

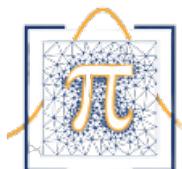
UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

Facultad de Ciencias

Maestría en Matemática



Modelado espacial bayesiano de niveles de retorno de precipitaciones extremas en Honduras



TESIS

QUE PARA OBTENER EL TÍTULO DE

MÁSTER EN MATEMÁTICA CON ORIENTACIÓN EN ESTADÍSTICA
MATEMÁTICA

PRESENTA

Sonia Marisol Méndez Ramos

Asesor: Cristian Cruz

HONDURAS

2022

DEDICATORIA

A mi abuelita, mi angelito y mayor inspiración. Te amo y te recuerdo siempre.

Agradecimientos

Quiero agradecer primeramente a Dios por siempre iluminar mi mente y guiar mi camino para poder concluir esta tesis, por darme fortaleza para nunca desistir y prestarme vida para poder alcanzar este logro académico.

A mis padres Pablo Méndez y Martha Ramos por ser un pilar fundamental en mi vida, son mi motor y mayor motivación. Gracias por su amor, esfuerzos y dedicación para ayudarme a llegar hasta aquí y convertirme en lo que soy. Gracias por siempre creer en mi y apoyarme de manera incondicional en todas las etapas y momentos de mi vida.

A mis hermanos Fernando y Raquel por su cariño, por siempre estar presentes acompañándome y apoyándome en cada etapa de mi vida. A la hermana que la vida me regaló (Eyvilin) por sus consejos y su amistad, por ser un ejemplo de perseverancia y siempre motivarme. A mis amigos y familia por su cariño y palabras de motivación para siempre continuar dando lo mejor y nunca rendirme, me acompañan siempre en todos mis sueños y logros.

De manera especial quiero agradecer a mi asesor PhD. Cristian Cruz por siempre motivarme a continuar, por el compromiso, dedicación y paciencia guiándome a lo largo de todo este tiempo. Durante el desarrollo de este trabajo logre comprender muchas cosas que sin su apoyo no hubiesen sido posibles. Gracias por creer en que lo lograría, por siempre estar presente atendiendo mis dudas y orientándome desde lo más simple hasta lo más complejo. Gracias infinitas por ser parte fundamental de este logro.

Índice general

Índice de figuras	v
Índice de tablas	vi
1. Introducción	1
2. Elementos de Inferencia	3
2.1. Conceptos previos	3
2.2. Elementos de Inferencia	5
2.3. Estimación en la teoría clásica	8
2.4. Inferencia bayesiana	9
2.5. Métodos de Simulación	11
2.5.1. Método Monte Carlo	11
2.5.2. Métodos de Monte Carlo por Cadenas de Markov (MCMC)	12
2.6. Comparación de Modelos	16
2.7. Modelos Jerárquicos	17
3. Teoría de Valores Extremos	19
3.1. Preliminares	19
3.2. Antecedentes	20
3.3. Teoría de valores extremos	23
3.4. Niveles de retorno	30
3.5. Teoría de valores extremos multivariante	31

4. Estadística espacial y precipitación	34
4.1. Estadística espacial	34
4.1.1. Generalidades de la Estadística Espacial	35
4.1.2. Geoestadística	36
4.2. Precipitación	43
4.2.1. Generalidades	43
4.2.2. Altitud y precipitación media como covariables en precipitación	44
4.2.3. Cambio climático y su relación con los extremos en precipitación	46
4.2.4. Precipitación en Honduras	46
5. Descripción de la metodología	50
5.1. Preliminares	50
5.2. Modelo jerárquico para excedencia de umbrales	51
5.3. Modelos de tasas de excedencias	57
5.4. MCMC en los modelos	58
5.4.1. Etapa 1	60
5.4.2. Etapa 2	63
5.5. Inferencia en los modelos	66
5.6. Descripción del criterio de selección de modelos	67
5.6.1. Estimador Gamma desplazado	67
6. Análisis de resultados	69
6.1. Selección de Modelos	69
6.2. Resultados: Etapa 1	70
6.3. Resultados: Etapa 2	74
6.4. Resultados del modelo completo	77
6.4.1. Niveles de retorno a 5 años	79
6.4.2. Niveles de retorno a 25 años	79
6.4.3. Niveles de retorno a 50 años	80
7. Conclusiones y trabajos futuros	83
7.1. Conclusiones	83
7.2. Trabajos futuros	85
A. Tablas	86

Índice de figuras

3.1. Valores de precipitación diaria para Honduras de 1972 a 2012	22
3.2. Valores máximos de precipitación diaria por año en Honduras de 1972 a 2012	23
3.3. Densidad de las distribuciones estándar de valores extremos	26
3.4. Densidades de GPD para diferentes valores de ξ y $\sigma = 1$	29
4.1. Parámetros del variograma experimental	39
4.2. Comparación gráfica de modelos teóricos de variograma	42
4.3. Valores de precipitación promedio por estación	45
4.4. Precipitación histórica anual en Honduras	48
4.5. Precipitación mensual acumulada de referencia	49
5.1. Nube variografica asociada a los valores de excedencias máximas	55
5.2. Ajuste del variograma experimental de los datos mediante el modelo exponencial	56
6.1. Valores de ξ modelo 3	72
6.2. Valores de ϕ para el modelo 3 obtenidos en la estación 1	72
6.3. Valores de ϕ para el modelo 3 obtenidos en la estación 30	73
6.4. Valores de ϕ para el modelo 3 obtenidos en la estación 59	73
6.5. Valores de ζ para el modelo A	75
6.6. Mapa de estimaciones de σ para cada estación meteorológica	78
6.7. Mapa de niveles de retorno estimados a 5 años	80
6.8. Mapa de niveles de retorno estimados a 25 años	81
6.9. Mapa de niveles de retorno estimados a 50 años	82

Índice de tablas

6.1. Descripción de los modelos utilizados en la etapa 1 y sus valores de log-verosimilitud marginal asociada	74
6.2. Descripción de los modelos utilizados en la etapa 2 y sus valores de log-verosimilitud marginal asociados	76
6.3. Resumen de los modelos escogidos para cada etapa	77
6.4. Valores estimados de los parámetros para el modelo completo	77
A.1. Información General de las estaciones meteorológicas incluidas en el estudio	86
A.2. Valores de los niveles de retornos estimados a 5, 10, 15, 25, 50 y 100 años para las estaciones incluidas en el estudio . .	87
A.3. Intervalos de confianza del 95 % para los niveles de retorno a 25 años asociados a las estaciones incluidas en el estudio .	88

Capítulo 1

Introducción

La presencia de valores muy altos o muy bajos en un estudio implica que realicemos un abordaje diferente. Este análisis específico se puede realizar mediante la teoría de valores extremos, esta teoría es muy importante por su aplicación en áreas como ingeniería, meteorología, economía, entre otras. La precipitación es uno de los fenómenos meteorológicos donde la ocurrencia de valores extremos es cada vez más frecuente, esto como consecuencia del cambio climático.

En Honduras, hablando específicamente en términos de precipitación extrema, se han presentado fenómenos como ser los huracanes Fifi (1974), Mitch (1998), Eta e Iota (2020), estos últimos catalogados como el fenómeno natural más severo en los últimos 20 años. Todos afectaron fuertemente al país y dada la naturaleza de los mismos se consideran poco usuales.

El estudio de estos eventos extremos y hablar de la probabilidad de ocurrencia de los mismos se vuelve difícil, por la cantidad de datos que se poseen, además de que su comportamiento no es el mismo en todos los puntos, geográficamente hablando, ya que la precipitación como variable de interés tiene implícitos muchos factores entre ellos: latitud, altitud, la estación del año y el relieve.

Al observar la ocurrencia de este tipo de fenómenos nos surgen varias interrogantes, como por ejemplo: ¿En qué momento se puede presentar nue-

vamente un fenómeno de esta magnitud?, ¿Existe un periodo de tiempo específico donde estos fenómenos se repiten?. Poder dar respuesta a estas preguntas es muy importante.

Si bien es cierto, no es posible detener la ocurrencia de fenómenos extremos, conocer su impacto brinda la oportunidad de crear planes para mitigar las consecuencias de las ocurrencias de los mismos. Como una de las posibles respuestas a las interrogantes planteadas, surge la necesidad de estimar periodos de retorno, es decir el número de años que en promedio un evento considerado como extremo se igualará o superará, y además poder determinar la frecuencia con la que se espera que ocurra.

El objetivo de este trabajo es poder estimar niveles de retorno para diferentes periodos de tiempo usando estos valores para generar mapas de retorno para precipitación en Honduras.

El documento de tesis esta organizado de la siguiente forma, en el capítulo 2 se presentan los elementos de inferencia, enunciamos los principales conceptos correspondientes a la inferencia tanto clásica como bayesiana, describimos además los modelos jerárquicos y los métodos de simulación utilizados en nuestro trabajo. En el capítulo 3 abordamos los conceptos generales de la teoría de valores extremos univariados, incluyendo un poco de la teoría multivariante. Definimos la Distribución Pareto Generalizada que es muy importante en la modelización de nuestros datos, así como los niveles de retorno que son nuestro foco principal. En el capítulo 4 se describe de manera general la Estadística espacial. Por la naturaleza de nuestros datos nos enfocamos en la rama de la Geoestadística, que es la que nos brinda las herramientas para el análisis del proceso espacial latente. También se muestran algunos conceptos de precipitación importantes para el abordaje de nuestros datos. En el capítulo 5 se presenta el modelo completo y su abordaje mediante modelos jerárquicos. Se describe el método de simulación utilizado, se presenta además el criterio utilizado para escoger el mejor modelo. En el capítulo 6 se describen los distintos modelos propuestos para el abordaje de los datos y se resumen los modelos escogidos con sus respectivas estimaciones que posteriormente usaremos para determinar completamente los niveles de retorno. Finalmente se presentan las conclusiones y algunos trabajos futuros.

Capítulo 2

Elementos de Inferencia

2.1. Conceptos previos

En esta sección incluiremos las definiciones de las distribuciones de probabilidad que son importantes en el desarrollo de nuestro trabajo. Estas se tomaron de (Gamerman, Migon, and Louzada, 2015).

Definición 2.1. Distribución Bernoulli: Decimos que X tiene una distribución Bernoulli con probabilidad de ocurrencia θ , es decir $X \sim \text{Ber}(\theta)$, si su función de probabilidad esta dada por:

$$p(x|\theta) = \theta^x(1 - \theta)^{1-x}, \quad x = 0, 1,$$

con $\theta \in [0, 1]$. Esta distribución tiene media, $E(X) = \theta$ y varianza, $\text{Var}(X) = \theta(1 - \theta)$.

Definición 2.2. Distribución Binomial:

Decimos que X tiene una distribución binomial con parámetro n y probabilidad de éxito θ , lo denotamos como $X \sim \text{Bin}(n, \theta)$, si su función de probabilidad esta dada por:

$$p(x|n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n$$

para $n \geq 1$, $\theta \in [0, 1]$. Esta distribución tiene media, $E(X) = n\theta$ y varianza, $Var(X) = n\theta(1 - \theta)$.

Definición 2.3. Distribución Uniforme: Decimos que X tiene distribución uniforme con parámetros θ_1 y θ_2 , lo denotamos como $X \sim U(\theta_1, \theta_2)$, si su función de densidad esta dada por:

$$p(x|\theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq x \leq \theta_2.$$

Esta distribución tiene media, $E(X) = (\theta_1 + \theta_2)/2$ y varianza, $Var(X) = (\theta_2 - \theta_1)^2/12$.

Definición 2.4. Distribución Gamma: Decimos que X tiene una distribución Gamma con parámetros α y β , lo denotamos como $X \sim G(\alpha, \beta)$, si su función de densidad esta dada por:

$$p(x|\alpha, \beta) = kx^{\alpha-1}e^{-\beta x}, \quad x > 0$$

para $\alpha, \beta > 0$. La constante k esta dada por $k = \beta^\alpha/\Gamma(\alpha)$, esta distribución tiene media, $E(X) = \alpha/\beta$ y varianza, $Var(X) = \alpha/\beta^2$

Definición 2.5. Vector de medias

Sea $\mathbf{X} = (X_1, \dots, X_p)$ vector aleatorio, el vector de medias μ de \mathbf{X} , es $\mu = (\mu_1, \dots, \mu_p)$ con $\mu_i = E(X_i)$

Definición 2.6. Matriz de covarianzas

Sea $\mathbf{X} = (X_1, \dots, X_p)$ vector aleatorio, la matriz de covarianzas Σ de \mathbf{X} , es $\Sigma = [\sigma_{ij}]$ con $\sigma_{ij} = cov(X_i, X_j)$. Donde Σ es una matriz simétrica ya que $cov(X_i, X_j) = cov(X_j, X_i)$ para todo (i, j) .

Definición 2.7. Distribución normal:

Decimos que X tiene una distribución normal con media μ y varianza σ^2 , lo denotamos como $X \sim N(\mu, \sigma^2)$, si su función de densidad esta dada por:

$$p(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \quad x \in \mathbb{R}$$

para $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

Sea $X = (X_1, \dots, X_p)'$ decimos que tiene una distribución normal multivariante con vector de medias μ y matriz de varianza-covarianza Σ , la denotamos por $X \sim N(\mu, \Sigma)$, si su función de densidad esta dada:

$$(2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\}, \quad x \in \mathbb{R}^p$$

para $\mu \in \mathbb{R}^p$ y $\Sigma > 0$.

Modelo de regresión logística

El modelo de regresión logística también llamado modelo logit, es un modelo lineal generalizado muy usado, ya que nos permite asegurarnos de que la variable dependiente tome valores solamente en $[0, 1]$. En este modelo las variables respuesta Y_1, \dots, Y_n son independientes, $Y_i \sim Ber(p_i)$ se tiene entonces que $E(Y_i) = p_i = p(Y_i = 1)$. Este modelo se asume que p_i esta asociado a los x_i mediante:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta x_i \quad (2.1)$$

El modelo asume que el logaritmo de la razón de las probabilidades es una función lineal del predictor x . Se tiene a partir de la ecuación 2.1 que:

$$p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}.$$

(Casella and Berger, 2002)

2.2. Elementos de Inferencia

Suponga que se tienen observaciones de una población homogénea con distribución F_θ , determinada mediante θ que es desconocido. Estimar el valor de θ es importante para poder describir la población.

Definición 2.8. Función de verosimilitud:

La función de verosimilitud de θ es la función que asocia los valores de $p(x|\theta)$ a cada θ , se denota por $l(\theta; x)$ o en algunos casos se denota también por $l_x(\theta)$, $l_x(\theta|x)$ o $l(\theta)$. Se define como:

$$l(\cdot; x) : \Theta \rightarrow \mathbb{R}^+$$

$$\theta \rightarrow l(\theta; x) = p(x|\theta)$$

Sea (x_1, \dots, x_n) una muestra de la variable aleatoria \mathbf{X} , con x_1, \dots, x_n independientes, entonces la función de verosimilitud esta dada por

$$l(\theta; x_1, \dots, x_n) = \prod_{i=1}^n p(x_i|\theta) \quad (2.2)$$

Toda la información provista por un experimento \mathbf{X} es resumida mediante la función de verosimilitud.

Definición 2.9. Información Esperada de Fisher:

Sea \mathbf{X} un vector aleatorio, con función de probabilidad $p(\mathbf{x}|\theta)$. La medida de información esperada de Fisher se define como:

$$I(\theta) = -E_{\mathbf{X}|\theta} \left[\frac{\partial^2 \log(p(\mathbf{X}|\theta))}{\partial \theta^2} \right].$$

Si $\theta = (\theta_1, \dots, \theta_p)$ es un vector de p parámetros, entonces la matriz de información esperada de Fisher se define como:

$$I(\theta) = -E_{\mathbf{X}|\theta} \left[\frac{\partial^2 \log(p(\mathbf{X}|\theta))}{\partial \theta \partial \theta'} \right]_{p \times p}$$

con elementos $I_{ij}(\theta)$ dados por:

$$I_{ij}(\theta) = -E_{\mathbf{X}|\theta} \left[\frac{\partial^2 \log(p(\mathbf{X}|\theta))}{\partial \theta_i \partial \theta_j} \right]_{p \times p}, \quad i, j = 1, \dots, p.$$

Definición 2.10. Información observada de Fisher:

La información observada de Fisher se define como:

$$J_x(\theta) = -\frac{\partial^2 \log(p(x|\theta))'}{\partial \theta \partial \theta}$$

En resumen $I(\theta) = E[J(\theta)]$.

La información observada de Fisher es una medida de información local mientras que la información esperada es una medida de información global.

En (Gamerman, Migon, and Louzada, 2015) se nos presenta el teorema de Bayes de la siguiente manera:

Definición 2.11. Teorema de Bayes:

Sea H la información inicial del parámetro de interés, supongamos que se puede resumir en términos de probabilidad mediante $p(\theta|H)$, asuma que una muestra X es observada, se tiene entonces $p(X|\theta, H)$. Con la nueva información H cambia a H^* , $H^* = H \cap \{X = x\}$, de manera que H^* la nueva información obtenida luego de observar X hace que la información disponible H sea mas precisa. Ya que $H^* \subseteq H$, puede ser resumida en $p(\theta|x, H) = p(\theta|H^*)$ con

$$p(\theta|x, H) = \frac{p(\theta, x|H)}{p(x|H)} = \frac{p(x|\theta, H)p(\theta|H)}{p(x|H)}$$

donde $p(x|H) = \int_{\Theta} p(x, \theta|H)d\theta$.

Notemos que el denominador no depende de θ , entonces

$$p(\theta|x, H) \propto p(x|\theta, H)p(\theta|H),$$

al tener presente la proporcionalidad solo necesitamos conocer la distribución posteriori hasta una constante de normalización, de manera simplificada:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

esta la forma usual en la que se nos presenta en la teoría el teorema de Bayes.

El teorema de Bayes fue introducido por Thomas Bayes en 1763 y nos presenta la forma de actualizar la probabilidad de θ tomando como punto de partida $p(\theta)$ para llegar a $p(\theta|x)$, por esta razón se identifica a $p(\theta)$ como distribución a priori y a $p(\theta|x)$ como distribución a posteriori.

Definición 2.12. Función de verosimilitud marginal

Sean x los datos, $p(x|\theta)$ una función de verosimilitud de un modelo para x indexado por un parámetro θ y $p(\theta)$ la distribución a priori para θ . La verosimilitud marginal para x esta definida por:

$$p(x) = \int p(x|\theta)p(\theta)d\theta$$

A la función de verosimilitud marginal también se le conoce como constante de normalización.

2.3. Estimación en la teoría clásica

Para la construcción de estimadores puntuales, en la teoría clásica, se tienen 3 métodos que son los más usados según (Gamerman, Migon, and Louzada, 2015).

1. Método de máxima verosimilitud:

Este es el método de estimación más usado, como su nombre lo indica se basa en la función de verosimilitud. Considere la observación X , con función de densidad $p(x|\theta)$. El estimador máximo verosímil (MLE) $\hat{\theta}$ de θ , es el valor de $\theta \in \Theta$ que maximiza $l(\theta; X)$.

Siendo $\hat{\theta}$ el valor de θ más verosímil, dada la muestra observada, es decir $\hat{\theta}$ es el valor que asocia más probabilidad a la ocurrencia de X .

2. Método de los momentos: Asumimos que X_1, \dots, X_n son independientes e idénticamente distribuidas del modelo paramétrico $\{p(\cdot|\theta), \theta \in \Theta \subset \mathcal{R}^p\}$, suponemos que el k -ésimo momento de $p(x|\theta)$ existe y lo denotamos por $\mu_k = E[X^k|\theta]$ para $k = 1, 2, \dots$. Asumimos más aún que $\mu_k = h_k(\theta)$, para algunas funciones h_k dadas, para $k = 1, 2, \dots$

Denotamos por $\mu_k = h_k(\theta)$ es el k -ésimo momento muestral

$$\hat{\mu}_k = \frac{\sum_{i=1}^n x_i^k}{n} = \bar{X}^k.$$

El estimador de los momentos es $\mu_k = \hat{\mu}_k$. El método de los momentos consiste en igualar los momentos de la muestra a los correspondientes momentos de la población, lo que significa estimar μ_k mediante $\hat{\mu}_k = h_k(\hat{\theta})$ y

resolver para θ , para alguna elección de valores para k . En la práctica si θ es un parámetro p -dimensional, sólo se necesitan p ecuaciones y la elección más común de valores para k es $k = 1, 2, \dots, p$.

3. Método de mínimos cuadrados:

Asumimos que $Y = (Y_1, \dots, Y_n)$ es una muestra aleatoria tal que $E(Y_i|\theta) = f_i(\theta)$ y $V(Y_i|\theta) = \sigma^2$, cada Y_i puede reescribirse:

$$Y_i = f_i(\theta) + e_i,$$

donde $E(e_i) = 0$ y $V(e_i) = \sigma^2$, $i = 1, \dots, n$.

El criterio de estimación es minimizar la suma de los errores cuadrados,

$$S(\theta) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [Y_i - f_i(\theta)]^2.$$

Si lo escribimos como vector $f(\theta) = (f_1(\theta), \dots, f_n(\theta))'$, de manera que $S(\theta)$ se reescribiría como: $S(\theta) = (Y - f(\theta))'(Y - f(\theta))$ el valor de θ que minimiza $S(\theta)$ es llamado estimador de mínimos cuadrados ordinarios de θ , este se encuentra resolviendo la ecuación $\partial S(\theta)/\partial \theta = 0$.

2.4. Inferencia bayesiana

En el enfoque bayesiano las probabilidades son tratadas como creencias, mientras que en el enfoque clásico son tratadas como frecuencias. Nuestro objetivo es poder estimar parámetros, en la inferencia bayesiana los parámetros son tomados como variables aleatorias con una distribución de probabilidad asociada.

Denotamos por θ al parámetro de interés que es desconocido, pero del cual se posee cierta cantidad específica de información. Esta información específica acerca de θ se pudo obtener por un estudio realizado anteriormente, por la experiencia, producto de una investigación o por la opinión de un experto. De alguna manera esa información inicial de la que se dispone debe ser cuantificada, para ello se le asocia a θ una distribución a priori $p(\theta)$. Donde $p(\theta)$ representa la creencia subjetiva inicial respecto a θ , previo a la observación de los datos.

Sean:

- x : Las observaciones de un problema dado, con densidad $p(x|\theta)$.
- θ : Parámetros, cantidades desconocidas usadas para especificar una distribución de x .
- $p(\theta)$: Distribución a priori para θ .

Algunos tipos de distribuciones a priori que podemos asociar al parámetro de interés son:

- Prioris conjugadas
- Prioris no informativas
- Prioris de referencia
- Prioris Jerárquicas
- Prioris independientes
- Prioris impropias

Estimación en la teoría bayesiana

Para realizar estimaciones en el análisis bayesiano seguimos los siguientes pasos:

- Tomamos un parámetro desconocido θ , con densidad a priori $p(\theta)$.
- Actualizamos la distribución a priori mediante el teorema de Bayes descrito en la definición 2.11 y obtenemos la distribución a posteriori: $p(\theta|\mathbf{X})$.
- Como estimador de θ tomamos la media a posteriori, $E(\theta|\mathbf{X})$.
- La varianza a posteriori, $Var(\theta|\mathbf{X})$ se usa como medida de dispersión para θ .

2.5. Métodos de Simulación

Enunciaremos algunos de los métodos computacionales más usados para generar aleatoriamente muestras de una distribución de interés, en el caso Bayesiano de la distribución a posteriori de θ , $p(\theta|\mathbf{X})$. Como resultado tendremos una muestra que cumple con las especificaciones deseadas, para poder observar sus características y a partir de ella obtener estimaciones de los parámetros de interés.

Algunos métodos son:

- Métodos de Monte Carlo
- Monte Carlo con muestreo por importancia
- Métodos de Remuestreo
 - ✓ Jackknife
 - ✓ Bootstrap

Todos estos métodos son abordados en diversidad de textos. En (Infante, A., and Urquiola, 2010) podemos encontrar las descripciones de los métodos utilizados en este trabajo y que se muestran a continuación.

2.5.1. Método Monte Carlo

Consiste en escribir una integral deseada como un valor esperado respecto a alguna distribución de probabilidad. Suponga que se quiere calcular

$$I = \int_a^b g(\theta)d(\theta) = \int_a^b [(b-a)g(\theta)] \frac{1}{b-a} d\theta = E_{U(a,b)}[(b-a)g(\theta)].$$

Para hacer esta aproximación se usa el método de los momentos

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n (b-a)g(\theta_i)$$

donde $\theta_1, \dots, \theta_n$ es una muestra aleatoria seleccionada de la distribución uniforme $U(a, b)$. Un algoritmo para la construcción de \hat{I} puede ser descrito siguiendo los siguientes pasos:

1. Genere $\theta_1, \dots, \theta_n$ de la distribución $U(a, b)$.
2. Calcule $g(\theta_1), \dots, g(\theta_n)$.
3. Obtenga la media muestral: $\bar{g} = \left(\frac{1}{n}\right) \sum_{i=1}^n g(\theta_i)$.
4. Determine $\hat{I} = (b - a)\bar{g}$.

Para generalizarlo, sea $I = E_p[g(\theta)]$ el valor esperado de $g(\theta)$ con respecto a una distribución con densidad $p(\theta)$. Entonces se tiene:

$$I = \int_a^b g(\theta)d\theta = \int_a^b \frac{g(\theta)}{p(\theta)}p(\theta)d(\theta) \approx \frac{1}{n} \sum_{i=1}^n \frac{g(\theta_i)}{p(\theta_i)}$$

con $\theta_1, \dots, \theta_n \sim p(\theta)$.

2.5.2. Métodos de Monte Carlo por Cadenas de Markov (MCMC)

El objetivo principal es la estimación de $I = \int h(x)f(x)dx$, la idea principal de estos métodos es construir una cadena de Markov a partir de la cual sea fácil simular y cuya distribución de equilibrio este dada por la distribución de interés, pueden ser aplicados a problemas de dimensiones altas. Algunos de estos métodos son:

- Muestreador Gibbs
- Metrópolis-Hastings
- Algoritmo E-M
- Monte Carlo Hamiltoniano

El algoritmo E-M y el algoritmo Monte Carlo Hamiltoniano solamente los enunciamos, a continuación describimos los métodos usados en nuestro trabajo.

Muestreador Gibbs

El muestreador de Gibbs es una forma de simular distribuciones multivariadas, se basa en la capacidad de las distribuciones condicionales. Es

apropiado cuando el muestreo de las distribuciones a posteriori conjunta no es posible.

Sea $\theta = (\theta_1, \dots, \theta_d)^T$, $p(\theta)$ la densidad de interés. Las condicionales completas son las que se utilizan para simular y están dadas por:

$$p(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d) = p(\theta_i | \theta_{-i}) = p_i(\theta_i), \quad i = 1, \dots, d.$$

Donde $p(\theta_i | \theta_{-i})$ es la distribución condicional de θ_i dados todos los demás parámetros de θ (excluyendo a θ_i).

El procedimiento para el algoritmo es el siguiente:

1. Para $j=1$. Se define el estado inicial de la cadena $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})^T$
2. Se calcula el nuevo valor $\theta^{(j)}$ a partir de $\theta^{(j-1)}$

$$\begin{aligned} \theta_1^{(j)} &\sim p\left(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}\right) \\ \theta_2^{(j)} &\sim p\left(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}\right) \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ \theta_d^{(j)} &\sim p\left(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}\right) \end{aligned}$$

3. Se aumenta el contador de j a $j+1$ y se repite el paso 2

Con este proceso se construye una cadena de Markov homogénea, ya que el valor simulado en cada iteración depende solo del valor de la iteración anterior. Se verifica también que $p(\theta)$ es una distribución estacionaria.

Definición 2.13. Distribución estacionaria:

Sea X_n , $n \geq 1$, una cadena de Markov con matriz de transición P . Una distribución de probabilidad π es estacionaria o invariante para X_n , si:

$$\pi = \pi P$$

Algoritmo de Metropolis-Hastings

En un algoritmo MCMC el estado actual de la cadena es: $\theta_1^{(j)}, \dots, \theta_d^{(j)}$ y el objetivo es simular $\theta_1^{(j+1)}$, el procedimiento es:

1. Proponga un valor candidato θ_1^{prop} , generado a partir de una distribución arbitraria con densidad

$$q\left(\theta_1^{prop} | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)$$

2. El próximo valor en la cadena $\theta_1^{(j+1)}$ se toma teniendo en cuenta lo siguiente:

$$\theta_1^{(j+1)} = \begin{cases} \theta_1^{prop} & \text{con probabilidad } p \\ \theta_1^{(j)} & \text{con probabilidad } 1 - p \end{cases}$$

defina:

$$r = \min \left\{ 1, \frac{p\left(\theta_1^{prop} | \theta_2^{(j)}, \dots, \theta_d^{(j)}\right) q\left(\theta_1^{(j)} | \theta_1^{prop}, \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)}{p\left(\theta_1^{(j)} | \theta_2^{(j)}, \dots, \theta_d^{(j)}\right) q\left(\theta_1^{prop} | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)} \right\}$$

con $p\left(\theta_1^{prop} | \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)$ la densidad a posteriori condicional de θ_1 en $\theta_1 = \theta_1^{prop}$ y de manera análoga para $p\left(\theta_1^{(j)} | \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)$ para implementar este paso se toma un valor $u \sim U(0, 1)$ y $\theta_1^{(j+1)} = \theta_1^{prop}$, si $u < r$ y $\theta_1^{(j+1)} = \theta_1^{(j)}$ en otro caso

Algunas observaciones acerca del algoritmo:

- El generador candidato $q\left(\theta_1^{prop} | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_d^{(j)}\right)$ es arbitrario
- Este método presenta ventaja respecto al muestreador de Gibbs ya que no se requiere conocer todas las distribuciones a posteriori condicionales completas. Solo se necesita conocer las condicionales hasta una constante de proporcionalidad.

Uno de los casos particulares del Algoritmo de Metropolis-Hastings es el de **Caminata aleatoria (algoritmo Metropolis con incrementos normales)**. Sea $q(\theta_1^{prop} | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_d^{(j)})$ la densidad de una normal $N(\theta_1^j, v)$. La simetría de la distribución de generador del candidato significa que los términos que involucran a $q(\theta_1^{prop} | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_d^{(j)})$ se cancelan en la fórmula de la probabilidad de aceptación, es decir:

$$r = \min \left\{ 1, \frac{p(\theta_1^{prop} | \theta_2^{(j)}, \dots, \theta_d^{(j)})}{p(\theta_1^j | \theta_2^{(j)}, \dots, \theta_d^{(j)})} \right\}.$$

La varianza de la distribución del generador del candidato es muy importante en las propiedades de mezcla del algoritmo. Si la varianza es grande entonces la propuesta varía mucho, así que la probabilidad de aceptación es baja. Si la varianza es pequeña entonces la probabilidad de aceptación es alta pero la propuesta varía poco. Generalmente la varianza v se escoge por ensayo y error, exigiendo una probabilidad de aceptación alrededor del 30 %.

Consideraciones acerca del MCMC

1. Es muy importante evaluar la convergencia de los algoritmos, pero se pueden generar problemas al trabajar con dimensiones altas.
2. Las cadenas deben ejecutarse varias veces, para verificar que la salida de las distintas cadenas es similar a pesar de que usen valores diferentes.
3. Deben ejecutarse cadenas largas durante grandes periodos de tiempo.
4. Se debe tener extremo cuidado con el uso de prioris impropias. En caso de utilizar prioris impropias siempre se debe verificar que la distribución a posteriori conjunta es propia, de lo contrario los resultados pueden no ser confiables.

Definición 2.14. Distribución a priori impropia

Una distribución a priori $p(\theta)$ es impropia si:

$$\int_{\Theta} p(\theta) d(\theta) = \infty$$

2.6. Comparación de Modelos

En (Liu and Wasserman, 2014) se enuncia la comparación de modelos y el factor de Bayes de la siguiente manera:

Sea $\mathcal{D}_n = \{X_1, \dots, X_n\}$ los datos, consideramos K modelos $\mathcal{M}_1, \dots, \mathcal{M}_K$ a cada modelo se le asigna una probabilidad a priori $\pi_j = p(\mathcal{M}_j)$ para el modelo \mathcal{M}_j y una priori $p_j(\theta_j|\mathcal{M}_j)$ a los parámetros θ_j bajo el modelo \mathcal{M}_j .

La distribución de probabilidad a posteriori del modelo \mathcal{M}_j condicionada a los datos \mathcal{D}_n es:

$$p(\mathcal{M}_j|\mathcal{D}_n) = \frac{p(\mathcal{D}_n|\mathcal{M}_j)\pi_j}{p(\mathcal{D}_n)} = \frac{p(\mathcal{D}_n|\mathcal{M}_j)\pi_j}{\sum_{k=1}^K p(\mathcal{D}_n|\mathcal{M}_k)\pi_k}$$

donde $p(\mathcal{D}_n|\mathcal{M}_j) = \int L_j(\theta_j)p_j(\theta_j)d\theta_j$, L_j es la función de verosimilitud para el j . De manera que

$$\frac{p(\mathcal{M}_j|\mathcal{D}_n)}{p(\mathcal{M}_k|\mathcal{D}_n)} = \frac{p(\mathcal{D}_n|\mathcal{M}_j)\pi_j}{p(\mathcal{D}_n|\mathcal{M}_k)\pi_k}. \quad (2.3)$$

Para escoger el modelo \mathcal{M}_j o \mathcal{M}_k observamos el lado derecho de la ecuación 2.3, si es mayor que 1 se escoge el modelo \mathcal{M}_j en otro caso, se escoge el modelo \mathcal{M}_k .

Definición 2.15. *Factor de Bayes*

El factor de Bayes entre los modelos \mathcal{M}_j y \mathcal{M}_k se define como

$$BF(\mathcal{D}_n) = \frac{p(\mathcal{D}_n|\mathcal{M}_j)}{p(\mathcal{D}_n|\mathcal{M}_k)} = \frac{\int L_j(\theta_j)p_j(\theta_j)d\theta_j}{\int L_k(\theta_k)p_k(\theta_k)d\theta_k}. \quad (2.4)$$

donde $p(\mathcal{D}_n|\mathcal{M}_j)$ es la verosimilitud marginal para el modelo \mathcal{M}_j .

El factor de Bayes puede ser visto como una alternativa a las pruebas de hipótesis clásicas. La comparación de modelos bayesianos es un método de selección de modelos que se basa en la ecuación 2.4.

2.7. Modelos Jerárquicos

Los modelos jerárquicos nos permiten modelar estadísticamente procesos más complejos y analizar mejor la relación de dependencia entre los parámetros involucrados en cada etapa. Un modelo jerárquico es un modelo estadístico multinivel que permite la incorporación de mayor cantidad de información dentro de un modelo. Un modelo jerárquico típico tiene la forma siguiente:

$$\begin{aligned}\lambda &\propto p(\lambda) \\ \theta_1, \dots, \theta_n | \lambda &\propto p(\theta | \lambda) \\ x_i | \theta_i &\propto p(x_i | \theta_i), \quad i = 1, \dots, n\end{aligned}$$

Es decir, dados los datos observados $x = (x_1, x_2, \dots, x_n)$ y el vector de parámetros desconocidos $\theta = (\theta_1, \dots, \theta_k)$, para especificar la distribución $p(x|\theta)$, suponemos que θ tiene distribución a priori $p(\theta|\lambda)$ donde λ es un vector de hiperparámetros. Si λ es conocido, la inferencia sobre θ se basa en la distribución a posteriori de la siguiente forma:

$$p(\theta|x, \lambda) = \frac{p(x, \theta|\lambda)}{p(x|\lambda)} = \frac{p(x, \theta|\lambda)}{\int p(x, \theta|\lambda)d\theta} = \frac{l(x, \theta)p(\theta|\lambda)}{\int l(x|\theta)p(\theta|\lambda)d\theta} \quad (2.5)$$

pero sabemos que en la practica λ no es conocida, entonces se requiere de una segunda etapa, utilizando la distribución $h(\lambda)$, la ecuación 2.5 se transforma en:

$$p(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{\int l(x|\theta)p(\theta|\lambda)h(\lambda)d\lambda}{\int \int l(x|\theta)p(\theta|\lambda)h(\lambda)d\theta d\lambda} \quad (2.6)$$

Por otra parte el modelar los θ_i como aleatorios (en lugar de efectos fijos), nos permite inducir una estructura de correlación entre ellos, podría ser correlación espacial, por ejemplo, esta estructura de correlación también puede estar presente en los datos observados y es importante modelarla. Es por eso que los métodos bayesianos jerárquicos son ampliamente utilizados en el análisis de datos espaciales, como se muestra con detalle en (Banerjee, Carlin, and Gelfand, 2003).

CAPÍTULO 2. ELEMENTOS DE INFERENCIA

En el siguiente capítulo se aborda de manera general la teoría de valores extremos de la cual los niveles de retorno son parte. Se presentan los conceptos fundamentales de esta teoría utilizados en el desarrollo de este trabajo.

Capítulo 3

Teoría de Valores Extremos

3.1. Preliminares

La teoría de valores extremos es la única disciplina estadística que desarrolla técnicas y modelos para describir lo inusual en lugar de lo habitual, se enfoca en aquellos valores que podríamos considerar como extremos o poco usuales, ya sea por que son considerados muy altos o muy bajos respecto a los demás. Estos valores extremos se consideran incluso como raros, comparados con lo que de cierta manera se considera normal, por lo que podría ocurrir que la cantidad de datos disponibles para realizar estudios de este tipo sean pocos.

Para poder definir que valores son considerados extremos se determina un valor especial que nos sirve de frontera, entre los valores considerados extremos y aquellos que no lo son, llamaremos a este valor de **umbral**. Todos aquellos valores que estén por encima o por debajo del umbral, dependiendo lo que nos interese estudiar, máximos o mínimos respectivamente, son los que se convierten en nuestro objeto de estudio.

Lo interesante de la teoría de valores extremos (*Extreme Value Theory*: EVT por sus siglas en inglés) son los diversos campos de aplicación como: ingeniería, biología, economía, meteorología, oceanografía, entre otros. En el campo de la economía se usan para estudiar pérdidas extremas o ga-

nancias muy altas, si hablamos de biología y del cambio climático que es un tema que nos concierne y preocupa a todos, notamos que hay eventos climáticos extremos denotados también como catástrofes climatológicas y atmosféricas como ser: sequía, huracanes, incendios, inundaciones, donde la teoría de valores extremos es utilizada.

Dichos eventos se consideran poco comunes comparado con aquello que se cataloga como normal, los mismos tienen implicaciones muy fuertes en términos económicos y pérdidas de recursos en general, ocurren cada vez con más frecuencia debido al cambio climático y su impacto en nuestro entorno es preocupante, porque si bien es cierto las consecuencias asociadas a dichos fenómenos no se pueden controlar o evitar, el hecho de poseer información al respecto o incluso poder tener predicciones permitiría tomar medidas que al menos mitiguen sus efectos.

3.2. Antecedentes

El origen de la teoría de los valores extremos como tal no es fácil de ubicar, es tema de interés desde hace mucho tiempo, muestra de ello son la diversidad de textos disponibles. Se ha estudiado desde el siglo XX, aunque según Coles (Coles, 2001) fue hasta 1950 que se propuso una metodología seria para modelar este tipo de sucesos. Se asocian las primeras aplicaciones al campo de la ingeniería, debido a que en sus diseños siempre se buscan estructuras que soporten las distintas fuerzas que impacten sobre las mismas.

Dentro de la teoría formal se puede ubicar el primer teorema fundamental con Fisher-Tippett (Fisher and Tippett, 1928) y Gnedenko (Gnedenko, 1943), que caracteriza la distribución asintótica del máximo. El segundo teorema fundamental con Pickands (Pickands III, 1975) y Haan (Balkema and Haan, 1974) en el que se describe el comportamiento de las colas de una distribución asociándolo a la distribución de valores generalizados.

Existen 2 tipos de modelos para el abordaje de los valores extremos.

1. Máximo por bloques:

En lugar de analizar todos los datos, se toman bloques de datos de

igual longitud. Para cada bloque se calcula el máximo de acuerdo a lo que se desea analizar generando una serie de máximos, convirtiéndose estos en el objeto de interés. Si los bloques se toman por años se generaría una serie de máximos anuales, de manera análoga si nos interesa estudiar el comportamiento del mínimo.

2. Método POT:

Método de picos sobre el umbral (Peaks Over Threshold) en este modelo se trabaja con valores de excedencias, se escoge un umbral u adecuado y se calculan las excedencias, el objeto de interés son todos aquellos valores que superan el mismo. El método de valores sobre un umbral fue propuesto inicialmente en el área de Hidrología y luego formalizado por R.L. Smith en base a resultados obtenidos previamente por Pickands.

La distribución de Pareto generalizada se utiliza como aproximación, para la distribución límite de las excedencias de umbrales. Al usar este método podría ocurrir que si escogemos un valor de umbral muy grande, obtendremos pocas excedencias y por ende estimadores de alta varianza, por otro lado si el umbral es demasiado pequeño los estimadores que obtenemos son sesgados.

En la practica una manera de selección de un umbral u óptimo es basándose en la linealidad de la función media de excesos $e(u)$ para la GPD, se escoge el valor máximo para el cual el gráfico de vida media residual muestre un comportamiento lineal.

El primer método es el más tradicional se utiliza en muestras grandes, con observaciones idénticamente distribuidas. Este tipo de modelo asume que los valores máximos obtenidos de cada bloque pueden ser ajustados mediante la distribución Generalizada de Valores Extremos (GEV).

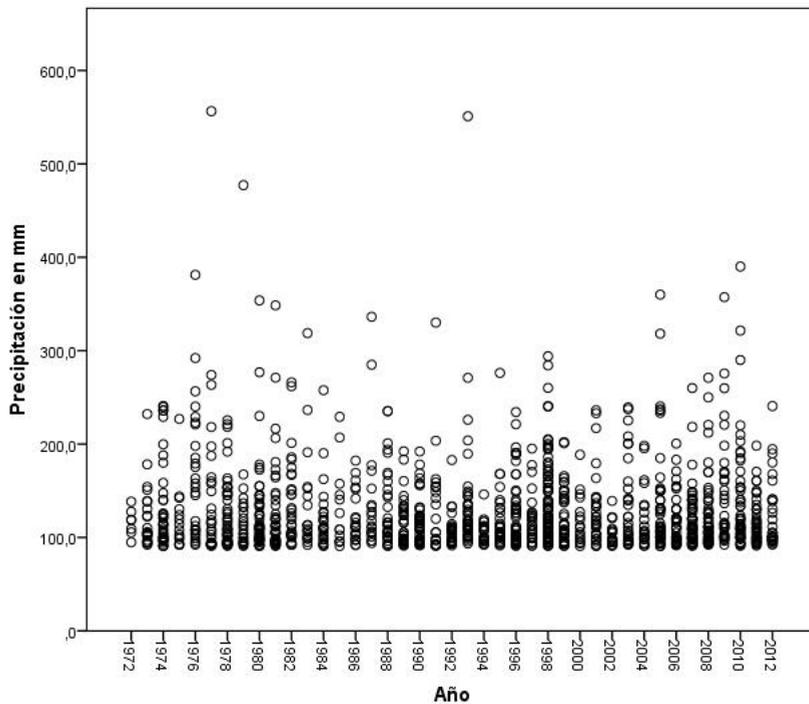
El segundo método trabaja con todos los datos, la elección del umbral óptimo es muy importante. Un umbral óptimo nos asegura buenos ajustes mediante la Distribución de Pareto Generalizada (GPD).

Para ilustrar un poco lo antes mencionado, en la Figura 3.1 se muestran los valores de precipitación diaria en Honduras, para los años de 1972 a 2012

CAPÍTULO 3. TEORÍA DE VALORES EXTREMOS

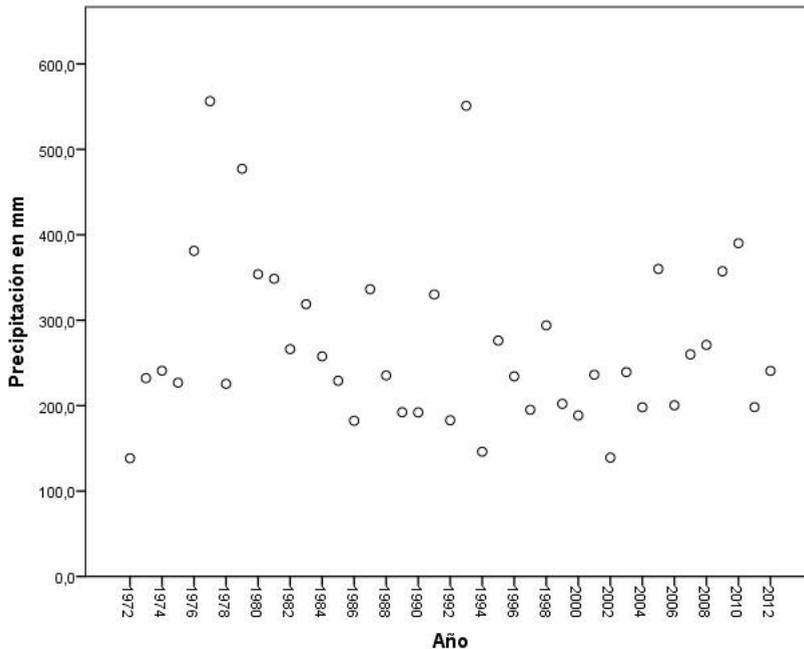
estos valores medidos en milímetros (mm). Se puede observar la presencia de valores de precipitación altos que son los que nos interesa estudiar y describir.

Figura 3.1: Valores de precipitación diaria para Honduras de 1972 a 2012



De acuerdo al método seleccionado se trabaja con todos los datos o solo con algunos, si se escoge usar el método de máximo por bloques el número de observaciones se reduce.

Figura 3.2: Valores máximos de precipitación diaria por año en Honduras de 1972 a 2012



En la Figura 3.2, se muestran únicamente los valores máximos de precipitación diaria en Honduras para cada año de 1972 a 2012. A partir de los valores de precipitación diaria mostrados en la Figura 3.1, se toma para cada año solamente el valor máximo de precipitación.

3.3. Teoría de valores extremos

Una forma de abordar la teoría de valores extremos es haciendo uso de la teoría de distribuciones asintóticas, de manera análoga a como se utiliza el teorema del limite central, con la diferencia que el enfoque esta vez es en los valores extremos en lugar de los valores medios.

Sea X_1, \dots, X_n una sucesión de variables aleatorias independientes e idénti-

camente distribuidas, el estudio de valores extremos se centra en el comportamiento de $M_n = \max(X_1, \dots, X_n)$, donde los X_i corresponden a valores de un proceso medidos en una escala de tiempo regular ya sea hora, día o mes y M_n es el valor máximo sobre todas las observaciones. Nos centraremos únicamente en el comportamiento del máximo sin perder de vista que también podría interesarnos el comportamiento de $m_n = \min(X_1, \dots, X_n)$, todo lo descrito a continuación para el máximo puede ser aplicado para describir el comportamiento del mínimo mediante la relación

$$\max(X_1, \dots, X_n) = -\min(-X_1, \dots, -X_n).$$

Para las variables X_i con función de distribución F idéntica pero desconocida se tiene que:

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z) \dots P(X_n \leq z) \\ &= (F(z))^n \end{aligned} \tag{3.1}$$

La distribución de F es desconocida y no es fácil de aproximar, observando los datos podríamos estimarla, pero al tener pocas observaciones esta no es una solución factible. La propuesta para obtener estimaciones apropiadas es construir resultados análogos a la distribución muestral de medias, recordemos que las conclusiones para la media se justifican por el teorema del límite central y la teoría asintótica.

Nos interesa el comportamiento de F^n cuando $n \rightarrow \infty$, es decir el comportamiento asintótico del máximo, la clave es encontrar una estandarización adecuada para el máximo seleccionando sucesiones a_n (medida de dispersión) y b_n (medida de tendencia central) adecuadas, de manera que exista en el límite una distribución asintótica de M_n^* , que denotaremos como máximo normalizado con $M_n^* = \frac{M_n - b_n}{a_n}$.

Primer Teorema Fundamental

Como mencionamos anteriormente la teoría formal de los valores extremos se basa en 2 teoremas fundamentales. Mediante el primer teorema fundamental se caracteriza el comportamiento asintótico de M_n^* , en este se

enuncia que todas las posibles distribuciones límites del máximo normalizado pertenecen a una de las 3 familias de distribuciones listadas y descritas en el mismo.

Teorema 3.3.1. Teorema de Fisher-Tippett

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

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z) \text{ cuando } n \rightarrow \infty, \quad (3.2)$$

donde G es una función de distribución no degenerada, entonces G pertenece a una de las siguientes familias:

I. Gumbel:

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, -\infty < z < \infty$$

II. Fréchet:

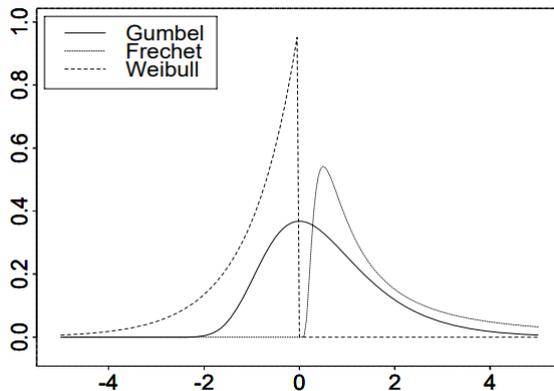
$$G(z) = \begin{cases} 0, & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases}$$

III. Weibull:

$$G(z) = \begin{cases} \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z < b \\ 1, & z \geq b \end{cases}$$

para parámetros $a > 0$, b y $\alpha > 0$.

Figura 3.3: Densidad de las distribuciones estándar de valores extremos



Se toma $\alpha = 1$ para las distribuciones de Fréchet y Weibull

Fuente: (Embrechts, Klüppelberg, and T., 1994)

Este es el primer teorema fundamental dentro de la teoría de valores extremos, nos presenta las leyes para el comportamiento límite del máximo. Este se enuncia en el capítulo 3 de (Embrechts, Klüppelberg, and T., 1994) que corresponde a las fluctuaciones del máximo en la teoría de valores extremos, este teorema menciona que sin importar la distribución de $F(x)$ el comportamiento límite del máximo, siempre que este límite exista, pertenece a una de estas 3 familias conocidas como **Distribuciones de Valores Extremos**. La demostración de este teorema se basa en los resultados de Gnedenko y Kolmogorov sobre teoremas límite para sumas de variables aleatorias independientes y se puede encontrar en (Gnedenko and Kolmogorov, 1954).

Cada familia posee un parámetro de locación b , un parámetro de escala a y en el caso de las familias de Fréchet y Weibull un parámetro de forma α . De manera resumida

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \approx G(z)$$

para n suficientemente grande.

Estas 3 familias de distribuciones pueden describirse como una sola reparametrizando $\xi = \frac{1}{\alpha}$, obteniendo la siguiente función de distribución:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right]\right\} \quad (3.3)$$

esta es la distribución de valores generalizada (GEV, por sus siglas en ingles).

Definición 3.1. Si existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ tales que:

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

para una función de distribución no degenerada G , entonces G es un miembro de la familia GEV

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right]\right\} \quad (3.4)$$

definida en $\{z : 1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}} > 0\}$ con $-\infty < \mu < \infty$, $-\infty < \xi < \infty$ y $\sigma > 0$.

Segundo Teorema Fundamental

Si se toma como objeto de estudio todos aquellos valores que exceden un umbral fijo (excedencias), estos valores quedan caracterizados por la Distribución de Pareto Generalizada (GPD).

Definición 3.2. Sea X una v.a con distribución F y extremo derecho X_F . Para $u < X_F$ fijo decimos que ha ocurrido una excedencia de u si $X > u$.

La distribución de X_u esta dada por:

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad x \geq 0$$

A la función $e(u) = E((X - u) | X > u)$ se le conoce como la función media de exceso, es posible encontrar un valor apropiado del umbral haciendo un gráfico de la función media de excesos empírica.

Definición 3.3. *La Distribución de Pareto Generalizada introducida por Pickands en 1975 es la familia bi-parámetrica de funciones de distribución dada por:*

$$G(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(\frac{-x}{\sigma}\right), & \xi = 0 \end{cases} \quad (3.5)$$

La GPD queda descrita por:

- σ : Parámetro de escala
- ξ : Parámetro de forma

Con $\xi < 0$, para $0 < x < -\frac{\sigma}{\xi}$ y $x > 0$, para $\xi > 0$.

Basado en estos resultados el modelo GPD surge como el modelo de referencia para modelar colas. Se tiene:

Frechet, para $\xi > 0$

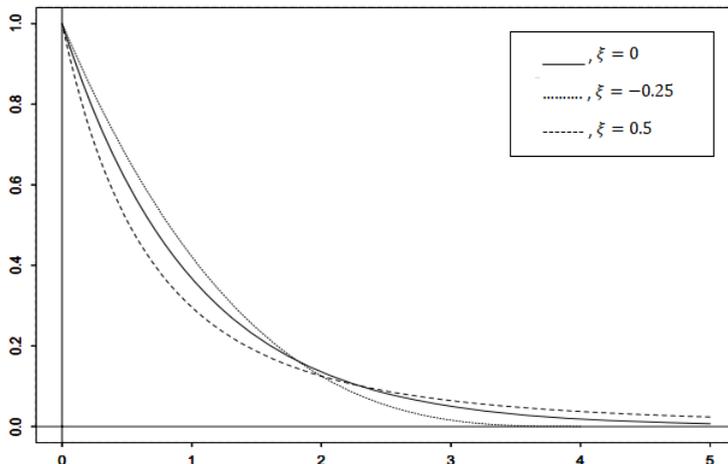
Gumbel, para $\xi = 0$.

Weibull, para $\xi < 0$

El parámetro de forma nos permite también clasificar las colas de una distribución de la siguiente manera:

- Cola ligera: $\xi < 0$
- Colas exponenciales o normales : $\xi = 0$
- Cola pesada: $\xi > 0$

Figura 3.4: Densidades de GPD para diferentes valores de ξ y $\sigma = 1$



Fuente: (Embrechts, Klüppelberg, and T., 1994)

Proposición 3.1. Sea F una función de distribución de la familia GPD, $F = G(\cdot; \xi, \sigma)$ para algún $\xi \in \mathbb{R}$ y $\sigma > 0$, entonces

$$F_u(x) = G(x; \xi, \sigma + \xi u);$$

con $0 \leq x \leq \infty$, si $\xi \leq 0$ y $0 \leq x \leq -\frac{\sigma}{\xi} - u$ si $\xi < 0$.

La proposición 3.1 muestra que la familia GPD es cerrada bajo transformaciones, y específicamente bajo la transformación de la distribución de excesos.

Teorema 3.3.2. Existe una función $\sigma(u)$ tal que

$$\lim_{x_F \rightarrow u} \left(\sup_{0 \leq x \leq X_F - u} |F_u(x) - G(x; \xi, \sigma(u))| \right) = 0,$$

$\xi \in \mathbb{R}$ y $X_F \leq \infty$ denota el límite derecho de la distribución F .

El segundo Teorema Fundamental, Teorema de Pickands-Balkema-de Haan enuncia que toda distribución tiene un umbral adecuado de manera que la

función de distribución de excesos esta tan cerca de una distribución GPD como queramos, un umbral adecuado también llamado óptimo es aquel que cumple las condiciones de un buen ajuste a una GPD para las excedencias, a estas excedencias se les denota en algunos casos como colas.

3.4. Niveles de retorno

Generalmente cuando estudiamos valores extremos, estos van relacionados con una aplicación y por ende se busca darle a estos valores una interpretación. Con la finalidad de relacionar la teoría de valores extremos con sus aplicaciones, es conveniente hablar de cuantiles o de manera análoga de niveles de retorno, más que de las estimaciones de parámetros como tal.

En el caso específico de los fenómenos extremos, es importante disponer de un valor que permita cuantificar la probabilidad de ocurrencia de los mismos, para un periodo de tiempo determinado. La teoría de valores extremos proporciona un enlace entre los datos tomados por hora o día y periodos de tiempo más largos, como lo son los niveles de retorno que pueden ser tomados por periodos de 5, 10, 15 o más años. Aunque el nivel de retorno tiene una forma cerrada, es una función no lineal de los parámetros GPD y la probabilidad de excedencia.

Además los niveles de retorno son cantidades climatológicas asociados con precipitación, que se comportan de manera diferente para diferentes puntos geográficos, es decir hay una componente espacial implícita que debe ser tomada en cuenta.

De las reglas de probabilidad básica, siguiendo la línea de trabajo de (Cooley, Nychka, and Naveau, 2007) para el abordaje de los niveles de retorno, tenemos la ecuación 3.5

$$P(Z > z + u) = \zeta_u \left(1 + \xi \frac{z}{\sigma_u} \right)^{-\frac{1}{\xi}}, \text{ con } \zeta_u = P(Z > u).$$

Es conveniente trabajar los niveles de retorno en años, sea n_y el número de observaciones tomadas en un año, se obtiene el nivel de retorno a r años

resolviendo la ecuación $P(Z > z_r) = \frac{1}{rn_y}$ para z_r . De manera que

$$z_r = u + \frac{\sigma_u}{\xi} \left[(rn_y \zeta_u)^\xi - 1 \right] \quad (3.6)$$

Para lograr estimar los niveles de retorno es necesario obtener las estimaciones de los parámetros y sustituirlas en la ecuación 3.6. En resumen los niveles de retorno Z_r son valores de precipitación que en promedio serán alcanzados o excedidos con probabilidad $1/Z_r$ al menos una vez en r años.

3.5. Teoría de valores extremos multivariante

Si tenemos dos o más procesos asociados a valores extremos estos se podrían modelar individualmente usando técnicas univariantes, pero muchas veces más que ese comportamiento individual o independiente nos interesa estudiar las relaciones que puedan existir entre dichos procesos. De manera que para medir estas relaciones los métodos de bloques máximos, selección de umbrales y distribuciones enunciados en la teoría de valores extremos univariada tiene su análogo en la teoría de valores extremos multivariante.

Sean $X_i = (X_{i,1}, \dots, X_{i,p})$, con $i = 1, \dots, n$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos con función de distribución conjunta F y marginales F_1, \dots, F_p .

El vector de máximos denotado por M para el caso multivariante se define como:

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

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

Caso bivariado

Coles en (Coles, 2001) caracteriza de manera asintótica la distribución de valores extremos, a partir de la función de distribución conjunta de

$$M^* = (M_1/n; M_2/n).$$

Teorema 3.5.1. *Sea $M^* = (M_1^*, M_2^*)$ con $(X_{i,1}, X_{i,2})$, $i = 1, \dots, n$ vectores independientes con distribución marginal Fréchet. Entonces si para $n \rightarrow \infty$,*

$$P(M_1^* \geq x_1, M_2^* \geq x_2) \xrightarrow{d} G(x_1, x_2), \quad (3.7)$$

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

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

donde

$$V(x_1, x_2) = 2 \int_0^1 \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) dH(w) \quad (3.8)$$

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

$$\int_0^1 w dH(w) = 1/2. \quad (3.9)$$

La familia de distribuciones que surgen como límites de la ecuación 3.7 se denomina la clase de distribuciones bivariantes de valor extremo. El teorema 3.5.1 implica que esta clase tiene correspondencia uno a uno con el conjunto de funciones de distribución H en $[0, 1]$ que satisface la ecuación 3.9. Si H es diferenciable con densidad h la integral dada en la ecuación 3.8 es simplemente

$$V(x_1, x_2) = 2 \int_0^1 \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) h(w) dw.$$

Sin embargo las distribuciones de valores extremos bivariantes también se generan con medidas H que no son diferenciable

Esta distribución engloba a infinitas familias de distribuciones, distinto a lo que ocurría en el caso univariado en donde la convergencia llevaba únicamente a 3 posibles familias de distribuciones.

Caso Multivariado

Es una generalización del caso bivariado que se basa en la pertenencia a un dominio de atracción.

Definición 3.4. Dominio de atracción: Sea $\{X_n, n \geq 1\}$ una sucesión de variables i.i.d. con función de distribución común F y $M_n = \max\{X_i, 1 \leq i \leq n\}$ decimos que la función F esta en el dominio de atracción de la distribución de valor extremo G definida en la ecuación 3.2, es decir $F \in \mathcal{D}(G)$, si existen constantes $a_n > 0$, $b_n \in \mathbb{R}$ tales que:

$$F^n(a_n x + b_n) = p(M_n \leq a_n x + b_n) \rightarrow G(x), \quad x \in \mathbb{R}.$$

Sea $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^d$, si $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p})$, $i = 1, \dots, n$ son vectores aleatorios independientes e idénticamente distribuidos de dimensión p con función de distribución F , se puede considerar la existencia de sucesiones $(a_n)_n$ y $(b_n)_n$ en \mathbb{R}^p , con $a_{n,j} > 0$ y $b_{n,j} > 0 \in \mathbb{R}$, para todo $j = 1, \dots, p$ y una función de distribución G con marginales no degeneradas tales que, cuando $n \rightarrow \infty$,

$$P\left(\left(\max_i \{X_i - b_n\} / a_n \leq x\right) = F^n(a_n \mathbf{x} + b_n) \rightarrow G(x),\right.$$

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

Lo que se menciono es solo un resumen general de la teoría de valores extremos multivariados a manera de enunciar e introducir la misma. Todo lo descrito en la teoría univariada tiene su análogo en la teoría multivariada, puede encontrarse más detalles en (Coles, 2001).

En el siguiente capítulo hablamos sobre los conceptos básicos de la estadística espacial y la precipitación como fenómeno climatológico que puede ser abordado desde el punto de vista espacial.

Capítulo 4

Estadística espacial y precipitación

4.1. Estadística espacial

Es parte de la naturaleza humana el hecho de que al estudiar un fenómeno o querer conocer más acerca de algo aparezcan de manera implícita algunas interrogantes: ¿Cómo?, ¿Cuándo?, pero también ¿Dónde?. En base a la necesidad de dar respuestas a estas interrogantes surge la estadística espacial, ya que deseamos usar toda la información que poseemos y ese ¿Dónde? puede aportar mucho a nuestro análisis. Ubicarnos espacialmente nos da información más específica acerca de lo que queremos saber.

La estadística espacial es una rama muy importante de la estadística, sus aplicaciones van desde el estudio de fenómenos de carácter biológico, físico, social y económico entre otros. Todo aquello donde la ubicación espacial es un componente relevante que aporta información de interés, está íntimamente ligado a la estadística espacial.

Como parte de la estadística espacial encontramos la geoestadística, que se enfoca en la estimación de fenómenos naturales haciendo uso de la teoría de funciones aleatorias. En nuestro caso estamos interesados en abordar el fenómeno de precipitación, que es un fenómeno natural. Sabemos que

de acuerdo al lugar donde nos ubiquemos las precipitaciones pueden ser mayores o menores, esperando que en lugares cercanos los niveles de las mismas sean parecidos. Esa ubicación espacial presente en dicho fenómeno debe ser tomada en cuenta en los modelos.

A cada valor de precipitación se le asocia una ubicación geográfica. Dada la naturaleza del fenómeno de estudio nuestra creencia a priori es que a mayor distancia geográfica, más diferencia hay entre los valores de precipitación medidos, mientras que a medida que la distancia se acorta los valores de precipitación serán más cercanos. Esta creencia inicial se debe verificar analizando nuestros datos, tomando en cuenta la componente espacial intrínseca a los mismos.

4.1.1. Generalidades de la Estadística Espacial

Modelo General

Sea x un punto o localidad en \mathcal{R}^d . Si x varia sobre un conjunto $D \subset \mathcal{R}^d$, con $d = 1, 2, 3$ se genera un proceso estocástico

$$\{Z(x); x \in D\}$$

cuyas realizaciones denotaremos como:

$$\{z(x); x \in D\}$$

los datos pueden ser discretos o continuos.

Tipos de datos espaciales

Como se describe en (Cressie, 1993) tenemos 3 tipos de datos espaciales:

1. Datos continuos espaciales:

- \mathcal{R} es un subconjunto del plano del área positiva (2-D) o del volumen (3-D).
- $Z(x)$ es una variable aleatoria en cada una de las infinitas ubicaciones continuas $x \in \mathcal{R}$.

2. Datos de área:

- $\mathcal{R} = \{x_1, x_2, \dots, x_n\}$ es una red regular o irregular en el plano.
- $Z(x_i)$ es una variable aleatoria en cada lugar x_i , $i = 1, 2, \dots, n$.

3. Datos de proceso de puntos espaciales:

- $\mathcal{R} = \{x_1, x_2, \dots, x_n\}$ es una colección aleatoria de puntos en el plano.
- $Z(x)$ no está especificada o es una variable aleatoria en un lugar $x \in \mathcal{R}$.

Clasificación de la estadística espacial

Para cada tipo de dato espacial se tiene una rama asociada dentro de la estadística espacial:

1. **Geoestadística:** comprende procesos estocásticos cuyos valores pueden ser conocidos en todos los puntos del área de estudio.
2. **Datos de red o de áreas:** se refiere a la distribución de eventos cuya localización se asocia a zonas delimitadas por polígonos.
3. **Patrón de puntos espaciales:** Conjunto de puntos distribuidos en un área fija, cuya ubicación fue generada por un mecanismo estocástico.

4.1.2. Geoestadística

Origen y definición

Georges Matheron reconocido como el padre de la Geoestadística acuñó este término en los años 60, enfocándolo en el campo de la geología por sus trabajos desarrollados en minería. Formalizó y generalizó matemáticamente un conjunto de técnicas desarrolladas por D. G. Krige en 1941, que utilizaban la correlación espacial para hacer predicciones en la evaluación de reservas de las minas de oro en Sudáfrica.

Matheron define Geoestadística en (Matheron, 2019) como “la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales”.

La Geoestadística es la rama de la estadística espacial que se especializa en el análisis y la modelación de la variabilidad espacial en ciencias de la tierra. Su objeto de estudio es el análisis y la predicción de fenómenos en espacio y/o tiempo.

Algunas de sus aplicaciones son:

- **Mapeo o estimación**
 - ✓ Modelos digitales y superficiales de terreno
 - ✓ Caracterización geoquímica
 - ✓ Meteorología
 - ✓ Geomarketing
 - ✓ Contaminación ambiental
 - ✓ Epidemiología
- **Caracterización de la incertidumbre en la estimación**
 - ✓ Muestreo óptimo
 - ✓ Riesgo ambiental
 - ✓ Estimación de errores

Elementos importantes

Definición 4.1. Variables regionalizadas(VR):

Toda variable espacial que presente estructura de correlación, es una variable regionalizada y su estudio se realiza desde un enfoque probabilístico en el campo de las funciones aleatorias o procesos estocásticos.

Dadas las variables aleatorias $Z(x_i)$, suponemos que se encuentran correlacionadas es decir que el valor en cada punto o ubicación depende del valor en otras ubicaciones, teniendo que a mayor proximidad mayor dependencia. Tiene sentido entonces querer construir una estadística para pares de puntos separados por distancia fija h , que permita describir el comportamiento de la variable regionalizada según la distancia entre pares de puntos.

Definición 4.2. Variograma:

La función variograma se define como la varianza de los valores de la variable $Z(x)$, para todos los pares de puntos separados por una distancia (h):

$$\text{Var}\{Z(x+h) - Z(x)\} = E\{[Z(x+h) - Z(x)]^2\} = 2\gamma(h).$$

Esta función es muy importante en el campo de la geoestadística, podría decir que es un parámetro determinante dentro de la misma.

Definición 4.3. Semivariograma:

La función semivariograma $\gamma(h)$ queda definida como:

$$\gamma(h) = \frac{1}{2}E\{[Z(x+h) - Z(x)]^2\}.$$

El semivariograma es una característica de disimilitud entre dos ubicaciones. Permite caracterizar la autocorrelación espacial en el proceso, esto la convierte en la principal herramienta para poder describir el comportamiento de la variación espacial del fenómeno de estudio.

Para realizar un análisis geoestadístico el primer paso es hacer un análisis exploratorio de los datos, para poder observar la dispersión de los mismos. El análisis exploratorio se realiza mediante el semivariograma experimental, este nos da una idea a priori de que modelo escoger.

A partir de la fórmula del semivariograma se define el semivariograma experimental, este es un estimador insesgado del variograma y se define como:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i+h) - Z(x_i)]^2,$$

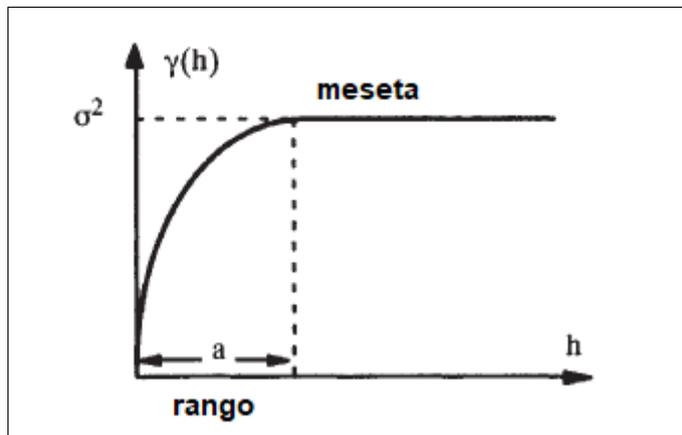
donde $N(h)$ es el número de pares con distancia h .

Posteriormente el variograma experimental debe ser ajustado a un modelo teórico adecuado que se define mediante los siguientes parámetros:

1. **Efecto pepita o nugget:** definido por Matheron (Matheron, 2019) es el valor del variograma para $h = 0$. En teoría debería ser cero, pero debido a errores de medida o ausencia de puntos separados a distancias menores de una distancia mínima, el variograma experimental no se anula.
2. **Alcance o rango:** es el valor de h a partir del cual $\gamma(h)$ se vuelve estacionario.
3. **Meseta:** es el valor máximo del semivariograma que indica la varianza espacial total, es la suma del efecto pepita más la varianza espacial.

En la Figura 4.1 se presenta el ejemplo de un variograma experimental, señalando los parámetros que lo determinan. En este caso el efecto nugget no está presente, se puede observar que para $h = 0$, $\gamma(h) = 0$. Posee meseta y rango ya que $\gamma(h)$ se vuelve estacionario en $h = a$. Cabe recalcar que no todos los variogramas experimentales alcanzan el valor de la meseta.

Figura 4.1: Parámetros del variograma experimental



Fuente: (Armstrong, 1998)

El objetivo ahora es encontrar un modelo teórico que permita la caracterización óptima de la autocorrelación espacial, en función de la distancia,

de manera que se puedan obtener buenas estimaciones de los parámetros que describen al variograma experimental y en caso de ser necesario hacer predicciones con buenos ajustes.

Supuestos

- Estacionariedad de segundo orden:

$$E(Z(x_i)) = \mu, \quad \text{cov}(Z(x_i), Z(x_j)) = C(x_i - x_j).$$

- Isotropia:

$$\text{cov}(Z(x_i), Z(x_j)) = C(\|x_i - x_j\|).$$

- Estacionariedad intrínseca:

$$E(Z(x_i)) = \mu, \quad \text{var}(Z(x_i) - Z(x_j)) = 2\gamma(x_i - x_j).$$

- Estacionariedad implica estacionariedad intrínseca

$$\gamma(x_i - x_j) = C(0) - C(x_i - x_j).$$

- Estacionariedad intrínseca no necesariamente implica estacionariedad.

Medidas de aleatoriedad

Algunas veces nos interesa medir el grado de aleatoriedad que presentan los datos. Para este fin en (Sierra, 2014) se define el valor $E = C_0/C_t$, donde:

- C_0 : corresponde al valor del efecto pepita.
- C : es el valor de la meseta.
- C_t : es el valor de la varianza espacial, $C_t = C - C_0$.

Este se se mide según el siguiente criterio:

- $E < 0.15$, indica aleatoriedad baja.
- $0.15 \leq E \leq 0.30$, indica aleatoriedad media.
- $E > 0.30$, indica aleatoriedad alta.

Por su parte $IED = C_0/(C_t + C_0)$ es el Índice de Dependencia Espacial y nos dice cual es la contribución de la parte aleatoria de la dependencia espacial respecto de la dependencia total. Cuanto menor sea el IDE, mayor es el grado de dependencia espacial.

Modelos teóricos de variograma

A partir del variograma experimental se deben ajustar modelos matemáticos. El modelo teórico del variograma debe satisfacer ciertas condiciones, para no terminar con una varianza negativa para el valor estimado, por lo tanto lo mejor es ajustar el variograma experimental a ciertos variogramas modelo que cumplan con estas propiedades.

Para el ajuste del variograma experimental se puede escoger entre los siguientes modelos teóricos que en (Giraldo, 2002), se definen de la siguiente manera:

- **Modelo esférico:** Este modelo corresponde a los fenómenos más frecuentemente observados y por ello es ampliamente utilizado. El comportamiento es de una tendencia casi lineal hasta que se alcanza el rango en donde se estabiliza el fenómeno (meseta). Las ecuaciones correspondientes son:

$$\gamma(h) = C \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right], \quad \text{si } h < a.$$

$$\gamma(h) = C, \quad \text{si } h > a,$$

donde C es el valor de la meseta, h la distancia y a el rango.

- **Modelo exponencial:** Tiene un comportamiento muy similar al esférico, la diferencia con aquel radica en que nunca se alcanza el valor de la meseta, pues su rango práctico a llega sólo a 95 por ciento. La ecuación correspondiente es:

$$\gamma(h) = C \left[1 - \exp \left(-\frac{3h}{a} \right) \right]. \quad (4.1)$$

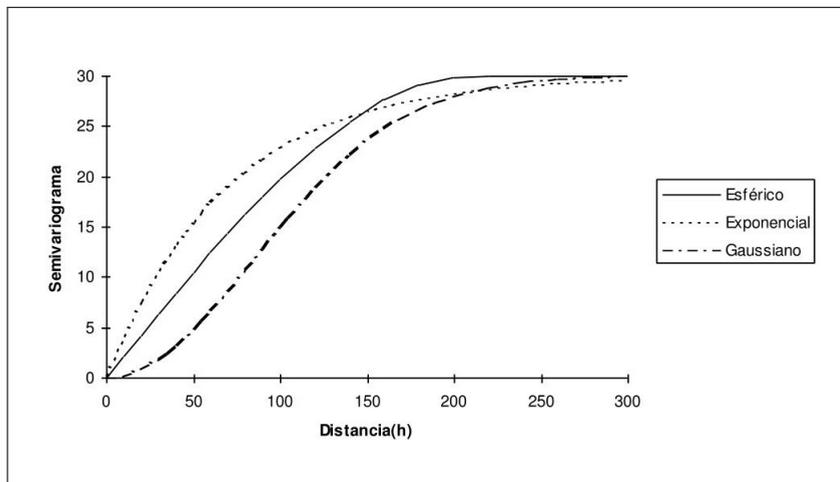
- **Modelo gaussiano:** Se utiliza cuando el fenómeno de estudio es

extremadamente continuo, su rango práctico es a . La ecuación correspondiente es:

$$\gamma(h) = C \left[1 - \exp\left(-\frac{3h^2}{a^2}\right) \right].$$

En la Figura 4.2, se presenta una comparación gráfica de los modelos teóricos del variograma experimental. Se puede observar que en este ejemplo en particular se tiene un rango, en el caso del modelo esférico, y un rango efectivo en el caso de los modelos exponencial y gaussiano. El valor que corresponde a la meseta es 30, por otra parte el 95 % de la meseta es igual a 28.5. Este variograma no posee efecto nugget.

Figura 4.2: Comparación gráfica de modelos teóricos de variograma



Fuente: (Giraldo, 2002)

En el modelado generalmente se supone que los fenómenos son isotrópicos, es decir, que el comportamiento es el mismo en cualquier dirección. Sin embargo no siempre sucede así, en esos casos se requiere el modelado de variogramas direccionales, en donde además de la distancia y la tolerancia de la distancia, se debe definir una tolerancia angular.

Procesos Gaussianos

En el contexto espacial un modelo estocástico subyacente común, es que todos los datos vienen de una distribución conjunta Gaussiana, cuya estructura depende de las ubicaciones de los datos y para describirlos basta con estimar los parámetros del modelo Gaussiano.

Estos procesos son importantes porque las predicciones y estimaciones son en teoría más fáciles de realizar, además poseen la ventaja de poder utilizar la teoría asintótica y la aplicación del Teorema del limite Central.

4.2. Precipitación

Definición 4.4. Latitud: *La latitud es la distancia desde un punto determinado del planeta a cualquier punto del Ecuador.*

Definición 4.5. Longitud: *La longitud es la distancia la distancia desde un punto determinado del planeta a cualquier punto del meridiano de Greenwich o meridiano 0.*

El sistema de coordenadas que componen la latitud y longitud, es conocido como Sistema de Coordenadas Geográficas y expresa sus valores en grados, minutos y segundos, acompañados de una letra que indica la orientación cardinal dentro del globo: Norte, Sur, Este u Oeste.

Definición 4.6. Altitud: *Es la distancia de un punto en relación al nivel del mar.*

Definición 4.7. Precipitación: *Se conoce como precipitación, a la cantidad de agua que cae a la superficie terrestre proveniente de la humedad atmosférica, ya sea en estado líquido (llovizna y lluvia) o en estado sólido (nieve, granizo).*

4.2.1. Generalidades

La precipitación es uno de los procesos meteorológicos más importantes, por lo que es fundamental poder cuantificar su intensidad y su comportamiento a través de los días, meses y años. El poseer un registro de las cantidades de precipitación a través del tiempo es útil en diversas áreas

como la agronomía, la ingeniería, la biología y en resumen para la vida en general.

El instrumento que se utiliza para medir la precipitación, se denomina pluviómetro. El volumen de lluvia recogida se mide en litros por metro cuadrado (l/m^2), o lo que es lo mismo, en milímetros ($mm.$).

Las lluvias según su intensidad se clasifican en:

- **Débil:** intensidad ≤ 2 $mm/hora$.
- **Moderada:** 2 $mm/hora$ < intensidad ≤ 15 $mm/hora$.
- **Fuerte:** 15 $mm/hora$ < intensidad ≤ 30 $mm/hora$.
- **Muy fuerte:** 30 $mm/hora$ < intensidad ≤ 60 $mm/hora$.
- **Torrenciales:** intensidad > 60 $mm/hora$.

4.2.2. Altitud y precipitación media como covariables en precipitación

En nuestros modelos se habla de incluir covariables que influyen en los valores de precipitación. La precipitación como fenómeno está relacionados con muchas variables, en nuestro caso nos centramos en 2 de ellas:

- Altitud
- Precipitación media

Altitud

La altitud es una característica geográfica, influye en la temperatura y por ende en la precipitación. A medida que la altitud va aumentando la temperatura tiende a disminuir, esto se debe a que en las zonas con menor altitud el aire es más denso. Por otra parte en las zonas con mayor altitud, a medida que las masas de aire ascienden, favorecen a la condensación del agua y la producción de lluvias.

Tenemos datos tomados de estaciones meteorológicas, ubicadas en distintos puntos geográficos del país. Cada estación posee un valor de altitud asociado, sabemos que a medida que la altitud aumenta, aumentan también los

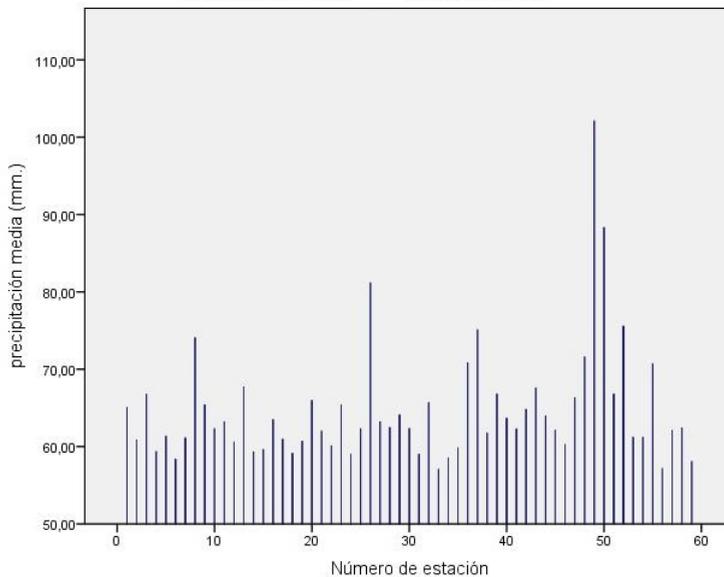
valores de precipitación, este efecto será medido e incluido en los modelos a considerar.

Precipitación media

Los valores de precipitación diaria, presentan variaciones en el tiempo debido a la estación del año, la temperatura, entre otros. En algunos meses los valores de precipitación tienden a ser más altos, lo mismo ocurre para las distintas estaciones.

Las estaciones meteorológicas se ubican en distintos puntos geográficos, al ser la precipitación un fenómeno espacial, se presentan variaciones en sus valores debido a la ubicación. Puede ocurrir que estaciones meteorológicas con un promedio de precipitación más alto, presenten este mismo comportamiento a través del tiempo. Este efecto será tomado en cuenta en los modelos para medir si en realidad el hecho de incluir o no esta variable, resulta ser significativo.

Figura 4.3: Valores de precipitación promedio por estación



En la Figura 4.3, se muestran los valores de precipitación media para cada una de las 59 estaciones involucradas en el análisis. Los valores corresponden al periodo de 1972-2012.

4.2.3. Cambio climático y su relación con los extremos en precipitación

Nuestro interés en los valores extremos de precipitación, nos lleva a mencionar un poco de como estos fenómenos están íntimamente ligados con el cambio climático. Según algunos de los resultados obtenidos en modelos desarrollados por el grupo de expertos internacional, que estudia el cambio climático, llamado Panel Intergubernamental para el Cambio Climático (IPCC es su acrónimo en inglés) presentados en su informe del 2022 (IPCC, 2022), se puede enunciar que en términos de precipitación debido al cambio climático:

- Se espera que a lo largo del siglo XXI, las sequías e inundaciones podrían ser más frecuentes y prolongadas.
- Se esperan incrementos en los actuales niveles de evaporación y precipitación.
- Un aumento muy probable de las precipitaciones intensas en regiones donde la precipitación es mayor.
- El suministro de agua potable podría peligrar en diversas zonas y algunas epidemias podrían extenderse más fácilmente.

Todos estos cambios son claramente eventos extremos en precipitación y pueden ser abordados mediante la teoría de valores extremos, con la finalidad de aportar información para la creación de modelos que busquen la mitigación mediante la gestión de riesgos como apoyo a la toma de decisiones de gobiernos y de la población en general.

4.2.4. Precipitación en Honduras

Honduras es un país muy montañoso, debido a esto la distribución de los valores de precipitación es variable. El viento asciende por las montañas

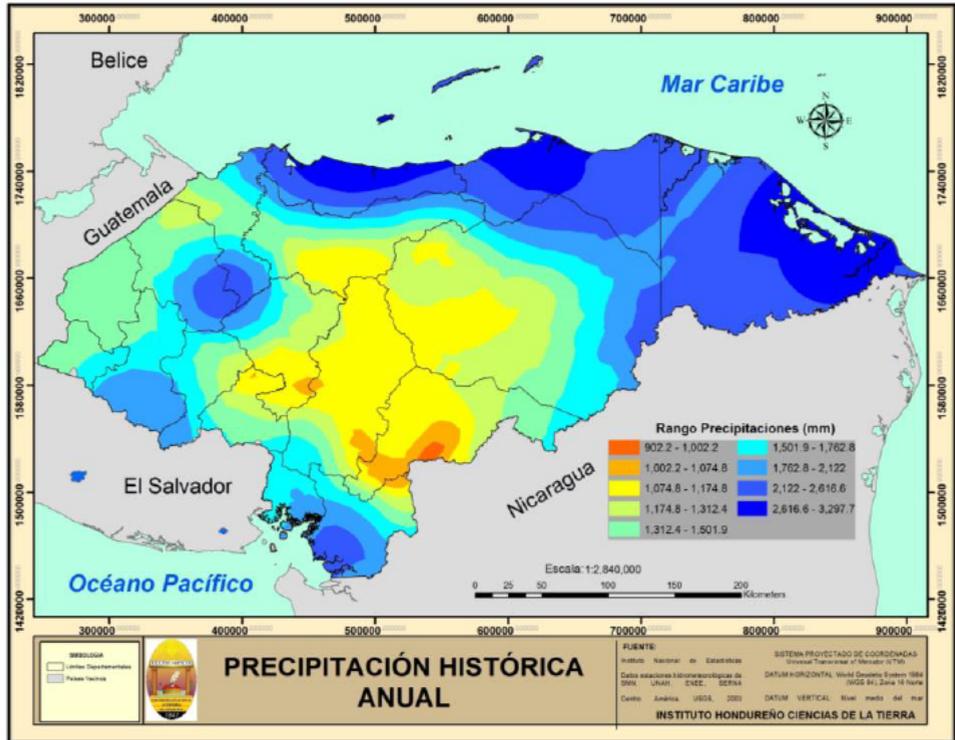
por efecto mecánico, este se enfría cargado de humedad luego se condensa, dando lugar a fuertes lluvias. A sotavento de los vientos Alisios del Noreste, la lluvia disminuye ostensiblemente, al recalentarse adiabáticamente en su continuo descenso (lluvias de 800 milímetros anuales). Mientras que en ambas fachadas marítimas (Caribe y Pacífico) las lluvias que provoca el aire tropical marítimo son cuantiosas, por encima de los 2500 mm.

Por otra parte, Honduras es influenciado por el fenómeno de El Niño y La Niña, el cual tiene un impacto muy importante en la distribución de las lluvias tanto espacial como temporal, presentando un fuerte déficit de lluvia en las regiones sur, suroccidente, oriente y central en su etapa cálida. Este comportamiento cambia con condiciones frías (La Niña) o neutras, donde la precipitación puede reflejarse entre el promedio normal o arriba de lo normal en la mayor parte del país así se enuncia en (Capel, 1994). Para encontrar más detalles respecto a la climatología y gestión de riesgo en Honduras puede consultar (Rodríguez, Kawas, Elvir, Torres, Wiese, Cardona, Canales, Gomez, Mejia, Sevilla, Mendoza, and Castillo, 2012).

En el mapa mostrado en la Figura 4.4 se presentan los valores de precipitación promedio anual en Honduras. El color azul corresponde a los valores de precipitación promedio anual más altos, estos se ubican en su mayoría en las regiones nororiental y sur podemos identificar también algunos valores en la parte sur de los departamentos de Cortés y Santa Bárbara. Los valores más bajos en su mayoría se ubican en la parte central de Honduras. En este mapa se puede observar la variabilidad presente en los valores de precipitación promedio para todo el país a lo largo de los años y como incluso en zonas cercanas el comportamiento de los valores de precipitación es diferente.

En nuestro estudio no se analizan todos los datos disponibles, ya que en Honduras hay meses donde los valores de precipitación son bajos. Los valores de precipitación diaria correspondientes a estos meses no son incluidos en los modelos, en términos generales no son excedencias, son valores que quedan por debajo del umbral. Los valores que se incluyen son los que corresponden a los meses de mayo a noviembre que son los que consideramos los meses más lluviosos en Honduras tal como se ilustra en la Figura 4.5.

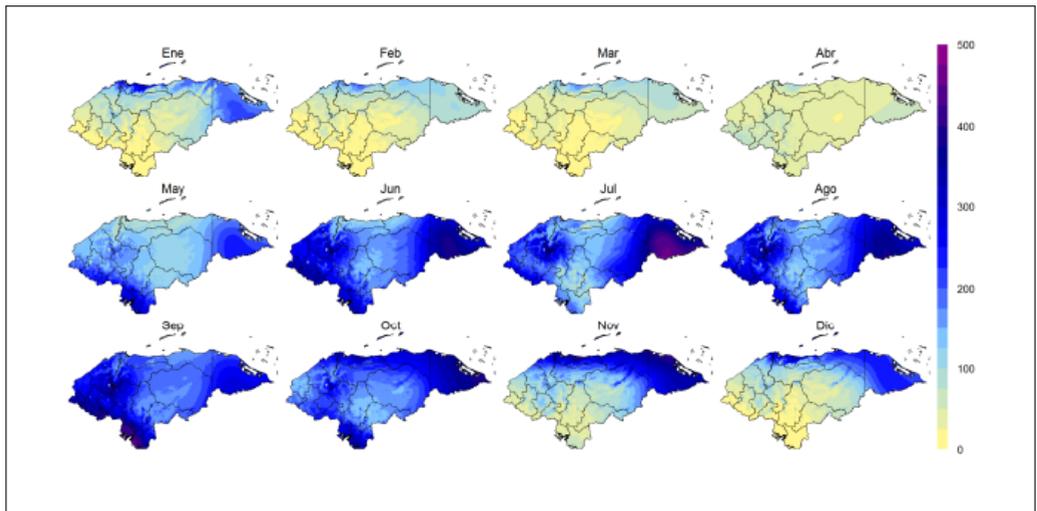
Figura 4.4: Precipitación histórica anual en Honduras



Fuente: (Rodríguez, Kawas, Elvir, Torres, Wiese, Cardona, Canales, Gomez, Mejia, Sevilla, Mendoza, and Castillo, 2012)

Todos estos elementos particulares de la precipitación en Honduras son muy importantes porque están relacionados directamente con los resultados obtenidos y las conclusiones de este trabajo. En el capítulo 5, describimos los modelos propuestos y la metodología para poder hacer las estimaciones de los niveles de retorno de precipitación y se presenta el criterio usado para seleccionar los mejores modelos.

Figura 4.5: Precipitación mensual acumulada de referencia



Mapas derivados de la interpolación de datos de estaciones meteorológicas de diversas fuentes. Los valores están dados en mm/mes.

Fuente: (Navarro, Llano, Obando, and D. Córdoba, 2018)

Capítulo 5

Descripción de la metodología

5.1. Preliminares

Nuestra finalidad es poder estimar distintos niveles de retorno para la precipitación diaria. Para lograr alcanzar el objetivo planteado, necesitamos estimar los parámetros de la GPD y las tasas de excedencia; como se describió en la ecuación 3.6, ya que los niveles de retorno son función de dichos parámetros.

Basándonos en el modelo usado en (Cooley, Nychka, and Naveau, 2007), artículo usado como base para el desarrollo de este trabajo. El modelo completo se aborda en 2 etapas:

- **Etapa 1:** asociada a los parámetros de la Distribución Pareto generalizada (GPD).
- **Etapa 2:** asociada a las tasas de excedencia.

A cada etapa le asociamos un modelo jerárquico, como el descrito en la sección 2.7. Estos modelos asociados a cada etapa son independientes entre sí, debido a la naturaleza de los parámetros. Los números asociados a cada etapa son usados solo para distinguirlas, el orden para abordarlas es

indistinto.

Cada modelo jerárquico esta compuesto por 3 niveles:

- **Nivel 1:** Nivel de los datos.
- **Nivel 2:** Nivel del proceso latente que controla la precipitación climatológica extrema.
- **Nivel 3:** Nivel de las distribuciones a priori de los parámetros involucrados en el proceso latente.

La inferencia sobre los parámetros θ , dados los datos de las estaciones meteorológicas $\mathbf{Z}(\mathbf{x})$, se realiza haciendo uso del Teorema de Bayes descrito en la definición 2.11 :

$$p(\theta|\mathbf{Z}(\mathbf{x})) \propto p(\mathbf{Z}(\mathbf{x})|\theta)p(\theta). \quad (5.1)$$

La distribución condicional de nuestros modelos jerárquicos queda descrita por:

$$p(\theta|\mathbf{Z}(\mathbf{x})) \propto p_1(\mathbf{Z}(\mathbf{x})|\theta_1)p_2(\theta_1|\theta_2)p_3(\theta_2), \quad (5.2)$$

donde p_j es la densidad asociada con el nivel j , en nuestro caso $j = 1, 2, 3$.

Para la construcción del modelo completo abordamos primero la etapa 1, el modelo jerárquico asociado se describe en las sección 5.2 y posteriormente la etapa 2, cuyo modelo es descrito en la sección 5.3.

5.2. Modelo jerárquico para excedencia de umbrales

Este es el modelo correspondiente a la primera etapa del modelo completo, consta de 3 niveles que se describen a continuación:

Nivel 1

Nivel base del modelo jerárquico asociado a la excedencia de umbrales, descrito por la GPD definida en la ecuación 3.5. Dadas las ubicaciones de

las s estaciones meteorológicas x_1, \dots, x_s denotamos los parámetros de la GPD, asociados a cada estación por:

$$\begin{aligned}\sigma_u &= [\sigma_u(x_1), \dots, \sigma_u(x_n)]^T, \\ \xi &= [\xi(x_1), \dots, \xi(x_n)]^T.\end{aligned}$$

Reparametrizamos la GPD haciendo $\phi = \log(\sigma_u)$, de esta manera permitimos que el parámetro ϕ tome valores positivos y negativos.

Sea $Z_k(\mathbf{x}_i)$ la k -ésima cantidad de precipitación diaria que excede el umbral u y fue registrada en la estación meteorológica \mathbf{x}_i con $i = 1, \dots, s$. Dado que $Z_k(\mathbf{x}_i)$ es una excedencia, asumimos que puede ser descrita mediante una GPD cuyos parámetros dependen de la ubicación de la estación de manera que:

$$p\{Z_k(\mathbf{x}_i) - u > z | Z_k(\mathbf{x}_i) > u\} = \left(1 + \frac{\xi(\mathbf{x}_i)z}{\exp\phi(\mathbf{x}_i)}\right)^{-1/\xi(\mathbf{x}_i)} \quad (5.3)$$

derivamos la función de distribución asociada a la ecuación 5.3 para obtener la función de densidad:

$$p_1(Z_k(\mathbf{x})|\theta_1) = \prod_{i=1}^s \prod_{k=1}^{n_i} \frac{1}{\exp\phi(\mathbf{x}_i)} \left(1 + \frac{\xi(\mathbf{x}_i)z}{\exp\phi(\mathbf{x}_i)}\right)^{-1/\xi(\mathbf{x}_i)-1}, \quad (5.4)$$

donde $\theta_1 = [\phi, \xi]^T$.

La ecuación 5.4 nos da la primera parte de la ecuación 5.2 asociada al nivel 1 de la etapa 1.

Nivel 2

En este nivel se busca describir la relación existente entre el nivel de los datos, con la orografía y climatología de las estaciones meteorológicas que son las que caracterizan el proceso espacial latente.

Esperaríamos que de acuerdo a la ubicación de las estaciones meteorológicas, se presenten variaciones en los valores extremos de precipitación diaria, debido a factores como ser el clima, la altitud, la temperatura, la ubicación geográfica como tal, entre otros. Las variaciones que logremos identificar,

no pueden ser explicadas completamente por las covariables y sus funciones. Para lograr capturar de forma eficiente ese efecto utilizamos métodos de la estadística espacial.

Sea $\phi(\mathbf{x})$ el parámetro log-transformado de escala y $\xi(\mathbf{x})$ el parámetro de forma del proceso. Estos parámetros son tratados como variables aleatorias, cada uno tiene una distribución a priori asociada.

Para describir el comportamiento de $\phi(\mathbf{x})$, tomando en cuenta su relación con el proceso espacial latente, hacemos uso de métodos estándar en geoestadística descritos en la subsección 4.1.2. Modelamos $\phi(\mathbf{x})$ mediante un proceso Gaussiano con:

$$\text{media: } E[\phi(\mathbf{x})] = \mu_\phi(\mathbf{x}),$$

$$\text{covarianza: } \text{cov}(\phi(\mathbf{x}), \phi(\mathbf{x}')) = k_\phi(\mathbf{x}, \mathbf{x}').$$

La media es función de los parámetros α_ϕ y las covariables:

$$\mu_\phi(\mathbf{x}) = f_\phi(\alpha_\phi, \text{covariables}). \quad (5.5)$$

La función f podemos modificarla para lograr incorporar relaciones entre las diferentes covariables involucradas en el modelo.

La covarianza es función de la distancia entre las estaciones meteorológicas y los parámetros β_ϕ , está dada por:

$$k_\phi(\mathbf{x}, \mathbf{x}') = \beta_{\phi,0} \times \exp(-\beta_{\phi,1} \times \|\mathbf{x} - \mathbf{x}'\|) \quad (5.6)$$

que corresponde al modelo teórico de un variograma exponencial definido por la ecuación 4.1.

Tenemos que:

- $\beta_{\phi,0}$: meseta.
- $\beta_{\phi,1}$: rango.

Nuestro modelo de covarianza asume que el proceso es isotrópico y estacionario, es poco probable detectar cualquier anisotropía o no estacionariedad debido a el número de estaciones meteorológicas que estamos considerando.

Escogemos como modelo teórico al modelo exponencial por su simplicidad y porque no asume ningún nivel de suavidad en el proceso latente.

Con respecto a $\xi(\mathbf{x})$ de manera similar a lo que se describe en (Cooley, Nychka, and Naveau, 2007), no sabemos como se comporta este parámetro ni su sensibilidad a las variables regionalizadas. Es difícil estimar $\xi(\mathbf{x})$, por esta razón lo tratamos como un valor único, sin tomar en cuenta las variaciones debido a la ubicación de las estaciones meteorológicas añadiendo posteriormente mas complejidad hasta lograr un buen ajuste. De manera que el parámetro de forma es modelado como un valor único para todas las estaciones, con distribución a priori $U(-\infty, \infty)$.

1. Un valor único para todas las estaciones, con distribución a priori $U(-\infty, \infty)$.
2. Como un proceso gaussiano con estructura similar a la priori para $\phi(\mathbf{x})$.

Tomando en cuenta la información que la ubicación de las estaciones nos proporciona acerca de los datos y del proceso espacial latente, tenemos:

$$p_2(\theta_1|\theta_2) = \frac{1}{\sqrt{(2\pi)^s|\Sigma|}} \exp \left[-\frac{1}{2}(\phi - \mu)^T \Sigma^{-1} (\phi - \mu) \right] \times p_\xi(\xi|\theta_\xi) \quad (5.7)$$

donde μ es un vector definido por la ecuación 5.5 evaluado en las covariables de la estación \mathbf{x}_i , Σ es la matriz de covarianza generada por la ecuación 5.6 en la ubicación de las estaciones, p_ξ viene de la distribución a priori escogida para ξ con parámetro θ_ξ , $\theta_1 = [\phi, \xi]^T$ y $\theta_2 = [\alpha_\phi, \beta_\phi, \theta_\phi]^T$.

La ecuación 5.7 nos proporciona la segunda parte del modelo completo, descrito en la ecuación 5.2, asociada al nivel 2 de esta primera etapa.

Nivel 3

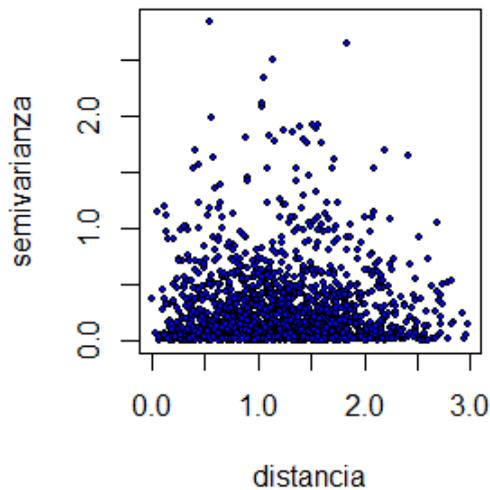
Nivel de las distribuciones a priori, en este le asignamos las distribuciones a priori a todos los parámetros involucrados en el modelo. Tenemos que α_ϕ , β_ϕ y θ_ϕ son los parámetros caracterizan el proceso latente, asumimos que cada parámetro en esta etapa es independiente de los otros.

No poseemos información a priori de como el parámetro ϕ se relaciona con las covariables, por esa razón escogemos prioris no informativas para los

parámetros de regresión α_ϕ , asumimos que $\alpha_{\phi,i} \sim U(-\infty, \infty)$ para todos los modelos.

Los parámetros β_ϕ son los que describen la estructura espacial del parámetro de escala log-transformado de la GPD, para estos es aconsejable usar prioris informativas. Queremos estimar estos parámetros en nuestros datos, para ese fin utilizamos el paquete GeoR que nos proporciona herramientas para el análisis de datos geoestadísticos en R.

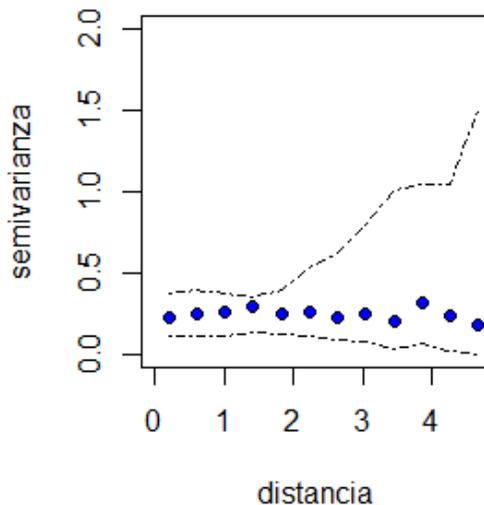
Figura 5.1: Nube variografica asociada a los valores de excedencias máximas



En la Figura 5.1, se muestra la nube de variogramas empíricos asociados a los datos. A cada estación le asociamos el valor del logaritmo de excedencia máxima, medida en centímetros, observamos que la distancia máxima entre dichos valores asociados a las estaciones es de 20.09 cm., la mayoría de los valores de la semivarianza oscilan entre 0 y 4.48 cm. aproximadamente.

A nuestro variograma empírico le asociamos un modelo teórico, con el fin de poder estimar los parámetros de covarianza, ajustamos el variograma experimental mediante un modelo exponencial, tomando en consideración el efecto nugget.

Figura 5.2: Ajuste del variograma experimental de los datos mediante el modelo exponencial



En la Figura 5.2, las líneas punteadas representan las envolventes correspondientes al modelo exponencial y los puntos corresponden al variograma experimental omnidireccional. Basándonos en este gráfico y retomando que $\beta_{\phi,0}$ es el parámetro asociado a la meseta, según los valores observados, tenemos que $\beta_{\phi,0}$ toma valores entre 0.01 y 0.4 aproximadamente de igual manera el parámetro asociado al rango $\beta_{\phi,1}$. Se puede observar en el variograma que tenemos presente el efecto nugget

Con la información obtenida del variograma exponencial definimos las distribuciones a priori de la siguiente manera:

$$\beta_{\phi,0} \sim U(0.01, 0.4),$$

$$\beta_{\phi,1} \sim U(0.01, 1.99).$$

Respecto al parámetro de forma $\xi(x)$, solo le asociamos distribuciones a priori cuando este es modelado mediante un proceso Gaussiano, en este caso $\xi(x)$ tiene coeficientes α_{ξ} y parámetros espaciales β_{ξ} . Para las prioris

de α_ξ tomamos $U(-\infty, \infty)$ y usamos información empírica para determinar priors apropiadas para β_ξ .

5.3. Modelos de tasas de excedencias

Segunda etapa de nuestro modelo completo, esta es independiente de la primera etapa. Para estimar completamente los niveles de retorno, debemos incluir un parámetro adicional que denotaremos como ζ_u . Donde ζ_u representa la tasa a la que un grupo de observaciones supera el umbral u . Las tasas de excedencia asociadas a las estaciones meteorológicas las denotamos por:

$$\zeta_u = [\zeta_u(x_1), \dots, \zeta_u(x_n)]^T.$$

Inicialmente nuestros datos están desagrupados, no tenemos la probabilidad de que una observación exceda el umbral u . En nuestros datos tenemos de momento que ζ_u es la probabilidad que una observación sea un máximo de grupo. Por la forma en como están divididos nuestros datos y dada la elección previa del umbral, ζ_u varia para cada estación y puede ser modelada espacialmente.

Para escoger el valor del umbral tomamos el percentil 97 para cada una de las estaciones meteorológicas incluidas en el estudio y con ellos calculamos el promedio. El valor de umbral escogido fue de 45mm. (4.5 cm) para todas las estaciones. La tasa de excedencia ζ_u cambia de acuerdo a la estación que se tome, por ello debe ser modelada espacialmente. Denotamos al parámetro de la tasa de excedencia en la estación \mathbf{x} por $\zeta(\mathbf{x})$.

De manera análoga a lo desarrollado en la etapa 1, para poder obtener inferencias sobre ζ_u utilizamos un modelo jerárquico con 3 niveles, teniendo siempre en cuenta que hay un proceso espacial latente que influye en las probabilidades de excedencia y cuyo efecto debe ser incluido y modelado.

Nivel de los datos

Primer nivel, el nivel base de esta etapa, para cada estación calculamos el número de excedencias y lo denotamos como n_i . El número de excedencias

total para cada estación sigue una distribución binomial $\text{Bin}(m_i, \zeta(\mathbf{x}_i))$, con:

- m_i : número total de observaciones por estación.
- $\zeta(\mathbf{x}_i)$: probabilidad de ser un grupo máximo.

Nivel del proceso

Segundo nivel del modelo jerárquico, en este es donde modelamos el proceso espacial latente, lo trabajamos de manera similar a como se trabajo el parámetro $\phi(\mathbf{x})$ en la GPD del nivel 1 en la etapa 1. Siguiendo la línea de (Diggle and Tawn, 1998) transformamos nuestro parámetro $\zeta(\mathbf{x})$ usando una transformación logit, esta se describe mediante la ecuación 2.1, para modelar este parámetro transformado como un proceso Gaussiano con $E[\zeta(\mathbf{x})] = \mu_\zeta(\mathbf{x})$ y $\text{cov}(\zeta(\mathbf{x}), \zeta(\mathbf{x}')) = k_\zeta(\mathbf{x}, \mathbf{x}')$.

Donde $\mu_\zeta(\mathbf{x})$ es función de α_ζ y la información de las covariables, la covarianza es función de los β_ζ obtenidos a partir del variograma exponencial.

Nivel de las prioris

Tercer nivel de nuestra segunda etapa, nivel donde asociamos las distribuciones a priori asociadas a los parámetros involucrados. Usaremos prioris no informativas para los parámetros de regresión, $\alpha_\zeta \sim U(-\infty, \infty)$. Basándonos en el variograma mostrado en la Figura 5.2, tomamos $\beta_{\zeta,0} \sim U[0.01, 0.4]$ y $\beta_{\zeta,1} \sim U[0.01, 1.99]$.

5.4. MCMC en los modelos

Para obtener muestras en ambas etapas simulamos la distribución posterior haciendo uso de los algoritmo MCMC: Metrópolis Hastings y Muestreador de Gibbs descritos en la subsección 2.5.2.

Para cada parámetro involucrado calculamos las distribuciones condicionales completas, ya que son las que nos permitirán simular muestras para los mismos, sin perder de vista que estamos usando modelos jerárquicos y los

mismos nos marcan el orden en el cual simulamos los parámetros. Generamos las propuestas usando caminata aleatoria y propuestas independientes con el objetivo de alcanzar tasas de aceptación adecuadas.

Para acelerar la convergencia, cuando actualizamos ϕ , usamos el estimador de máxima verosimilitud de los parámetros de la GPD de cada estación para obtener un candidato adecuado. Sea $\hat{\phi}$ el estimador de máxima verosimilitud del parámetro ϕ de la GPD. Por las propiedades asintóticas del estimador de máxima verosimilitud,

$$\hat{\phi} \approx \phi + \epsilon,$$

donde $\epsilon \sim MVN(\mu, \Sigma)$ y μ es definida por la ecuación 5.5 y Σ por la ecuación 5.6.

Podemos denotar la distribución conjunta de $\hat{\phi}$ y ϕ por:

$$\begin{pmatrix} \hat{\phi} \\ \phi \end{pmatrix} = MVN \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{bmatrix} \Sigma + \mathcal{I}^{-1} & \Sigma \\ \Sigma & \Sigma \end{bmatrix} \right) \quad (5.8)$$

la distribución condicional es:

$$\phi | \hat{\phi} \sim MVN(\mu + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(\hat{\phi} - \mu), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma). \quad (5.9)$$

Usamos la ecuación 5.9 como la densidad del candidato en nuestro paso de Metropolis Hastings. Esta elección se hace porque en muestras grandes la inferencia del estimador de máxima verosimilitud, debería ser cercana a la posteriori Bayesiana. El muestreo de la distribución para el estimador máximo verosímil debería proporcionarnos una buena distribución candidata para esta parte de la posterior.

Después del parámetro ϕ , actualizamos la media α_ϕ y β_ϕ . La densidad candidata de Metropolis Hastings para α_ϕ y β_ϕ son implementadas como caminatas aleatorias.

Para el parámetro de forma, repetimos el proceso, actualizamos ξ como una caminata aleatoria cuando lo modelamos, ya sea como un valor único o como un par de valores y de la misma manera que con ϕ cuando lo

modelamos espacialmente. La tasa de excedencia se maneja de manera análoga.

Las simulaciones las realizamos usando el programa estadístico R. En cada simulación realizamos 100,000 iteraciones, posteriormente realizamos un *burn-in* de 1,000 iteraciones. Las cadenas resultantes mostraban dependencia, por lo que para reducir la misma aplicamos un *thinning* promedio de 30. Después de haber realizado *burn-in* y el *thinning* obtuvimos muestras finales de 3,300 valores aproximadamente.

Sean $Z(x_i)$ los datos de las estaciones meteorológicas, con $\{x_1, x_2, \dots, x_s\}$ las ubicaciones de las s estaciones, en nuestro caso $s = 59$. Por simplicidad al especificar los modelos denotaremos $Z(x_i)$ mediante Z .

Para cada estación calculamos: $\phi(x_i)$ y $\xi(x_i)$.

Para poder hacer el análisis bayesiano y poder aplicar los algoritmos MCMC, es necesario calcular las distribuciones condicionales completas para poder simular las distribuciones a posteriori por cada modelo. A continuación, presentamos las condicionales completas para cada modelo:

5.4.1. Etapa 1

Modelo 0

$$\phi(\mathbf{x}_i) = \phi$$

$$\xi(\mathbf{x}_i) = \xi$$

$$\xi|Z, \phi \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1}$$

$$\phi|Z, \xi \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1}$$

Donde n_i es el número de observaciones que son excedencias para cada una de las estaciones meteorológicas involucradas en el estudio con $i = 1, \dots, 59$.

Para este primer modelo donde ambos parámetros son tratados como constantes, simulamos primero ξ dado que este parámetro presenta sensibilidad. Con la estimación obtenida para ξ simulamos ϕ .

Modelo 1

$$\phi = \alpha_0 + \epsilon_\phi$$

$$\xi(\mathbf{x}_i) = \xi$$

El parámetro ϕ es definido mediante un modelo de regresión que depende de α_0 . Debemos simular primero α_0 para obtener los valores de ϕ asociados a cada estación. Como en este modelo ya se incluye la componente espacial, debemos simular β_0 y β_1 que son los parámetros que modelan esta componente. El orden para simular las muestras asociadas a cada parámetro es análoga a la del modelo anterior.

$$\phi|\hat{\phi} \sim MVN(\mu + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(\hat{\phi} - \mu), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma),$$

$$\alpha_0|Z, \phi, \xi, \beta_0, \beta_1 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1} * \exp[-1/2(\phi - \mu)^T \Sigma^{-1}(\phi - \mu)],$$

$$\begin{aligned} \beta_0, \beta_1|Z, \phi, \xi, \alpha_0 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1} * \frac{1}{|\Sigma|^{1/2}} \exp[-1/2(\phi - \mu)^T \Sigma^{-1}(\phi - \mu)] \\ * \mathbb{1}_{[0.01, 0.4]}(\beta_0) * \mathbb{1}_{[0.01, 1.99]}(\beta_1). \end{aligned}$$

Para los parámetros β_0 y β_1 usamos valores de propuestas con distribución uniforme, así garantizamos que los valores simulados se mantengan en los intervalos que definimos, basándonos en el variograma presentado en la Figura 5.2. Para los demás parámetros usamos valores de propuesta utilizando caminata aleatoria, con los valores de las varianzas de estos controlamos las tasas de aceptación, siempre asegurándonos de obtener convergencia en los parámetros.

Finalmente simulamos de manera independiente ξ , usamos caminata aleatoria para los valores de propuesta. Por la forma en como lo definimos, este se comporta de manera constante para todas las estaciones meteorológicas.

$$\xi|Z, \phi, \alpha_0, \beta_0, \beta_1 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1}.$$

En este modelo tenemos que $\mu = E[\alpha_0 + \epsilon_\phi]$, $\mu = \alpha_0$. En cada iteración calculamos el valor de μ asociado a cada una de las estaciones meteorológicas.

Modelo 2

$$\phi = \alpha_0 + \alpha_1 * \mathbf{mp} + \epsilon_\phi$$

$$\xi(\mathbf{x}_i) = \xi$$

El parámetro ϕ ahora se describe mediante un modelo de regresión con parámetros α_0 y α_1 , para estimar ϕ necesitamos simular estos valores primero. Por la forma en como definimos ϕ para cada estación obtendremos valores diferentes, debido a la influencia de los valores de α_0 y α_1 . El orden en el que simulamos y las distribuciones condicionales completas asociadas a cada parámetro involucrado en este modelo se muestran a continuación.

$$\phi|\hat{\phi} \sim MVN(\mu + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(\hat{\phi} - \mu), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma),$$

$$\alpha_0|Z, \phi, \xi, \beta_0, \beta_1 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1} * \exp[-1/2(\phi - \mu)^T \Sigma^{-1}(\phi - \mu)],$$

$$\alpha_1|Z, \phi, \xi, \beta_0, \beta_1, \alpha_0 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1} * \exp[-1/2(\phi - \alpha)^T \Sigma^{-1}(\phi - \alpha)].$$

En α_1 utilizamos también valores de propuesta con caminata aleatoria, para los demás parámetros se mantienen las condiciones descritas en el modelo anterior.

$$\beta_0, \beta_1 | Z, \phi, \xi, \alpha_0 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1} * \frac{1}{|\Sigma|^{1/2}} \exp[-1/2(\phi-\mu)^T \Sigma^{-1}(\phi-\mu)] \\ * \mathbb{1}_{[0.01,0.4]}(\beta_0) * \mathbb{1}_{[0.01,1.99]}(\beta_1).$$

Simulamos de manera independiente ξ , usando caminata aleatoria para los valores de propuesta. Este parámetro se comporta de manera constante en todas las estaciones.

$$\xi | Z, \phi, \alpha_0, \beta_0, \beta_1 \propto \prod_{i=1}^{59} \prod_{j=1}^{n_i} \frac{1}{e^\phi} \left(1 + \frac{\xi z_{i,j}}{e^\phi}\right)^{-\frac{1}{\xi}-1}.$$

En este modelo tenemos que $\mu = E[\alpha_0 + \alpha_1 * mp + \epsilon_\phi]$, $\mu = \alpha_0 + \alpha_1 * \mathbf{mp}$. Para cada iteración calculamos el valor de μ asociado a cada una de las estaciones.

Modelo 3

$$\phi = \alpha_0 + \alpha_1 * \mathbf{altitud} + \epsilon_\phi \\ \xi(\mathbf{x}_i) = \xi$$

En este modelo μ se define como: $\mu = E[\alpha_0 + \alpha_1 * \mathbf{altitud} + \epsilon_\phi]$, $\mu = \alpha_0 + \alpha_1 * \mathbf{altitud}$. El valor de μ cambia para cada estación meteorológica producto de la influencia de los parámetros α_0 , α_1 y de la altitud. Se calcula el valor de μ para cada estación.

Todas las condicionales completas y demás especificaciones se calculan de igual manera que en el modelo 2.

5.4.2. Etapa 2

Modelo A

$$\zeta = \zeta$$

$$\zeta | Z \propto \zeta^{\sum_{i=1}^s n_i} (1 - \zeta)^{\sum_{i=1}^s (m_i - n_i)} \mathbb{1}_{[0,1]}(\zeta).$$

Donde Z corresponde a los datos de precipitación diaria que son excedencias. El número de excedencias total asociada a cada estación sigue una distribución binomial que describimos en la Sección 5.3.

Calculamos para cada estación el número de excedencias n_i . En este primer modelo tomamos las tasas de excedencia como un valor que se mantiene constante para todas las estaciones.

Modelo B

$$\zeta = \alpha_0 + \epsilon_\zeta$$

En este segundo modelo para las tasas de excedencia se toma en cuenta la componente espacial, por esa razón debemos simular valores para los parámetros β_0 y β_1 . La forma de abordaje de este modelo es similar a la forma de abordar ϕ en los modelos de la etapa 1. Se definen las tasas de excedencia mediante un modelo de regresión lineal con parámetro α_0 .

Sea $g(\zeta) = \ln\left(\frac{\zeta}{1-\zeta}\right)$. El orden para simular en cada iteración y las distribuciones condicionales completas se muestran a continuación.

$$g(\zeta)|g(\hat{\zeta}) \sim MVN(\mu + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(g(\hat{\zeta}) - \mu), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma),$$

$$\alpha_0|Z, g(\zeta) \propto \left(\frac{e^\mu}{1+e^\mu}\right)^{\sum n_i} \left(1 - \left(\frac{e^\mu}{1+e^\mu}\right)\right)^{\sum_{i=1}^s (m_i - n_i)} \\ * \exp[-1/2(g(\zeta) - \mu)^T \Sigma^{-1}(g(\zeta) - \mu)].$$

Para los valores de propuesta en $g(\zeta)$ y α_0 utilizamos caminata aleatoria. Con los valores de varianza asociados a los valores de propuesta ajustamos las tasas de aceptación para lograr también la convergencia.

$$\beta_0, \beta_1|Z, g(\zeta), \alpha_0 \propto \frac{1}{|\Sigma|^{1/2}} \exp[-1/2(g(\zeta) - \mu)^T \Sigma^{-1}(g(\zeta) - \mu)] \\ * \mathbb{1}_{[0.01, 0.4]}(\beta_0) * \mathbb{1}_{[0.01, 1.99]}(\beta_1).$$

Tenemos que β_0 y β_1 son los parámetros que modelan la componente espacial asociada a las tasas de excedencia. Para garantizar que estos valores

se mantengan en los rangos que definimos para ellos, utilizamos valores de propuesta uniforme.

En cada iteración calculamos $\mu = \alpha_0$, este valor es diferente para cada estación debido a la influencia de α_0 .

Modelo C

$$\zeta = \alpha_0 + \alpha_1 * mp + \epsilon_\zeta$$

En este modelo se incluye también la componente espacial asociada a las tasas de excedencia. El parámetro asociado a las tasas de excedencia ζ se describe mediante un modelo de regresión con parámetros α_0 y α_1 . El orden para simular en cada iteración y las condicionales completas se muestra a continuación:

$$\text{Sea } g(\zeta) = \ln\left(\frac{\zeta}{1-\zeta}\right).$$

$$g(\zeta)|g(\hat{\zeta}) \sim MVN(\mu + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(g(\hat{\zeta}) - \mu), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma),$$

$$\begin{aligned} \alpha_0|Z, g(\zeta) \propto & \left(\frac{e^\mu}{1+e^\mu}\right)^{\sum n_i} \left(1 - \left(\frac{e^\mu}{1+e^\mu}\right)\right)^{\sum_{i=1}^s (m_i - n_i)} \\ & * \exp[-1/2(g(\zeta) - \mu)^T \Sigma^{-1}(g(\zeta) - \mu)], \end{aligned}$$

$$\begin{aligned} \alpha_1|Z, g(\zeta), \alpha_0 \propto & \left(\frac{e^\mu}{1+e^\mu}\right)^{\sum n_i} \left(1 - \left(\frac{e^\mu}{1+e^\mu}\right)\right)^{\sum_{i=1}^s (m_i - n_i)} \\ & * \exp[-1/2(g(\zeta) - \mu)^T \Sigma^{-1}(g(\zeta) - \mu)]. \end{aligned}$$

Para los valores de propuesta en $g(\zeta)$, α_0 y α_1 utilizamos caminata aleatoria. Con los valores de varianza asociados a los valores de propuesta ajustamos las tasas de aceptación.

$$\beta_0, \beta_1|Z, g(\zeta), \alpha_0 \propto \frac{1}{|\Sigma|^{1/2}} \exp[-1/2(g(\zeta) - \mu)^T \Sigma^{-1}(g(\zeta) - \mu)]$$

$$* \mathbb{1}_{[0.01,0.4]}(\beta_0) * \mathbb{1}_{[0.01,1.99]}(\beta_1).$$

Para β_0 y β_1 parámetros que modelan la componente espacial en este modelo, utilizamos valores de propuesta uniforme.

En cada iteración calculamos la media, $E[\zeta] = \alpha_0 + \alpha_1 * mp$, $\mu = \alpha_0 + \alpha_1 * mp$, este valor es diferente para cada estación debido a la influencia de α_0 , α_1 y la precipitación media.

Modelo D

$$\zeta = \alpha_0 + \alpha_1 * altitud + \epsilon_\zeta$$

Todas las condicionales se calculan de igual manera que en el modelo anterior. En este modelo que es análogo al modelo anterior tenemos ahora que $\mu = \alpha_0 + \alpha_1 * altitud$, calculamos este valor para cada iteración y para cada estación ya que ahora tenemos cambios en μ debido a la influencia de α_0 , α_1 y la altitud.

5.5. Inferencia en los modelos

Mencionamos anteriormente que los niveles de retorno son función de los parámetros de la GPD, $\phi(x)$, $\xi(x)$ y del parámetro de las tasas de excedencia, $\zeta(x)$. De manera que para poder estimar los niveles de retorno asociados a cada estación basta con conocer los parámetros antes mencionados.

Mediante el algoritmo MCMC obtenemos muestras para los parámetros de la GPD, de igual manera para los parámetros encargados de modelar la componente espacial β_ϕ y del parámetro asociado a las tasas de excedencia. La media a posteriori es usada como estimador puntual para cada uno de los parámetros. Las medias puntuales y los cuantiles son usados para mostrar los mapas de niveles de retorno.

5.6. Descripción del criterio de selección de modelos

La verosimilitud marginal descrita en la definición 2.12, es una cantidad importante usada en la prueba y comparación de modelos desde el enfoque bayesiano como se mostró en la sección 2.6, ya que el factor de Bayes para la comparación de modelos es un cociente de verosimilitudes marginales.

En (Raftery, Newton, Satagopan, and Krivitsky, 2007) se muestra que $p(x)$ puede expresarse como una esperanza con respecto a la distribución posterior del parámetro, motivando así una estimación basada en una muestra de Monte Carlo de la distribución posterior. Por el teorema de Bayes tenemos que:

$$\frac{1}{p(x)} = \int \frac{p(\theta|x)}{p(x|\theta)} d\theta = E\left\{\frac{1}{p(x|\theta)} \middle| x\right\}. \quad (5.10)$$

La ecuación 5.10 llamada identidad de la media armónica nos muestra como la verosimilitud marginal $p(x)$, puede ser aproximada por la media armónica simple de la verosimilitud, mediante

$$\hat{\pi}_{HM}(x) = \left[\frac{1}{B} \sum_{t=1}^B \frac{1}{p(x|\theta^t)} \right]^{-1}, \quad (5.11)$$

basado en B muestras $\theta^1, \theta^2, \dots, \theta^B$, de la distribución posteriori de $p(\theta|x)$. Esta muestra puede ser obtenida a través de la implementación de un MCMC estándar.

5.6.1. Estimador Gamma desplazado

Este método se basa en el hecho de que la distribución a posteriori de la log-verosimilitud es aproximadamente una distribución gamma. Esto conduce a un estimador de la máxima verosimilitud alcanzable, y también un estimador del número efectivo de parámetros que es sencillo de calcular a partir de las log-verosimilitudes, independientemente de la parametrización del modelo.

Este criterio se describe a detalle en (Raftery, Newton, Satagopan, and Krivitsky, 2007). El enfoque es utilizado para estabilizar la estimación

de la media armónica. Si MCMC se utiliza para simular a partir de la posteriori, suponemos que la muestra obtenida ha pasado por un proceso de *thinning*, de tal manera que tenemos una sucesión aproximadamente independiente de log-verosimilitudes, $l_t : t = 1, \dots, B$. Basándonos en que asintóticamente a medida que la cantidad de datos subyacentes a las las verosimilitudes aumenta hasta el infinito, la distribución posteriori de las log-verosimilitudes viene dada por $l_{max} - l_t \sim \text{Gamma}(\alpha, 1)$, donde l_{max} es la máxima log-verosimilitud alcanzable, y $\alpha = d/2$. Donde d es la dimensión del parámetro θ , es decir, el número de parámetros del modelo subyacente. La log-verosimilitud marginal esta dada por:

$$\log\pi(y) = \log E[e^{-l_t}|y] = l_{max} + \alpha \log(1 - \alpha).$$

Los mejores modelos son aquellos con el mayor valor de log-verosimilitud marginal, esto permite seleccionar el modelo que realiza el mejor ajuste de los datos escogiendo así el modelo que hace a los datos observados los más probables.

En el capítulo 6, presentamos los modelos escogidos y las estimaciones de los niveles de retorno para precipitación mediante mapas de retornos.

Capítulo 6

Análisis de resultados

6.1. Selección de Modelos

Este estudio está siendo desarrollado para Honduras tomando valores de precipitación diaria de 59 estaciones meteorológicas. Agradecemos a la Unidad de Meteorología del Instituto Hondureño de Ciencias de la Tierra (IH-CIT), por su valiosa colaboración al compartirnos los datos para poder desarrollar este trabajo. De la data completa de valores de precipitación diaria se seleccionaron las estaciones que tienen datos desde 1972 hasta 2012, para cada estación seleccionada tenemos registros de precipitación diarias de 41 años.

Las demás estaciones no se tomaron en cuenta porque no poseían la cantidad de registros completos para estos años. La información general de las estaciones meteorológicas escogidas se incluye en el Apéndice A.1.

Definimos un rango de fechas como temporada lluviosa, tomamos los valores de precipitación diaria para las fechas del 1 de mayo al 30 de noviembre, por ser estos los meses en los que se identifica la mayor cantidad de precipitación en Honduras, como se ilustra en la Figura 4.5. Para cada una de las estaciones meteorológicas se cuenta con 214 registros tomados por año, es decir 8774 registros para cada una de las 59 estaciones meteorológicas escogidas.

Escogimos el umbral haciendo uso de gráficos de vida media residual, una vez escogido el umbral se calcularon las excedencias para cada estación. En este punto cada estación tiene un numero diferente de registros, ahora nos interesan únicamente aquellos valores que exceden el umbral.

6.2. Resultados: Etapa 1

Presentaremos primero los modelos para la excedencia de umbrales. Para la primera etapa de nuestro modelo completo se describen 4 modelos, mostrando los resultados del modelo escogido.

Modelo 0:

$$\begin{aligned}\phi &= \phi \\ \xi &= \xi\end{aligned}$$

Este es el modelo base, suponemos en este primer modelo que ambos parámetros se compartan de igual manera para todas las estaciones meteorológicas. El componente espacial y las covariables no son tomados en cuenta. Si este modelo fuese escogido, concluiríamos que los valores de precipitación diaria no se ven influenciados por la ubicación y que no se presentan en ellos variaciones producto de la altitud y la precipitación media.

Modelo 1:

$$\begin{aligned}\phi &= \alpha_0 + \epsilon_\phi \\ \xi &= \xi\end{aligned}$$

Este es el primer modelo trabajado en el espacio latitud/longitud. El modelo base es modificado para tomar en cuenta las ubicaciones de las estaciones meteorológicas. Se aborda el parámetro ϕ como función de α_0 , en este caso para cada estación tendremos un valor de ϕ distinto, asumimos que ξ se mantiene constante para todas las estaciones.

Al escoger este modelo concluiríamos que el parámetro ϕ presenta variaciones en su valor, debido a la ubicación de la estación que estamos trabajando. Los resultados evidenciaron que hay presencia de cola pesada en nuestros

datos, ξ toma valores entre 0.09 y 0.15, por otra parte el valor de ϕ varia dependiendo la estación que escojamos.

Modelo 2:

$$\phi = \alpha_0 + \alpha_1(mp) + \epsilon_\phi$$

$$\xi = \xi$$

El modelo es trabajado en el espacio latitud/longitud, se busca evidenciar si al incluir la covariable precipitación media (mean precipitation en inglés que abreviaremos como **mp**), obtenemos un mejor ajuste para los valores de precipitación diaria.

Para cada estación meteorológica calculamos la precipitación media para incluirla en los modelos. El parámetro ϕ varia para cada estación meteorológica, se define mediante una regresión lineal con coeficientes α_0 y α_1 . Asumimos que ξ se mantiene constante para todas las estaciones.

Los resultados obtenidos para ξ son iguales que en el modelo anterior, similarmente en ϕ se presentan variaciones para cada estación, esta vez producto de la precipitación diaria media y la ubicación de las estaciones meteorológicas.

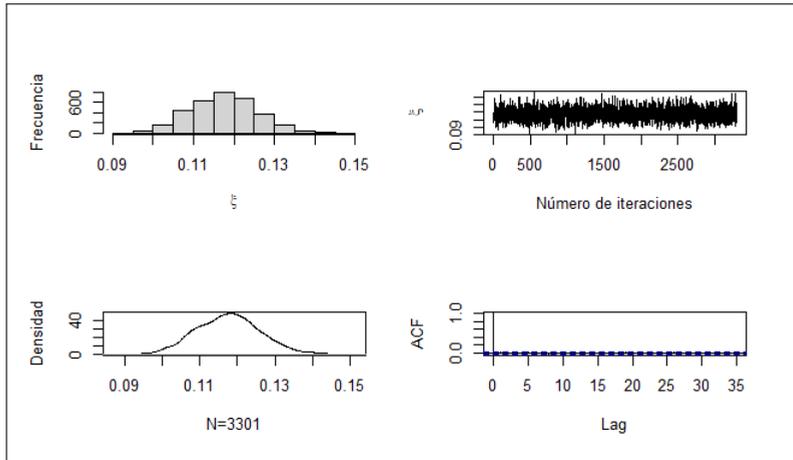
Modelo 3:

$$\phi = \alpha_0 + \alpha_1(altitud) + \epsilon_\phi$$

$$\xi = \xi$$

Modelo trabajado también en el espacio latitud/longitud, en este se incluye la covariable altitud para verificar si es significativa y si su aporte hace que obtengamos un modelo más preciso para nuestro datos. Los resultados obtenidos se muestran a continuación:

Figura 6.1: Valores de ξ modelo 3



En la Figura 6.1, se presentan los gráficos asociados a ξ de la muestra final para el modelo 3. Se evidencia la presencia de colas pesadas, estos valores se mantienen para todas las estaciones. Se puede observar la convergencia del parámetro y como después de realizar el refinamiento a la muestra no se observa dependencia entre las observaciones.

Figura 6.2: Valores de ϕ para el modelo 3 obtenidos en la estación 1

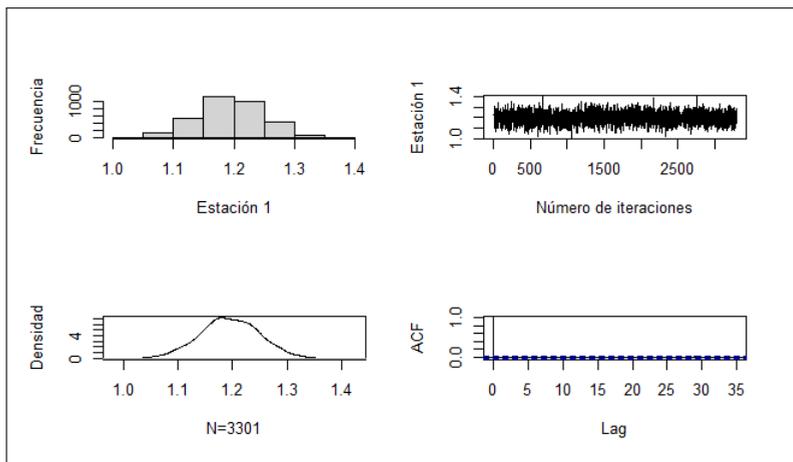


Figura 6.3: Valores de ϕ para el modelo 3 obtenidos en la estación 30

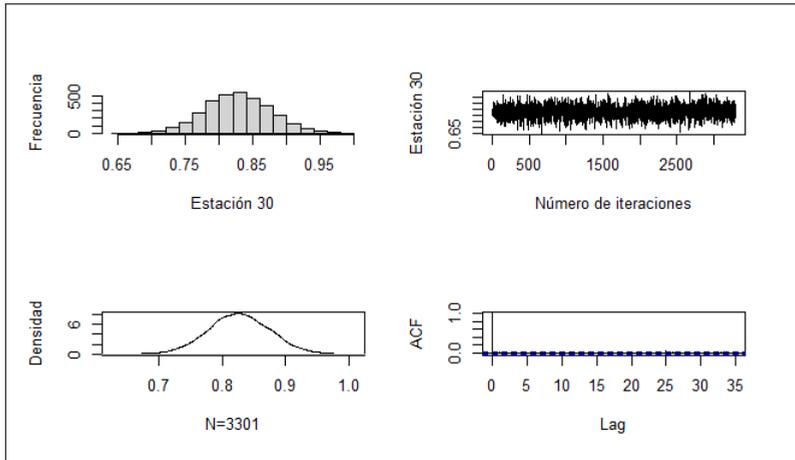
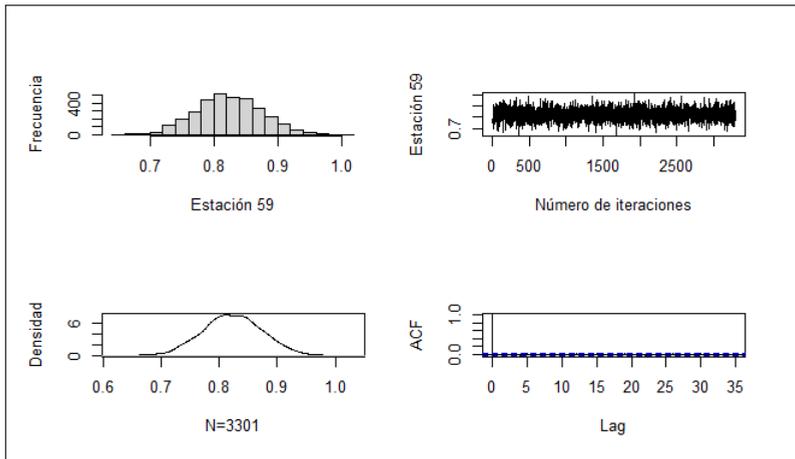


Figura 6.4: Valores de ϕ para el modelo 3 obtenidos en la estación 59



En las Figuras 6.2, 6.3 y 6.4, se muestran los gráficos resultantes de ϕ , en las muestras obtenidas para el modelo 3, asociado a las estaciones meteorológicas 1, 30 y 59 respectivamente. El parámetro de forma presenta variaciones en sus valores, esta variación se debe a que la covariable alti-

tud es significativa en este modelo. Podemos decir que las variaciones en la altitud de las estaciones meteorológicas se ven reflejadas en el parámetro de escala de la GPD que ajusta nuestros valores de excedencias.

En la Tabla 6.1, se muestra el resumen de los modelos usados en esta primera etapa, acompañados del valor obtenido mediante el criterio de la log-verosimilitud marginal, descrito en la subsección 5.6.1 (recordemos que $\phi = \log\sigma_u$). El modelo resaltado es el modelo escogido para esta etapa:

Tabla 6.1: Descripción de los modelos utilizados en la etapa 1 y sus valores de log-verosimilitud marginal asociada

Modelo	Log-verosimilitud marginal
Modelo 0	
$\phi = \phi$	-42623.61
$\xi = \xi$	
Modelo 1	
$\phi = \alpha_0 + \epsilon_\phi$	-32001.55
$\xi = \xi$	
Modelo 2	
$\phi = \alpha_0 + \alpha_1(mp) + \epsilon_\phi$	-31921.52
$\xi = \xi$	
Modelo 3	
$\phi = \alpha_0 + \alpha_1(altitud) + \epsilon_\phi$	-31901.19
$\xi = \xi$	

6.3. Resultados: Etapa 2

En esta etapa se muestran los modelos asociados a las tasas de excedencia, usamos 5 modelos que se describen a continuación.

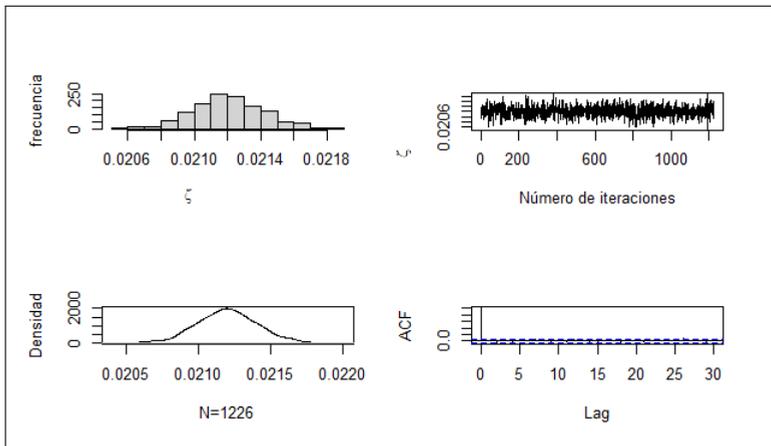
Modelo A:

$$\zeta = \zeta$$

La tasa de excedencia es modelada como un parámetro que se comporta de manera uniforme para todas las estaciones. En la Figura 6.5, podemos

observar que bajo este modelo ζ toma valores entre 0.0206 y 0.0218, se evidencia también la convergencia del parámetro. En esta muestra final que se obtiene al realizar el refinamiento a la muestra inicial no hay dependencia en las observaciones.

Figura 6.5: Valores de ζ para el modelo A



Modelo B:

$$\zeta = \alpha_0 + \epsilon_\zeta$$

Se modela ζ mediante una regresión lineal con coeficiente α_0 , en este modelo para cada estación se obtiene una tasa de excedencia diferente, debido a la ubicación de las estaciones. Los valores de las tasas asociadas a las distintas estaciones, según las muestras resultantes, presentan pequeñas variaciones entre si.

Modelo C:

$$\zeta = \alpha_0 + \alpha_1(mp) + \epsilon_\zeta$$

Se incluye el efecto de la precipitación media y la ubicación en el parámetro ζ mediante la regresión lineal, las conclusiones son las mismas que en el modelo anterior.

Modelo D:

$$\zeta = \alpha_0 + \alpha_1(altitud) + \epsilon_\zeta$$

Las variaciones de los valores de ζ asociados a las distintas estaciones debido a la altitud, son pequeñas de manera similar a lo ocurrido los modelos anteriores.

Modelo E:

$$\zeta = \alpha_0 + \alpha_1(mp) + \alpha_2(altitud) + \epsilon_\zeta$$

En este modelo se incluyen las covariables, precipitación media y altitud, mediante una regresión con coeficientes α_0 y α_1 . La cantidad de información agregada por este modelo es mucho mayor, pero los resultados obtenidos para ζ siguen siendo bastante similares a los de los modelos descritos anteriormente.

Los resultados obtenidos para esta etapa se resumen en la Tabla 6.2, se muestran todos los modelos considerados con su valor de log-verosimilitud asociado. El modelo resaltado es el modelo que se escogió para esta etapa, en conclusión tenemos que las covariables usadas en esta etapa no fueron significativas.

Todos los modelos se comportan de manera similar siendo el modelo base, un modelo practico que describe bien el comportamiento de las tasas de excedencia asociadas a las estaciones.

Tabla 6.2: Descripción de los modelos utilizados en la etapa 2 y sus valores de log-verosimilitud marginal asociados

Modelo	Log-verosimilitud marginal
Modelo A $\zeta = \zeta$	-24.54897
Modelo B $\zeta = \alpha_0 + \epsilon_\zeta$	-30.39126
Modelo C $\zeta = \alpha_0 + \alpha_1(mp) + \epsilon_\zeta$	-30.37409
Modelo D $\zeta = \alpha_0 + \alpha_1(altitud) + \epsilon_\zeta$	-30.68142
Modelo E $\zeta = \alpha_0 + \alpha_1(mp) + \alpha_2(altitud) + \epsilon_\zeta$	-32.36031

6.4. Resultados del modelo completo

Los modelos que mejor se ajustan a nuestros datos, son aquellos con el mayor valor de log-verosimilitud marginal. Los modelos escogidos para cada etapa se describen en la Tabla 6.3.

Tabla 6.3: Resumen de los modelos escogidos para cada etapa

Etapa 1
Modelo 3:
$\phi = \alpha_0 + \alpha_1(\textit{altitud}) + \epsilon_\phi$
$\xi = \xi$
Etapa 2
Modelo A:
$\zeta = \zeta$

Estimaciones de los parámetros

Para estimar los niveles de retorno tomamos los resultados de los modelos escogidos en las 2 etapas, para las estimaciones puntuales calculamos la media a posteriori de los parámetros. Los resultados obtenidos para ξ y ζ se muestran en la Tabla 6.4 , podemos observar que $\xi = 0.1176$ y $\xi > 0$, lo que evidencia la presencia de colas pesadas.

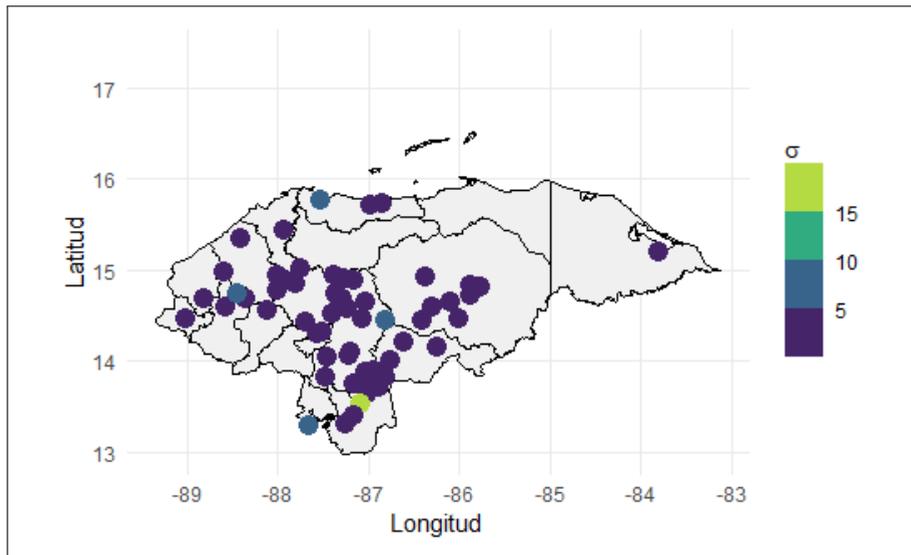
Tabla 6.4: Valores estimados de los parámetros para el modelo completo

Parámetro	Estimación
ξ	0.1176
ζ	0.0212

Las estimaciones para el parámetro de forma se hacen para cada estación, recordemos que $\phi = \log(\sigma)$, de manera que $\sigma = e^\phi$. Los valores de las estimaciones puntuales para cada una de las estaciones se muestran en la Figura 6.6. σ toma valores que van desde 0 a 20cm., es decir valores entre 0 y 200 milímetros.

En la mayoría de las estaciones se observa un comportamiento similar, σ toma valores entre 0 y 50 milímetros. Observamos algunos valores que se encuentran dispersos en el mapa, para los que σ toma valores entre 50 y 100 milímetros. Podemos identificar además una estación que toma los valores más altos, entre 150 y 200 milímetros, esta estación es la de Liure, ubicada en el departamento del Paraíso.

Figura 6.6: Mapa de estimaciones de σ para cada estación meteorológica



Estimaciones de los niveles de retornos

Realizamos las estimaciones para los niveles de retorno a 5, 10, 15, 25 y 100 años, los valores asociados a cada una de las 59 estaciones se presentan en el apéndice A.2. Se muestran a continuación los valores de los niveles de retorno estimados a 5, 25 y 50 años mediante mapas de niveles de retornos. La posición de los puntos en el mapa corresponden a la ubicación geográfica de las estaciones meteorológicas escogidas. Tenemos en estos mapas escalas que nos permiten ubicar los niveles de valores de retorno como bajos o altos dependiendo del color asociado a cada punto, los valores en la escala están dados en centímetros.

A medida que tomamos periodos de tiempo más largos los valores asociados a los niveles de retorno van aumentando, de manera que la probabilidad de ser alcanzados o excedidos va disminuyendo.

6.4.1. Niveles de retorno a 5 años

En el mapa mostrado en las Figura 6.7, podemos observar que los niveles de retorno a 5 años toman valores que van de 0 a 100 centímetros.

En la mayoría de las estaciones incluidas en el estudio el comportamiento para los valores de los niveles de retorno a 5 años es bastante similar, hay variaciones entre ellos debido a la altitud de las estaciones. Los niveles de retorno de precipitación toman valores de 0 a 200 milímetros en la mayoría de las estaciones.

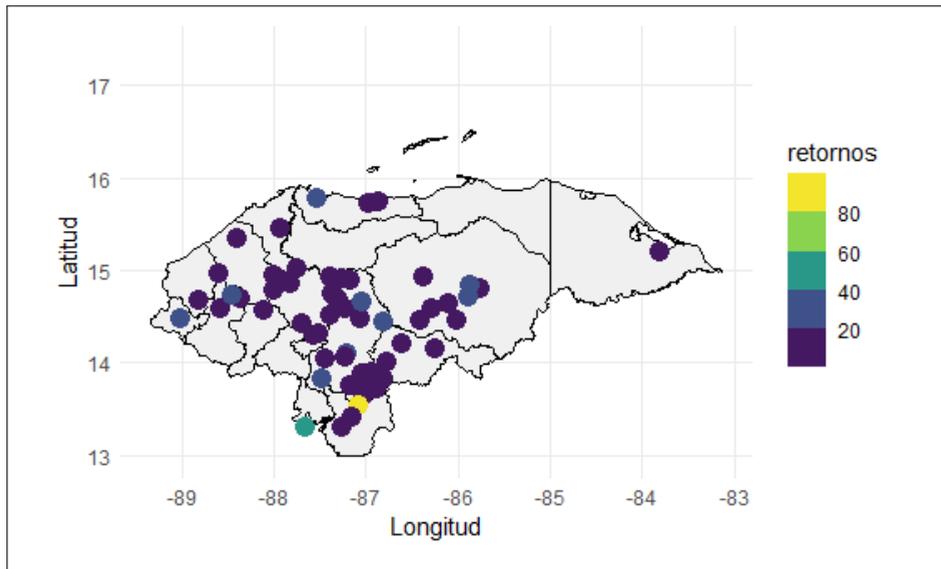
Dos estaciones en particular resaltan en el mapa, la primera que toma valores entre los 400 y 600 milímetros, esta estación es la de Amapala ubicada en el departamento de Valle. La segunda es la estación que presenta los valores más altos, la estación de Liure ubicada en el departamento de El Paraíso, sus valores estimados de nivel de retorno están entre los 800 y 1000 milímetros. En esta estación en particular es probable que ese valor de 826 milímetros, en promedio sea alcanzado o excedido con probabilidad $1/826$ al menos una vez en 5 años.

6.4.2. Niveles de retorno a 25 años

Los valores estimados para los niveles de retorno a 25 años se muestran en la Figura 6.8, estos van de 0 a 150 centímetros. El comportamiento descrito para los niveles de retorno a 5 años se mantiene, en la mayoría de las estaciones tenemos valores de niveles de retorno similares que van de 0 a 250 milímetros.

Tenemos 3 valores que destacan de los demás, dos de ellos con valores en el rango de 500 a 750 milímetros estos corresponden a las estaciones de El Tablón ubicada en el departamento de Lempira cuyo nivel de retorno estimado es de 571 milímetros. El otro punto corresponde a la estación de Amapala en este se estimo un valor de 717 milímetros. El tercer punto ubicado en la escala más alta, con valores de niveles de precipitación entre

Figura 6.7: Mapa de niveles de retorno estimados a 5 años



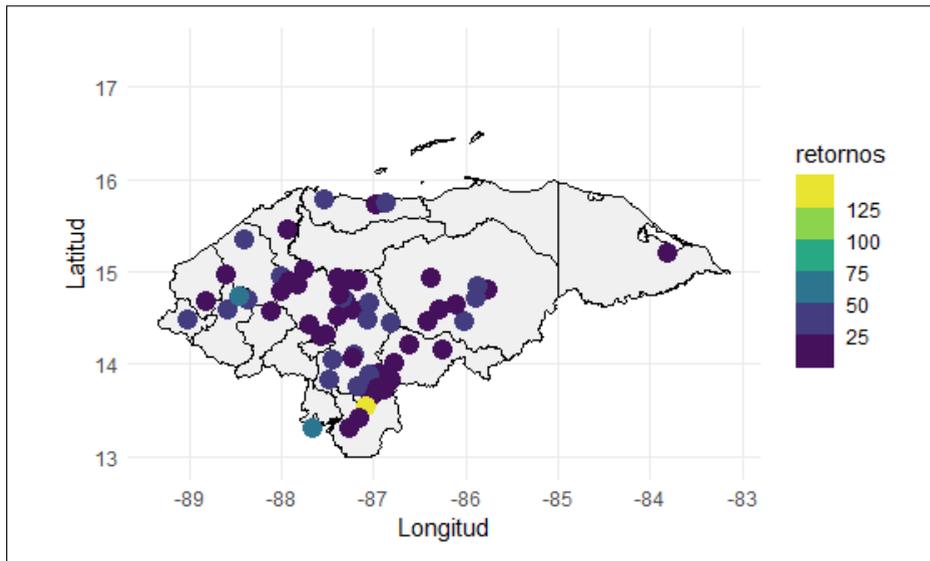
1250 y 1500 milímetros, corresponden a la estación de Liure ubicada en el departamento de El Paraíso.

6.4.3. Niveles de retorno a 50 años

En la Figura 6.9, se muestran los resultados obtenidos para las estimaciones de los niveles de retorno estimados a 50 años. Podemos ver en los cambios en las escalas que a medida que aumentan los años que tomamos para los niveles de retorno estas van aumentando, en este caso van de 0 a 160 centímetros.

El comportamiento descrito en el mapa para los niveles de retorno a 50 años, es bastante similar al ilustrado en los mapas anteriores. Para la mayoría de las estaciones tenemos valores de niveles de retorno entre 0 y 400 milímetros. Identificamos 2 puntos que resaltan por estar ubicados en las escalas más altas. El valor asociado a la estación de Amapala tomando valores entre los 800 y los 1200 milímetros y por otra parte la estación de Liure con valores de niveles de retorno entre los 1200 y 1600 milímetros.

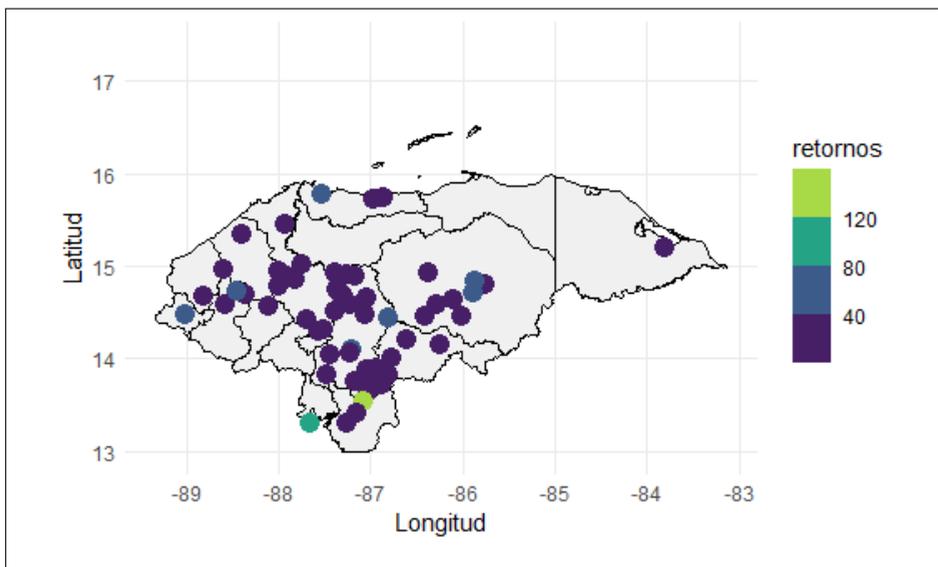
Figura 6.8: Mapa de niveles de retorno estimados a 25 años



Queremos hacer énfasis en que un valor de nivel de retorno más alto no hace a una estación mas propensa a inundaciones. Al tener presencia de valores de niveles de retorno tan altos, es menos probable que ese valor sea alcanzado o excedido. Los valores de niveles de retorno asociados a cada una de las estaciones involucradas en este estudio y para las diferentes ubicaciones geográficas en general, presentan variaciones porque como hemos mencionado a lo largo de este trabajo, la precipitación no es un fenómeno con un comportamiento constante. Los valores de precipitación cambian, de igual manera los niveles de retorno, debido a la ubicación de la estación meteorológica de interés y mas aún como lo describen nuestros modelos, debido a la altitud.

En el Apéndice A.3, se presentan los intervalos del 95% de confianza para los niveles de retorno a 25 años asociados a las estaciones meteorológicas incluidas en este estudio. Los intervalos de confianza más amplios corresponden a estaciones identificadas con el #49, que esta ubicada en Liure y la #50, ubicada en Amapala (que ya habíamos identificado antes por poseer las estimaciones puntuales más altas).

Figura 6.9: Mapa de niveles de retorno estimados a 50 años



Capítulo 7

Conclusiones y trabajos futuros

7.1. Conclusiones

- El parámetro de forma ξ es un parámetro que es altamente sensible. Obtener convergencia en este parámetro fue difícil, por esa razón solamente se trabajo bajo el supuesto de que se comporta de manera constante para todas las estaciones.
- En la etapa 1, un cuarto modelo a considerar por la forma en como se describieron los demás modelos es aquel donde se incluyan ambas covariables en ϕ , es decir:

$$\phi = \alpha_0 + \alpha_1(mp) + \alpha_2(altitud) + \epsilon_\phi.$$

Este modelo fue trabajado en ambas etapas, pero en la etapa 1 no se logro obtener convergencia en algunos de los parámetros involucrados por lo que no se incluye como modelo propuesto.

- Los modelos escogidos en cada etapa son aquellos con mayor valor de log-verosimilitud marginal. Para los datos trabajados la covariable de elevación resulto ser significativa, lo que nos dice que los valores

de precipitación para estas estaciones escogidas en Honduras presentan variaciones entre sí producto de su elevación. Por otra parte con las tasas de excedencia no se evidencia que las covariables tengan incidencia en el cálculo de las mismas, estas se mantienen constantes para todas las estaciones.

- Tenemos pocas estaciones que cumplen con los requerimientos para este estudio, ya sea porque la cantidad de datos no es suficiente o por la calidad de los mismos. Hay zonas bastantes grandes del país que no se alcanzan a cubrir, hacer interpolaciones en estas zonas nos llevaría a tomar conclusiones erróneas por eso se presentan únicamente las estimaciones puntuales asociadas a las estaciones escogidas.
- Sería muy interesante poder incluir la temperatura y la humedad relativa como covariables y ver su efecto en las estimaciones, pero no tenemos acceso a esta información para nuestros datos por eso se decidió usar la precipitación media y la altitud que son de las que poseemos información.
- La mayoría de las estaciones incluidas en el estudio tienen un comportamiento similar en cuanto a los niveles de retorno se refiere, se ubican en su mayoría en la escala más baja esto tomando en cuenta sus ubicaciones y la altitud como covariable influyendo en los valores de precipitación diaria. Esto tiene sentido ya que ocurren 2 cosas con las ubicaciones de las estaciones escogidas para este estudio, la mayoría se ubican geográficamente cerca y solo estamos tomando en cuenta la altitud de las estaciones y segundo la mayoría de las estaciones se encuentran en zonas con valores de precipitación bajos.
- A pesar de que los valores estimados para los niveles de retorno son parecidos en la mayoría de las estaciones hay variaciones presentes en ellos producto de su ubicación geográfica y su altitud, en la mayoría de las estaciones los valores para los niveles de retorno son bajos pero es necesario entender que un valor de retorno alto no necesariamente debe ser asociado con mayor probabilidad de inundación, únicamente significa que la probabilidad de que ese nivel de precipitación sea excedido en esta estación es más alto que en las demás.

7.2. Trabajos futuros

- Estimar en nuestros datos distintos valores de umbrales utilizando métodos tanto clásicos como bayesianos, incluyendo entre ellos el análisis de cópulas. Para cada valor de umbral propuesto obtener las estimaciones de los parámetros. Obtener las estimaciones de los niveles de retorno, haciendo comparaciones entre ellos que permitan medir la variación de los mismos debido a la elección de un umbral determinado y escoger el método que nos proporciona las mejores estimaciones.
- La precipitación como fenómeno climatológico esta fuertemente relacionado con variables como temperatura y humedad relativa. Estas covariables pueden ser significativas en nuestros modelos haciendo las estimaciones más precisas, por lo que queremos incluir modelos donde estas variables sean tomadas en cuenta.
- Debido a que el número de estaciones que cumplen con nuestros requerimientos para este estudio eran pocas, no pudimos presentar estimaciones para todo el país. Pretendemos agregar más estaciones tomando un periodo de tiempo más extenso para hacer el estudio más preciso, esto nos da la posibilidad de replicarlo en otras zonas, haciendo predicciones y generando mapas con valores continuos para todo Honduras
- En este trabajo nos enfocamos en valores de precipitación extrema (máximos en precipitación), pero sabemos que también es importante estimar valores mínimos en precipitación, por lo que aspiramos a adecuar este trabajo enfocándolo en esos valores mínimos y poder hacer estimaciones encaminadas al análisis de sequías.

Apéndice A

Tablas

Tabla A.1: Información General de las estaciones meteorológicas incluidas en el estudio

Código	Nombre de la estación	Departamento
1	Quimistán	Santa Bárbara
2	El Cajón	Cortés
3	El Jaral	Cortés
4	Victoria	Yoro
5	Flores	Comayagua
6	El Coyolar	Comayagua
7	Agua Caliente	Francisco Morazán
8	La Labor	Ocotepeque
9	Gracias Lempira	Lempira
10	Pito Solo	Comayagua
11	La Ermita	Francisco Morazán
12	Agua Caliente F,M,	Francisco Morazán
13	Santa Rita	Yoro
14	Playitas	Comayagua
15	Santa Elena	Cortés
16	San Nicolás	Comayagua
17	Sulaco	Yoro
18	Vallecillo	Francisco Morazán
19	Esquías	Comayagua
20	San Ignacio	Francisco Morazán
21	Ulapa	Santa Bárbara
22	Marale	Francisco Morazán
23	Gualtaya	Copán
24	Palmital	Comayagua
25	San José de la Montaña	Lempira
26	El Tablón	Lempira
27	El Nispero	Santa Bárbara
28	La Unión Atlántida	Atlántida
29	Cayetano	Olancho
30	Guayabillas	Olancho
31	Santa María Guyambre	El Paraíso
32	San Francisco de Becerra	Olancho
33	Las Mesetas	Olancho
34	El Vijao	Olancho

Continúa en la página siguiente.

APÉNDICE A. TABLAS

Código	Nombre de la estación	Departamento
35	Manto	Olancho
36	San Pedro Catacamas	Olancho
37	Río Abajo	Francisco Morazán
38	Sabana Grande	Francisco Morazán
39	Reitoca	Francisco Morazán
40	Lepaterique	Francisco Morazán
41	Potreros El Paraíso	El Paraíso
42	Guinope	El Paraíso
43	Maraita	Francisco Morazán
44	San Antonio de Flores	El Paraíso
45	San Lucas	El Paraíso
46	Texiguat	El Paraíso
47	Nueva Armenia	Francisco Morazán
48	Colonia 21 de Octubre	Francisco Morazán
49	Liure	El Paraíso
50	Amapala	Valle
51	La Ceiba	Atlántida
52	Tela	Atlántida
53	La Mesa	Cortés
54	Puerto Lempira	Gracias a Dios
55	Catacamas	Olancho
56	Tegucigalpa	Francisco Morazán
57	Choluteca	Choluteca
58	Las Ánimas	El Paraíso
59	Oropoli	El Paraíso

Tabla A.2: Valores de los niveles de retornos estimados a 5, 10, 15, 25, 50 y 100 años para las estaciones incluidas en el estudio

Código	Años					
	5	10	15	25	50	100
1	18.06	21.61	23.82	26.76	31.04	35.69
2	16.50	19.63	21.59	24.19	27.98	32.09
3	18.84	22.59	24.93	28.03	32.56	37.48
4	15.02	17.76	19.48	21.76	25.08	28.68
5	15.78	18.72	20.56	23.00	26.56	30.42
6	14.85	17.55	19.24	21.49	24.75	28.30
7	14.47	17.07	18.70	20.86	24.00	27.42
8	24.67	29.94	33.23	37.60	43.97	50.88
9	17.89	21.39	23.57	26.48	30.70	35.29
10	15.62	18.53	20.34	22.75	26.26	30.08
11	17.32	20.67	22.75	25.53	29.58	33.97
12	15.33	18.16	19.92	22.27	25.69	29.40
13	19.56	23.49	25.95	29.21	33.96	39.12
14	14.67	17.33	18.99	21.19	24.40	27.89
15	13.91	16.37	17.91	19.95	22.92	26.14
16	17.27	20.60	22.69	25.45	29.48	33.86
17	13.62	16.00	17.48	19.46	22.34	25.46
18	13.80	16.23	17.75	19.76	22.70	25.89
19	14.79	17.48	19.16	21.39	24.64	28.17
20	20.50	24.68	27.29	30.76	35.81	41.30
21	15.25	18.06	19.81	22.14	25.54	29.22
22	12.72	14.87	16.21	17.99	20.59	23.41
23	16.42	19.53	21.47	24.05	27.81	31.90
24	14.22	16.75	18.34	20.44	23.51	26.84
25	17.16	20.47	22.53	25.28	29.28	33.62
26	36.56	44.93	50.16	57.11	67.23	78.22
27	16.99	20.25	22.29	25.00	28.94	33.22
28	14.14	16.66	18.23	20.32	23.37	26.67

Continúa en la página siguiente.

APÉNDICE A. TABLAS

Código	Años					
	5	10	15	25	50	100
29	18.87	22.62	24.97	28.08	32.62	37.54
30	13.89	16.34	17.87	19.90	22.87	26.08
31	14.32	16.89	18.49	20.62	23.72	27.08
32	16.62	19.78	21.76	24.38	28.21	32.36
33	13.13	15.38	16.78	18.65	21.38	24.33
34	13.12	15.37	16.78	18.64	21.37	24.32
35	14.82	17.52	19.20	21.43	24.69	28.23
36	23.40	28.33	31.41	35.51	41.47	47.95
37	25.22	30.63	34.01	38.50	45.04	52.14
38	15.76	18.70	20.53	22.97	26.53	30.38
39	21.09	25.42	28.12	31.72	36.96	42.64
40	17.83	21.31	23.48	26.37	30.57	35.14
41	15.39	18.24	20.01	22.37	25.81	29.55
42	16.37	19.47	21.41	23.98	27.73	31.79
43	18.79	22.52	24.85	27.94	32.45	37.35
44	16.81	20.03	22.04	24.71	28.59	32.81
45	15.39	18.23	20.00	22.36	25.80	29.53
46	16.27	19.34	21.26	23.81	27.53	31.56
47	19.26	23.11	25.52	28.72	33.38	38.43
48	23.06	27.90	30.93	34.95	40.81	47.17
49	82.62	103.03	115.76	132.69	157.35	184.12
50	45.50	56.21	62.89	71.77	84.71	98.76
51	19.99	24.04	26.56	29.92	34.81	40.12
52	27.71	33.77	37.56	42.59	49.92	57.87
53	16.91	20.16	22.18	24.87	28.79	33.04
54	16.43	19.55	21.49	24.08	27.85	31.94
55	23.75	28.78	31.92	36.09	42.17	48.76
56	13.16	15.42	16.83	18.70	21.44	24.40
57	16.66	19.84	21.82	24.45	28.29	32.46
58	15.41	18.26	20.04	22.40	25.85	29.59
59	13.85	16.29	17.81	19.84	22.79	26.00

Tabla A.3: Intervalos de confianza del 95% para los niveles de retorno a 25 años asociados a las estaciones incluidas en el estudio

Número de estación	Límite inferior	Límite superior
1	24.23	29.36
2	21.81	26.77
3	26.36	29.92
4	19.84	23.85
5	20.71	25.43
6	19.78	23.31
7	18.96	22.81
8	34.84	40.54
9	23.99	29.27
10	20.79	24.9
11	23.11	28.12
12	20.45	24.35
13	27.43	31.11
14	19.41	23.1
15	18.3	21.68
16	23.74	27.36
17	18.1	20.9
18	18.15	21.52
19	19.89	22.99
20	27.05	35.2
21	20.57	23.84
22	16.95	19.05
23	22.45	25.75

Continúa en la página siguiente.

APÉNDICE A. TABLAS

Número de estación	Límite inferior	Límite superior
24	18.78	22.24
25	23.24	27.53
26	52.05	62.79
27	22.61	27.52
28	18.62	22.26
29	25.22	31.33
30	18.36	21.58
31	19.38	21.96
32	21.94	26.99
33	17.06	20.32
34	17	20.46
35	19.35	23.77
36	32.67	38.58
37	36	41.2
38	21.13	24.95
39	28.17	35.78
40	23.82	29.19
41	20.65	24.3
42	22.13	26.01
43	25.88	30.19
44	22.82	26.72
45	20.8	23.97
46	21.52	26.34
47	26.23	31.3
48	32.35	37.88
49	116.16	151.36
50	64.76	79.44
51	26.66	33.7
52	39.66	45.78
53	22.26	27.83
54	21.21	27.22
55	33.41	39.15
56	17.06	20.44
57	22.03	27.24
58	20.61	24.37
59	18.21	21.6

Bibliografía

- M. Armstrong. *Basic Linear Geostatistics*. Springer Springer Berlin Heidelberg, 1998.
- A. Balkema and L. Haan. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.
- S. Banerjee, B. Carlin, and A. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.
- J. Capel. Aspectos geográficos de la pluviometría de honduras. *Papeles de geografía*, 20:9–26, 1994.
- G. Casella and R. Berger. *Statistical Inference*. Duxbury, 2002.
- S. Coles. *An Introduction to Statistical Modeling of Extrem Values*. Springer series in statistics, London, 2001.
- D. Cooley, D. Nychka, and P. Naveau. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Associations*, 102:824–840, 2007.
- N. Cressie. *Statistics for Spatial Data*. Wiley, 1993.
- P. Diggle and J. Tawn. Model-based geostatistics. *The Applied Statistics*, 47:299–350, 1998.
- P. Embrechts, C. Klüppelberg, and M. T. *Modelling Extremal Events for Insurance and Finance*. UTB, Stuttgart, 1994.

BIBLIOGRAFÍA

- FECYT. Meteorología y climatología ;unidad didáctica : Semana de la ciencia y la tecnología 2004, 2007.
- R. Fisher and Tippett. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceeding of the Cambridge Philosophical Society*, 24:180–190, 1928.
- D. Gamerman, H. Migon, and F. Louzada. *Statistical Inference An Integrated Approach*. Chapman and Hall Texts in Statistical Science Series, Boca Raton, 2015.
- R. Giraldo. Introducción a la geoestadística-teoría y aplicación. Technical report, Universidad Nacional de Colombia, 2002.
- B. V. Gnedenko. Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44:423–453, 1943.
- B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, 1954.
- S. Infante, H. A., and A. Urquiola. Introducción a la inferencia bayesiana. Technical report, Universidad de Carabobo, 2010.
- IPCC. Climate change 2022: Impacts, adaptation, and vulnerability. Technical report, Cambridge University Press, 2022.
- H. Liu and L. Wasserman. *Statistical Machine Learning*. 2014.
- G. Matheron. Matheron’s theory of regionalised variables, 2019.
- I. Mochales. *Modelos Estadísticos para Valores Extremos y sus Aplicaciones*. Tesis de doctorado, Universidad Autónoma de Barcelona, Barcelona, 2013.
- F. Navarro, C. and Monserrate, L. Llano, Obando, and J. D. Córdoba. Desarrollo de los escenarios climáticos de Honduras y módulo académico de capacitación. Technical report, Centro Internacional de Agricultura Tropical (CIAT); Programa de las Naciones Unidas para el Desarrollo (PNUD); Dirección Nacional de Cambio Climático de MiAmbiente, 2018.

BIBLIOGRAFÍA

- J. Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.
- A. Raftery, M. A. Newton, J. M. Satagopan, and P. N. Krivitsky. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian statistics*, 8:1–45, 2007.
- M. Rodriguez, N. Kawas, O. Elvir, L. Torres, K. Wiese, A. Cardona, C. Canales, G. Gomez, J. Mejia, N. Sevilla, L. Mendoza, and P. Castillo. *Atlas Climático y de Gestión de Riesgo de Honduras*. 2012.
- E. Sierra. Herramientas geoestadísticas en la construcción de modelos digitales de elevación. 2014.