

```
5 datos2009 <- data.frame(IZT=IZT2009.CO.M[,5],MER=MER2009.CO.M[5],PED=PED2009.
  CO.M[,5],SAG=SAG2009.CO.M[,5],SUR=SUR2009.CO.M[,5],TLA=TLA2009.CO.M[,5],TLI
  =TLI2009.CO.M[,5],UIZ=UIZ2009.CO.M[,5],VIF=VIF2009.CO.M[,5])
6
7 datos2010 <- data.frame(IZT=IZT2010.CO.M[,5],MER=MER2010.CO.M[,5],PED=PED2010.
  CO.M[,5],SAG=SAG2010.CO.M[,5],SUR=SUR2010.CO.M[,5],TLA=TLA2010.CO.M[,5],TLI
  =TLI2010.CO.M[,5],UIZ=UIZ2010.CO.M[,5],VIF=VIF2010.CO.M[,5])
8
9 datos2011 <- data.frame(IZT=IZT2011.CO.M[,5],MER=MER2011.CO.M[,5],PED=PED2011.
  CO.M[,5],SAG=SAG2011.CO.M[,5],SUR=SUR2011.CO.M[,5],TLA=TLA2011.CO.M[,5],TLI
  =TLI2011.CO.M[,5],UIZ=UIZ2011.CO.M[,5],VIF=VIF2011.CO.M[,5])
10
11 datos2012 <- data.frame(IZT=IZT2012.CO.M[,5],MER=MER2012.CO.M[,5],PED=PED2012.
  CO.M[,5],SAG=SAG2012.CO.M[,5],SUR=SUR2012.CO.M[,5],TLA=TLA2012.CO.M[,5],TLI
  =TLI2012.CO.M[,5],UIZ=UIZ2012.CO.M[,5],VIF=VIF2012.CO.M[,5])
12
13 datos2013 <- data.frame(IZT=IZT2013.CO.M[,5],MER=MER2013.CO.M[,5],PED=PED2013.
  CO.M[,5],SAG=SAG2013.CO.M[,5],SUR=SUR2013.CO.M[,5],TLA=TLA2013.CO.M[,5],TLI
  =TLI2013.CO.M[,5],UIZ=UIZ2013.CO.M[,5],VIF=VIF2013.CO.M[,5])
14
15 datos2014 <- data.frame(IZT=IZT2014.CO.M[,2],MER=MER2014.CO.M[,2],PED=PED2014.
  CO.M[,2],SAG=SAG2014.CO.M[,2],SUR=SUR2014.CO.M[,2],TLA=TLA2014.CO.M[,2],TLI
  =TLI2014.CO.M[,2],UIZ=UIZ2014.CO.M[,2],VIF=VIF2014.CO.M[,2])
16
17 datos2015 <- data.frame(IZT=IZT2015.CO.M[,2],MER=MER2015.CO.M[,2],PED=PED2015.
  CO.M[,2],SAG=SAG2015.CO.M[,2],SUR=SUR2015.CO.M[,2],TLA=TLA2015.CO.M[,2],TLI
  =TLI2015.CO.M[,2],UIZ=UIZ2015.CO.M[,2],VIF=VIF2015.CO.M[,2])
```

Se procede a transformar los datos para que tengan distribución Fréchet unitaria y se ajustan los modelos de Smith y Schlather.

```
1 setwd("/Users/jcjim/Documents/Doctorado/AnalisisCO/")
2 library(SpatialExtremes)
3 library(extRemes)
4 n <- 9
5 M <- matrix(0,n,3)
6 M1 <- c()
7 for(i in 1:n){
8   fit <- fevd(datos2011[,i])
9   M1[i] <- is.fixedfevd(fit)
10  fit <- summary(fit)
11  M[i,] <- fit$par
12 }
13
14 loc <- M[,1]
15 scal <- M[,2]
16 form <- M[,3]
17
```

Referencias

```
18 #Transforma los datos a que tenga distribucion Frechet unitarias
19 for(i in 1:n)
20 datos2011[,i] <- gev2frech(datos2011[,i],loc[i],scal[i],form[i])
21 datos2011 <- as.matrix(datos2011)
22
23 #Se ajustar el modelo de Smith
24 fitt11S <- fitmaxstab(datos2011, coord, cov.mod="gauss",method = "BFGS")
25
26 #Se ajusta el modelo de Schlather,
27 fitt11.w <- fitmaxstab(datos2011, coord, cov.mod="whitmat",method = "BFGS")
28 fitt11.c <- fitmaxstab(datos2011, coord, cov.mod="cauchy",method ="SANN")
29 fitt11.p <- fitmaxstab(datos2011, coord, cov.mod="powexp")
30 fitt11.b <- fitmaxstab(datos2011, coord, cov.mod="bessel")
31 fitt11.cg <- fitmaxstab(datos2011, coord, cov.mod="caugen",method = "SANN")
32 #Se calcula el TIC para comparar los modelos que se ajustaron
33 c(TIC(fitt11.w),TIC(fitt11.c),TIC(fitt11.p),TIC(fitt11.b),TIC(fitt11.cg))
```

En el siguiente se definen las tendencias espaciales y se hacen los mapas predictivos que se presentaron en el [Capítulo2](#).

```
1 loc.form <- y ~ lon + lat
2 scale.form <- y ~ lon + lat
3 shape.form <- y ~ 1
4
5 M1 <- fitmaxstab(datos2012, coord, "powexp", loc.form,scale.form,shape.form)
6 M2 <- fitmaxstab(datos2012, coord, "powexp", loc.form, scale.form, shape.form,
   smooth=2)
7
8 x <- seq(-99.45,-98.6, length = 100)
9 y <- seq(19,19.82, length = 100)
10
11 map(M2, x,y, param = "loc",xlab="lon",ylab="lat",main=expression(mu))
12 points(corde1,pch=".")
13 lines(corde,col=4,lwd=3)
14 points(cordmex[,2:3],pch=".")
15 points(coord,pch=17, col="red")
16
17 map(M2, x, y, param = "scale",xlab="lon",ylab="lat",main=expression(sigma))
18 points(corde1,pch=".")
19 lines(corde,col=4,lwd=3)
20 points(cordmex[,2:3],pch=".")
21 points(coord,pch=17, col="red")
22
23 map(M2, x, y, param = "quant", ret.per = 5,xlab="lon",ylab="lat",main="Nivel
   de retorno")
24 points(corde1,pch=".")
25 lines(corde,col=4,lwd=3)
26 points(cordmex[,2:3],pch=".")
```

Referencias

```
27 points(coord,pch=17, col="red")
28
29 M0 <- fitspatgev(datos2012, coord, loc.form, scale.form, shape.form)
30 x <- seq(-99.45,-98.6, length = 100)
31 y <- seq(19,19.82, length = 100)
32 grid <- expand.grid(x, y); colnames(grid) <- c("lon", "lat")
33
34 ans <- predict(M0, newdata=grid, ret.per= 2)$Q2 #Para el nivel de retorno a 2
    años
35 ans1 <- predict(M0, newdata=grid, ret.per= 4)$Q4 #Para el nivel de retorno a 4
    años
36 ans2 <- predict(M0, newdata=grid, ret.per= 5)$Q5 #Para el nivel de retorno a 5
    años
37
38 plot(x,y,"n",xlab="lon", ylab="lat") #~
39 image(x , y, matrix(ans, 100), add = TRUE, col = cm.colors(64),asp=1)
40 contour(x , y , matrix(ans, 100), add = TRUE);
41 points(corde1,pch=".")
42 lines(corde,col=4,lwd=3)
43 points(cordmex[,2:3],pch=".")
44 points(coord,pch=17, col="red")
45
46 plot(x,y,"n",xlab="lon", ylab="lat")
47 image(x , y, matrix(ans1, 100), add = TRUE, col = cm.colors(64))
48 contour(x , y , matrix(ans1, 100), add = TRUE);
49 points(corde1,pch=".")
50 lines(corde,col=4,lwd=3)
51 points(cordmex[,2:3],pch=".")
52 points(coord,pch=17, col="red")
53
54 plot(x,y,"n",xlab="lon", ylab="lat")#, main="5 años")
55 image(x , y, matrix(ans2, 100), add = TRUE, col = cm.colors(64))
56 contour(x , y , matrix(ans2, 100), add = TRUE, xlab= "lon", ylab= "lat")
57 points(corde1,pch=".")
58 lines(corde,col=4,lwd=3)
59 points(cordmex[,2:3],pch=".")
60 points(coord,pch=17, col="red")
```

Finalmente se usa la rutina `latent` que se encuentra en el paquete `SpatialExtremes` para generar las cadenas que contienen las distribuciones a posterior para los parámetros del proceso latente que se describe en el [Capítulo 3](#).

```
1 hyper <- list()
2 hyper$betaMeans <- list(loc = rep(0, 4), scale = rep(0, 4), shape = 0)
3 hyper$betaIcov <- list(loc = diag(rep(1/1000, 4)), scale = diag(rep(1/1000, 4)
    ), shape = 1/10)
4 hyper$sills <- list(loc = c(2, 10), scale = c(2,3), shape = c(2,1))
5 hyper$ranges <- list(loc = c(4,2), scale = c(4,2), shape = c(4,2))
```

Referencias

```
6 hyper$smooths <- list(loc = c(1,1), scale = c(1,1), shape = c(1,1))
7 prop <- list(gev = c(10, 5, 3), ranges = c(6, 3, 2), smooths = rep(0, 3))
8 start <- list(sills = c(50, 10, 0.5), ranges = c(10, 10, 10), smooths = c(1,
  1, 1), beta = list(loc = c(100, 0, 0, 0), scale = c(50, 0, 0, 0), shape =
  0.001))
9
10 cadena <- latent(datos8.15, coord, "powexp", loc.form, scale.form, shape.form,
  hyper = hyper, prop = prop, start = start, n = 10000, burn.in = 5000, thin
  = 10)
```

Para hacer los mapas se modifíco el código del paquete, esto para poder hacer dentro del mapa nuestra región de estudio, el código modificado se muestra a continuación.

```
1 setwd("/Users/jcjim/Documents/Doctorado/AnálisisCO/2012")
2
3 dyn.load("SpatialExtremes.so")
4
5 cuan <- function(p, loc = 1, scale = 1, shape = 1, lower.tail = TRUE){
6   if((min(p, na.rm = TRUE) <= 0) || (max(p, na.rm = TRUE) >= 1))
7     stop("'p' must contain probabilities in (0,1)")
8   if (min(scale) < 0)
9     warning("There are some invalid scale GEV parameters")
10  if (length(p) != 1)
11    stop("invalid p")
12  if (!lower.tail)
13    p <- 1 - p
14  n <- length(loc)
15  ans <- .C("gev", as.double(p), as.integer(n), as.double(loc),
16          as.double(scale), as.double(shape), quant = double(n),
17          PACKAGE = "SpatialExtremes")$quant
18  return(ans)
19 }
20
21 dyn.load("SpatialExtremes.so")
22
23 mapa.latente.mod <- function (fitted, x, y, covariates = NULL, param = "quant
  ", ret.per = 100,
24   col = terrain.colors(64), plot.contour = TRUE, fun = mean,
25   level = 0.95, show.data = TRUE, control = list(nlines = 500),
26   ...)
27 {
28   if (!(param %in% c("loc", "scale", "shape", "quant")))
29     stop("'param' should be one of 'loc', 'scale', 'shape' or 'quant'")
30   if (ncol(fitted$coord) > 2)
31     stop("It's not possible to use this function when the coordinate space has
  a dimension > 2")
32   if (is.null(covariates) && !is.null(fitted$marg.cov))
33     stop("Your model seems to make use of covariate but you supplied none")
```

Referencias

```
34 if (!is.null(covariates)) {
35   if (missing(x) || missing(y))
36     stop("if 'covariates' is supplied 'x' and 'y' must be too")
37   if (!is.array(covariates))
38     stop("'covariates' must be an array - see the example")
39   covariates.names <- dimnames(covariates)[[3]]
40   model.names <- colnames(fitted$marg.cov)
41   if (is.null(model.names))
42     stop("Your fitted model doesn't seem to make use of covariates")
43   if (!all(model.names %in% covariates.names))
44     stop("Some required covariates are missing. Please check")
45   dim.cov <- dim(covariates)
46   if ((dim.cov[1] != length(x)) || (dim.cov[2] != length(y)))
47     stop("'covariates' doesn't match with 'x' and 'y'")
48 }
49 else covariates.names <- NULL
50 if (missing(x)) {
51   x.range <- range(fitted$coord[, 1])
52   x <- seq(x.range[1], x.range[2], length = 100)
53 }
54 if (missing(y)) {
55   y.range <- range(fitted$coord[, 2])
56   y <- seq(y.range[1], y.range[2], length = 100)
57 }
58 n.x <- length(x)
59 n.y <- length(y)
60 n.chain <- nrow(fitted$chain.loc)
61 n.loccoeff <- ncol(fitted$loc.dsgn.mat)
62 n.scalecoeff <- ncol(fitted$scale.dsgn.mat)
63 n.shapecoeff <- ncol(fitted$shape.dsgn.mat)
64 ans <- array(NA, c(n.x, n.y, n.chain))
65 new.data <- matrix(NA, nrow = n.x * n.y, ncol = 2 + length(covariates.names)
66 )
67 for (i in 1:n.x) new.data[(n.y * (i - 1) + 1):(n.y * i),] <- cbind(x[i], y,
68   covariates[i, , ])
69 colnames(new.data) <- c(colnames(fitted$coord), covariates.names)
70 loc.dsgn.mat <- modeldef(new.data, fitted$loc.form)$dsgn.mat
71 scale.dsgn.mat <- modeldef(new.data, fitted$scale.form)$dsgn.mat
72 shape.dsgn.mat <- modeldef(new.data, fitted$shape.form)$dsgn.mat
73 if (param == "loc") {
74   for (i in 1:n.chain) {
75     loc <- matrix(loc.dsgn.mat %*% fitted$chain.loc[i,1:n.loccoeff], n.x, n
76     .y, byrow = TRUE)
77     res <- as.numeric(fitted$chain.loc[i, -(1:(n.loccoeff + 3))]) -
78     fitted$loc.dsgn.mat %*% fitted$chain.loc[i,1:n.loccoeff])
79     ans[, , i] <- loc + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod =
80     fitted$cov.mod[1], sill = fitted$chain.loc[i, "sill"], range = fitted$chain
```

Referencias

```
.loc[i, "range"], smooth = fitted$chain.loc[i, "smooth"], grid = TRUE,
control = control)$cond.sim
76   }
77 }
78 else if (param == "scale") {
79   for (i in 1:n.chain) {
80     scale <- matrix(scale.dsgn.mat %*% fitted$chain.scale[i,1:n.scalecoeff
81   ], n.x, n.y, byrow = TRUE)
82     res <- as.numeric(fitted$chain.scale[i, -(1:(n.scalecoeff + 3))] -
83   fitted$scale.dsgn.mat %*% fitted$chain.scale[i,1:n.scalecoeff])
84     n.trials <- 1
85     flag <- FALSE
86     while (!flag) {
87       scale <- scale + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod =
88   fitted$cov.mod[2], sill = fitted$chain.scale[i, "sill"], range =
89   fitted$chain.scale[i, "range"], smooth = fitted$chain.shape[i, "smooth"],
90   grid = TRUE, control = control)$cond.sim
91     flag <- all(scale > 0)
92     n.trials <- n.trials + 1
93     if (n.trials >= 100) {
94       ans[, , i] <- NA
95       break
96     }
97   }
98   ans[, , i] <- scale
99 }
100 }
101 else if (param == "shape") {
102   for (i in 1:n.chain) {
103     shape <- matrix(shape.dsgn.mat %*% fitted$chain.shape[i,1:n.shapecoeff
104   ], n.x, n.y, byrow = TRUE)
105     res <- as.numeric(fitted$chain.shape[i, -(1:(n.shapecoeff + 3))] -
106   fitted$shape.dsgn.mat %*% fitted$chain.shape[i,1:n.shapecoeff])
107     ans[, , i] <- shape + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod
108   = fitted$cov.mod[1], sill = fitted$chain.shape[i, "sill"], range =
109   fitted$chain.shape[i, "range"], smooth = fitted$chain.shape[i, "smooth"],
110   grid = TRUE, control = control)$cond.sim
111   }
112 }
113 else {
114   for (i in 1:n.chain) {
115     loc <- matrix(loc.dsgn.mat %*% fitted$chain.loc[i,1:n.loccoeff], n.x, n
116   .y, byrow = TRUE)
117     res <- as.numeric(fitted$chain.loc[i, -(1:(n.loccoeff + 3))] -
118   fitted$loc.dsgn.mat %*% fitted$chain.loc[i,1:n.loccoeff])
119     loc <- loc + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod =
120   fitted$cov.mod[1], sill = fitted$chain.loc[i, "sill"], range = fitted$chain.
```

Referencias

```
loc[i, "range"], smooth = fitted$chain.loc[i, "smooth"], grid = TRUE,
control = control)$cond.sim
108   flag <- FALSE
109   scale <- matrix(scale.dsgn.mat %*% fitted$chain.scale[i,1:n.scalecoeff
], n.x, n.y, byrow = TRUE)
110   res <- as.numeric(fitted$chain.scale[i, -(1:(n.scalecoeff + 3))] -
fitted$scale.dsgn.mat %*% fitted$chain.scale[i,1:n.scalecoeff])
111   n.trials <- 1
112   while (!flag) {
113     scale <- scale + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod =
fitted$cov.mod[2], sill = fitted$chain.scale[i, "sill"], range =
fitted$chain.scale[i, "range"], smooth = fitted$chain.shape[i, "smooth"],
grid = TRUE, control = control)$cond.sim
114     flag <- all(scale > 0)
115     n.trials <- n.trials + 1
116     if (n.trials >= 100)
117       break
118   }
119   if (n.trials < 100) {
120     shape <- matrix(shape.dsgn.mat %*% fitted$chain.shape[i,1:n.
shapecoeff], n.x, n.y, byrow = TRUE)
121     res <- as.numeric(fitted$chain.shape[i, -(1:(n.shapecoeff + 3))] -
fitted$shape.dsgn.mat %*% fitted$chain.shape[i,1:n.shapecoeff])
122     shape <- shape + condrgp(1, cbind(x, y), fitted$coord, res, cov.mod =
fitted$cov.mod[1], sill = fitted$chain.shape[i, "sill"], range =
fitted$chain.shape[i, "range"], smooth = fitted$chain.shape[i, "smooth"],
grid = TRUE, control = control)$cond.sim
123     ans[, , i] <- cuan(1 - 1/ret.per, loc, scale,shape)
124   }
125   else ans[, , i] <- NA
126 }
127 }
128 prob.low <- (1 - level)/2
129 prob.up <- 1 - prob.low
130 post.sum <- ci.low <- ci.up <- matrix(NA, n.x, n.y)
131 for (i in 1:n.x) {
132   post.sum[i, ] <- apply(ans[i, , ], 1, fun, na.rm = TRUE)
133   ci.low[i, ] <- apply(ans[i, , ], 1, quantile, prob.low, na.rm = TRUE)
134   ci.up[i, ] <- apply(ans[i, , ], 1, quantile, prob.up, na.rm = TRUE)
135 }
136 op <- par(no.readonly = TRUE)
137 layout(matrix(1:2, 1), widths = rep(c(1, 0.2), 1))
138 for (i in 1:3) {
139   mar <- c(4,4,1,0.5)
140   par(mar = mar)
141   dummy <- switch(as.character(i), '1' = ci.low, '2' = post.sum, '3' = ci.up
)
```

Referencias

```
142 breaks <- seq(min(dummy), max(dummy), length = length(col) + 1)
143 image(x, y, dummy, ..., col = col, breaks = breaks ,xlab="lon",ylab="lat",
      axes =TRUE)
144 if (plot.contour)
145     contour(x, y, dummy, add = TRUE)
146 if (show.data)
147     points(fitted$coord,pch=17, col="red")
148 points(cordmextodas[,2:3],pch=".") #Hace los puntos de los municipios
149 points(corde1[,2:3],col=4,pch=".") #Hace la grafica de DF
150 lines(corde[,2:3],col=4,lwd=2)
151 #Para los nombres de las estaciones para las gráficas anteriores
152 text(coord[1,1]+0.02,coord[1,2],"IZT")
153 text(coord[2,1]+0.02,coord[2,2],"MER")
154 text(coord[7,1]+0.02,coord[7,2],"TLI")
155 text(coord[9,1]+0.02,coord[9,2],"VIF")
156 text(coord[8,1]+0.02,coord[8,2],"UIZ")
157 text(coord[5,1]+0.02,coord[5,2],"SUR")
158 text(coord[3,1]+0.02,coord[3,2],"PED")
159 text(coord[4,1]+0.02,coord[4,2],"SAG")
160 text(coord[6,1]+0.01,coord[6,2]+0.02,"TLA")
161 mar <- c(4, 0.2, 1, 4)
162 par(las = 1, pty = "m", mar = mar)
163 plot.new()
164 plot.window(xlim = c(0, 1), ylim = range(breaks), xaxs = "i", yaxs = "i")
165 rect(0, breaks[-length(breaks)], 1, breaks[-1], col = col, border = NA)
166 axis(4, at = pretty(dummy))
167 box()
168 }
169 on.exit(par(op))
170 invisible(list(x = x, y = y, post.sum = post.sum, ci.low = ci.low, ci.up =
      ci.up))
171 }
```
