



# COLEGIO DE POSTGRUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

**CAMPUS MONTECILLO**

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

**Regresión espacial de precipitaciones extremas en el  
Estado de Tabasco**

Rebeca Alejandra Pérez Figueroa

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2012

---

La presente tesis titulada: **REGRESION ESPACIAL DE PRECIPITACIONES EXTREMAS EN EL ESTADO DE TABASCO**, realizada por la alumna: **REBECA ALEJANDRA PEREZ FIGUEROA**, bajo la dirección del consejo particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## **MAESTRA EN CIENCIAS**

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA**

**ESTADÍSTICA**

### **CONSEJO PARTICULAR**

**CONSEJERO**

  
\_\_\_\_\_  
**Dr. Sergio Pérez Elizalde**

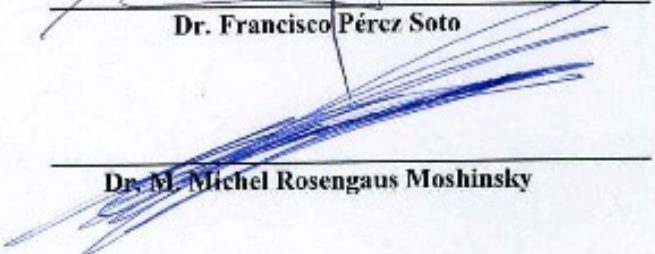
**ASESOR**

  
\_\_\_\_\_  
**Dr. Javier Suárez Espinosa**

**ASESOR**

  
\_\_\_\_\_  
**Dr. Francisco Pérez Soto**

**ASESOR**

  
\_\_\_\_\_  
**Dr. M. Michel Rosengaus Moshinsky**

**Montecillo, Texcoco, Estado de México, Mayo de 2012**

# **Regresión espacial de precipitaciones extremas en el Estado de Tabasco**

Rebeca Alejandra Pérez Figueroa

Colegio de Postgraduados, 2012

Las precipitaciones extremas en el estado de Tabasco, México, causan pérdidas económicas y serios estragos a los ecosistemas cada año. Esta investigación se basa en el desarrollo de un análisis de eventos de precipitación extrema en Tabasco empleando la información en la base de datos MAYA que comprende observaciones de lluvia diarias para nodos-geográficamente equidistantes. El principal objetivo es proponer un modelo de regresión espacial para datos de precipitaciones extremas con el fin de estimar períodos de retorno vía un modelo jerárquico Bayesiano, y proveer de mapas de riesgo basados en el modelo ajustado así como en la distribución predictiva de precipitación extrema. Se encontró que la localización geográfica incrementa la exactitud en la determinación de niveles de retorno al compartir la información para el análisis.

**Palabras clave: Bayesiano, regresión espacial, lluvia extrema.**

# Spatial regression of extreme precipitations in the State of Tabasco

Rebeca Alejandra Pérez Figueroa

Colegio de Postgraduados, 2012

Extreme precipitations in the state of Tabasco, Mexico, causes economic lose and ecosystem damages every year. This research is aimed to develop an analysis of extreme rainfall occurrences in Tabasco using the information in the data base MAYA which comprises daily rainfall observations on equidistant nodes. The main objective was to propose a spatial regression model for extreme rainfall data in order to estimate return periods via a Bayesian hierarchical model, providing risk maps based on the fitted model as well as on the predictive distribution of extreme precipitation. It was found that the geographical localization improves the accuracy on the determination of return levels by sharing information between neighbor sources of data.

**Key words:** Bayesian, spatial regression, extreme rainfall.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios de postgrado.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Al Dr. Sergio Pérez Elizalde por su tiempo dedicado a la elaboración de la tesis, por su guía y apoyo.

Al Dr. Javier Suárez Espinosa por su colaboración en este proyecto.

Al Dr. Francisco Pérez Soto por su completa disposición y apoyo para este proyecto de tesis.

Al Dr. Michel Rosengaus Moshinsky del Servicio Meteorológico Nacional por su invaluable apoyo.

## DEDICATORIA

A mis **padres:** Por haberme dado la vida en principio; por todo el amor, dedicación, educación, mucha paciencia y todas las grandes cosas que he aprendido a su lado. Es muy cierto que gracias a ustedes hoy estoy sumando un éxito más a mi vida.

A mis **hermanas:** A ustedes les agradezco el invaluable tiempo y apoyo que he recibido de su parte con los años. Somos compañeras de vida, de aventuras, de travesuras pero también de experiencias de vida.

A mi **esposo:** Por que a tu lado, se añade una definición de amor a la vida; pero también por la paciencia y el inmesurable apoyo que me has brindado.

A mi : Che giova a te, core mio lésser amato?

# Indice

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>4</b>
<b>3. Materiales y métodos</b>	<b>5</b>
3.1. Base de datos . . . . .	5
3.2. Teoría de Valores Extremos . . . . .	7
3.3. Distribución de Valores Extremos Generalizada . . . . .	7
3.4. Niveles y períodos de Retorno . . . . .	9
3.5. Análisis Espacial . . . . .	10
3.5.1. Mapeo geoestadístico . . . . .	10
3.5.2. Variables ambientales . . . . .	11
3.5.3. Modelos de predicción espacial . . . . .	11
3.5.4. Modelo de correlación . . . . .	15
3.6. Inferencia Bayesiana . . . . .	16
3.6.1. Estimación puntual y por intervalo . . . . .	17
3.6.2. Distribución a priori . . . . .	18
3.6.3. Distribución a posteriori . . . . .	20

3.6.4. Distribución predictiva . . . . .	20
3.6.5. Computación Bayesiana . . . . .	21
3.7. Selección del mejor modelo . . . . .	23
3.7.1. Criterio de Información de la Devianza (DIC) . . . . .	23
3.8. Análisis de convergencia . . . . .	25
3.8.1. Prueba de convergencia de Gelman y Rubin . . . . .	25
<b>4. Modelación de precipitaciones extremas en el estado de Tabasco</b>	<b>27</b>
4.1. Base de datos de precipitación histórica . . . . .	27
4.2. Regresión espacial . . . . .	28
4.2.1. El modelo estadístico . . . . .	28
4.2.2. Función de verosimilitud . . . . .	29
4.2.3. Proceso espacial latente . . . . .	29
4.2.4. Distribuciones <i>a priori</i> para los parámetros del modelo . . . . .	30
4.2.5. Modelo de correlación y selección del mejor modelo . . . . .	31
4.3. Niveles de retorno . . . . .	37
4.3.1. Niveles de retorno basados en el modelo espacial ajustado . . . . .	37
4.3.2. Niveles de retorno basado en la función de distribución predictiva de Valores Extremos . . . . .	38
<b>5. Conclusiones</b>	<b>40</b>
<b>Referencias</b>	<b>41</b>
<b>Anexos</b>	<b>43</b>



# Indice de tablas

4.1. Comparación de los modelos empleados en la regresión espacial . . . . .	32
4.2. Resumen <i>a posteriori</i> de los $\beta$ 's del parámetro localidad e intervalo de credibilidad al 95 % para el modelo 3 . . . . .	34
4.3. Resumen <i>a posteriori</i> de los $\beta$ 's del parámetro escala e intervalo de credibilidad al 95 % para el modelo 3 . . . . .	34
4.4. Resumen <i>a posteriori</i> del parámetro forma e intervalo de credibilidad al 95 % para el modelo 3 . . . . .	35

# Índice de figuras

3.1. Región geográfica a estudiar . . . . .	6
3.2. Estadística espacial y sus tres subáreas de estudio . . . . .	10
3.3. Predicción espacial es el proceso de estimar el valor de características para sitios desconocidos en base a observaciones existentes: (a) esquema de espacio horizontal, (b) valores de una variable de interés en un espacio unidimensional . . . . .	12
3.4. Predicción espacial implica la aplicación de un algoritmo de predicción en un arreglo de nodos (predicción espacial punto a punto) . . . . .	13
3.5. Forma ideal de un variograma donde se muestran los parámetros nugget, sill y range . . . . .	15
4.1. Convergencia de las iteraciones de los parámetros de regresión para el $\mu$ de la GEV . . . . .	33
4.2. Convergencia de las iteraciones de los parámetros de regresión para el $\sigma$ de la GEV . . . . .	34
4.3. Convergencia de las iteraciones del parámetro de regresión para $\xi$ de la GEV . . . . .	35
4.4. Densidades a posteriori estimadas de los parámetros de regresión para la localidad . . . . .	35
4.5. Densidades a posteriori estimadas de los parámetros de regresión para la escala . . . . .	36
4.6. Densidad a posteriori estimada del parámetro de regresión para la forma . . . . .	37

## Índice de figuras

---

4.7. Niveles de retorno para 5 años de período de retorno obtenidos con el ajuste de la GEV y con la función predictiva . . . . .	38
4.8. Comparación de los niveles de retorno para 10 años de período de retorno obtenidas con la regresión espacial y con la función predictiva . . . . .	39
4.9. Comparación de los niveles de retorno para 100 años de período de retorno obtenidas con la regresión espacial y con la función predictiva . . . . .	39

# Capítulo 1

## Introducción

Definiendo al clima como una serie de condiciones meteorológicas que caracterizan a una región, debemos tener en cuenta que el clima es muy dinámico, razón por la cual se tratan de prever las condiciones climáticas de algún lugar con el fin de prevenir a la población sobre posibles excentricidades del clima.

Llamamos eventos extremos a aquellos que presentan un valor probabilístico muy bajo (o muy alto) de ocurrir. Debido a que las técnicas estadísticas clásicas han fallado al estudiar este tipo de sucesos a causa de la poca información que se posee acerca de dichos eventos, se ha recurrido a la teoría de valores extremos (EVT por sus siglas en inglés), que ofrece una teoría en específico para este tipo de datos.

El país se divide por regiones, realizando dicha división en base a las características propias de cada zona. En particular, el estado de Tabasco y parte del estado de Chiapas se encuentran en la Región Hidrológica número 30 llamada Grijalva-Usumacinta, que a su vez se subdivide en varias subcuencas. Para fines de este trabajo, se empleó un el área geográfica rectangular comprendida entre las coordenadas  $17.2^\circ$  a  $19^\circ$  de latitud norte y de  $-94.6^\circ$  a  $-91^\circ$  de longitud oeste que incluye el territorio del estado de Tabasco, que es la región de estudio.

El estado de Tabasco es uno de los estados de la república mexicana sobre los cuales las lluvias extremas han incidido más a lo largo de los años, causando con esto daños a la población y pérdidas económicas. Tabasco se localiza en la zona del trópico, es una entidad federativa ubicada geográficamente entre los  $17^\circ 15'$  y  $18^\circ 39'$  latitud norte, y entre los  $91^\circ 00'$  y  $94^\circ 17'$  longitud oeste, presentando temperaturas relativamente altas.

En el estado llueve la mayor parte del año, delimitando a la época lluviosa del mes de junio al mes de marzo. Las lluvias se intensifican hacia el verano cuando se presentan lluvias intensas, mientras que en el otoño se presentan los nortes, tormentas acompañadas de fuertes vientos provenientes del Golfo de México.

## 1. Introducción

---

En particular para este estado destacan registros de inundaciones en 1975, 1990, 1999 y 2007, siendo ésta última la que produjo el mayor número de desastres para el estado.

De acuerdo con estudios realizados por diferentes instituciones como el IMTA y el Instituto de Ingeniería de la UNAM, algunas de las causas que destacan en el problema de inundaciones en Tabasco, es el crecimiento acelerado de la población comprendido en el período de 1975 a 1985, lo que se tradujo en asentamientos de la población en zonas inundables, incluyendo el crecimiento de la capital Villahermosa hacia las márgenes de los ríos, por lo cual, cuando se llegan a presentar lluvias extremas, los problemas antes mencionados desencadenan un mal desempeño de las obras hidráulicas naturales y no naturales, teniendo como consecuencia inundaciones.

Se tienen registros de inundaciones considerables en 1975, 1990, 1999 y 2007. Aún así, la inundación que más catástrofe ha causado en el territorio de Tabasco fue la de 2007, en la que según registros del IMTA (cita) se originaron precipitaciones extraordinarias mayores a los 400 milímetros en tan sólo 24 horas debidas a la presencia de dos frentes fríos y precipitaciones de cerca de 1000 mm en tres días (28 al 30 de octubre de 2007) en la cuenca media del río Grijalva y de los ríos de la Sierra, aunados a los escurrimientos generados de precipitaciones previas de los días 11, 12, 23 y 24 de octubre sobre la cuenca.

El objetivo del presente trabajo es el análisis de precipitaciones extremas, lo que permitirá realizar mapas de riesgo de precipitaciones para el estado de Tabasco, que en los últimos años ha padecido inundaciones con alto impacto socioeconómico, debidas entre otras causas, a precipitaciones extremas.

Con esto se busca proporcionar estimaciones del riesgo de precipitaciones extremas futuras en esa entidad federativa. Se espera que dicha información sea de ayuda para estudios posteriores y/o para la implementación de planes de prevención.

En esta investigación se plantea una regresión espacial bajo el supuesto de que existe dependencia en las observaciones de precipitación extrema de lugares contiguos. Se asume que variables como el tiempo, la distancia de cada ubicación al mar, así como la ubicación geográfica tienen un efecto significativo en la explicación de dichas precipitaciones.

Una forma de medida común y relativamente fácil de entender los eventos extremos es el nivel de retorno, y esta es la medida que reportan los atlas de precipitación. El  $r$ -ésimo nivel de retorno es excedido con probabilidad  $1/r$  cada  $r$  años.

En este trabajo se muestran estimaciones para los períodos de retorno a 5, 10 y 100 años. Se emplearon máximos anuales debido a la poca información disponible, obteniendo los niveles de retorno, con los cuales se calcularon los mapas de riesgo.

El presente trabajo se estructura de la siguiente manera: en el capítulo 2 se mencionan los objetivos de la investigación. En el capítulo 3 se describe la teoría de valores extremos así como la teoría geoestadística que fue clave fundamental para el estudio. En el capítulo

4 se presentan los resultados, teniendo finalmente los capítulos 5 y 6, que contienen las conclusiones del trabajo y anexos del mismo, respectivamente.

# Capítulo 2

## Objetivos

- Realizar un análisis estadístico espacial de la precipitación extrema en el estado de Tabasco.
- Obtener mapas de isolíneas de niveles de retorno de precipitación.

# Capítulo 3

## Materiales y métodos

En este capítulo se hará mención de la metodología a seguir, es decir, de la teoría de valores extremos a través de la precipitación de la región mencionada, empleada para estudiar variables de eventos extremos, como lo son en este caso las precipitaciones.

### 3.1. Base de datos

El método tradicional de generar los registros de observaciones de las variables climáticas en las denominadas estaciones meteorológicas eran en papel, de las cuales se hacían compendios, tal es el ejemplo de las normales climatológicas en las que se podían consultar las principales variables meteorológicas tales como velocidad de los vientos, precipitación, temperaturas máximas, mínimas y promedio, radiación solar, velocidad y dirección de los vientos, entre otras. Dichos compendios presentaban la falta de datos, atribuible al factor humano, o a la imposibilidad de realizar la lectura en los instrumentos cuando se suscitaba algún evento meteorológico inusual, lo que presentaba un problema para aquellas personas que recurrían a éstas bases de datos para realizar estudios o análisis de cualquier índole. Poco después la base de datos se compiló en el software manejador de datos CLICOM (en sus primeras versiones), presentando aún las inconveniencias antes mencionadas.

Actualmente se tienen estaciones automáticas, las cuales monitorean las variables climáticas que solían hacerse de forma manual. Dichas estaciones realizan mediciones en tiempo real, lo que es una gran ventaja para meteorólogos y, en cuanto a la confiabilidad de las bases de datos, información fidedigna y completa.

Dada la gran importancia de tener información completa en áreas como la hidrología y el pronóstico del tiempo, el Servicio Meteorológico Nacional (SMN) se dio a la tarea de

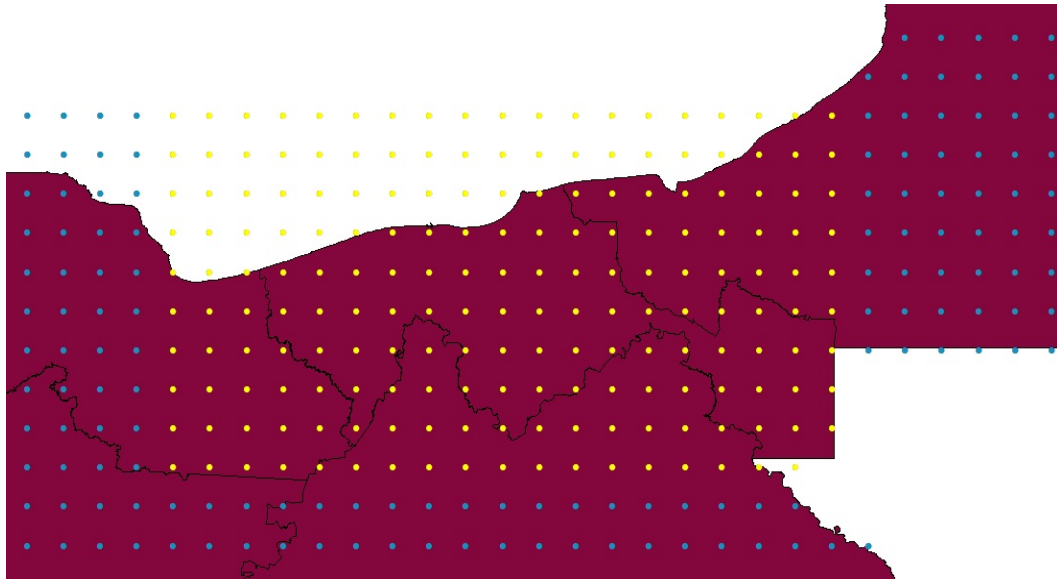


### 3.1. Base de datos

---

crear una base de datos que no contuviera espacios en blanco, surgiendo así MAYA en su primera versión, que fue elaborada por una consultora de la Organización Meteorológica Mundial a solicitud del Servicio Meteorológico Nacional. Esta base reúne datos de temperatura y precipitación diarias. La región que abarca ésta base de datos va desde  $-117.2^\circ$  a  $-86.0^\circ$  de longitud oeste y, de  $14.0^\circ$  a  $33.0^\circ$  latitud norte. Es una interpolación que se realizó con el método de Kriging, que no tiene falta de información y se tienen en total 15072 nodos, de los cuales 4542 están sobre territorio nacional y 10530 se encuentran sobre zonas marinas (que no presentan valores).

En el presente trabajo se emplearon precipitaciones máximas anuales del estado de Tabasco, para lo cual de la base de datos completa MAYA v1.0, solo se empleó la información de la variable precipitación, depurando ésta hasta solo tener la información del polígono de interés comprendido entre las coordenadas  $17.2^\circ$  a  $19^\circ$  de latitud norte y de  $-94.6^\circ$  a  $-91^\circ$  de longitud oeste. Esto significa que se empleó la información de 135 nodos del total de la base de datos, como se muestra en la Figura 3.1



**Figura 3.1:** Región geográfica a estudiar

Debido a que la información contenida en MAYA no son datos puntuales, si no datos interpolados en una maya regular. Los resultados obtenidos con el análisis serán de la misma forma no representativos de algún punto (ubicación de los nodos), sino de una región del orden de los 20 km alrededor de los nodos.

### 3.2. Teoría de Valores Extremos

En la actualidad se ha visto más frecuente el uso de eventos que por su ocurrencia suelen llamarse extremos, es decir, que son eventos con poca probabilidad de ocurrencia. Generalmente llamamos valores extremos a aquellos valores de eventos de extraña ocurrencia, llámense precipitaciones, temperaturas, heladas, radiaciones, en el área de la meteorología. Aún así, la gran relevancia de su estudio se basa en las consecuencias que acarrearán y su impacto en diversos ámbitos.

La teoría de valores extremos (EVT por sus siglas en inglés) es la rama de la probabilidad y la estadística dedicada a caracterizar el comportamiento de observaciones atípicas (Cooley, 2005). La teoría de valores extremos tiene sus inicios a principios del siglo XX. Algunas de las primeras aplicaciones de ésta teoría fue a mediados del siglo pasado para el análisis de emisiones radioactivas, resistencia de materiales, análisis de flujos, análisis sísmicos y análisis pluviales, por mencionar algunas aplicaciones tempranas (Kotz, 2000)(?).

El objetivo tradicional de la estadística es describir la mayor parte de la distribución, por lo que, la teoría de valores extremos se interesa sólo en describir el comportamiento de valores muy altos (o muy bajos). Es por eso que son de interés los valores aislados para el análisis de valores extremos y, los datos que describen el comportamiento típico de la variable aleatoria, son descartados. Aún con grandes cantidades de datos para un estudio, cuando se estudian extremos, siempre se cuentan con datos insuficientes (Cooley, 2005).

Las técnicas estadísticas más comunes fallan al proporcionar análisis confiables acerca de la cola de una distribución. La media y la varianza de la muestra no son de ayuda al describir observaciones aisladas. SI las técnicas de máxima verosimilitud se basan en conjunto de datos completo, no modelan el extremo de la distribución con exactitud. Por ello, para describir la cola de una distribución existen diferentes técnicas estadísticas (Cooley, 2005).

La teoría de valores extremos para datos univariados se basa en argumentos asintóticos que conducen a la distribución de valores extremos generalizada (GEV) (Cooley, 2005). Además provee el vínculo entre datos registrados en un marco de tiempo diario (o cada hora) y cantidades de mayor longitud de escala, tales como los niveles de retorno.

### 3.3. Distribución de Valores Extremos Generalizada

La distribución de valores extremos es la distribución límite de una serie de variables aleatorias normalizadas, independientes e idénticamente distribuidas (iid).

### 3.3. Distribución de Valores Extremos Generalizada

---

La distribución de valores extremos es una herramienta muy útil para modelar valores extremos. De manera análoga a como la distribución normal se emplea para modelar sumas finitas de variables aleatorias independientes e idénticamente distribuidas de acuerdo a los resultados asintóticos del Teorema Central del Límite. De la misma forma, la distribución de valores extremos generalizada puede utilizarse para modelar máximos de un número finito de variables aleatorias independientes e idénticamente distribuidas, como consecuencia de resultados asintóticos. Cabe mencionar que la función de distribución, la función de densidad y la función cuantil de la GEV presentan una forma cerrada. Esto permite estimar los parámetros numéricamente por medio de los métodos de máxima verosimilitud y momentos.

Supongamos que  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes e idénticamente distribuidas con función de distribución  $F$  y sea  $M_n = \max(X_1, X_2, \dots, X_n)$ , entonces

$$\lim_{n \rightarrow \infty} P(M_n \leq x) = \lim_{n \rightarrow \infty} F^n(x) = 0 \quad (3.1)$$

para cualquier  $x$  con  $F(x) < 1$ .

Si existen unas constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$ , que cumplen la función de parámetros de escala y forma, entonces estandarizando y cuando  $n$  tiende a infinito

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{d} G(x) \quad (3.2)$$

es decir,  $(M_n - b_n)/a_n$  converge en distribución a  $G$ , donde  $G$  es una función de distribución límite no degenerada, entonces  $G$  puede ser una de las tres familias de funciones de distribución de valores extremos

$$Gumbel : \Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (3.3)$$

$$Fréchet : \Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases} \quad (3.4)$$

$$Weibull : \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad (3.5)$$

con  $\alpha > 0$  en las familias Fréchet y Weibull. Las tres familias de distribuciones anteriores

### 3.4. Niveles y períodos de Retorno

---

pueden considerarse casos particulares de la distribución de valores extremos generalizada (GEV). Haciendo una reparametrización  $\xi = 1/\alpha$ , la función de distribución de valores extremos generalizada es

$$G(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \quad (3.6)$$

donde  $\alpha > 0$ ,  $-\infty < \mu < \infty$ ,  $1 + \xi(x - \mu)/\sigma > 0$ ,  $x_+ = \text{máx}(x, 0)$ .

La función de densidad es de la forma

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}-1} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (3.7)$$

para  $1 + \xi(x - \mu)/\sigma > 0$

Los parámetros  $\mu, \sigma, \xi$  son de localidad, escala y forma, respectivamente.

### 3.4. Niveles y períodos de Retorno

Es de interés en el análisis de eventos extremos, la recurrencia de dichos eventos en el tiempo. Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias *iid* con función de distribución continua  $F$  y  $\mu \in R$  un umbral dado. Considerando el evento  $X_i > u$  como un evento extremo y definiendo el primer tiempo de excedencia

$$T(u) = \text{mín}(i \geq 1 : X_i > u) \quad (3.8)$$

Se requiere saber  $E[T(u)]$ , tomando  $p = 1 - F(u)$  se tiene que

$$P(T(u) = k) = (1 - p)^{k-1} p, k = 1, 2, \dots, \quad (3.9)$$

de modo que  $T(u)$  es una variable geométrica y  $E[T(u)] = p^{-1}$ . Siendo  $u$  **nivel de retorno** con un **período de retorno**  $p^{-1}$  para los eventos  $\{X_i > u\}$ , donde  $u$  es el cuantil de  $F$  con un nivel de probabilidad  $1 - p$ .

## 3.5. Análisis Espacial

### 3.5.1. Mapeo geoestadístico

Algunas variables de interés en el estudio de la naturaleza conllevan una referencia espacio temporal. Esta referencia es determinada por al menos cuatro parámetros

- Ubicación geográfica (latitud y longitud o proyecciones cartesianas X,Y);
- Altura sobre el nivel del mar (altitud);
- Tiempo de la medida (año, mes, día, hora, minuto, etc.);
- Soporte espacio temporal (tamaño de los bloques del material asociado con las medidas; intervalo de tiempo de dichas mediciones);

La geoestadística es una parte de la estadística especializada en el análisis e interpretación de datos geográficamente referenciados. Cressie (1991) considera que la geoestadística es una de las tres áreas especializadas en el análisis de datos espaciales (Hengl, 2009), las otras dos son el análisis de patrones puntuales (también llamados "procesos puntuales") y el análisis de datos tipo área (objetos polígonos) como se muestra en la Figura 3.2



**Figura 3.2:** Estadística espacial y sus tres subáreas de estudio

Para otros, la estadística espacial es un proceso de extraer resúmenes de información de datos espaciales y compararlos con modelos teóricos que explican como se originan y desarrollan patrones espaciales. Dado que la geoestadística evolucionó en la industria de

### 3.5. Análisis Espacial

---

la minería, por un largo tiempo significó estadística aplicada a la geología. Desde entonces, las técnicas geoestadísticas han encontrado aplicación en numerosas áreas, empezando por el mapeo de suelos, meteorología, ecología, oceanografía, geoquímica, epidemiología, geografía humana, entre otras.

Entonces, la geoestadística de nuestros tiempos puede ser definida como una rama de la estadística que se especializa en el análisis e interpretación de datos espacialmente referenciados, pero enfocándose en componentes continuos inherentes (elementos espaciales).

La predicción espacial o interpolación espacial se enfoca en predecir valores de la variable de estudio en toda el área de interés, que típicamente resulta en imágenes o mapas. En la geoestadística, para el caso de un kriging ordinario, por ejemplo, la interpolación corresponde a casos donde la ubicación estimada se encuentra rodeada por las ubicaciones de la muestra y se encuentra en un entorno de auto-correlación espacial.

#### 3.5.2. Variables ambientales

Estas variables son medidas descriptivas de diferentes elementos del ambiente. Las variables ambientales pueden pertenecer a diferentes dominios, desde la biología (distribución de las especies y medidas de biodiversidad), la ciencia del suelo (tipos y propiedades del suelo), la ciencia de la vegetación (especies de plantas, tipos de cobertura vegetal), climatología (variables climáticas en la superficie, áreas por encima y por debajo del nivel del suelo), hidrología (condiciones y cantidades de agua), etc. La información de estas variables es generalmente colectada a través de un muestreo; la información de dichas muestras se emplea para producir mapas mostrando la distribución de los datos en un área. La precisión y actualización de estos mapas de elementos del ambiente representa un papel importante para la toma de decisiones, evaluación de tierras, degradación ambiental, escasez de agua, entre otros.

El primer paso para llevar a cabo un mapa geoestadístico confiable de variables ambientales es entender primeramente las fuentes de variabilidad de los datos. Ésta variabilidad es resultado de procesos estocásticos y determinísticos. En otras palabras, la variabilidad en los datos es una suma de dos componentes: la variación espacial natural y un ruido inherente  $\varepsilon$ , éste último principalmente debido a los errores de medición.

#### 3.5.3. Modelos de predicción espacial

En una situación ideal la variabilidad de las variables ambientales está determinada por un conjunto finito de valores que siguen alguna ley física conocida. Si se conoce el algoritmo, los valores de las variables de interés se pueden predecir con exactitud. En la realidad, la relación entre la característica de interés y el ambiente físico es tan

### 3.5. Análisis Espacial

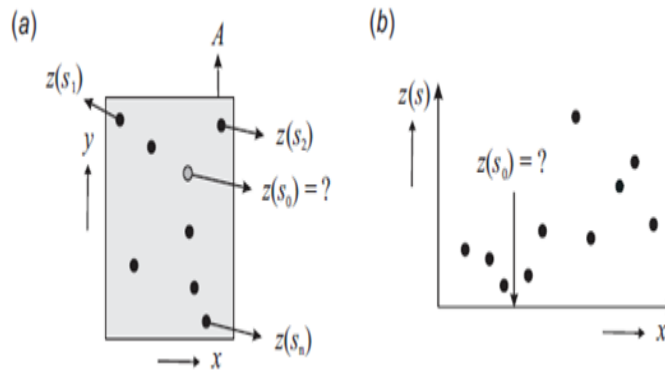
complejo que no se puede modelar de manera exacta. Ésto puede ser por cosas que no sabemos como falta de información y, una posibilidad es proponer un modelo usando la información conocida de la variable de interés a través de una estimación indirecta o no determinística.

Definiendo el problema usando notación matemática, se tiene una serie de observaciones de una variable de interés (también conocida como la variable respuesta)  $Z$  que se denota  $z(s_1), z(s_2), \dots, z(s_n)$  donde  $s_i = (x_i, y_i)$  es alguna ubicación,  $x_i$  y  $y_i$  sus coordenadas en el espacio geográfico y  $n$  es el número de observaciones. El dominio geográfico de interés se denota por  $A$ . Nos enfrentamos con información muestreada ( $z(s_n)$ ), que son realizaciones de un proceso ( $Z = Z(s), \forall s \in A$ ) (Hengl, 2009).

Suponiendo que las muestras son representativas y consistentes, el valor de la variable de interés para una nueva ubicación puede ser obtenido usando un modelo de predicción espacial. En términos estadísticos, un modelo de predicción espacial genera realizaciones  $\hat{s}_0$  de la variable de interés de acuerdo a lo siguiente

$$(\hat{s}_0) = E \{Z | z(s_i), q_k(s_0), \gamma(h), s \in A\} \quad (3.10)$$

donde  $z(s_i)$  es la base de datos de entrada,  $\gamma(h)$  es el modelo de covarianza definiendo la estructura del modelo de autocorrelación espacial (Figura 3.3),  $q_k(s_0)$  una lista de predictores determinísticos también conocidos como covariables o variables explicativas, información que necesita estar disponible para cada ubicación en  $A$ . En otras palabras, el modelo de predicción espacial envuelve una lista de procedimientos para generar valores de interés dada la calibración de los datos y el dominio espacial (Hengl, 2009).



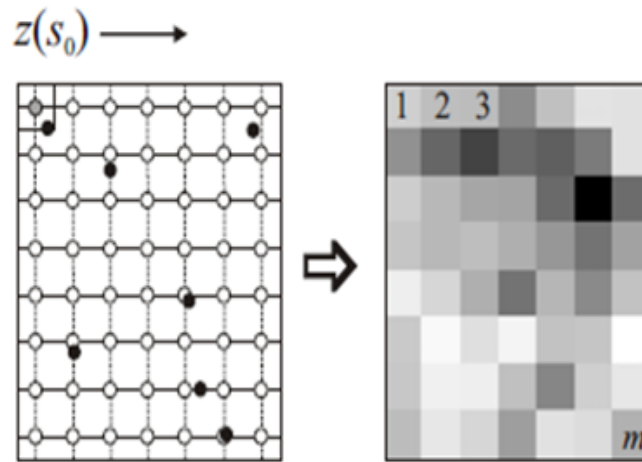
**Figura 3.3:** Predicción espacial es el proceso de estimar el valor de características para sitios desconocidos en base a observaciones existentes: (a) esquema de espacio horizontal, (b) valores de una variable de interés en un espacio unidimensional

En términos raster (técnicas gráficas que emplean arreglos matriciales de pixeles) de Sistemas de Información Geográfica (GIS por sus siglas en inglés), el dominio espacial

### 3.5. Análisis Espacial

---

de interés es una matriz rectangular, un arreglo de filas por columnas de una malla de nodos sobre el dominio de interés



**Figura 3.4:** Predicción espacial implica la aplicación de un algoritmo de predicción en un arreglo de nodos (predicción espacial punto a punto)

$$\mathbf{z} = \mathbf{z}(s_j), j = 1, \dots, m; s_j \in \mathbf{A}$$

donde  $\mathbf{z}$  es el arreglo de los datos,  $z(s_j)$  es el valor del nodo  $s_j$  en la malla, y  $m$  es el número total de nodos en la malla.

Los modelos de predicción espacial pueden ser clasificados de acuerdo a la experiencia contenida en el área de estudio de los análisis estadísticos

- Modelos determinísticos
- Modelos estadísticos lineales
- Sistemas expertos

**Modelos estadísticos o probabilísticos lineales:** En el caso de éstos modelos, los parámetros se estiman siguiendo la teoría probabilística. Las predicciones son acompañadas por una estimación del error de predicción. Existen al menos cuatro grupos de modelos estadísticos lineales

- Kriging (geoestadística plana)
- Correlación ambiental



### 3.5. Análisis Espacial

---

- Modelos con enfoque Bayesiano
- Modelos híbridos

Los modelos geoestadísticos son modelos utilizados generalmente para datos espacialmente referenciados en un dominio espacial continuo. Este es un método general que puede aplicarse a datos respuesta normales y no normales en una variedad de estudios ecológicos (Hengl, 2009).

Sean  $\mathbf{y} = (\mathbf{y}(\mathbf{s}_1), \dots, \mathbf{y}(\mathbf{s}_n))'$ , un conjunto de observaciones espacialmente referenciadas y cuantificables, donde  $s$  denota alguna ubicación en el dominio espacial  $A$  (usualmente una región bidimensional). En este modelo las variables respuesta ( $y_s$ ) se asumen independientes y con distribución de Valores Extremos Generalizada (GEV). Aunado a esto, cada parámetro de la distribución de valores extremos se propone que siga un modelo lineal ligando la variación espacial a través de un proceso espacial ( $z_s$ ) y el vector de covariables ( $\mathbf{x}(\mathbf{s})$ ). De acuerdo con Johnson y Hoeting (2011), el proceso geoestadístico  $z = (z(s_1), \dots, z(s_n))$  se asume que tiene distribución normal  $N(z|X\beta, S)$ , donde  $X$  es una matriz de covariables con cada fila asociada al sitio  $s_i$   $i = 1, \dots, n$ . La matriz de covarianza de este proceso se define por la función espacial de covarianza de la forma

$$\begin{aligned} Cov\{z(s), z(s+h)\} &= \sigma^2 \rho(h'\Phi h) \\ Var\{z(s)\} &= \sigma^2 + \tau^2 \end{aligned} \quad (3.11)$$

donde  $\rho$  es una función de correlación isotrópica,  $\Phi$  es una matriz positiva definida de  $2 \times 2$ ,  $\sigma^2$  es el parámetro *sill* y  $\tau^2$  es el parámetro *nugget*. La función de correlación  $\rho(\cdot)$  puede tomar varias formas, las más típicas son la exponencial, Matérn y funciones de correlación esférica como se muestra en la Figura 3.5.

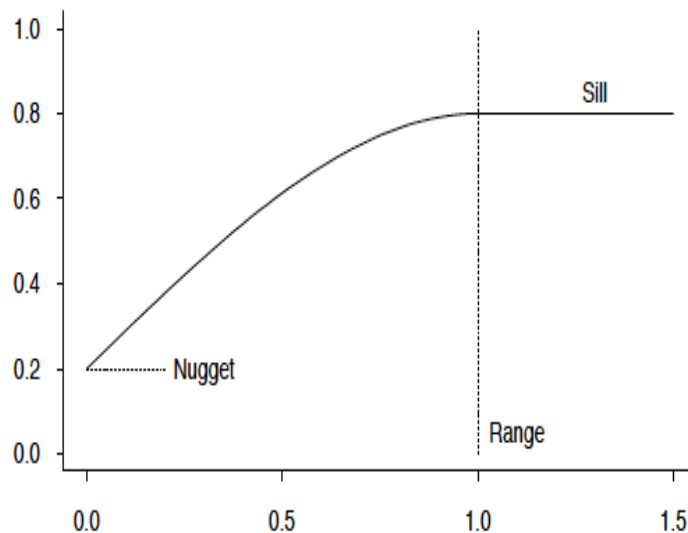
La función conjunta de las observaciones  $\mathbf{y}$  y el proceso espacial  $\mathbf{z}$  está dado por

$$p(y, z | \beta, \xi, X) = p(y|z) N(z|X\beta, S) \quad (3.12)$$

donde  $\xi$  es un vector de elementos de  $\Phi$ . Por lo tanto, la verosimilitud para el modelo de regresión geoestadístico estará dado por

$$L(\beta, \xi | y, X) = \int_z p(y|z) N(z|X\beta, S) dz \quad (3.13)$$

### 3.5. Análisis Espacial



**Figura 3.5:** Forma ideal de un variograma donde se muestran los parámetros nugget, sill y range

#### 3.5.4. Modelo de correlación

Sea la función  $cov(Z(s_1), Z(s_2)) = C(s_1 - s_2)$  para todo  $s_1, s_2 \in A$  un covariograma, donde  $C(\cdot)$  es la función de covarianza (nótese que también se le llama función de autocovarianza por los analistas de series de tiempo). Dado que  $C(0) > 0$ , sea

$$\rho(h) = C(h) / C(0) \quad (3.14)$$

un correlograma (también llamado función de autocorrelación). Esta cantidad es usada tradicionalmente por los analistas de series de tiempo para diagnosticar no estacionariedad, determinando el tipo de dependencia estacionaria para el ajuste del modelo.

Considere la relación

$$\text{var} \{Z(s_1) - Z(s_2)\} = \text{var} \{Z(s_1)\} + \text{var} \{Z(s_2)\} - 2\text{cov} \{Z(s_1), Z(s_2)\} \quad (3.15)$$

Si  $Z(\cdot)$  tiene estacionariedad de segundo orden, de la ecuación (3.15)

$$\text{var} (Z(s_1) - Z(s_2)) = 2 \{C(0) - C(s_1 - s_2)\} \quad (3.16)$$

, lo que implicaría que  $Z(\cdot)$  es intrínsecamente estacionaria, teniendo

$$2\gamma(h) = 2(C(0) - C(h)) \quad (3.17)$$

### 3.6. Inferencia Bayesiana

---

Si  $C(h) \rightarrow 0$ , así como  $\|h\| \rightarrow \infty$ , entonces  $2\gamma(h) \rightarrow 2C(0)$ . La cantidad  $C(0)$  se denomina la meseta del semivariograma (?).

#### Familia de modelos de correlación Whittle-Matérn

La clase de funciones de covarianza de Whittle-Matérn se recomienda para la modelación de covarianza espacial, ésta es de la forma:

$$\rho(t) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{vt}\phi)^\nu K_\nu(2\sqrt{vt}\phi) , \text{ para } t > 0 \quad (3.18)$$

donde  $K_\nu$  (?) es la función modificada de Bessel de orden  $\nu$ . Los parámetros  $\phi$  y  $\nu$  representan un parámetro de escala y un parámetro de suavidad (Banerjee *et al.*, 2004).

El uso de la función de covarianza de Matérn permite a los datos informar sobre  $\nu$ . Es posible obtener información acerca del proceso de suavizamiento a pesar de haber observado el proceso en solo un número finito de ubicaciones (Banerjee *et al.*, 2004).

#### Modelo de correlación exponencial

Hay funciones de correlación isotrópicas que son válidas en todas las dimensiones. Si asumimos que la correlación espacial es función principalmente de la distancia  $d_{ii}$  entre dos ubicaciones  $s_i$  (Banerjee *et al.*, 2004). Existen varias opciones para la  $Cov(Z(s_1), Z(s_2))$ , entre ellas la de potencia exponencial, que es de la forma

$$\rho(t) = \sigma^2 \exp(-|\varphi t|^p); \text{ para } t > 0 \quad (3.19)$$

### 3.6. Inferencia Bayesiana

Cuando se modelan datos observados así como cantidades desconocidas como variables aleatorias, el enfoque Bayesiano provee al análisis estadístico un vínculo de cohesión al combinar modelos para datos complejos e información externa u opiniones expertas. En este enfoque, además de especificar el modelo de la distribución  $f(\mathbf{y} | \theta)$  para los datos observados  $\mathbf{y} = (y_1, \dots, y_n)'$  dado un vector de parámetros desconocidos  $\theta = (\theta_1, \dots, \theta_k)$ , dando a  $\theta$  una distribución a priori  $\pi(\theta | \lambda)$ , donde  $\lambda$  es un vector de hiperparámetros. Si  $\lambda$  es conocido, la inferencia acerca de  $\theta$  es basada en su distribución posteriori (Banerjee *et al.*, 2004).

### 3.6. Inferencia Bayesiana

---

$$p(\theta|y, \lambda) = \frac{p(y, \theta|\lambda)}{p(y|\lambda)} = \frac{p(y, \theta|\lambda)}{\int p(y, \theta|\lambda) d\theta} = \frac{f(y|\theta)\pi(\theta|\lambda)}{\int f(y|\theta)\pi(\theta|\lambda) d\theta} \quad (3.20)$$

Nótese que la contribución de los datos (en la forma de la verosimilitud de  $f$ ) y el conocimiento externo u opinión (en la forma de la a priori  $\pi$ ) en la distribución a posteriori final. Dado que en la práctica  $\lambda$  será desconocida, una distribución en una segunda etapa (o hyperpriori)  $h(\lambda)$  será requerida y la función anterior será reemplazada por

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{\int f(y|\theta)\pi(\theta|\lambda)h(\lambda) d\lambda}{\int \int f(y|\theta)\pi(\theta|\lambda)h(\lambda) d\theta d\lambda} \quad (3.21)$$

Al modelar los  $\theta_i$  como variables aleatorias (en lugar de ser valores fijos), es posible introducir estructuras de correlación específicas a éstas nuevas variables, y como consecuencia a los datos observados  $y_i$  también. Los métodos jerárquicos Bayesianos permiten una variedad de aplicaciones en el análisis de datos espaciales.

Mientras que la computación asociada con los métodos Bayesianos puede ser de grandes dimensiones, la inferencia subsecuente es relativamente inmediata, especialmente en el caso de estimaciones. Esto es, porque una vez que se ha calculado (o estimado) la distribución a posteriori, la inferencia se obtiene al realizar un resumen acerca de esta distribución, dado que por el teorema de Bayes ésta resume todo lo que se sabe acerca de los parámetros del modelo a partir de los datos.

#### 3.6.1. Estimación puntual y por intervalo

Una de las razones por las cuales el enfoque bayesiano resulta muy útil es la posibilidad de definir estimadores óptimos, sin restricciones como las de los estimadores insesgados en el enfoque frecuentista. El estimador de Bayes considera que una función de pérdida asociada a un estimador  $\delta$  del parámetro  $\theta$ , siendo ésta  $L(\theta, \delta)$ , habiendo funciones de pérdida conocidas que son

- Función de pérdida cuadrática

$$L(\theta, \delta) = (\theta - \delta)^2 \quad (3.22)$$

- Función de pérdida absoluta

$$L(\theta, \delta) = |\theta - \delta| \quad (3.23)$$

### 3.6. Inferencia Bayesiana

---

Supongamos que  $\theta$  es univariado. Dada la distribución a posteriori  $p(\theta | y)$ , es una estimación puntual Bayesiana de  $\theta$ . Tres opciones conocidas representan la media posterior,

$$\hat{\theta} = E(\theta | y), \quad (3.24)$$

la mediana posterior,

$$\hat{\theta} : \int_{-\infty}^{\hat{\theta}} p(\theta | y) d\theta = 0.5, \quad (3.25)$$

y la moda posterior,

$$\hat{\theta} : p(\hat{\theta} | y) = \sup_{\theta} p(\theta | y), \quad (3.26)$$

Nótese que el último estimador es típicamente el más sencillo de calcular dado que no requiere de ninguna integración, se puede reemplazar  $p(\theta | y)$  por su forma no estandarizada  $f(y | \theta)p(\theta)$  y obtener la misma respuesta (ya que éstas dos sólo difieren en un factor multiplicativo  $m(y)$  que no depende de  $\theta$ ). En efecto, si la posterior existe dada una a priori unitaria  $p(\theta) = 1$ , entonces la moda posterior es el estimador de máxima verosimilitud (MLE).

La distribución posterior permite estipular probabilidades directas no solo acerca de la mediana, sino de cualquier cuantil. Por ejemplo, supongamos que podemos encontrar los cuantiles  $\alpha/2$  y  $(1 - \alpha/2)$  de  $p(\theta | y)$ , esto es, los puntos  $q_L$  y  $q_U$  tales que

$$\int_{-\infty}^{q_L} p(\theta | y) d\theta = \alpha/2 \quad (3.27)$$

$$\text{y } \int_{q_U}^{\infty} p(\theta | y) d\theta = 1 - \alpha/2 \quad (3.28)$$

Entonces es claro  $P(q_L < \theta < q_U | y) = 1 - \alpha$ , la probabilidad de que  $\theta$  se encuentre en  $(q_L, q_U)$  es de  $100(1 - \alpha)\%$ .

#### 3.6.2. Distribución a priori

El enfoque Bayesiano como ya se describió, ofrece un vínculo entre los datos y el conocimiento previo que se posea. Entonces, la complicación que se tiene con las distribuciones a priori  $\pi(\theta)$  es el problema para elegir las en la práctica, por lo que a menudo se emplean familias conjugadas de densidades a priori para facilitar los cálculos.

### 3.6. Inferencia Bayesiana

---

#### Distribución a priori no informativa de Jeffreys

En muchas ocasiones no existe información a priori confiable acerca de los parámetros o se pretende realizar inferencia basada sólo en los datos. La regla de Jeffreys es una técnica para obtener densidades a priori y es útil cuando no se tiene conocimiento previo de los parámetros de  $f(x|\theta)$ .

La técnica desarrollada por Jeffreys ofrece una alternativa invariante ante transformaciones uno a uno, pero puede conducir a paradojas en la presencia de parámetros de ruido (Souparno y Mallick, 2005). La distribución a priori de Jeffreys para el caso multiparamétrico es

$$f^J(\theta) \propto |I(\theta)|^{1/2} \quad (3.29)$$

donde  $|\cdot|$  implica el determinante y  $I(\theta)$  es la matriz de información de Fisher, la cual se expresa

$$I(\theta)_{ij} = -E_{X|\theta} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(\underline{X}|\theta) \right) \quad (3.30)$$

La densidad a priori de Jeffreys es proporcional a la raíz cuadrada positiva del determinante de la matriz de información esperada de Fisher.

#### Funciones impropias

Una propiedad básica de una función de densidad  $f(x)$  es que integra sobre su rango a 1. Si  $h(x)$  es una función uniforme en la línea real, esto es,  $h(x) = c$ ,  $-\infty < x < \infty$ ,  $c > 0$  entonces es una función impropia, ya que la integral

$$\int_{-\infty}^{\infty} h(x) dx = c \int_{-\infty}^{\infty} dx \quad (3.31)$$

no existe, no importa el valor de  $c$ .

En algunos casos, una función a priori impropia se puede combinar con la verosimilitud para dar una densidad propia.

## 3.6. Inferencia Bayesiana

---

### 3.6.3. Distribución a posteriori

Se señaló que el análisis Bayesiano se basa en la inferencia de la densidad a posteriori, sin embargo, algunos de los problemas que se llegan a presentar en el cálculo del estimador de Bayes son que la densidad a posteriori no está disponible en forma cerrada, debido a que la densidad a priori no presente forma analítica, también puede presentarse que el cálculo de  $E[g(\theta) | \underline{X} = \underline{x}]$  no se puede realizar analíticamente y el manejo de otras funciones de pérdida diferentes al Error Cuadrático.

Se espera que la distribución a posteriori incorpore información proveniente de los datos, y ésta tendrá menor varianza que la distribución a priori al conjuntarse el conocimiento de los datos y de la información a priori.

La distribución de probabilidad a posterior contiene toda la información actual acerca del parámetro  $\theta$ . Comúnmente los resúmenes de localización usados son la media, la mediana y la moda de la distribución; la variación es generalmente resumida por la desviación estándar, el rango intercuantil y otros cuantiles.

Mucha de la práctica recae en el uso de aproximaciones normales al aplicar transformaciones que simetrizan a  $\theta$ , y en este caso la media y la desviación estándar juegan el papel importante. La moda es importante en estrategias computacionales para problemas mas complejos porque facilita el calcular la media o la mediana.

Cuando la distribución posterior presenta una forma cerrada, la media, mediana y la desviación estándar de la distribución posterior tienen también una forma cerrada ([Gelman et al., 2004](#)).

En las últimas décadas han surgido rutinas computacionales en la práctica Bayesiana que han dado solución a dichos problemas analíticos, tales como las Cadenas de Markov Monte Carlo (MCMC) y aproximaciones numéricas, debido a su habilidad para realizar inferencia de distribuciones a posteriori de grandes dimensiones, cuyo objetivo es producir una *muestra* de la distribución a posteriori. MCMC es un método conocido como Metrópolis que fue propuesto primeramente por Metrópolis, Rosenbluth, Teller and Teller (?), que luego fue generalizado por Hastings en 1970. El método de Metrópolis-Hastings (M-H) ha sido empleado extensivamente en la física ([Chib y Greenberg, 1995](#)). Algunos ejemplos de rutinas MCMC son el Muestreo de Gibbs ([Casella y George, 1992](#)) y el algoritmo Metrópolis Hastings (MH) ([Chib y Greenberg, 1995](#)).

### 3.6.4. Distribución predictiva

Una vez obtenida la densidad a posteriori, nos interesa obtener la distribución de la respuesta  $Y$  en un nuevo valor  $s_o$  dado un vector de covariables asociadas  $x(s_o)$ ; este paso

### 3.6. Inferencia Bayesiana

---

en la predicción es el "kriging" bayesiano. Denotando el valor desconocido en ese nuevo punto por  $Y(s_0)$  y usando las notaciones  $Y_0 \equiv Y(s_0)$  por conveniencia, la solución en la teoría Bayesiana establece encontrar la distribución predictiva

$$p(y_0 | \mathbf{y}, X, \mathbf{x}_0) = \int p(y_0, \theta | \mathbf{y}, X, \mathbf{x}_0) d\theta = \int p(y_0 | \mathbf{y}, \theta, \mathbf{x}_0) p(\theta | \mathbf{y}, X) d\theta \quad (3.32)$$

donde  $p(y_0 | \mathbf{y}, \theta, \mathbf{x}_0)$  es la distribución condicional de  $Y_0$  dados los datos originales  $\mathbf{Y}$ .

Supongamos que se genera una muestra a posteriori  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(G)}$  proveniente de la distribución posterior  $p(\theta | \mathbf{y}, X)$ . Entonces de la integral predictiva anterior puede ser calculada mediante simulación Monte Carlo de la forma

$$\hat{p}(y_0 | y, X, x_0) = \frac{1}{G} \sum_{g=1}^G p(y_0 | y, \theta^{(g)}, x_0) \quad (3.33)$$

En la práctica se utiliza simulación para generar cada  $\theta^{(g)}, y_0^{(g)} \sim p(y_0 | \mathbf{y}, \theta^{(g)}, \mathbf{x}_0)$ . La colección  $\{y_0^{(1)}, y_0^{(2)}, \dots, y_0^{(G)}\}$  es una muestra de la densidad predictiva posterior.

#### 3.6.5. Computación Bayesiana

##### Muestreo de Gibbs

Suponemos que el modelo contiene  $k$  parámetros,  $\theta = (\theta_1, \dots, \theta_k)'$ . Para implementar el muestreo de Gibbs primero se debe asumir que se pueden generar muestras de cada una de las distribuciones condicionales completas  $\{p(\theta_i | \theta_{j \neq i}, \mathbf{y}), i = 1, \dots, k\}$  en el modelo. Dichas muestras deben poder obtenerse directamente (si las distribuciones condicionales fueran conocidas) o indirectamente (vía muestreo de aceptación rechazo).

**Muestreo de Gibbs:** Para  $(t \in 1 : T)$ , repetir:

**Paso 1:** Generar  $\theta_1^{(t)}$  de  $p(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{y})$

**Paso 2:** Generar  $\theta_2^{(t)}$  de  $p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{y})$

⋮

**Paso k:** Generar  $\theta_k^{(t)}$  de  $p(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{y})$



### 3.6. Inferencia Bayesiana

---

Bajo condiciones regulares, que se cumplen generalmente para la mayoría de los modelos estadísticos, se puede mostrar que la muestra obtenida en la iteración  $t$ ,  $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$ , converge en distribución a la verdadera distribución conjunta posterior  $p(\theta_1, \dots, \theta_k \mid \mathbf{y})$  (Banerjee *et al.*, 2004).

#### El algoritmo Metrópolis-Hastings

El muestreo de Gibbs es sencillo de aplicar pero requiere de la habilidad para poder muestrear de cada una de las distribuciones condicionales  $p(\theta_i \mid \theta_{j \neq i}, \mathbf{y})$  completas. Desafortunadamente, cuando la distribución a priori  $p(\theta)$  y la verosimilitud  $f(\mathbf{y} \mid \theta)$  no son conjugadas, una o más de las distribuciones condicionales puede no tener una forma cerrada. Aún en esta situación,  $p(\theta_i \mid \theta_{j \neq i}, \mathbf{y})$  existe excepto por una constante de proporcionalidad, dado que es proporcional a la porción de  $f(\mathbf{y} \mid \theta) p(\theta)$  que implica a  $\theta_i$ .

El algoritmo de Metrópolis (o algoritmo Metrópolis-Hastings) es un algoritmo de aceptación-rechazo que ataca este problema, dado que requiere sólo una función proporcional a la distribución de la cual se desea muestrear, requiriendo un paso de aceptación de una densidad candidata particular.

Como se mencionó, el objetivo del algoritmo Metrópolis, en este caso, es la generación de condicionales (típicamente univariadas) completas. Supongamos que se requiere generar de una distribución conjunta a posteriori  $p(\theta \mid y) \propto h(\theta) \equiv f(\mathbf{y} \mid \theta)p(\theta)$ . Comenzamos por especificar una densidad candidata  $q(\theta^* \mid \theta^{(t-1)})$  que es una función de densidad válida para cada posible valor de la variable condicionante  $\theta^{(t-1)}$ , y satisface  $q(\theta^* \mid \theta^{(t-1)}) = q(\theta^{(t-1)} \mid \theta^*)$ , por lo tanto  $q$  es simétrica en sus argumentos. Dado un valor inicial  $\theta^{(0)}$  en la iteración  $t = 0$  el algoritmo procede como se muestra:

**Algoritmo Metrópolis-Hastings:** Para  $(t \in 1 : T)$ ,

1. Generar  $\theta^*$  de  $q(\cdot \mid \theta^{(t-1)})$
2. Calcular la razón  $r = h(\theta^*)/h(\theta^{(t-1)}) = \exp[\log h(\theta^*) - \log h(\theta^{(t-1)})]$
3. Si  $r \geq 1$ , entonces  $\theta^{(t)} = \theta^*$ ;

$$\text{Si } r < 1, \text{ entonces } \theta^{(t)} = \begin{cases} \theta^* & \text{con probabilidad } r \\ \theta^{(t-1)} & \text{con probabilidad } 1 - r \end{cases} .$$

Entonces bajo condiciones generales, como las que sustentan el muestreo de Gibbs, un valor  $\theta^{(t)}$  converge en distribución a un valor de la densidad posterior verdadera  $p(\theta \mid \mathbf{y})$ . La convergencia al utilizar Metrópolis se esperaría que fuera más lenta que con un muestreo de Gibbs regular.

## 3.7. Selección del mejor modelo

### 3.7.1. Criterio de Información de la Devianza (DIC)

El progreso en la computación bayesiana debido a los métodos MCMC ha hecho posible el ajustar modelos estadísticos complejos lo que implica el determinar el modelo que mejor se ajuste de una gran cantidad de posibles candidatos. Por lo tanto, se ha hecho cada vez más necesario e importante el desarrollar criterios que ayuden en la selección del modelo más eficiente.

Una propuesta hecha por Spiegelhalter, Best y Carlin (2006) es el Deviance Information Criterion (DIC), una versión o generalización bayesiana del muy conocido Criterio de Información de Akaike (AIC), el cual se relaciona además con el Criterio de Información Bayesiana (BIC) (Celeux, 2006)(?).

El DIC se basa en la distribución a posteriori de la log-verosimilitud. Identifica el número de parámetros en los modelos, lo que es requerido para el cálculo del Criterio de Información Bayesiana y el Criterio de Información de Akaike.

Asumamos en general, que la distribución condicional de los datos  $\mathbf{y} = (y_1, \dots, y_n)'$ , depende de un vector de parámetros  $\theta$  con dimensión  $p$ . Desde un punto de vista frecuentista, la valoración del modelo se basa en la *desvianza*, la diferencia en la log-verosimilitud entre los modelos ajustado y el saturado. El modelo saturado se refiere al modelo con tantos parámetros como observaciones, lo que conlleva a un ajuste perfecto de los datos. La distribución posterior de la desvianza clásica para un modelo bayesiano se define por

$$D(\theta) = -2\log f(y | \theta) + 2\log h(y) \quad (3.34)$$

donde  $\log h(y)$  es un término de estandarización que esta en función de los datos. Basado en la distribución posterior de  $D(\theta)$ , el DIC consiste de dos componentes: un término que determina la bondad de ajuste y el otro término que penaliza el incremento de la complejidad del modelo.

Entre mayor ajuste existe del modelo a los datos, mayores son los valores de la verosimilitud.  $\bar{D}$  se define como dos veces la log-verosimilitud. El segundo componente cuantifica la complejidad del modelo mediante el *número efectivo de parámetros*,  $p_D$ , definido como la diferencia entre la media posterior de la desviación y la desviación evaluada en la media posterior de los parámetros:

### 3.7. Selección del mejor modelo

---

$$\begin{aligned}
 p_D &= E_{\theta|y} [D] - D(E_{\theta|y} [\theta]) \\
 &= \bar{D} - D(\bar{\theta}) \\
 &= E_{\theta|y} [-2 \log f(y|\theta) + 2 \log f(y|\bar{\theta}(y))]
 \end{aligned} \tag{3.35}$$

De la expresión anterior puede notarse que  $p_D$  es el término que penaliza para verosimilitudes aproximadamente normales y para información a priori difusa. En esta situación, la media posterior iguala el estimador de máxima verosimilitud (MLE)  $\hat{\theta}$ , y la distribución a posterior de  $\theta$  es bien aproximada por una distribución normal multivaridada con media igual a el estimador de máxima verosimilitud

$$\theta|y \approx N\left(\hat{\theta}, -\frac{\partial^2 \theta}{\partial \theta^2} \Big|_{\theta=\hat{\theta}}\right) \tag{3.36}$$

Representando la desviación a través de una expansión de Taylor de segundo orden alrededor del estimador de máxima verosimilitud, el término de primer orden desaparece y la desviación es aproximada por la suma de la desviación en el estimador de máxima verosimilitud y un término cuadrático que sigue una distribución  $\chi^2$  con  $p$  grados de libertad. Al obtener la esperanza, el número efectivo de parámetros iguala a  $p$ , el verdadero número de parámetros.

El *Criterio de Información de la Devianza* (DIC) se define entonces como la suma de ambos componentes antes mencionados

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}) = D(\bar{\theta}) + 2p_D. \tag{3.37}$$

En una situación clásica, el DIC puede ser visto como una generalización del conocido Criterio de Información de Akaike (AIC), el cual se define por  $AIC = D(\hat{\theta}) + 2p$  donde  $\hat{\theta}$  es el estimador de máxima verosimilitud (MLE) del vector de parámetros y siendo  $p$  el número de parámetros. Para modelos no jerárquicos,  $p \approx p_D$ ,  $\hat{\theta} \approx \bar{\theta}$  y  $DIC \approx AIC$ .

## 3.8. Análisis de convergencia

### 3.8.1. Prueba de convergencia de Gelman y Rubin

Se ha hablado de las ventajas que proveen los métodos MCMC para poder simular de distribuciones complejas y generalmente multivariadas, al generar una cadena de Markov con la densidad objetivo como densidad estacionaria. Por lo general se generan  $m > 1$  cadenas de simulaciones de longitud  $n$ , que pueden ser, aunque no necesariamente, independientes.

Debido a la falta de técnicas para calcular o estimar convergencia de las cadenas de Markov, se necesita llevar a cabo algún tipo de análisis estadístico para evaluar la convergencia. Estos procedimientos llamados diagnósticos de convergencia, tienen dos categorías: aquellos basados únicamente en los valores de salida de la simulación y, los que emplean información adicional acerca de la densidad objetivo (Brooks, 1998).

Gelman and Rubin (1992) propusieron utilizar un algoritmo simple que calculara la moda para localizar regiones de alta densidad y muestrear de una mezcla de distribuciones  $t$  ubicadas en éstas modas para generar valores iniciales apropiados.

Sea  $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$  un vector de parámetros con una distribución final  $p(\theta)$  y se interesa en todos los componentes  $\theta_k$ . La evaluación de convergencia se realiza para cada componente  $\theta_k$ . Sea  $\theta$  cualquier  $\theta_k$  y supóngase que se tienen  $m > 1$  cadenas paralelas de longitud  $n$ , siendo  $\theta_{kj}$  cada  $\theta$  de las  $m$  cadenas de longitud  $n$  con  $j = 1, 2, \dots, n$ ,  $k = 1, 2, \dots, m$ , y se obtiene la varianza entre cadenas  $B/n$ , y  $W$  la varianza dentro de las cadenas, de la siguiente forma

$$\begin{aligned}\frac{B}{n} &= \frac{1}{m-1} \sum_{k=1}^m (\tilde{\theta}_k - \tilde{\theta})^2 \\ W &= \frac{1}{m(n-1)} \sum_{k=1}^m \sum_{j=1}^n (\theta_{kj} - \tilde{\theta}_i)^2\end{aligned}\tag{3.38}$$

Teniendo esto, se calcula la varianza como un promedio ponderado de  $W$  y  $B$

$$\hat{\sigma}^2 = \frac{n-1}{n}W + \frac{B}{n}\tag{3.39}$$

siendo un estimador insesgado de la varianza  $\sigma^2$  y, como estimador de la varianza a posteriori se tiene  $\hat{V} = \hat{\sigma}^2 + \frac{B}{mn}$ .

### 3.8. Análisis de convergencia

---

Para monitorear la convergencia de la cadena, se calcula la razón de varianzas que es llamado *factor de reducción de escala*

$$R = \frac{\hat{V}}{\sigma^2} \quad (3.40)$$

Debido a que el denominador de  $R$  no es conocido, se puede estimar de los datos, por lo tanto, se puede sobreestimar a  $R$  al subestimar  $\sigma^2$  y  $W$ , por lo que

$$\hat{R} = \frac{\hat{V}}{W} \quad (3.41)$$

que puede interpretarse de la siguiente forma: si  $\hat{R}$  se acerca a 1, las cadenas están sobrepuestas y proviene de la misma distribución; si  $\hat{R}$  mayor a 1 implicaría que con mayores simulaciones se pueden mejorar la inferencia sobre  $\theta$  (Brooks, 1998).

# Capítulo 4

## Modelación de precipitaciones extremas en el estado de Tabasco

En este capítulo se describirá la manipulación de los datos de acuerdo a la metodología del capítulo anterior, se presentan y describen los resultados en base a los objetivos propuestos en el primer capítulo.

### 4.1. Base de datos de precipitación histórica

El Servicio Meteorológico Nacional (SMN) que forma parte de la Comisión Nacional del Agua (CONAGUA), lleva a cabo el análisis de variables climatológicas para su medición y el pronóstico del tiempo atmosférico como de eventos meteorológicos a nivel nacional.

El Sistema Meteorológico Nacional cuenta a lo largo del país con mas de 5000 estaciones meteorológicas, información con lo cual se creó la base de datos MAYA v1.0. Dicha base de datos divide al territorio mexicano en una cuadrícula de áreas imaginarias de 0.2 de grado por 0.2 de grado tanto en latitud como en longitud.

Se cuenta con información de precipitaciones diarias de 40 años, esto es, de 1961 a 2000. Por lo tanto, dentro de la información contenida en MAYA v1.0 se tomo la precipitación diaria en nodo por nodo, lo que significa que es la información de precipitación de 14600 días por cada nodo de la cuadrícula en la que se divide el país. Una vez que se tuvo esto, en Arcview se llevó a cabo la separación de los nodos contenidos en el estado de Tabasco dado que la ubicación de los nodos no obedece a la separación política de los estados, los cuales son un total de 135 nodos.

Una vez hecho lo anterior, se realizó el filtro de los datos con el objetivo de obtener

## 4.2. Regresión espacial

---

los máximos anuales de todos los nodos ubicados en el área del estado de Tabasco para posteriormente delimitar específicamente el área de estudio.

En un principio se pensó en que el área de estudio se concentrara mediante el área de las cuencas del río Grijalva y Usumacinta pero debido a que dichos ríos abarcan tanto gran parte del territorio del Tabasco como de Chiapas, por lo que se optó en reducir el área de interés.

Después de estudiar varias opciones de región óptima se llegó a la decisión de tomar a toda el área comprendida por el estado de Tabasco en una forma rectangular que cubriera las colindancias que tiene el estado con Veracruz, Chiapas y Campeche con el fin de que la predicción de los niveles de retorno fueran más reales al incorporar información aledaña.

## 4.2. Regresión espacial

Aunque las herramientas para modelar valores extremos univariados están bien desarrolladas, la extensión a modelos de datos espaciales extremos es un área de amplio estudio. Ésta aplicación está enfocada al análisis de la variación de la precipitación extrema a lo largo del espacio, no en la estructura multivariada de eventos individuales de precipitación (Cooley, 2005).

Con el fin de entender el proceso latente que conlleva la dependencia climatológica de precipitaciones extremas, el enfoque Bayesiano integra todos los datos en un solo modelo. Un modelo espacial no requiere de que los datos se normalicen y por el contrario permite cambios en los parámetros para registrar diferencias en los datos (Cooley, 2005).

Se asume para la región que algunas covariables podrían proveer información importante al describir el proceso espacial latente de los datos extremos. Considérense dos covariables disponibles: distancia al mar y el tiempo promedio en el intervalo de años estudiado. Es factible que la distancia al mar tenga influencia significativa en el comportamiento climatológico de la lluvia extrema. De la misma forma, se supone que el tiempo explique en parte la ocurrencia de precipitaciones extremas.

### 4.2.1. El modelo estadístico

Existen tres niveles en el modelo jerárquico a emplear. El primero, modela los datos (máximos anuales). El segundo nivel modela el proceso latente que conlleva la precipitación extrema para la región. Finalmente, el tercer nivel consiste en las distribuciones a priori de los parámetros que controlan el proceso latente.

## 4.2. Regresión espacial

---

El proceso latente es de interés primario dado que los niveles de retorno son cantidades climatológicas. El nivel de retorno  $z(x)$  (3.9) es función de  $\mu(x), \sigma(x), \xi(x)$  y se lleva a cabo a través de estos parámetros.

### 4.2.2. Función de verosimilitud

Una distribución de valores extremos (GEV) conforma el primer nivel del modelo jerárquico para los datos de máximos anuales de 40 años, donde la función de densidad sería

$$f(\underline{x} | \underline{\mu}, \underline{\sigma}, \underline{\xi}) = \prod_{i=1}^s \prod_{k=1}^{n_i} \frac{1}{\sigma_k} \left[ 1 + \xi_k \left( \frac{x_i - \mu_k}{\sigma_k} \right) \right]^{-\frac{1}{\xi_k} - 1} \exp \left\{ - \left[ 1 + \xi_k \left( \frac{x_i - \mu_k}{\sigma_k} \right)^{-\frac{1}{\xi_k}} \right] \right\} \quad (4.1)$$

De esta forma, teniendo una matriz de 135 columnas y 40 filas, donde las columnas indican las ubicaciones, y las filas los máximos de los 40 años de información.

### 4.2.3. Proceso espacial latente

En el segundo nivel del modelo jerárquico, se caracteriza al proceso latente con base en una estructura que relaciona los parámetros de los datos con las covariables de la región consideradas.

Este método puede ser aplicado independientemente a un número finito de ubicaciones y las estimaciones obtenidas servirán para obtener una estimación del grado de variabilidad u homogeneidad espacial en los datos extremos (Casson y Coles, 1999). Esto requiere la especificación de un modelo espacial, por lo que una opción es la de proponer relaciones lineales para los parámetros  $(\mu, \sigma, \xi)$ , en términos de la ubicación espacial así como de las covariables que se cree tienen efecto significativo en el fenómeno de la precipitación extrema.

Se tiene que para cada  $x \in R^d$  se supone que  $Y(x)$  tiene distribución GEV cuyos parámetros  $\{\mu(x), \sigma(x), \xi(x)\}$  varían ligeramente para  $x \in R^d$  de acuerdo a un proceso estocástico  $S(x)$ . Se asume que los parámetros de la GEV son procesos gaussianos mutuamente independientes. Por ejemplo, para el parámetro de localidad  $\mu(x)$ , se tiene que

$$\mu(x) = f_{\mu(x)}(x; \beta_{\mu}) + S_{\mu}(x; \alpha_{\mu}, \lambda_{\mu}, \kappa_{\mu}) \quad (4.2)$$



## 4.2. Regresión espacial

---

donde  $f_\mu$  es una función determinística dependiendo de parámetros de regresión  $\beta_\mu$  y  $S_\mu$  proceso gaussiano estacionario con media y cero con función de covarianza caracterizada por los parámetros sill  $\alpha_\mu$ , rango  $\lambda_\mu$  y parámetros de forma  $\kappa_\mu$ . Para los parámetros de escala  $\sigma(x)$  y de forma  $\xi(x)$  se emplean planteamientos similares. Donde las funciones determinísticas son funciones lineales de la forma

$$f_\mu(\underline{x}; \beta) = \underline{x}'\beta \quad (4.3)$$

Por lo tanto, las variables quedarían definidas en base al modelo anterior de la siguiente manera

$$\begin{aligned} \mu(z) &= \beta_0 + \beta_1 lon + \beta_2 lat + \beta_3 dist + \beta_4 t \\ \log(\sigma)(z) &= \beta_0 + \beta_1 lon + \beta_2 lat + \beta_3 dist + \beta_4 t \\ \xi(z) &= 1 \end{aligned} \quad (4.4)$$

donde las covariables son la distancia al mar y tiempo promedio en el intervalo de tiempo analizado.

Este es el modelo en general; como se mostrará más adelante, de acuerdo al modelo analizado será el número de covariables que se tengan, y, por tanto, el número de parámetros en cada modelo. Dichas especificaciones se muestran y detallan en la siguiente sección.

### 4.2.4. Distribuciones *a priori* para los parámetros del modelo

En el tercer nivel de la jerarquía, se asignan distribuciones a priori a los parámetros que caracterizan el proceso latente. Asumimos que cada parámetro en esta etapa es independiente de los demás.

Para los parámetros  $\beta$  de regresión que describen la estructura espacial se supone tienen distribución normal multivariada  $\beta_\mu \sim N_p(\underline{\mu}, \Sigma)$ , el parámetro de escala  $\sigma$  tiene distribución similar.

El proceso estocástico es Gaussiano estacionario con media cero, teniendo las funciones de covarianza exponencial y Matérn, donde el parámetro sill de ambas funciones tiene distribución Gamma inversa para los tres parámetros de la distribución GEV. La función de covarianza para los otros dos parámetros de la distribución GEV  $\sigma$  y  $\psi$  toma la misma forma que la descrita para  $\mu$ . Entonces se tiene:

## 4.2. Regresión espacial

---

$$\pi(\alpha_\mu) \sim \frac{\beta^\alpha}{\Gamma(\alpha)} \alpha_\mu^{-\alpha-1} \exp\left\{-\frac{\beta}{\alpha_\mu}\right\} \quad (4.5)$$

El parámetro de rango de la función de covarianza del proceso estocástico tiene distribución Gamma

$$\pi(\lambda_\mu) \sim \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda_\mu^{\alpha-1} \exp\left\{-\frac{\lambda_\mu}{\beta}\right\} \quad (4.6)$$

Y el parámetro de suavizamiento de la función de covarianza tiene de la misma forma distribución Gamma

$$\pi(\kappa_\mu) \sim \frac{1}{\Gamma(\alpha)\beta^\alpha} \kappa_\mu^{\alpha-1} \exp\left\{-\frac{\kappa_\mu}{\beta}\right\} \quad (4.7)$$

Por lo tanto, realizando las operaciones correspondientes de la verosimilitud de la densidad de los datos multiplicada por las distribuciones a priori de cada parametro:

La densidad posterior quedaría

$$p(\Theta | \underline{x}) \propto f(\underline{x} | \underline{\mu}, \underline{\sigma}, \underline{\xi}) N_p(\beta_\mu, \Sigma) N_p(\beta_\sigma, \Sigma) N_p(\beta_\xi, \Sigma) \pi_\mu(\alpha) \times \\ \pi_\sigma(\alpha) \pi_\xi(\alpha) \pi_\mu(\lambda) \pi_\sigma(\lambda) \pi_\xi(\lambda) \pi_\mu(\kappa) \pi_\sigma(\kappa) \pi_\xi(\kappa) \quad (4.8)$$

Una vez que el modelo ha sido explicado, se generaron cadenas con un tamaño de 100000 iteraciones en el proceso latente con su respectivo burn-in, teniendo finalmente cadenas de 10000 iteraciones como estimaciones de la distribución posterior, empleando en cada modelo las covariables respectivas.

### 4.2.5. Modelo de correlación y selección del mejor modelo

Para seleccionar los modelos, se empleó el Criterio de Información de la Devianza (DIC). EL DIC produce una medida del ajuste del modelo  $\bar{D}$  y otra medida de complejidad del modelo  $p_D$  y los suma para obtener un valor final. Un valor menor implica un mejor modelo.

## 4.2. Regresión espacial

---

La selección de las covariables se hizo bajo el supuesto de que dichas variables pudieran tener un efecto significativo en la predicción del comportamiento de la precipitación en el área de estudio. Dado que el estado de Tabasco se encuentra en el Golfo de México, tanto la cercanía al mar así como el tiempo promedio en el intervalo de años estudiado pueden tener relevancia.

De entre los seis modelos analizados, el modelo 1 presentó un DIC cercano al del modelo 3 y se podría decir que la covariable tiempo podría ser no significativa.

El modelo 2, en el cual la única diferencia con el modelo 1 es la función de correlación, (Whittle-Matérn), presenta un DIC alto en relación al modelo 3, por lo cual puede concluirse que ambos son equivalentes.

En el modelo 4, se elimina la covariable tiempo y se emplea el modelo de correlación de Whittle-Matérn; así, se obtuvo un valor de DIC más alto que en el modelo tres, por lo que hay evidencia que la variable tiempo promedio no tiene efecto significativo en el nivel de precipitación.

En el modelo 5, además de no contener la variable distancia, la función de correlación utilizada es la power exponencial. Para éste modelo, el DIC es el más alto de los seis modelos. Finalmente, en el modelo 6 que tampoco considera la covariable distancia pero si la la función de correlación de Whittle-Matérn, se obtuvo un valor del DIC más alto que el obtenido con el modelo 3.

La Tabla 4.1 muestra los modelos empleados y sus correspondientes valores del DIC.

**Tabla 4.1:** Comparación de los modelos empleados en la regresión espacial

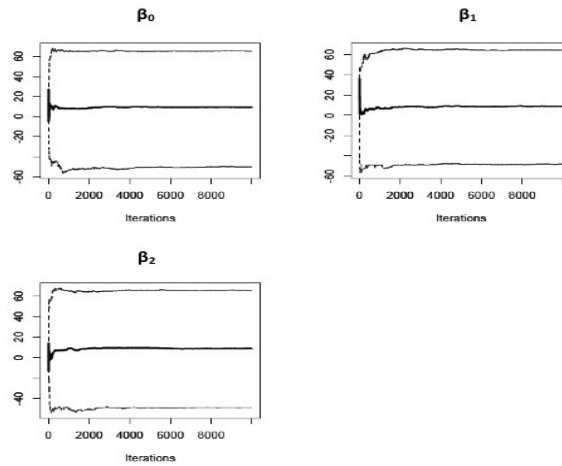
Modelo	Covariables	M. correlación	DIC
1	lat, lon, t, dist	powexp	52289.78
2	lat, lon, t, dist	Whittle-Matérn	52290.32
3	lat,lon, dist	powexp	52282.39
4	lat, lon, dist	Whittle-Matérn	52289.80
5	lat, lon, t	powexp	53481.39
6	lat, lon, t	Whittle-Matérn	52287.69

De entre los seis modelos que se muestran en la Tabla 4.1, aquel que tiene un *Deviance Information Criterion* (DIC) más bajo es el tercero, siendo las covariables incluidas, la latitud, longitud y distancia al mar, obteniendo un  $DIC = 52282.39$ .

Por otra parte, se llevó a cabo el diagnóstico de Gelman y Rubin así como el diagnóstico de Heidelberg y Welch para comprobar la convergencia de las iteraciones para los parámetros de regresión de los modelos lineales de cada parámetro de la GEV, lo que se ilustra en las siguientes imágenes

## 4.2. Regresión espacial

---

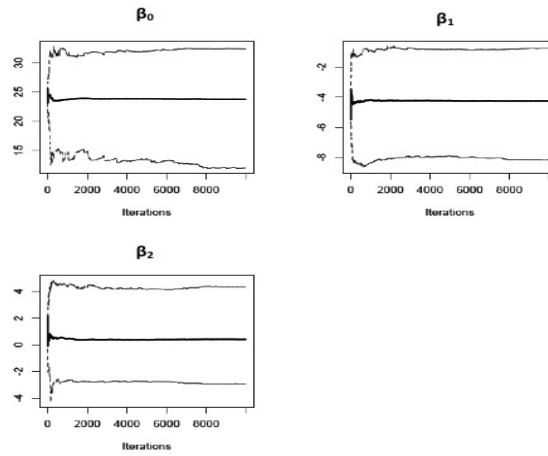


**Figura 4.1:** Convergencia de las iteraciones de los parámetros de regresión para el  $\mu$  de la GEV

Combinando la información de precipitación con las estimaciones de los parámetros de la Distribución de Valores Extremos, se calcula el  $n$ -ésimo período de retorno, que es el nivel de precipitación excedido cada año con probabilidad  $1/n$ , derivado de la función de distribución inversa acumulada de la GEV. Una vez que se escogió el modelo que presentó el menor valor del DIC, es aquel con el cual se realizarán los mapas para los períodos de retorno de 5, 10 y 100 años, los cuales se mostrarán en la siguiente sección.

Se presenta un resumen de la distribución de los parámetros  $\mu, \sigma, \xi$  a posteriori que incluyen media y desviación estándar por nodo, de los cuales a continuación está el del parámetro de localidad, los demás se incluyen en los anexos. Se muestran las densidades posteriores para los parámetros  $(\mu, \sigma, \xi)$ , la que se muestra a continuación es para el parámetro de localidad, la de los dos parámetros restantes se hallan en la sección de anexos. Y se incluye un resumen *a posteriori* de los coeficientes de regresión para cada parámetro de la GEV, así como un intervalo de credibilidad para cada uno de ellos.

## 4.2. Regresión espacial



**Figura 4.2:** Convergencia de las iteraciones de los parámetros de regresión para el  $\sigma$  de la GEV

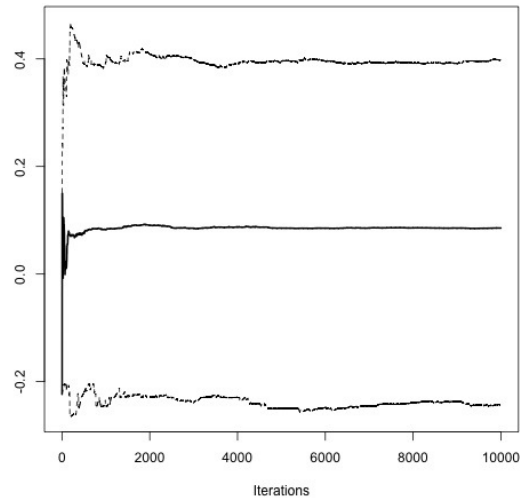
**Tabla 4.2:** Resumen *a posteriori* de los  $\beta$ 's del parámetro localidad e intervalo de credibilidad al 95% para el modelo 3

	Media	IC 95 %
$\beta_{lon}$	17.80	(-41.12 , 69.74)
$\beta_{lat}$	-9.55	(-26.02 , 8.04 )
$\beta_{dist}$	0.9025	(-16.42 , 17.33)

**Tabla 4.3:** Resumen *a posteriori* de los  $\beta$ 's del parámetro escala e intervalo de credibilidad al 95% para el modelo 3

	Media	IC 95 %
$\beta_{lon}$	23.34	(12.73 , 32.81)
$\beta_{lat}$	-4.30	(-8.15 , -0.7627)
$\beta_{dist}$	0.4802	(-3.02 , 4.177)

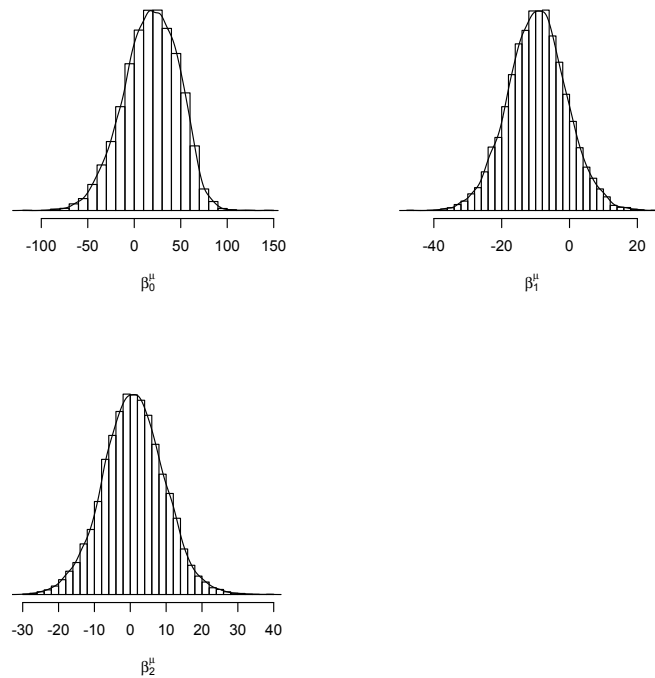
## 4.2. Regresión espacial



**Figura 4.3:** Convergencia de las iteraciones del parámetro de regresión para  $\xi$  de la GEV

**Tabla 4.4:** Resumen *a posteriori* del parámetro forma e intervalo de credibilidad al 95 % para el modelo 3

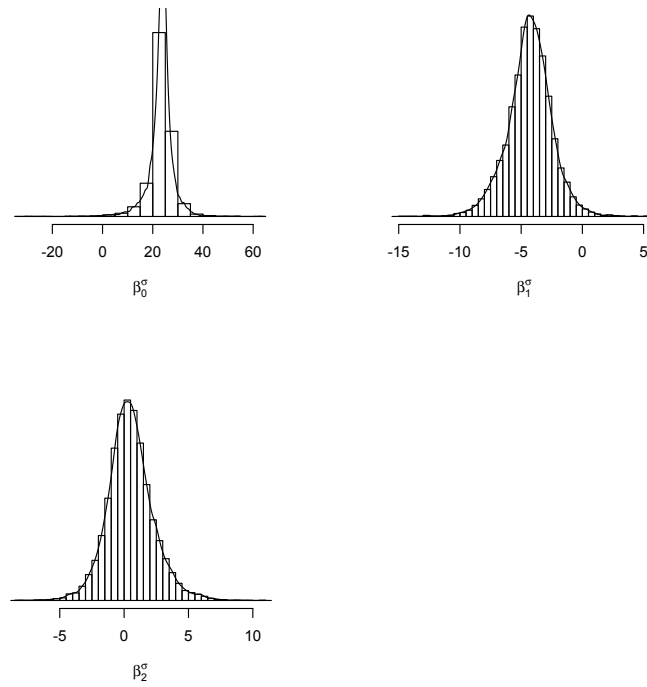
	Media	IC 95 %
$\beta_0$	0.08144	(-0.2454 , 0.3969)



**Figura 4.4:** Densidades a posteriori estimadas de los parámetros de regresión para la localidad

## 4.2. Regresión espacial

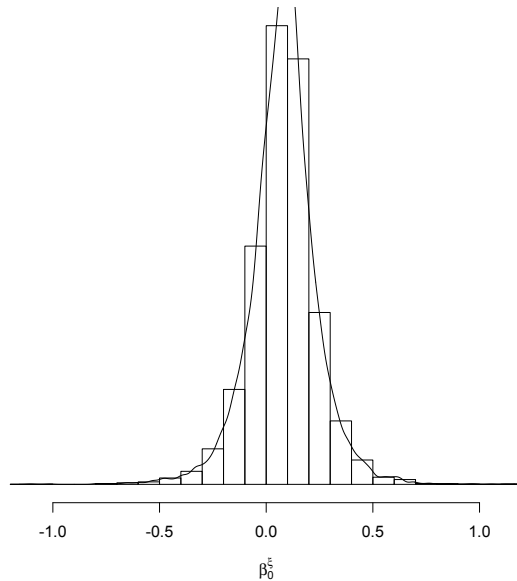
---



**Figura 4.5:** Densidades a posteriori estimadas de los parámetros de regresión para la escala

### 4.3. Niveles de retorno

---



**Figura 4.6:** Densidad a posteriori estimada del parámetro de regresión para la forma

## 4.3. Niveles de retorno

### 4.3.1. Niveles de retorno basados en el modelo espacial ajustado

Como resultado de este trabajo se obtuvieron estimaciones de niveles de precipitación, como se mencionó anteriormente para tres períodos de retorno, que son cinco, diez y cien años respectivamente.

Esto quiere decir que las estimaciones realizadas para los tres períodos de retorno, son los valores de precipitación probable a presentarse representados por las curvas de nivel que se muestran en las gráficas, a lo largo del territorio del estado de Tabasco.

De las principales razones de estudio de extremos es indagar acerca de sus posibles causas de estos eventos, con el objetivo de realizar estimaciones a futuro que puedan ser de ayuda en la prevención de posibles futuros desastres. Esto es de gran importancia para el estado de Tabasco debido a sus recientes y históricas inundaciones, donde la precipitación es un factor que influye de gran manera para tales eventos. Dada la ubicación geográfica del estado de Tabasco en la cuenca del río Grijalba-Usumacinta, Villahermosa su capital es



### 4.3. Niveles de retorno

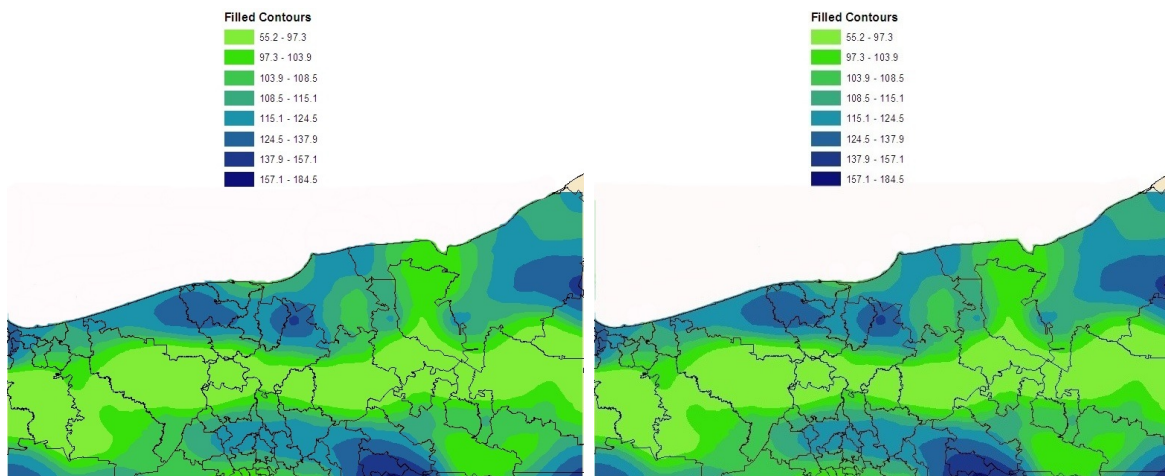
de las más afectadas por éstas catástrofes climatológicas dejan a su paso una gran pérdida económica, de vivienda, agrícola.

En las siguientes figuras se muestra una comparación de los niveles de retorno obtenidos con los ajustados de la GEV y con la función de distribución predictiva.

#### 4.3.2. Niveles de retorno basado en la función de distribución predictiva de Valores Extremos

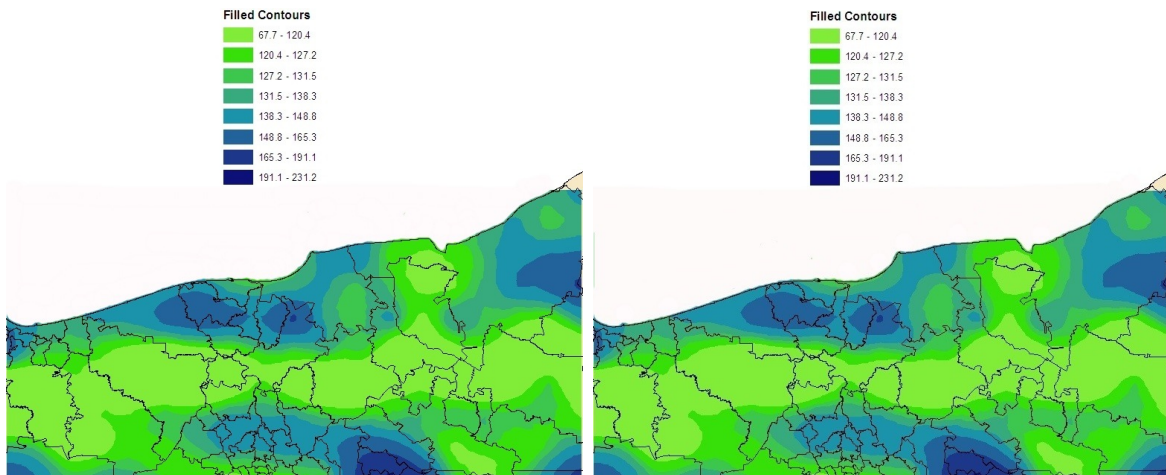
Una vez teniendo la distribución de los parámetros a posteriori de la GEV, se computó la función de distribución predictiva generando observaciones aleatorias de ésta distribución, con los cuales se calcularon los niveles de retorno para 5, 10 y 100 años de período de retorno. Éstas cantidades presentan en general valores más altos que las precipitaciones obtenidas con la regresión espacial, se muestran a continuación los mapas con las curvas de nivel y, en los anexos se muestra la tabla con las estimaciones de precipitación por ambos métodos.

A continuación se muestran los comparativos de los niveles de retorno para los diferentes períodos de retorno

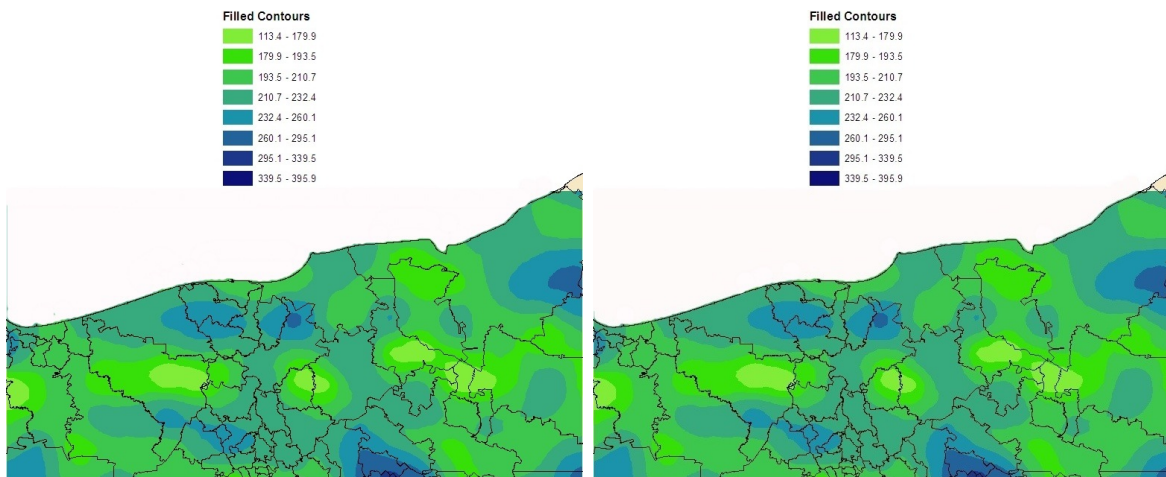


**Figura 4.7:** Niveles de retorno para 5 años de período de retorno obtenidos con el ajuste de la GEV (izquierda) y con la función predictiva a posteriori (derecha)

### 4.3. Niveles de retorno



**Figura 4.8:** Niveles de retorno para 10 años de período de retorno obtenidos con el ajuste de la GEV (izquierda) y con la función predictiva a posteriori (derecha)



**Figura 4.9:** Niveles de retorno para 100 años de período de retorno obtenidos con el ajuste de la GEV (izquierda) y con la función predictiva a posteriori (derecha)

# Capítulo 5

## Conclusiones

- La motivación de haber realizado el estudio con el enfoque Bayesiano como se menciona en el trabajo, nos permite conjuntar la información de precipitación que se tiene, además del conocimiento previo que se posee en relación a éste tipo de eventos.
- Los análisis hidrológicos tradicionales emplean el enfoque frecuentista para el cálculo de los niveles de retorno, en los que sólo se emplean los datos históricos. Esto deja de lado información muy valiosa para este tipo de estudios, que la estadística Bayesiana utiliza como información que ayuda fuertemente a realizar la inferencia a posteriori para eventos futuros.
- La geoestadística ofrece la ventaja de analizar e interpretar información correlacionada geográficamente. Esto permitió analizar la información de precipitación registrada en los nodos, mediante su ubicación geográfica (latitud y longitud), con base en que esta variable varía de acuerdo a su ubicación espacial.
- El objetivo del trabajo fue el de obtener los niveles de precipitación para tres períodos de retorno, que fueron 5, 10 y 100 años para obtener así las isocurvas de precipitación correspondientes a cada período, gráficas que se mostraron en el gráfico de la sección anterior. Esto implica que los valores en las figuras 4.7, 4.8 y 4.9, son valores nodales probables a presentarse en períodos de tiempo a 5, 10 y 100 años respectivamente. De las tres figuras, puede denotarse que las zonas con mayores valores de precipitación son aquellas colindantes con el estado de Chiapas, que es la zona aguas arriba de la cuenca del Grijalva; así como la ciudad de Villahermosa que se encuentra en la desembocadura del río Grijalva y, en general a lo largo de toda la costa del estado de Tabasco, esto debido a la colindancia con el Golfo de México.
- Para este trabajo, debido a la interpolación que conlleva previamente la información de precipitación, esto es, que los datos no son datos puntuales por nodo, sino que son interpolaciones en un área de 20 kilómetros a la redonda del nodo. Por lo tanto,

## 5. Conclusiones

---

la interpolación de los extremos sobreestimarán los extremos verdaderos del área; en trabajos futuros se podría eliminar la interpolación de la precipitación y con esto obtener un estudio más puntual.

- Se realizaron dos mapas de riesgo. Los primeros fueron el resultado de la estimación realizada con el ajuste de la distribución generalizada de valores extremos; el segundo, las estimaciones se obtuvieron directamente de la distribución predictiva a posteriori, que fueron relativamente mayores a las estimaciones con el ajuste de la GEV.
- En estudios futuros sería muy útil el empleo de un umbral, esto es establecer un límite a partir del cual la precipitación represente un riesgo altamente potencial de desastre para alguna zona en particular. Con esto, en lugar de emplear un solo dato por año, se emplearía una mayor cantidad de información con un umbral establecido. De la misma forma, otro tipo de estudio importante para zonas como Tabasco que se encuentran en zonas costeras, es el análisis por temporadas, esto es, realizar un análisis de los fenómenos de precipitación en las dos temporadas más notorias que son la invierno-primavera y verano-otoño.
- La utilidad de esta información radica en su variable implementación a varios campos de conocimiento, desde la proyección de obras hidráulicas, planes de prevención, estudios de vulnerabilidad, la planeación de calendarios de riego, diseño de sistemas de captación de agua de lluvias, entre otras aplicaciones.

# Referencias

- Banerjee, S., Carlin, B. P. y Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.
- Casella, G. y George, E. I. (1992). Explaining the Gibbs Sampler. *American Statistical Association*, 46, 167–174.
- Casson, E. y Coles, S. (1999). Spatial Regression Models for extremes. *Journal of Extremes*, 1, 4, 449–468.
- Chib, S. y Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistical Association*, 49, 327–335.
- Cooley, D. S. (2005). *Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances*. Tesis Doctoral, University of Colorado, Department of Applied Mathematics.
- Gelman, A., Carlin, J. B., Stern, H. S. y Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC.
- Hengl, T. (2009). *A practical guide to Geostatistical Mapping*. Google.
- Johnson, D. S. y Hoeting, J. A. (2011). Bayesian multimodel inference for Geostatistical regression models. *Plos One*, 6, 11.
- Souparno, G. y Mallick, B. K. (2005). A hierarchical bayesian spatio-temporal model for extreme precipitation events. Department of Statistics, Texas A&M University.
- Spiegelhalter, D. J., Best, N. G. y Carlin, B. P. (2006). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, 583–640.
- Vaca, J. G. L. (2005). *Estimadores Bayesianos en la distribucion de Valores Extremos tipo Gumbel*. Proyecto Fin de Carrera, Colegio de Postgraduados.

# Anexos

## Anexo A: Niveles de retorno con las funciones predictiva y regresión espacial

Se muestran los valores de las precipitaciones obtenidas (láminas en mm acumuladas en 24 hrs) mediante la función predictiva de la distribución Generalizada de Valores extremos (izquierda), y los niveles de retorno obtenidos mediante la regresión espacial (derecha).

Sitio	5 años	10 años	100 años	5 años	10 años	100 años
1	184.123	220.929	354.471	164.854	195.102	306.463
2	121.398	138.460	219.844	115.448	134.723	204.079
3	90.454	108.659	179.266	91.786	109.007	170.218
4	108.759	127.741	177.258	107.340	126.414	189.916
5	184.491	211.245	323.932	169.792	199.437	308.430
6	184.225	231.239	395.885	168.996	201.800	329.632
7	156.538	190.231	300.758	154.510	185.028	301.353
8	91.427	107.648	180.332	96.191	115.689	187.579
9	107.719	126.254	208.756	100.853	120.311	190.651
10	91.768	111.354	173.486	89.427	107.128	169.992
11	84.657	101.841	151.459	84.181	100.562	158.912
12	107.569	124.367	183.928	109.853	129.259	192.745
13	105.668	125.088	194.346	106.073	126.156	192.387
14	115.064	138.795	213.424	112.276	133.393	203.859
15	126.960	149.923	223.541	126.863	150.789	236.037
16	113.855	130.321	187.451	127.803	151.402	237.562
17	134.422	154.814	265.777	136.240	159.854	247.417
18	154.078	187.724	295.510	134.300	158.856	251.482
19	113.688	135.384	223.788	104.507	126.364	211.565
20	102.647	122.139	207.152	103.672	124.280	203.067
21	86.783	104.299	173.953	91.220	111.300	186.421
22	105.426	129.940	215.152	102.307	121.977	194.989
23	80.212	96.797	160.932	84.352	102.035	168.218
24	101.639	123.851	184.225	109.137	131.569	207.557

25	126.293	153.375	231.018	124.866	149.101	233.906
26	141.948	171.083	293.434	134.591	160.234	252.614
27	129.304	153.409	218.803	135.762	161.384	255.315
28	129.013	153.176	233.042	125.170	151.436	247.813
29	117.575	145.547	244.037	121.807	147.519	241.861
30	130.970	156.835	279.079	122.158	147.201	239.349
31	108.063	127.779	234.334	109.458	131.687	214.375
32	93.260	109.724	163.239	101.904	121.727	197.469
33	103.218	124.795	198.277	97.568	118.064	197.502
34	83.466	100.446	179.101	87.839	107.484	183.553
35	77.539	96.269	158.197	84.542	104.406	182.561
36	106.677	132.777	263.271	95.918	118.743	210.020
37	109.031	128.871	196.221	108.316	131.090	209.613
38	97.729	114.983	182.834	105.403	127.870	207.537
39	109.340	133.293	204.287	113.908	138.117	227.097
40	112.963	136.127	218.254	119.061	145.337	244.254
41	121.585	145.692	213.298	126.789	153.411	255.242
42	118.314	145.368	215.455	127.010	153.376	254.539
43	124.832	142.437	233.723	128.994	153.408	246.604
44	103.882	126.197	220.040	105.831	128.607	218.013
45	104.985	127.691	232.850	99.364	121.371	209.355
46	111.669	131.008	201.461	114.854	141.167	237.712
47	120.194	148.223	234.041	117.768	144.010	243.968
48	110.403	133.250	213.494	115.491	141.550	243.975
49	117.973	150.389	226.515	115.017	141.636	247.742
50	108.485	135.489	242.972	112.902	139.015	244.122
51	122.142	153.573	291.158	114.418	140.309	245.666
52	95.397	118.100	230.802	94.689	117.726	213.184
53	111.438	135.343	251.920	107.114	132.667	240.484
54	73.507	95.679	189.464	68.431	86.420	157.632
55	73.507	95.423	181.899	73.897	93.360	173.790
56	64.423	79.911	172.215	69.191	88.567	177.482
57	62.776	77.257	156.983	71.342	91.308	186.540
58	114.813	151.299	273.380	98.292	125.188	244.011
59	71.891	87.743	176.932	79.670	100.262	186.043
60	73.475	86.237	143.682	74.018	87.769	137.157
61	64.916	78.249	133.153	67.279	80.034	126.458
62	92.581	118.891	229.647	90.446	112.960	204.785
63	77.042	103.583	209.725	77.000	98.848	194.232
64	68.548	90.944	231.503	65.571	84.098	170.892
65	62.389	77.525	135.921	68.042	86.930	173.954
66	76.484	94.190	147.678	76.164	96.774	182.338
67	96.836	128.304	251.339	87.897	111.368	210.217
68	72.549	89.263	185.480	70.643	90.187	178.483
69	58.176	71.660	124.041	60.492	75.537	143.479
70	55.238	67.676	134.670	67.434	84.731	157.392
71	82.754	98.332	148.862	81.661	96.761	150.596

---

72	75.906	88.461	135.374	77.242	91.049	140.071
73	110.033	130.143	204.302	108.499	128.558	200.348
74	76.481	94.578	179.106	74.654	92.048	162.071
75	56.996	68.831	113.364	64.670	79.559	144.285
76	74.920	91.577	180.554	75.553	94.726	176.473
77	73.908	90.332	161.124	78.802	96.923	169.091
78	91.310	113.262	183.134	89.573	108.730	182.890
79	102.331	122.991	192.096	105.283	125.017	198.068
80	93.237	110.713	177.278	99.092	118.581	188.864
81	69.320	80.969	132.427	72.456	86.076	134.355
82	78.809	91.853	134.441	81.729	96.900	150.750
83	115.819	141.120	265.805	116.225	139.218	223.514
84	132.973	152.237	240.972	131.635	153.134	232.336
85	97.539	111.642	179.065	102.873	122.169	196.266
86	102.550	130.992	232.462	96.345	117.803	203.396
87	84.222	106.195	187.839	85.953	106.710	194.678
88	86.264	111.357	223.214	87.110	108.975	204.892
89	75.580	95.191	179.962	82.306	103.333	194.467
90	90.084	110.367	212.891	91.474	112.812	202.462
91	100.539	122.386	200.117	99.440	121.114	208.239
92	102.866	121.552	187.696	113.455	134.382	214.021
93	162.872	193.052	298.703	150.301	175.202	266.417
94	123.471	151.565	242.562	126.072	149.494	235.379
95	89.805	104.731	179.890	93.980	112.285	177.390
96	101.972	121.101	195.547	98.746	118.343	188.231
97	98.018	113.940	177.490	121.476	145.502	233.721
98	146.107	174.830	264.110	146.568	173.908	275.409
99	78.618	100.111	189.089	89.447	111.496	205.947
100	142.404	175.159	276.332	136.027	163.502	263.817
101	95.011	118.289	187.651	97.342	117.945	191.185
102	95.840	117.951	185.573	101.945	124.834	207.857
103	178.513	217.832	353.499	162.024	192.991	306.528
104	120.752	147.072	228.924	121.797	149.813	255.897
105	145.590	182.426	296.084	130.988	161.695	279.386
106	145.491	176.424	282.887	148.358	177.816	284.627
107	146.763	178.908	262.338	144.523	170.332	260.389
108	120.006	143.968	226.351	122.814	146.145	226.710
109	127.183	147.422	215.907	121.453	144.217	221.561
110	120.838	140.247	210.130	125.103	146.858	220.261
111	168.978	194.350	325.159	155.360	183.085	284.972
112	128.675	153.882	293.393	128.026	156.559	263.012
113	142.424	170.269	272.534	131.546	158.037	254.949
114	117.096	135.935	210.091	122.767	145.172	224.917
115	92.583	105.971	154.887	99.316	117.185	176.794
116	93.884	107.295	159.521	106.574	127.065	198.770
117	112.701	131.648	194.063	120.537	144.662	232.166
118	100.291	118.417	184.105	110.144	134.204	219.381



119	125.463	151.349	233.709	117.212	140.345	219.163
120	102.009	118.402	174.280	104.610	124.003	188.369
121	105.128	128.040	183.686	111.280	135.013	215.688
122	103.929	124.494	176.721	101.749	120.924	182.489
123	124.487	147.842	219.566	114.829	136.279	207.127
124	115.990	135.517	195.922	116.676	140.218	218.362
125	126.150	160.492	268.954	121.415	149.052	245.175
126	130.997	158.239	261.301	128.879	156.058	246.258
127	119.412	141.440	220.632	120.715	144.523	220.944
128	101.082	119.126	176.182	102.452	121.645	181.203
129	98.659	115.277	172.759	99.524	117.852	175.196
130	139.689	167.362	256.341	125.406	150.257	233.377
131	123.554	149.076	225.021	125.595	149.761	229.260
132	108.829	127.164	168.797	107.956	127.102	186.698
133	99.358	115.668	174.135	100.810	119.330	175.209
134	127.865	152.343	234.539	124.263	148.211	223.638
135	140.324	164.098	273.681	136.182	163.168	250.758

## Anexo B: Resumen de distribución a posteriori para el parámetro de localidad

	Mean	SD		Mean	SD		Mean	SD
Loc1	110.78	5.053	Loc46	67.67	4.722	Loc91	63	2.943
Loc2	80.45	3.69	Loc47	71.94	3.626	Loc92	76.91	2.978
Loc3	60.24	3.02	Loc48	70.95	3.621	Loc93	105.64	4.271
Loc4	70.73	3.581	Loc49	69.95	3.515	Loc94	84.1	3.407
Loc5	116.77	4.512	Loc50	69	3.368	Loc95	60.46	3.048
Loc6	112.63	7.072	Loc51	71.22	3.853	Loc96	62.92	2.992
Loc7	101.27	4.381	Loc52	56.72	3.009	Loc97	78.44	3.668
Loc8	61.39	2.975	Loc53	65.45	3.74	Loc98	97.99	3.892
Loc9	65.64	3.087	Loc54	37.69	3.116	Loc99	53.86	2.952
Loc10	56.97	3.034	Loc55	41.71	2.975	Loc100	86.63	4.249
Loc11	54.18	2.952	Loc56	39.35	2.531	Loc101	59.58	3.629
Loc12	72.16	3.252	Loc57	41.34	2.835	Loc102	60.58	3.963
Loc13	67.28	3.386	Loc58	55.8	5.117	Loc103	106.52	4.704
Loc14	71.82	3.645	Loc59	45.83	3.149	Loc104	72.67	4.735
Loc15	83.12	3.377	Loc60	48.99	2.418	Loc105	77.59	5.532
Loc16	85.39	3.519	Loc61	44.26	2.302	Loc106	95.15	4.625
Loc17	94.26	3.565	Loc62	52.93	3.097	Loc107	96.62	4.062
Loc18	91.16	4.305	Loc63	42.24	2.876	Loc108	79.19	4.004
Loc19	66.96	3.998	Loc64	37.43	2.438	Loc109	78.4	3.667
Loc20	67.8	3.023	Loc65	39.04	2.501	Loc110	83.75	3.954

Loc21	55.73	3.082	Loc66	42.2	2.9	Loc111	105.75	6.622
Loc22	67.34	3.469	Loc67	49.6	3.483	Loc112	77.48	4.07
Loc23	53.07	2.854	Loc68	40.24	2.498	Loc113	83.97	4.097
Loc24	66.63	3.801	Loc69	37.09	1.933	Loc114	81.76	3.731
Loc25	80	3.935	Loc70	39.19	2.409	Loc115	65.08	3.109
Loc26	88.07	4.338	Loc71	54.05	2.602	Loc116	68.61	3.492
Loc27	89.85	4.015	Loc72	51.93	2.295	Loc117	76.94	3.843
Loc28	78.13	3.763	Loc73	71.94	3.462	Loc118	65.9	3.951
Loc29	75.77	3.567	Loc74	45.42	2.928	Loc119	73.56	4.175
Loc30	77.41	3.892	Loc75	40.91	2.112	Loc120	67.33	3.125
Loc31	70.03	3.314	Loc76	44.56	2.601	Loc121	66.36	4.002
Loc32	67.38	2.977	Loc77	48.13	2.685	Loc122	64	3.286
Loc33	62.24	3.019	Loc78	56.52	2.885	Loc123	73.46	5.496
Loc34	53.95	2.729	Loc79	70.16	2.917	Loc124	71.42	3.637
Loc35	50.64	2.82	Loc80	63.76	3.064	Loc125	69.98	5.756
Loc36	57.38	3.782	Loc81	47.45	2.305	Loc126	76.56	4.927
Loc37	65.68	3.619	Loc82	53.91	2.454	Loc127	73.81	3.995
Loc38	64.17	3.528	Loc83	75.01	4.292	Loc128	63.68	4.174
Loc39	70.59	3.599	Loc84	93.23	3.268	Loc129	62.73	3.735
Loc40	72.83	3.67	Loc85	69.38	2.92	Loc130	77.85	5.177
Loc41	80.46	3.684	Loc86	60.07	3.33	Loc131	78.76	4.271
Loc42	81.23	3.611	Loc87	52.26	2.737	Loc132	69.37	3.59
Loc43	86.47	3.982	Loc88	52.4	3.018	Loc133	62.56	4.618
Loc44	66.91	3.109	Loc89	48.64	2.743	Loc134	76.39	4.372
Loc45	62.22	3.003	Loc90	56.64	2.759	Loc135	83.38	4.98

## Anexo C: Resumen de distribución a posteriori para el parámetro de escala

	Mean	SD		Mean	SD		Mean	SD
Scale1	33.45	2.398	Scale46	29.260	2.431	Scale91	21.600	1.913
Scale2	21.89	2.016	Scale47	27.900	2.239	Scale92	22.250	1.993
Scale3	19.85	1.692	Scale48	26.710	2.336	Scale93	27.690	2.088
Scale4	23.78	2.098	Scale49	26.850	2.160	Scale94	26.020	2.030
Scale5	32.82	2.443	Scale50	26.030	2.100	Scale95	21.080	1.754
Scale6	33.93	3.209	Scale51	25.480	2.014	Scale96	22.510	1.924
Scale7	32.38	2.448	Scale52	22.200	1.969	Scale97	26.670	2.362
Scale8	21.51	1.985	Scale53	24.180	2.201	Scale98	29.920	2.190
Scale9	21.98	1.754	Scale54	18.470	2.217	Scale99	20.500	2.141
Scale10	20.44	1.812	Scale55	18.870	1.873	Scale100	30.680	2.263
Scale11	18.88	1.841	Scale56	16.610	1.791	Scale101	23.780	2.172
Scale12	24.68	1.994	Scale57	16.410	2.284	Scale102	25.790	2.446

Scale13	25.31	1.973	Scale58	24.100	2.063	Scale103	34.410	2.375
Scale14	26.24	2.079	Scale59	19.750	2.397	Scale104	30.000	2.646
Scale15	27.49	2.096	Scale60	15.680	1.800	Scale105	32.420	2.649
Scale16	26.34	2.186	Scale61	14.350	1.793	Scale106	33.150	2.497
Scale17	25.88	2.134	Scale62	22.100	1.945	Scale107	30.410	2.207
Scale18	26.37	1.943	Scale63	19.820	1.975	Scale108	27.840	2.218
Scale19	22.59	2.178	Scale64	15.510	1.692	Scale109	27.690	2.028
Scale20	21.79	1.973	Scale65	16.120	1.867	Scale110	26.690	2.178
Scale21	21.79	1.905	Scale66	19.860	1.888	Scale111	30.710	2.675
Scale22	21.56	1.812	Scale67	22.230	1.960	Scale112	31.080	2.350
Scale23	19.22	1.964	Scale68	17.060	1.848	Scale113	29.540	2.199
Scale24	27.38	2.268	Scale69	13.120	1.534	Scale114	25.790	2.269
Scale25	28.44	2.110	Scale70	16.400	1.818	Scale115	22.210	2.065
Scale26	29.08	2.216	Scale71	17.350	1.580	Scale116	24.070	2.388
Scale27	28.45	2.350	Scale72	15.940	1.562	Scale117	27.180	2.415
Scale28	29.13	2.111	Scale73	22.920	1.806	Scale118	27.920	2.469
Scale29	28.51	2.010	Scale74	17.320	1.773	Scale119	28.030	1.987
Scale30	27.67	1.946	Scale75	13.570	1.595	Scale120	24.230	1.973
Scale31	24.26	2.012	Scale76	17.870	1.644	Scale121	28.910	2.345
Scale32	20.97	2.100	Scale77	18.270	1.778	Scale122	24.940	1.933
Scale33	21.31	1.856	Scale78	19.950	1.762	Scale123	26.950	2.485
Scale34	20.45	1.772	Scale79	21.670	1.921	Scale124	29.420	2.105
Scale35	20.31	1.893	Scale80	22.080	1.960	Scale125	32.720	2.535
Scale36	22.92	1.874	Scale81	15.750	1.627	Scale126	34.060	2.473
Scale37	27.24	2.147	Scale82	17.520	1.694	Scale127	31.010	2.208
Scale38	25.98	2.212	Scale83	25.550	2.054	Scale128	26.070	2.325
Scale39	26.82	2.223	Scale84	23.750	2.018	Scale129	24.630	2.171
Scale40	28.29	2.114	Scale85	20.300	2.034	Scale130	30.830	2.448
Scale41	28.14	2.181	Scale86	21.580	1.749	Scale131	30.630	2.560
Scale42	27.75	2.092	Scale87	19.480	1.768	Scale132	25.900	2.204
Scale43	25.84	2.096	Scale88	19.750	1.992	Scale133	26.110	2.670
Scale44	23.33	2.044	Scale89	19.280	2.036	Scale134	31.960	2.360
Scale45	22.07	1.981	Scale90	20.230	1.882	Scale135	34.770	2.558

## Anexo D: Resumen de distribución a posteriori para el parámetro de forma

	Mean	SD		Mean	SD		Mean	SD
Shape1	0.0979	0.0449	Shape46	0.094	0.049	Shape91	0.153	0.041
Shape2	0.0840	0.0432	Shape47	0.119	0.046	Shape92	0.119	0.043
Shape3	0.0754	0.0446	Shape48	0.138	0.047	Shape93	0.095	0.044
Shape4	0.0336	0.0464	Shape49	0.147	0.048	Shape94	0.096	0.043

Shape5	0.0976	0.0976	Shape50	0.153	0.047	Shape95	0.076	0.044
Shape6	0.1340	0.0435	Shape51	0.160	0.046	Shape96	0.077	0.046
Shape7	0.1200	0.0429	Shape52	0.171	0.046	Shape97	0.096	0.046
Shape8	0.0994	0.0421	Shape53	0.180	0.054	Shape98	0.103	0.043
Shape9	0.0866	0.0432	Shape54	0.134	0.072	Shape99	0.189	0.053
Shape10	0.0742	0.0472	Shape55	0.166	0.055	Shape100	0.092	0.045
Shape11	0.0749	0.0518	Shape56	0.232	0.048	Shape101	0.074	0.048
Shape12	0.0237	0.0414	Shape57	0.255	0.053	Shape102	0.087	0.048
Shape13	0.0281	0.0441	Shape58	0.209	0.050	Shape103	0.095	0.044
Shape14	0.0355	0.0448	Shape59	0.173	0.050	Shape104	0.115	0.047
Shape15	0.0775	0.0429	Shape60	0.081	0.049	Shape105	0.123	0.047
Shape16	0.0930	0.0427	Shape61	0.088	0.054	Shape106	0.088	0.043
Shape17	0.1027	0.0428	Shape62	0.161	0.046	Shape107	0.064	0.045
Shape18	0.1135	0.0422	Shape63	0.203	0.046	Shape108	0.057	0.045
Shape19	0.1341	0.0430	Shape64	0.246	0.047	Shape109	0.047	0.045
Shape20	0.1219	0.0432	Shape65	0.235	0.047	Shape110	0.042	0.047
Shape21	0.1077	0.0449	Shape66	0.170	0.052	Shape111	0.097	0.043
Shape22	0.1021	0.0472	Shape67	0.179	0.051	Shape112	0.106	0.046
Shape23	0.1062	0.0512	Shape68	0.223	0.048	Shape113	0.093	0.046
Shape24	0.0451	0.0440	Shape69	0.224	0.044	Shape114	0.076	0.044
Shape25	0.0665	0.0412	Shape70	0.179	0.045	Shape115	0.036	0.043
Shape26	0.0846	0.0428	Shape71	0.077	0.045	Shape116	0.066	0.045
Shape27	0.0959	0.0418	Shape72	0.075	0.044	Shape117	0.088	0.048
Shape28	0.0964	0.0414	Shape73	0.080	0.043	Shape118	0.072	0.049
Shape29	0.0965	0.0401	Shape74	0.153	0.045	Shape119	0.049	0.043
Shape30	0.0982	0.0401	Shape75	0.200	0.045	Shape120	0.033	0.040
Shape31	0.1047	0.0414	Shape76	0.188	0.045	Shape121	0.046	0.050
Shape32	0.1216	0.0420	Shape77	0.147	0.044	Shape122	0.012	0.042
Shape33	0.1308	0.0404	Shape78	0.129	0.042	Shape123	0.031	0.040
Shape34	0.1300	0.0402	Shape79	0.102	0.043	Shape124	0.033	0.042
Shape35	0.1395	0.0433	Shape80	0.085	0.043	Shape125	0.062	0.047
Shape36	0.1490	0.0464	Shape81	0.074	0.044	Shape126	0.031	0.050
Shape37	0.0557	0.0465	Shape82	0.075	0.044	Shape127	0.011	0.046
Shape38	0.0737	0.0455	Shape83	0.096	0.043	Shape128	-0.012	0.049
Shape39	0.0970	0.0436	Shape84	0.098	0.043	Shape129	-0.006	0.047
Shape40	0.1121	0.0426	Shape85	0.124	0.043	Shape130	0.037	0.048
Shape41	0.1221	0.0429	Shape86	0.148	0.042	Shape131	0.025	0.052
Shape42	0.1242	0.0416	Shape87	0.185	0.045	Shape132	-0.010	0.047
Shape43	0.1210	0.0404	Shape88	0.205	0.050	Shape133	-0.033	0.059
Shape44	0.1384	0.0409	Shape89	0.197	0.052	Shape134	-0.003	0.051
Shape45	0.1499	0.0413	Shape90	0.180	0.045	Shape135	0.016	0.053

## Anexo E: HPD para el parámetro localidad

	Lower	Upper		Lower	Upper		Lower	Upper
Loc 1	100.588	120.317	Loc 46	59.167	77.680	Loc 91	57.098	68.561
Loc 2	73.195	87.626	Loc 47	64.719	78.968	Loc 92	71.048	82.734
Loc 3	54.488	66.223	Loc 48	63.966	78.057	Loc 93	97.733	114.325
Loc 4	63.905	77.876	Loc 49	62.823	76.767	Loc 94	77.312	90.712
Loc 5	107.789	125.665	Loc 50	62.440	75.756	Loc 95	54.343	66.386
Loc 6	98.545	126.462	Loc 51	63.720	78.719	Loc 96	56.980	68.776
Loc 7	92.556	109.896	Loc 52	50.856	62.760	Loc 97	70.676	84.967
Loc 8	55.440	67.209	Loc 53	57.936	72.567	Loc 98	90.271	105.541
Loc 9	59.541	71.603	Loc 54	31.829	43.948	Loc 99	47.918	59.586
Loc 10	50.836	62.623	Loc 55	35.833	47.512	Loc 100	78.413	94.976
Loc 11	48.558	60.108	Loc 56	34.259	44.268	Loc 101	52.751	66.991
Loc 12	65.656	78.456	Loc 57	35.759	46.945	Loc 102	52.569	68.072
Loc 13	60.255	73.750	Loc 58	45.995	66.080	Loc 103	97.476	115.840
Loc 14	64.689	78.950	Loc 59	39.695	52.006	Loc 104	63.467	81.864
Loc 15	76.644	89.751	Loc 60	44.164	53.699	Loc 105	66.983	88.933
Loc 16	78.170	92.058	Loc 61	39.791	48.854	Loc 106	86.610	104.814
Loc 17	87.275	101.100	Loc 62	46.838	59.016	Loc 107	88.581	104.383
Loc 18	82.857	99.770	Loc 63	36.590	47.832	Loc 108	71.636	87.135
Loc 19	58.814	74.608	Loc 64	32.778	42.306	Loc 109	71.505	85.865
Loc 20	61.679	73.555	Loc 65	34.068	43.870	Loc 110	76.236	91.630
Loc 21	49.814	61.953	Loc 66	36.857	48.000	Loc 111	91.928	117.877
Loc 22	60.751	74.264	Loc 67	42.637	56.317	Loc 112	69.655	85.769
Loc 23	47.620	58.725	Loc 68	35.351	45.155	Loc 113	76.107	92.257
Loc 24	58.699	73.647	Loc 69	33.476	41.066	Loc 114	74.457	89.148
Loc 25	72.486	88.009	Loc 70	34.698	44.172	Loc 115	58.866	71.042
Loc 26	79.089	96.174	Loc 71	48.841	59.106	Loc 116	61.526	75.084
Loc 27	81.893	97.575	Loc 72	47.254	56.323	Loc 117	69.514	84.411
Loc 28	70.735	85.612	Loc 73	64.726	78.528	Loc 118	58.160	73.553
Loc 29	68.888	83.004	Loc 74	39.890	51.323	Loc 119	65.438	81.941
Loc 30	69.834	85.282	Loc 75	36.761	45.126	Loc 120	61.209	73.580
Loc 31	63.825	76.909	Loc 76	39.636	49.853	Loc 121	58.469	74.181
Loc 32	61.304	73.016	Loc 77	42.913	53.392	Loc 122	57.380	70.452
Loc 33	56.252	68.146	Loc 78	51.165	62.399	Loc 123	62.178	83.635
Loc 34	48.801	59.543	Loc 79	64.269	75.587	Loc 124	64.172	78.405
Loc 35	45.184	56.387	Loc 80	57.843	69.834	Loc 125	58.213	81.246
Loc 36	50.270	64.890	Loc 81	42.970	51.993	Loc 126	67.163	86.362
Loc 37	58.420	72.605	Loc 82	49.087	58.744	Loc 127	66.196	81.964
Loc 38	57.231	71.150	Loc 83	66.688	83.473	Loc 128	55.479	71.749
Loc 39	63.519	77.560	Loc 84	86.697	99.517	Loc 129	55.153	69.922
Loc 40	65.373	79.830	Loc 85	63.470	74.887	Loc 130	67.164	87.499
Loc 41	73.158	87.505	Loc 86	53.711	66.797	Loc 131	70.053	86.940
Loc 42	74.246	88.556	Loc 87	46.678	57.437	Loc 132	62.540	76.660
Loc 43	78.692	94.252	Loc 88	46.299	58.141	Loc 133	53.933	71.744
Loc 44	60.597	72.846	Loc 89	43.216	54.036	Loc 134	67.780	85.031

Loc 45	56.536	68.153	Loc 90	51.118	61.979	Loc 135	73.427	93.146
--------	--------	--------	--------	--------	--------	---------	--------	--------

## Anexo F: HPD para el parámetro escala

	Lower	Upper		Lower	Upper		Lower	Upper
Scale1	28.71	38.124	Scale46	24.79	34.271	Scale91	17.87	25.306
Scale2	18.12	25.971	Scale47	23.49	32.171	Scale92	18.44	26.244
Scale3	16.76	23.283	Scale48	22.00	31.205	Scale93	23.76	31.887
Scale4	19.83	28.020	Scale49	22.56	31.053	Scale94	22.20	30.203
Scale5	28.33	37.977	Scale50	21.97	30.121	Scale95	17.73	24.573
Scale6	27.51	40.124	Scale51	21.68	29.569	Scale96	18.87	26.355
Scale7	27.76	37.339	Scale52	18.36	26.078	Scale97	21.96	31.331
Scale8	17.55	25.267	Scale53	20.02	28.654	Scale98	25.72	34.283
Scale9	18.41	25.314	Scale54	14.26	22.774	Scale99	16.34	24.616
Scale10	16.94	23.915	Scale55	15.30	22.527	Scale100	26.39	35.191
Scale11	15.45	22.580	Scale56	13.12	20.182	Scale101	19.47	27.950
Scale12	20.88	28.755	Scale57	12.12	21.045	Scale102	20.97	30.609
Scale13	21.41	29.076	Scale58	20.00	28.087	Scale103	29.66	39.067
Scale14	22.05	30.174	Scale59	15.11	24.362	Scale104	24.89	35.245
Scale15	23.46	31.511	Scale60	12.38	19.423	Scale105	27.22	37.577
Scale16	22.04	30.626	Scale61	11.03	17.997	Scale106	28.32	38.081
Scale17	21.89	30.320	Scale62	18.32	25.937	Scale107	26.25	34.888
Scale18	22.63	30.230	Scale63	15.97	23.711	Scale108	23.52	32.124
Scale19	18.33	26.838	Scale64	12.19	18.747	Scale109	23.79	31.615
Scale20	18.00	25.698	Scale65	12.59	19.820	Scale110	22.59	31.029
Scale21	18.18	25.558	Scale66	16.13	23.487	Scale111	25.66	36.079
Scale22	18.24	25.254	Scale67	18.59	26.236	Scale112	26.54	35.667
Scale23	15.39	23.015	Scale68	13.56	20.797	Scale113	25.38	33.898
Scale24	23.00	31.739	Scale69	10.16	16.082	Scale114	21.15	29.989
Scale25	24.12	32.409	Scale70	12.99	20.094	Scale115	18.21	26.245
Scale26	24.66	33.365	Scale71	14.25	20.396	Scale116	19.53	28.846
Scale27	23.80	32.980	Scale72	12.89	19.032	Scale117	22.57	31.979
Scale28	25.10	33.282	Scale73	19.50	26.557	Scale118	23.19	32.876
Scale29	24.72	32.599	Scale74	13.91	20.773	Scale119	24.22	31.931
Scale30	23.78	31.405	Scale75	10.43	16.661	Scale120	20.27	28.036
Scale31	20.57	28.454	Scale76	14.85	21.225	Scale121	24.68	33.866
Scale32	16.86	25.153	Scale77	15.04	21.991	Scale122	21.19	28.793
Scale33	17.74	24.963	Scale78	16.50	23.303	Scale123	22.23	31.970
Scale34	17.10	24.068	Scale79	18.00	25.535	Scale124	25.21	33.452
Scale35	16.67	24.027	Scale80	18.40	26.092	Scale125	27.89	37.694

## Anexos

---

Scale36	19.32	26.644	Scale81	12.58	18.937	Scale126	29.58	39.155
Scale37	23.10	31.513	Scale82	14.46	21.039	Scale127	26.80	35.466
Scale38	21.58	30.217	Scale83	21.60	29.613	Scale128	21.49	30.519
Scale39	22.68	31.376	Scale84	19.78	27.690	Scale129	20.47	28.845
Scale40	24.21	32.521	Scale85	16.50	24.414	Scale130	26.35	36.040
Scale41	23.70	32.312	Scale86	18.22	25.039	Scale131	25.69	35.700
Scale42	23.64	31.858	Scale87	16.12	23.023	Scale132	21.84	30.474
Scale43	21.80	29.910	Scale88	15.86	23.608	Scale133	20.85	31.207
Scale44	19.39	27.312	Scale89	15.27	23.247	Scale134	27.48	36.701
Scale45	18.13	25.901	Scale90	16.79	24.196	Scale135	29.94	39.839