

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS  
DEPARTAMENTO DE MATEMÁTICAS**

**PROYECTO DE TITULACIÓN**

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN CALIDAD Y  
PRODUCTIVIDAD”**

**TEMA:**

ESTUDIO ESTADÍSTICO DE LAS MAGNITUDES SÍSMICAS DE LAS  
PROVINCIAS DE GUAYAS Y SANTA ELENA USANDO LA TEORÍA  
DE VALORES EXTREMOS

**AUTOR:**

DANNY GABRIEL MENDIETA PARRA

Guayaquil - Ecuador

2021

## RESUMEN

En esta investigación se presenta el modelaje estadístico de la peligrosidad sísmica de las provincias de Guayas y Santa Elena, aplicando la teoría de valores extremos, específicamente la distribución Generalizada de Valores Extremos (GEVD) y la Distribución Generalizada de Pareto (GPD), mediante el uso de los métodos de máxima verosimilitud (MLE) y L-Momentos en la estimación de los parámetros respectivos. Se realizó una caracterización del catálogo sísmico mediante Estadística Descriptiva, determinándose que la distribución de los datos tiene sesgo hacia la derecha, por lo que la distribución normal no es una buena candidata para su modelaje. Se realizó la estimación de los parámetros para los modelos GEVD y GPD por ambos métodos, máxima verosimilitud y L-Momentos, determinándose que el modelo de mejor ajuste fue el GEVD considerando que cada bloque máximo corresponde a cada año de observaciones resultando en un número de bloques de 35. Se demostró que el modelo de Gumbel podría también usarse para modelar los datos, sin embargo, limitado a períodos de retorno bajos, hasta 50 años, luego de ese tiempo, existe una distorsión significativa respecto a los otros modelos. La regresión de Poisson al igual que la regresión lineal simple se usaron para modelar la relación de Gutenberg-Richter entre la frecuencia de los sismos y sus magnitudes, demostrando que ambos métodos son apropiados para este fin. Los niveles de retorno se calcularon para diferentes periodos de tiempo, siendo los más confiables los del modelo GEVD con número de bloques de 35 y el método de L-Momentos el que mostró menor error estándar, generando intervalos de 95% de confianza más angostos que con el método de máxima verosimilitud .

## **ABSTRACT**

In this research, a statistical model of the seismic danger of the provinces of Guayas and Santa Elena is presented, applying the theory of extreme values, specifically, using the Generalized Extreme Value Distribution and Generalized Pareto Distribution, by applying the methods of maximum likelihood (MLE) and L-Moments, for the estimation of the respective parameters. Characterization of the seismic catalog was carried out using Descriptive Statistics, determining that the data distribution is skewed to the right so that the normal distribution is not a good candidate to model it. The estimation of the parameters for the GEVD and GPD models was performed by both methods, MLE and L-Moments. It was determined that the best-fitted model was the GEVD considering that each block maxima corresponds to each year of observations with a number of blocks of 35. It was shown that the Gumbel model could also be used to model the data, however, limited to low return periods, up to 50 years, where after this period, there exists a significant distortion compared to the other models. Poisson regression, as well as simple linear regression, were used to model the Gutenberg-Richter relationship between the earthquake frequency and its magnitudes, showing both methods are appropriate for this purpose. Return levels were calculated for different return periods, being the most reliable those from the GEVD model with block number of 35 and the L-Moments method being the one that showed the lowest standard error, providing narrower 95% confidence intervals than the ones calculated with the MLE method.

## **DEDICATORIA**

A mi pequeña Sarita, mi angelito en el cielo.

# AGRADECIMIENTO

Al creador de todo lo existente.

A mis padres, los únicos seres incondicionales a lo largo de toda mi vida.

A mi directora de Proyecto de Titulación, PhD. Sandra García y a los vocales PhD. Christian Galarza y Mgtr. Mario Solórzano por su invaluable contribución.

## DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.



---

Danny Gabriel Mendieta Parra

# TRIBUNAL DE GRADUACIÓN



---

Mgr. Francisco Moreira Villegas  
PRESIDENTE



---

Sandra García Bustos, Ph. D.  
DIRECTOR



---

Mgr. Mario Solórzano Carvajal  
VOCAL 1



---

Christian Galarza Morales, Ph. D.  
VOCAL 2

## ABREVIATURAS O SIGLAS

<b>GEVD</b>	Distribución Generalizada de Valores Extremos
<b>GPD</b>	Distribución Generalizada de Pareto
<b>INEC</b>	Instituto Ecuatoriano de Estadísticas y Censos
<b>M<sub>w</sub></b>	Escala Sismológica de Magnitud de Momento
<b>MLE</b>	Estimación por Máxima Verosimilitud
<b>AIC</b>	Criterio de Información de Akaike
<b>BIC</b>	Criterio de Información Bayesiano

# TABLA DE CONTENIDO

	Pág.
RESUMEN.....	II
ABSTRACT.....	III
DEDICATORIA.....	IV
AGRADECIMIENTO.....	V
DECLARACIÓN EXPRESA.....	VI
ABREVIATURAS O SIGLAS.....	VIII
TABLA DE CONTENIDO.....	IX
LISTADO DE FIGURAS.....	XI
LISTADO DE TABLAS.....	XIII
<b>CAPÍTULO 1</b>	
1. INTRODUCCIÓN.....	1
1.1. Antecedentes.....	1
1.2. Descripción del problema.....	3
1.3. Objetivos.....	4
Objetivo General.....	4
Objetivos específicos.....	4
1.4. Hipótesis.....	4
1.5. Alcance.....	5
<b>CAPÍTULO 2</b>	
2. MARCO TEÓRICO.....	6
2.1. Distribución Generalizada de Valores Extremos GEVD.....	8
2.1.1. Estimación de Parámetros para GEVD.....	9
2.1.2. Estimación de parámetros para GEVD por método de Máxima Verosimilitud MLE.....	9
2.1.3. Estimación de parámetros para GEVD por el método de L-Momentos.....	10
2.2. Distribución Generalizada de Pareto GPD.....	12
2.2.1. Estimación de parámetros para GPD.....	14
2.2.2. Estimación de parámetros para GPD por el método	

de Máxima Verosimilitud MLE. ....	14
2.2.3. Estimación de parámetros para GPD por el método de L-Momentos. ....	15
2.3. Criterios de selección de modelos.....	15
2.4. Test de razón de verosimilitud.....	16
<b>CAPÍTULO 3</b>	
3. METODOLOGÍA.....	17
3.1. Análisis descriptivo de los datos.....	17
3.2. Estimación de parámetros de la Distribución Generalizada de Valores Extremos (GEVD).....	22
3.2.1. Estimación de parámetros de GEVD usando el método de Máxima Verosimilitud (MLE). ....	22
3.2.2. Estimación de parámetros de GEVD usando el método de L-Momentos. ....	25
3.3. Estimación de parámetros de la Distribución Generalizada de Pareto (GPD). ....	27
3.3.1. Estimación de parámetros de GPD usando el método de Máxima Verosimilitud (MLE) ....	29
3.3.2. Estimación de parámetros de GPD usando el método de L-Momentos. ....	32
3.4. Modelo Gutenberg Richter: Regresión lineal y Regresión Poisson.....	34
<b>CAPÍTULO 4</b>	
4. RESULTADOS.....	37
4.1. Niveles de retorno: GEVD.....	38
4.2. Ajuste del modelo: GEVD.....	44
4.3. Niveles de retorno GPD.....	48
4.4. Ajuste del modelo: GPD.....	52
<b>CAPÍTULO 5</b>	
5. CONCLUSIONES Y RECOMENDACIONES.....	55
6. Referencias .....	59
7. Apéndices y anexos.....	61

## LISTADO DE FIGURAS

	Pág.
<b>Figura 3.1.</b> Gráfica de magnitud vs tiempo de ocurrencia de sismos en Ecuador.....	18
<b>Figura 3.2.</b> Histograma de densidad de sismos en Ecuador.....	18
<b>Figura 3.3.</b> Gráfica de magnitud vs tiempo de ocurrencia de sismos en Guayas y Santa Elena.....	19
<b>Figura 3.4.</b> Histograma de densidad de sismos en Guayas y Santa Elena.....	20
<b>Figura 3.5.</b> Diagrama de magnitud vs el tiempo de ocurrencia de sismos en Guayas y Santa Elena: medias y error estándar .....	21
<b>Figura 3.6.</b> Diagrama de Autocorrelaciones Parciales para sismos en Guayas y Santa Elena .....	21
<b>Figura 3.7.</b> Estimación del Parámetro de Forma $\xi$ con diferentes tamaños de Bloques de GEVD usando MLE.....	23
<b>Figura 3.8.</b> Estimación del Parámetro de Localización $\mu$ con diferentes tamaños de Bloques de GEVD usando MLE.....	24
<b>Figura 3.9.</b> Estimación del Parámetro de Escala $\sigma$ con diferentes tamaños de Bloques de GEVD usando MLE.....	25
<b>Figura 3.10.</b> Estimación del Parámetro de Forma $\xi$ con diferentes tamaños de Bloques de GEVD usando L-Momentos.....	26
<b>Figura 3.11.</b> Estimación del Parámetro de Localización $\mu$ con diferentes tamaños de Bloques de GEVD usando L-Momentos.....	26
<b>Figura 3.12.</b> Estimación del Parámetro de Escala $\sigma$ con diferentes tamaños de Bloques de GEVD usando L-Momentos.....	27
<b>Figura 3.13.</b> Gráfico de Excedencia de medias, selección de umbral $u$ .....	28
<b>Figura 3.14.</b> Gráfico de Excedencia de medias y parámetro de Forma $\xi$ , selección de umbral.....	29
<b>Figura 3.15.</b> Estimación del Parámetro de Forma con diferentes tamaños de Bloques de GPD usando MLE.....	30
<b>Figura 3.16.</b> Estimación del Parámetro de Escala con diferentes	

tamaños de Bloques de GPD usando MLE.....	30
<b>Figura 3.17.</b> Parámetro de Escala Re-parametrizado. GPD por MLE.....	32
<b>Figura 3.18.</b> Estimación del Parámetro de Forma con diferentes tamaños de Bloques de GPD usando L-momentos.....	33
<b>Figura 3.19.</b> Estimación del Parámetro de Escala con diferentes tamaños de Bloques de GPD usando L-momentos.....	33
<b>Figura 3.20.</b> Parámetro de Escala Re-parametrizado. GPD usando L-momentos.....	34
<b>Figura 3.21.</b> Modelo Gutenberg-Richter con las regresiones.....	36
<b>Figura 4.1.</b> Niveles de retorno con tamaño de bloques. GEVD por MLE y L-Momentos.....	39
<b>Figura 4.2.</b> Niveles de retorno. GEVD por MLE y L-momentos, $nb = 30$ .....	41
<b>Figura 4.3.</b> Niveles de retorno. GEVD por MLE y L-momentos, bloques por año de observación $nb = 35$ .....	43
<b>Figura 4.4.</b> Niveles de retorno. Gumbel por MLE, bloques por año de observación $nb = 35$ .....	44
<b>Figura 4.5.</b> Gráficos de probabilidad y de cuantiles GEVD por MLE, $nb = 30$ .....	46
<b>Figura 4.6.</b> Gráficos de probabilidad y de cuantiles. GEVD por L-Momentos, $nb = 30$ .....	46
<b>Figura 4.7.</b> Gráficos de probabilidad y de cuantiles. GEVD por MLE, $nb = 35$ .....	47
<b>Figura 4.8.</b> Gráficos de probabilidad y de cuantiles. GEVD por L-Momentos, $nb = 35$ .....	47
<b>Figura 4.9.</b> Niveles de retorno con diferentes umbrales. GPD por MLE y L-momentos.....	51
<b>Figura 4.10.</b> Niveles de retorno. GPD por MLE y L-momentos.....	51
<b>Figura 4.11.</b> Gráficos de probabilidad y de cuantiles. GPD por MLE.....	53
<b>Figura 4.12.</b> Gráficos de probabilidad y de cuantiles. GPD por L-Momentos.....	54

## LISTADO DE TABLAS

	Pág.
<b>Tabla 3.1.</b> Estimaciones del número de sismos con los modelos Lineal y Poisson de los sismos de Guayas y Santa Elena.....	35
<b>Tabla 4.1.</b> Resumen de estimación de parámetros.....	38
<b>Tabla 4.2.</b> Resumen de Niveles de Retorno Sismos Guayas y Santa Elena: GEVD y GPD.....	40

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1. Antecedentes

Los sismos y terremotos se producen por la liberación de energía de la parte interna de la Litósfera, provocando que las placas tectónicas de la tierra se muevan con diferentes intensidades dependiendo de la cantidad de energía liberada, lo que ocasiona movimientos fuertes en el suelo debido a la propagación de las ondas sísmicas que viajan por toda la superficie adyacente al epicentro.

La severidad del sismo va a depender de su magnitud: a mayor su magnitud, mayores los daños causan a la población. Los sismos de mayores magnitudes, a partir de 7 grados de la escala de Richter, se denominan terremotos. En la historia de la humanidad se han registrado grandes catástrofes asociadas a terremotos y sismos, como la ocurrida en Haití en 2010 con más de 300 mil víctimas mortales, cuya magnitud fue de 7.0  $M_w$  en la escala de Richter y es considerado el segundo más destructor, debajo del ocurrido en Shaanxi (China) en 1556 con más de 800 mil víctimas (Rivas, 2014). En nuestro país, recientemente en 2016, vivimos un terremoto de magnitud de 7.8  $M_w$  en la escala de Richter con un número de víctimas que no superó el millar y que, además dejó una destrucción de infraestructura que hasta ahora (5 años después) no se puede recuperar.

Dentro de los conceptos más importantes relacionados con sismología está el de peligrosidad sísmica que se define como la probabilidad de ocurrencia de un sismo con cierto valor de intensidad, durante un periodo de tiempo determinado y en un área específica (Benito, 1999). Los métodos para determinar la peligrosidad sísmica se clasifican en los siguientes grupos (Muñoz, 1989):

- **Deterministas:** Se refiere a los métodos que asumen que un evento sísmico en el futuro se producirá de manera similar a los del pasado; esta hipótesis está asociada al concepto de estacionariedad. Estas consideraciones ponen como límite superior a un evento sísmico posible en el futuro a los valores máximos obtenidos eventos del pasado.
- **Probabilistas:** Son métodos basados en las leyes estadísticas que definen los fenómenos sísmicos con los cuales se puede establecer el comportamiento de estos eventos dado el historial sísmico de una localidad. Con estos métodos se puede determinar la probabilidad de ocurrencia de movimientos telúricos de cierta intensidad en un periodo de tiempo determinado.

Para el área de la ciencia Sismológica, un ejemplo de un método o modelo probabilístico es el modelo de Valores Extremos. Específicamente, la teoría de Valores Extremos se usa en el análisis estadístico de eventos inusuales con varios campos de aplicación, por ejemplo, en la Hidrología, Economía, Estudios Actuariales y, por supuesto, Sismología.

Con relación a los trabajos asociados a este campo en Ecuador, está por ejemplo la investigación realizada por García et. al. (2018), donde para caracterizar las 5 zonas sísmicas del país, se usó la Teoría de Valores Extremos y también un análisis por regresión de Poisson, determinando que las zonas 1 y 5 fueron las más representativas en términos de recurrencia y magnitud sísmica, esto haciendo uso del método de máxima verosimilitud para estimar los parámetros (García S. N., 2021).

Un estudio más cercano al objetivo de esta investigación fue el realizado por (García S. N., 2018) donde se hace un análisis estadístico de la sismología de la Costa ecuatoriana con el uso de la

Teoría de Valores Extremos, específicamente aplicando la Distribución Generalizada de Valores Extremos y también con el modelo de Distribución Generalizada de Pareto, ambos empleando el método de máxima verosimilitud en la estimación de los parámetros, donde se determinó que la zona con mayor probabilidad de tener sismos de magnitud de al menos 7.7 Mw en los próximos 50 años se encuentra entre el norte de Manabí y el sur de Esmeraldas.

En las provincias de Guayas y Santa Elena no se han registrado sismos de magnitud tan altas como los presentados en el pasado en Manabí y Esmeraldas, sin embargo, es importante analizar la peligrosidad sísmica en las provincias en mención, utilizando el modelo de valores extremos para determinar qué tan frecuente puede ocurrir un sismo de gran magnitud aplicando un modelo probabilístico, justificado por la cantidad de habitantes que tiene este sector del país. De acuerdo con las proyecciones del INEC, para el 2020 se esperaba que las dos provincias concentren el 27.35% (4.8 Mill.) de la población del Ecuador (INEC, 2021).

## **1.2. Descripción del problema**

La ubicación geográfica del Ecuador lo califica como un país propenso a eventos sísmicos debido a que se encuentra entre dos placas tectónicas: Nazca y Sudamericana, además de las fallas que recorren internamente todo el país. Esto ha generado que en Ecuador se produzcan un promedio de 9 sismos mayores a 5 grados de magnitud en la escala de Richter al año desde 1925 aproximadamente, lo cual representa un verdadero peligro en la vida de muchos ecuatorianos.

Sismos corticales o superficiales (ocurren la placa continental a profundidades inferiores a los 60 km) de magnitudes moderadas a fuertes han sido documentados para la costa sur del Ecuador, especialmente para la ciudad de Guayaquil. Las áreas costeras del Golfo de Guayaquil pueden ser afectados por tsunamis o sismos de

origen local, generados desde la zona de subducción y también por deslizamientos submarinos donde las condiciones tectónicas y batimétricas conformarían potenciales daños físicos para las zonas turísticas de Santa Elena y áreas densamente pobladas y sectores industriales en desarrollo.

### **1.3. Objetivos.**

#### ***Objetivo General.***

Modelar estadísticamente la peligrosidad sísmica de las provincias de Guayas y Santa Elena mediante el método de Valores Extremos para poder estimar el promedio de ocurrencia de un sismo de gran magnitud.

#### ***Objetivos específicos.***

- Analizar el historial de sismos en las provincias Guayas y Santa Elena mediante Estadística Descriptiva para conocer su comportamiento.
- Modelar las máximas magnitudes de los sismos con la Teoría de Valores Extremos de manera conjunta en las provincias objeto de estudio, usando los métodos de máxima verosimilitud y L-Momentos.
- Aplicar la ley de Gutenberg-Richter para cuantificar la relación frecuencia-magnitud de la actividad sísmica de las provincias Guayas y Santa Elena, mediante Regresión Lineal (usando el método de mínimos cuadrados) y también con el uso de la regresión de Poisson.

### **1.4. Hipótesis.**

La aplicación de la Teoría de Valores Extremos permitirá la obtención de un modelo probabilístico para determinar la peligrosidad sísmica de las Provincias del Guayas y Santa Elena. Las dos distribuciones de valores extremos mayormente utilizadas, la Distribución Generalizada de Valores Extremos y la Distribución Generalizada de Pareto, son las más adecuadas para modelar la sismicidad de las provincias en

mención y por lo tanto proporcionarán los modelos adecuados para evaluar la peligrosidad sísmica.

#### **1.5. Alcance.**

Este proyecto consiste en un estudio y análisis de los sismos ocurridos en las provincias de Guayas y Santa Elena, desde el año 1906 hasta el 2017. Este trabajo no pretende establecer un modelo para predecir la fecha de ocurrencia de un sismo, pues eso no es posible, sino tener una idea del tiempo promedio de recurrencia de un sismo de gran magnitud.

El estudio tuvo una duración de seis meses y a continuación se presenta la metodología, desarrollo y resultados de la investigación.

# CAPÍTULO 2

## 2. MARCO TEÓRICO

El estudio estadístico de eventos extremos es muy necesario en varias aplicaciones prácticas, siendo gran parte para el estudio de fenómenos naturales como lo son los terremotos. Los sismos de gran magnitud, comúnmente se clasifican como atípicos, generalmente ocurren con una muy baja frecuencia en un determinado evento y siempre se ubican en los extremos de la distribución de probabilidad correspondiente. Comúnmente los eventos ubicados en el lado derecho de la cola de la distribución son de gran interés en los estudios, en el caso pertinente corresponden a los sismos de mayor magnitud. Uno de los métodos más usados para el modelaje de las colas de una distribución es la Teoría de Valores Extremos EVT por sus siglas en inglés, desarrollada en la década de 1950 pero con gran desarrollo reciente debido a la introducción de potentes computadoras. Estas han permitido el cálculo y el uso de métodos numéricos avanzados gracias al avance tecnológico. Las dos principales distribuciones límites de la EVT son: Distribución Generalizada de Valores Extremos (GEVD) y la Distribución Generalizada de Pareto (GPD).

El modelo en su forma general se enfoca en el comportamiento estadístico del estadístico de orden  $n$ :

$$M_n = \max\{X_1, X_2, \dots, X_n\},$$

es decir, el máximo, donde  $X_1, X_2, \dots, X_n$  es una secuencia de variables aleatorias independientes teniendo una función de distribución en común  $F$  (Coles, 2001). Aquí  $X_i, 1 \leq i \leq n$ , normalmente representa un proceso de interés con mediciones de magnitudes temporales en días, meses, años, etc. El valor de  $M_n$  es el máximo valor de la secuencia sujeta al intervalo de tiempo escogido. Aceptando que la función de distribución  $F$  es desconocida, se busca encontrar una función de distribución  $F^n$  que se

ajusta a los valores máximos. El procedimiento sigue la analogía usada en la teoría de límite central pero ahora para valores extremos. Este proceso tiene una dificultad debido a que se encuentra un límite degenerado en el comportamiento asintótico de la distribución de densidad. Esta dificultad se evita realizando una renormalización lineal de la variable  $M_n$ :

$$M_n^* = \frac{M_n - b_n}{a_n},$$

para secuencias de constantes reales  $\{a_n > 0\}$  y  $\{b_n\}$ . La elección apropiada de  $\{a_n\}$  y  $\{b_n\}$  estabiliza la localización y escala de  $M_n^*$  a medida que  $n$  crece, evitando así las dificultades que trae el uso de la variable  $M_n$ . Luego, se tiene que buscar las distribuciones límites para  $M_n^*$  con la selección apropiada de  $\{a_n\}$  y  $\{b_n\}$  (Coles, 2001).

El rango entero de distribuciones límites posibles para  $M_n^*$  están dadas por el Teorema de Tipo Extremos:

**Teorema 1:** *si existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tal que*

$$Pr\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \text{ tal que } n \rightarrow \infty$$

Donde  $G$  es una función de distribución no degenerada y por lo tanto  $G$  pertenece a una de las siguientes familias:

$$I(\text{Gumbel}): G(z) = \exp\left\{-\exp\left[-\left(\frac{M_n - b_n}{a}\right)\right]\right\}, \quad -\infty < z < \infty$$

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

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

Para  $a > 0$ ,  $b$  reales y en el caso de las familias II y III,  $\alpha > 0$ .

## 2.1. Distribución Generalizada de Valores Extremos (GEVD)

El determinar cuál de las tres familias de distribuciones anteriores tiene el mejor ajuste a un conjunto de datos en particular puede ser un proceso bastante complicado, sin embargo, las tres pueden combinarse para dar como resultado la familia de distribución de **Valores Extremos Generalizados**. Esta distribución se muestra en el teorema 3.1.1 en Coles (Coles, 2001) y se describe a continuación:

**Teorema 2:** *si existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tal que*

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

*cuando  $n \rightarrow \infty$ , siendo  $G$  una función de distribución no degenerada miembro de la familia GEVD*

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

$$-\infty < \mu < \infty, \sigma > 0 \text{ y } -\infty < \xi < \infty$$

definida en el conjunto  $\{z: 1 + \xi(z - \mu)/\sigma > 0\}$ , donde los parámetros satisfacen  $-\infty < \mu < \infty, \sigma > 0$  y  $-\infty < \xi < \infty$ . Los tres parámetros del modelo son: el parámetro de Localización  $\mu$ , el parámetro de Escala  $\sigma$  y el parámetro de Forma  $\xi$ . El parámetro de forma también se lo conoce como índice de cola (Coles, 2001). Esta distribución que combina las tres familias en el Teorema 1, hace que el ajuste de los datos sea más fácil permitiendo que los datos sean ajustados aplicando la GEVD directamente. La ecuación de la GEVD viene determinada por el parámetro de forma  $\xi$ , cuyo valor permitirá determinar la distribución límite. Si  $\xi > 0$ , se deberá usar la distribución de Fréchet; cuando  $\xi = 0$ , se escoge la distribución de Gumbel y si  $\xi < 0$  se toma en consideración la distribución de Weibull (García S. N., 2018). Se debe notar que, si el parámetro de forma  $\xi < 0$  la distribución está acotada y por lo tanto hay un valor máximo calculable de  $a$ . Esto no sucede para los demás

valores del parámetro de forma, resultando en distribuciones no acotadas (Aras, 2019).

La GEVD hace uso del método llamado *bloque máximo*. En el método de bloque máximo, los datos son distribuidos en bloques de igual tamaño, donde cada uno de estos bloques entrega un valor máximo que son usados en las predicciones de la distribución de los valores extremos. Escoger el tamaño de cada bloque es crítico, ya que bloques muy pequeños podrían producir una pobre representación debido a estacionalidad o tendencias en la serie tiempo de los datos, produciendo finalmente resultados sesgados; por otro lado, bloques muy grandes requieren de un número elevado de datos dejando, por lo tanto, pocos bloques máximos, resultando en mayor varianza.

#### **2.1.1. Estimación de Parámetros para GEVD**

Existen múltiples métodos para la estimación de los parámetros para la GEVD. El más popular es el de Máxima Verosimilitud debido a su adaptabilidad y a las propiedades estadísticas (Coles, 2001). El segundo método que será usado en este trabajo es el método de L-Momentos. Este último usa combinaciones lineales de estadísticos de orden para estimar los parámetros. El desarrollo teórico para el uso de este método fue desarrollado por (Hosking, 1990). Al final será de interés hacer la comparación de los dos métodos de estimación de parámetros.

#### **2.1.2. Estimación de parámetros para la GEVD por método de Máxima Verosimilitud.**

Usamos el método de Máxima Verosimilitud para estimar los tres parámetros de GEVD definida en el teorema 2, asumiendo que las observaciones son independientes y que tienen la misma función de distribución GEV en común. El logaritmo de la verosimilitud de la distribución distingue dos casos del parámetro de forma  $\xi = 0$  y  $\xi \neq 0$ . En el primer caso, la GEVD requiere de un tratamiento

separado usando el límite de Gumbel de la GEVD (Coles, 2001). Esto conduce al log-verosimilitud:

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left( \frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ - \left( \frac{z_i - \mu}{\sigma} \right) \right\}.$$

Para  $\xi \neq 0$  se obtiene el siguiente log-verosimilitud:

$$\begin{aligned} \ell(\mu, \sigma, \xi) = & -m \log \sigma - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^m \log \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right] \\ & - \sum_{i=1}^m \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}, \end{aligned}$$

si

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

Al maximizar las dos funciones anteriores en cada caso con respecto a los parámetros conduce a la obtención de los estimadores de los parámetros de la GEVD. A través de este proceso también se obtiene fácilmente la matriz de información de Fisher (Coles, 2001). Calculando la inversa de la matriz de información y evaluándola con las estimaciones obtenidas nos conduce a una estimación asintótica de la matriz de varianzas y covarianzas.

### 2.1.3. Estimación de parámetros para la GEVD por el método de L-Momentos

El método de los momentos tiene bastante tiempo de ser usado en Estadística, sin embargo, su aplicación en la estimación de parámetros no siempre genera resultados satisfactorios, por ejemplo, cuando las muestras son pequeñas. En este último escenario, los valores obtenidos pueden ser muy diferentes de los momentos que corresponden a las distribuciones de probabilidad de la que proviene la muestra. Las estimaciones de los parámetros de la distribución ajustada por el método de los momentos son

frecuentemente menos precisas que las obtenidas por otros métodos de estimación como, por ejemplo, el de máxima verosimilitud. El método alternativo que se describe a continuación se basa en lo que se denomina L-momentos. Son análogos a los momentos convencionales, pero pueden ser estimados a partir de combinaciones lineales de estadísticos de orden, i.e., por L-estadísticos. Entre las ventajas de L-momentos se tiene que son menos susceptibles a los sesgos en la estimación y se aproximan más a una distribución normal asintótica en muestras finitas. Los parámetros estimados resultan muchas veces más precisos en muestra pequeñas que aquellos resultantes de la estimación por máxima verosimilitud (Hosking, 1990).

Sea  $X$  una variable aleatoria de valor real con función de distribución acumulada entonces para una muestra de tamaño  $n$  tomada de la distribución de  $X$ , el L-Momento de  $X$  está dado por (Hosking, 1990):

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k,r}, \quad r = 1, 2, \dots \quad (1)$$

Donde  $X_{r-k,r}$  son los estadísticos de orden  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  y  $EX_{r-k,r}$  es el valor esperado de  $X_{r-k,r}$ . La L en “L-Momentos” enfatiza que  $\lambda_r$  es una función lineal de los estadísticos de orden esperados.

Los valores de los diferentes  $\lambda_r$  se obtienen a través del paquete R *Lmoments*. Los momentos se relacionan con los parámetros de la GEVD a través de las siguientes ecuaciones:

$$\lambda_1 = \mu - \frac{\sigma}{\xi} (1 - \Gamma(1 - \xi)), \quad (2)$$

$$\lambda_2 = \frac{-\sigma}{\xi} \left( (1 - 2^\xi) \Gamma(1 - \xi) \right), \quad (3)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = 2 \frac{1-3\xi}{1-2\xi} - 3 \quad (4)$$

Donde  $\Gamma$  es la función Gamma. En particular  $\lambda_1$ ,  $\lambda_2$  y  $\tau_3$  podrían considerarse como las medidas de *localización*, *escala* y *sesgo* respectivamente (Hosking, 1990). Hay que notar que (4) no se puede resolver analíticamente, por lo tanto, al usar del paquete R *RootSolve* se requiere un método de Newton para obtener la solución del parámetro de Forma  $\xi$ . Sustituyendo el resultado obtenido en (3) y luego en (2), se generan las estimaciones de los parámetros. Nuevamente, la media de los valores estimados son iguales a los parámetros. Para poder obtener la matriz de varianzas y covarianzas se necesita hacer una transformación respecto a las relaciones arriba citadas. En primer lugar, cada parámetro tiene que ser expresado en términos de  $\lambda_i$  para  $1 \leq i \leq 3$ . Luego, usando el método Delta, se obtiene el error del parámetro de forma estimado mediante la relación:

$$Var(\hat{\xi}) \approx \nabla \xi^T V \nabla \xi, \quad (5)$$

donde

$$\nabla \xi^T = \left[ \frac{\partial \xi}{\partial \lambda_1}, \frac{\partial \xi}{\partial \lambda_2}, \frac{\partial \xi}{\partial \lambda_3} \right], \quad (6)$$

y  $V$  la matriz de varianzas y covarianzas de  $\lambda_i$  para  $1 \leq i \leq 3$ , la cual se obtiene también con el paquete R *Lmoments*. Para obtener las estimaciones del error para los parámetros de localización y escala se sigue el mismo procedimiento (Hosking, 1990).

## 2.2. Distribución Generalizada de Pareto

En la descripción de la GEVD se mencionó que en cada bloque máximo se escoge el valor más alto, pero hay casos en que los bloques tienen más de un valor que pueden considerarse altos, mientras que otros no tienen ninguno. El problema con esta situación es que se podría perder información importante en el momento de determinar la distribución de

probabilidad de estos valores altos. Para evitar este problema, se puede escoger un valor arbitrario alto (umbral) y clasificar todas las observaciones que superan este umbral como valores extremos, asegurando de esta manera que todos los valores extremos formen parte del análisis. Esto da lugar al método de *Picos Sobre Umbrales*, POT por sus siglas en inglés, el cual estima los parámetros de la cola de una distribución usando todas las excedencias sobre el umbral fijado. Igual que en el método de GEVD, se debe tener presente que hay que hacer un balance entre la varianza y el sesgo al momento de fijar el umbral. Si se fija un umbral demasiado bajo, esto resultará en la inclusión de observaciones que realmente no son valores extremos. Por el contrario, si se fija un umbral demasiado alto, esto dejará insuficientes observaciones disponibles para realizar las inferencias apropiadas acerca de la distribución principal (Aras, 2019). La distribución de estas excedencias está relacionada a la GEVD resultando el siguiente teorema:

**Teorema 3:** *Sea  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes con una función de distribución común  $F$ , y sea*

$$M_n = \max\{X_1, \dots, X_n\}.$$

*Denotamos un término arbitrario en la secuencia  $X_i$  por  $X$ , y suponemos que  $F$  satisface el teorema 2, tal que para  $n$  grande,*

$$\Pr\{M_n \leq z\} \approx G(z),$$

donde

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

*para algún  $\mu, \sigma > 0$  y  $\xi$ . Entonces, para un valor suficientemente grande  $u$ , la función de distribución de  $(X - u)$ , condicionada en  $X > u$ , es aproximadamente*

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}},$$

definido en  $\{y: y > 0 \text{ y } (1 + \xi y)/\tilde{\sigma} > 0\}$ , donde

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

La familia de distribuciones definidas por  $H$  se denomina **Familia Generalizada de Pareto** (Coles, 2001). Nuevamente, solo si el parámetro de forma  $\xi < 0$ , la distribución es acotada y por lo tanto el valor máximo esperado puede ser calculado.

### 2.2.1. Estimación de parámetros para la GPD

Para la estimación de los parámetros de la GPD se procederá con los dos métodos citados en 2.1.1, Máxima Verosimilitud MLE y L-Momentos. Como se explicó anteriormente, ambos tienen sus ventajas particulares y al final se podrá realizar la comparación para determinar la efectividad de estimación.

### 2.2.2. Estimación de parámetros para GPD por el método de Máxima Verosimilitud

El primer paso para la estimación de los parámetros de la GPD, es determinar un umbral adecuado; la metodología para el cálculo del mismo se amplía en (Coles, 2001). Una vez que se tiene el umbral adecuado  $u$ , los parámetros para la GPD pueden ser estimados por MLE. Suponiendo que  $y_1, \dots, y_k$  son los  $k$  excesos sobre el umbral  $u$ , entonces para  $\xi \neq 0$  la función log-verosimilitud para la GPD se deriva del teorema 3 como:

$$\ell(\sigma, \xi) = -k \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \log \left(1 + \frac{\xi y_i}{\sigma}\right), \quad (7)$$

si

$$\left(1 + \frac{\xi y_i}{\sigma}\right) > 0, \text{ para } i = 1, \dots, k. \text{ De lo contrario } \ell(\sigma, \xi) = -\infty.$$

Si el parámetro de forma toma el valor de  $\xi = 0$ , entonces la distribución no es acotada. El log-verosimilitud para la GPD cuando  $\xi = 0$  está dada por:

$$\ell(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^k y_i. \quad (8)$$

Maximizando estas funciones, se obtiene la estimación de los parámetros. Además, la media y varianza de estos estimadores se calculan con el mismo proceso como se explicó anteriormente.

### 2.2.3. Estimación de parámetros para la GPD por el método de L-Momentos

Los momentos relativos a los parámetros de GPD se obtienen a través de las siguientes ecuaciones:

$$\lambda_1 = \mu + \frac{\sigma}{(1-\xi)} \quad (9)$$

$$\lambda_2 = \frac{\sigma}{(1-\xi)(2-\xi)} \quad (10)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \frac{1+\xi}{3-\xi}. \quad (11)$$

Dado que los valores de  $\lambda_r$  y su matriz de varianzas y covarianzas se obtienen mediante el uso del paquete R *Lmoments*, la ecuación (11) se puede resolver por método numéricos. Reemplazando la estimación obtenida en (10) y luego en (9), se calculan las estimaciones de los parámetros requeridos. Escribiendo los parámetros como funciones de los valores de L-Momentos, se puede calcular el error a través del método Delta, usando el mismo razonamiento de la sección previa. La matriz de varianzas y covarianzas para los estimadores de L-Momentos se obtienen con el uso del paquete R *Lmoments*, y con el paquete R *Rootsolve* se calculan los gradientes requeridos como en (6).

### 2.3. Criterios de selección de modelos

El Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) son índices que frecuentemente se usan para la selección de modelos estadísticos. Si  $k$  es el número de parámetros,  $\ell$  el valor de máxima verosimilitud del modelo y  $n$  el tamaño de muestra, el AIC y BIC se definen de la siguiente manera:

$$AIC = 2k - 2\ln(\ell) \quad (12)$$

$$BIC = -2\ln(\ell) + k. \quad (13)$$

El modelo que dé como resultado un valor más alto de la función de máxima verosimilitud es el que mejor ajuste ofrece, esto se traduce en menor valor de los índices de AIC y BIC (Montesinos, 2011). Por lo tanto, la regla que se usa para la selección es que el modelo de menor AIC o BIC es el mejor. De los dos índices, el BIC es más sensible al número de parámetros usados en el modelo y el tamaño de la muestra.

### 2.4. Test de razón de verosimilitud

Se trata de una prueba de hipótesis que permite probar dos modelos, uno con  $p$  parámetros y otro con  $p + 1$  parámetros con las siguientes hipótesis:

$$H_0 = \text{Parámetro adicional del segundo modelo es igual a cero}$$
$$H_1 = \text{Parámetro adicional del segundo modelo no es igual a cero.}$$

En este trabajo se usará esta prueba para determinar si el parámetro de forma  $\xi$  del modelo GEVD es cero, correspondiendo por lo tanto a un modelo de distribución Gumbel. Entonces, si no se rechaza la hipótesis nula, ( $p$ -valor  $> 0.05$ ), este parámetro sería insignificante confirmándose una distribución de Gumbel.

# CAPÍTULO 3

## 3. METODOLOGÍA

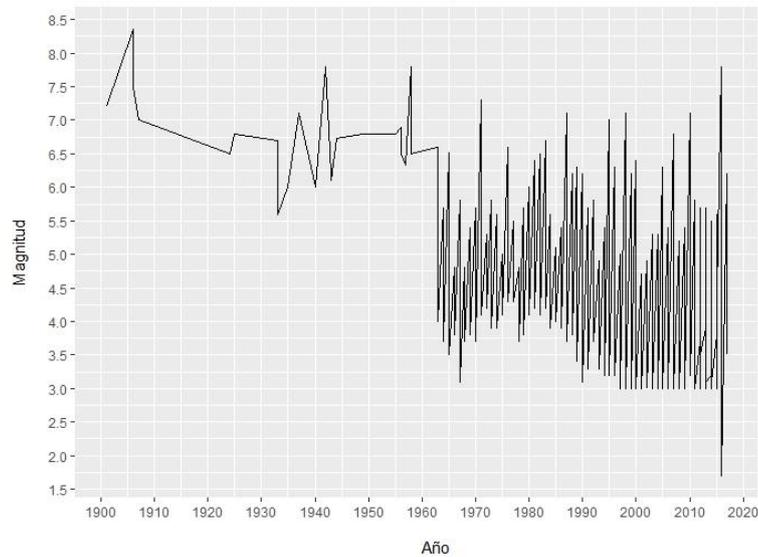
### 3.1. Análisis descriptivo de los datos

El set de datos disponibles para la actual investigación consta de 12531 observaciones de sismos registrados para todas las provincias del país, de las cuales 1083 pertenecen a las provincias de Guayas y Santa Elena; los periodos de tabulación corresponden a los años 1906 hasta 2017 a nivel nacional y desde 1943 hasta 2017 para las provincias de interés. A pesar de tener información sismológica tan temprana como de inicios del siglo XX, los datos disponibles hasta 1960 son poquísimos. La fuente de donde se obtuvo la información es el Instituto Geofísico de la Escuela Superior Politécnica Nacional (Instituto Geofísico, 2020). Además de las magnitudes ajustadas en unidades de  $M_w$  de los sismos en escala de Richter, también se cuenta con la profundidad (km), longitud y latitud del epicentro de los sismos, clasificada por provincias y por región.

Por otro lado, es importante resaltar que Santa Elena tiene vida jurídica como provincia del Ecuador desde el 7 de noviembre de 2007, anterior a esa fecha formaba parte de la provincia del Guayas, de tal manera que, en términos históricos al hablar de las dos provincias en el periodo de la tabulación de datos, se estará refiriendo fundamentalmente a la provincia del Guayas. De hecho, el número de sismos registrados como provincia de Santa Elena son 85 mientras que para Guayas son 998 en el periodo de estudio.

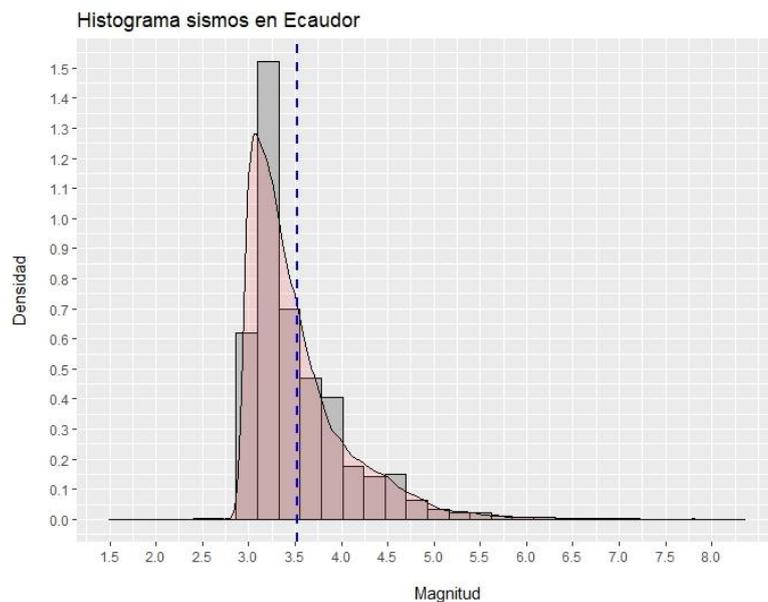
Realizando un gráfico de magnitud vs el tiempo de ocurrencia con los datos disponibles a nivel nacional se obtiene la figura 3.1, como se ha mencionado arriba, los datos registrados a partir de 1960 son más numerosos que los anteriores, es probable que la tecnología disponible antes de 1960 no fuera la adecuada como los modernos sistemas de medición de sismos. Además, no se muestran señales de estacionalidad de los datos.

**Figura 3.1. Gráfica de magnitud vs tiempo de ocurrencia de sismos en Ecuador**



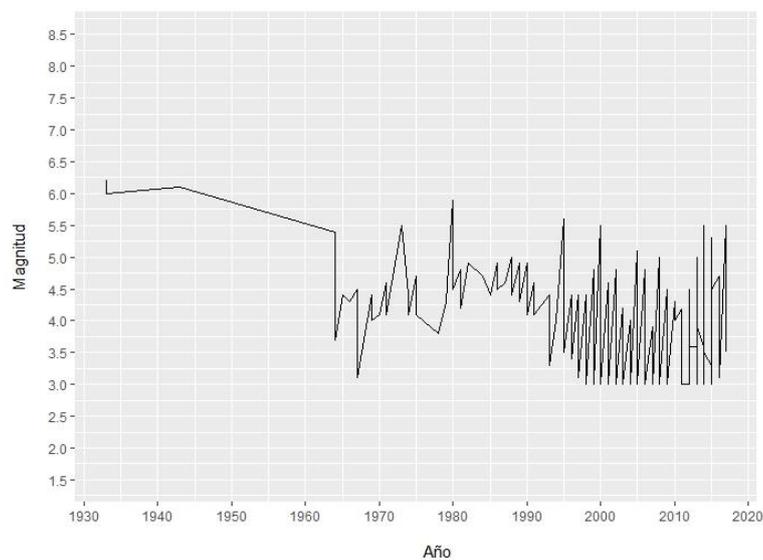
Los datos a nivel nacional tienen una media  $\mu = 3.53$  y una desviación estándar  $\sigma = 0.56$  en la magnitud, con valores máximos y mínimos de 8.35 y 1.70  $M_w$ . En la figura 3.2 se presenta el histograma con la densidad estimada de los eventos sísmicos. Esta gráfica brinda una definición a priori de la posible distribución de los datos. La primera observación es que la distribución es evidentemente sesgada hacia la derecha.

**Figura 3.2. Histograma de densidad de sismos en Ecuador**



El gráfico de magnitud vs el tiempo de ocurrencia para los datos solo de las provincias Guayas y Santa Elena, que se presenta en la figura 3.3, al igual que la 3.1, muestra poquísimos datos antes del año 1960. De manera idéntica no se observa una tendencia marcada de estacionalidad y tampoco viola la estacionariedad de los datos. No se tiene registro de actividad sísmica antes de 1930.

**Figura 3.3. Gráfica de magnitud vs tiempo de ocurrencia de sismos en Guayas y Santa Elena**

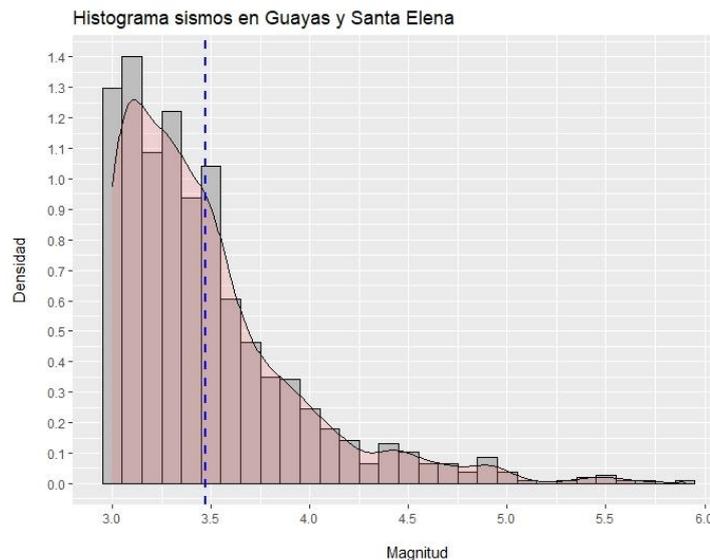


En la figura 3.4 se presentan el histograma para los sismos en las provincias objeto de estudio. La media de la magnitud de los sismos es de  $\mu = 3.49$  con una desviación estándar  $\sigma = 0.50$ , con valores máximos y mínimos 6.2 y 3.0  $M_w$ . Al igual que en el caso de los datos a nivel nacional, la distribución para Guayas y Santa Elena tiene un fuerte sesgo hacia la derecha indicando que sería matemáticamente incorrecto el tratamiento de los datos usando una distribución simétrica como la distribución normal. Por consiguiente, se requiere una distribución diferente para modelar la cola de manera correcta como ya se expuso en el apartado 2.

Además, el análisis de los datos arroja que el sismo de mayor magnitud se registró en el año 1933 con una magnitud de 6.20, que de acuerdo a la escala de los efectos de Richter (Seguí, 2021) se clasifica como un sismo que podría ocasionar daños severos en áreas muy pobladas (Magnitud entre 6.1

y 6.9), de hecho, en este grupo se registran 2 eventos que representan el 0.18% de las observaciones. Los sismos que podrían ocasionar daños ligeros a edificios (Magnitud entre 5.5 y 6.0) son 8 que representan el 0.74% de los casos. La mayoría de los eventos sísmicos han ocurrido con una magnitud menor a 5.5, que podrían causar daños menores o que no se sienten, representando el 99.08%.

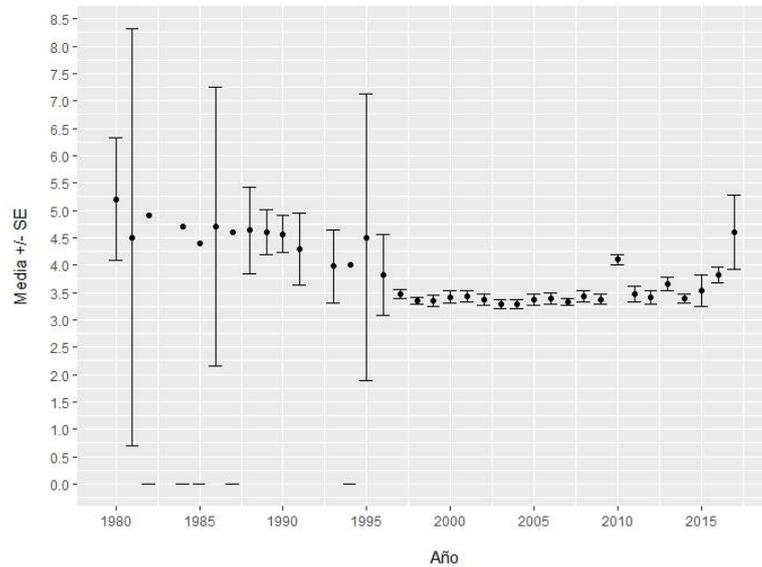
**Figura 3.4. Histograma de densidad de sismos en Guayas y Santa Elena**



En la figura 3.5 se puede ver un gráfico de magnitud vs el tiempo de ocurrencia con la media de sismos reportados por cada año. Las barras de los intervalos de confianza (de cuanto por ciento) del error se reducen drásticamente luego del año 1990, en parte debido al incremento de mediciones sísmicas en las últimas décadas. Los puntos que aparecen sin barra son años en los que hubo un evento reportado. De acuerdo con la observación, tampoco se muestra tendencia de estacionalidad por lo que implicaría una estacionariedad de los datos.

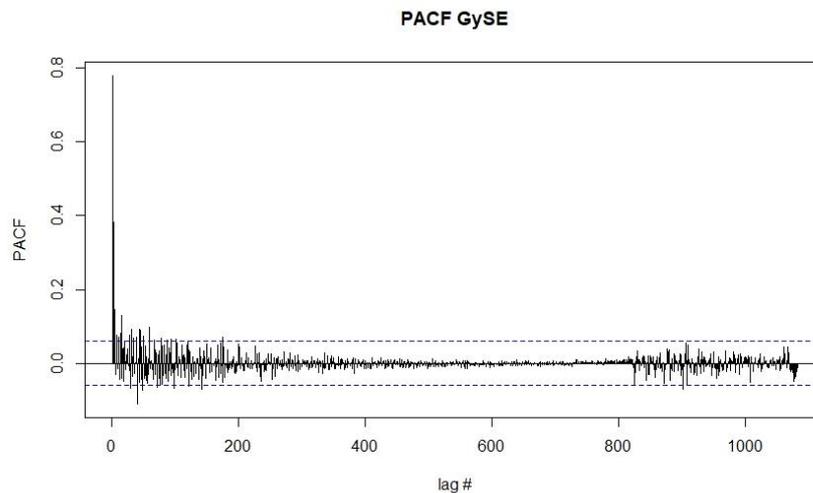
Una prueba que ayuda a determinar la estacionariedad es la Prueba Aumentada de Raíz Unitaria de Dickey-Fuller o ADF por sus siglas en inglés. La prueba se aplicó a todos los datos de las provincias Guayas y Santa Elena; para esto se usó el paquete de R *tseries*, donde el valor  $p$  obtenido fue de 0.039 por lo que se comprueba la estacionariedad de los datos.

**Figura 3.5. Diagrama de magnitud vs el tiempo de ocurrencia de sismos en Guayas y Santa Elena: medias y error estándar**



Con la construcción de la gráfica de la Función de Autocorrelaciones Parciales (PACF), se puede también verificar la estacionariedad de los datos, ya que se espera el PACF tienda a cero para cada *lag-time*, demostrando que no hay dependencia con el tiempo. Esta prueba se realizó con todos los datos y se presenta en la figura 3.6. En la gráfica se observan ciertas excedencias en la primera parte de los datos, pero en general se puede decir que no hay autocorrelaciones, y que las desviaciones pueden deberse al conocido ruido blanco.

**Figura 3.6. Diagrama de Autocorrelaciones Parciales para sismos en Guayas y Santa Elena**



### **3.2. Estimación de parámetros de la Distribución Generalizada de Valores Extremos**

Como se definió en 2.1, el modelo precisa que se determinen o más bien que se estimen sus parámetros, para los cual en este trabajo se aplicaron dos métodos, el de Máxima Verosimilitud y el método de L-Momentos que a continuación se describen para la información sísmológica de las provincias de Guayas y Santa Elena.

Los tres parámetros de este modelo son los de Forma  $\xi$ , Escala  $\sigma$  y Localización  $\mu$ . Anteriormente se mencionó que esta distribución se basa en el método de Bloque Máximo en donde cada bloque entrega un máximo valor de magnitud del sismo que luego son agrupados y con estos se procede a la estimación de los respectivos parámetros.

Es importante observar que el 98.2% de las mediciones reportadas en el catálogo de sismos corresponden al periodo comprendido desde 1980 hasta 2017, por lo que se tomarán las mediciones de este periodo para realizar las estimaciones de los modelos respectivos, teniendo presente lo reportado por García et. al. (García S. N., 2018).

#### **3.2.1. Estimación de parámetros de la GEVD usando el método de Máxima Verosimilitud**

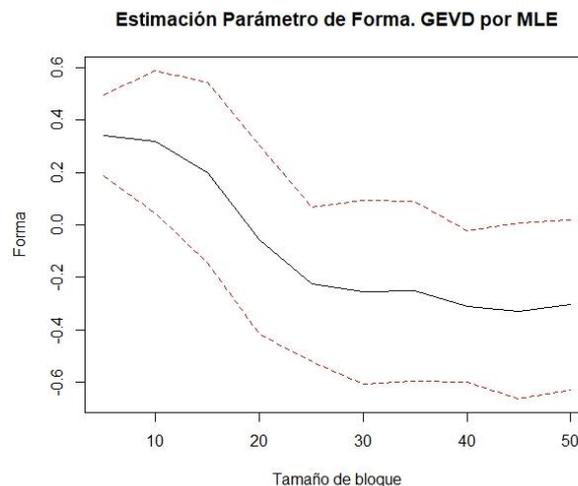
El proceso de estimación de los parámetros de la GEVD consiste en la clasificación de las observaciones en bloques de igual tamaño, para luego realizar el ajuste de la GEVD al conjunto resultante de los valores máximos de cada bloque, sin embargo, por pragmatismo, también se suele realizar bloques con observaciones correspondientes a un año completo de mediciones (Coles, 2001).

Como se mencionó antes, aunque se cuenta con una trazabilidad de sismos desde 1906, la mayor parte de los datos han sido colectados en los últimos años, por lo que el número de las incidencias por año no ha sido uniforme como es de esperarse ante

un fenómeno natural. Por ello, se ha realizado una distribución de los datos en bloques de igual tamaño sin tomar en consideración al año al que pertenecen, pero además considerando el criterio de bloques por año.

Para realizar una buena elección del número de bloques en los que se distribuirán los datos, se ha realizado la estimación de los parámetros con diferentes números de bloques, iniciando con 5 y llegando hasta 50, de cada uno de los cuales se extrae el valor máximo. Aplicando el paquete de *R extRemes*, se obtienen un total de 10 estimaciones de cada parámetro (véase Anexo A). Además, los errores estándares para las estimaciones se obtuvieron a través de la matriz de varianza y covarianza de los estimadores. Sus respectivos intervalos de confianza 95% también se muestra en el anexo A.

**Figura 3.7. Estimación del Parámetro de Forma  $\xi$  con diferentes tamaños de Bloques de la GEVD usando MLE**



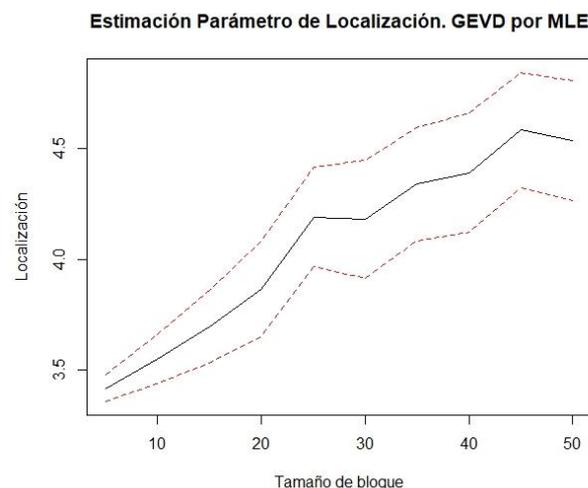
En la Figura 3.7, las estimaciones del parámetro de Forma  $\xi$  se han graficado vs. el tamaño de bloque respectivo, de acuerdo con la teoría, la estimación para este parámetro estaría entre un valor de 0 y -0.5, en donde los estimadores de Máxima Verosimilitud son regulares, es decir presentan propiedades asintóticas regulares

(Coles, 2001). Usando la normalidad asintótica de los estimadores, también se calcularon los errores estándares.

En el caso del parámetro de Forma, el tamaño de bloque adecuado estaría entre 25 y 35, con el menor error estándar en el bloque 25, sin embargo, hay que cuidar que el tamaño no sea muy pequeño para evitar sesgos en las estimaciones. El intervalo de confianza más pequeño se obtiene para un tamaño de bloque 30, perfilándose como el tamaño de bloque adecuado.

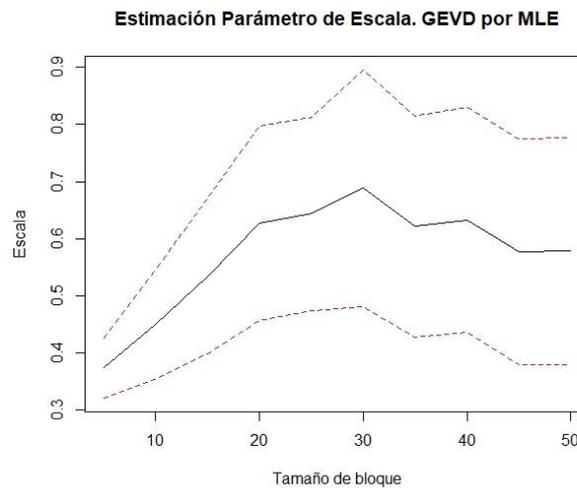
Siguiendo el mismo procedimiento se construyó la gráfica que se presenta en la figura 3.8 para la estimación del parámetro de Localización  $\mu$ . Se observa que, a diferencia del parámetro de Forma, cuya tendencia es decreciente, el parámetro de Localización es creciente, mostrando cierta estabilidad para los tamaños de bloques entre 25 y 30.

**Figura 3.8. Estimación del Parámetro de Localización  $\mu$  con diferentes tamaños de Bloques de la GEVD usando MLE**



En la figura 3.9 se presenta la estimación del parámetro Escala  $\sigma$ . Este no presenta una tendencia específica.

**Figura 3.9. Estimación del Parámetro de Escala  $\sigma$  con diferentes tamaños de Bloques de la GEVD usando MLE**



Usando el criterio que consiste en generar un bloque por cada año de registro de observaciones, se obtuvieron 35 bloques donde las estimaciones de los parámetros son: Localización = 4.554, Forma = -0.229 y Escala = 0.414.

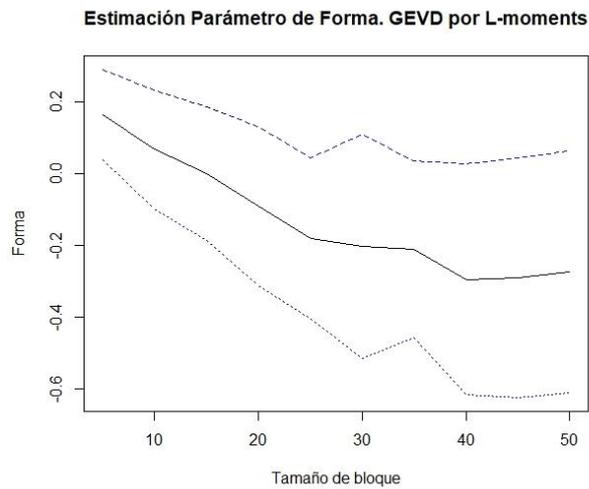
### 3.2.2. Estimación de parámetros de la GEVD usando el método de L-Momentos

El siguiente método para la estimación de los parámetros de la GEVD es el de L-Momentos. Las ecuaciones que conectan a los momentos estadísticos relacionados con los parámetros de interés son las igualdades 2, 3 y 4 de la sección 2.1.3. Como lo establece (Aras, 2019), el primer parámetro a obtener es el de Forma  $\xi$ , el cual se calcula mediante el uso del método de Newton-Raphson para obtener raíces con la ayuda del paquete de R *rootSolve*. Los parámetros de Escala  $\sigma$  y Localización  $\mu$  se obtienen por sustitución en las ecuaciones 1 y 2 (Aras, 2019). Igual que en la sección anterior, se realizó la estimación de los parámetros para diferentes números de bloques iniciando en 5 y finalizando en 50, con un total de 10 bloques.

En la figura 3.10 se recogen las estimaciones del parámetro de forma; la tendencia de la curva formada es muy similar a la obtenida

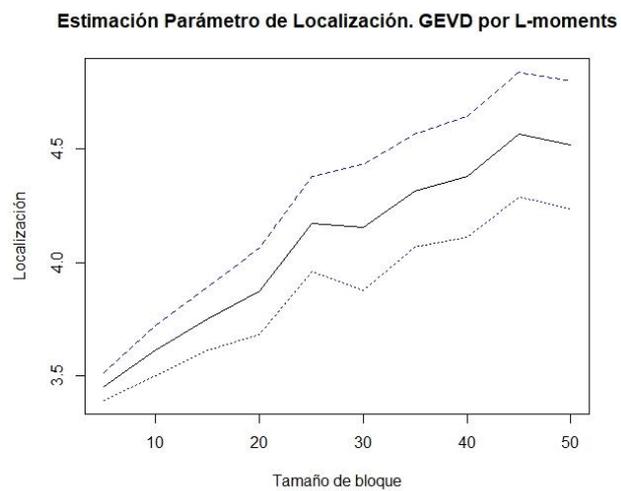
con la estimación por MLE. De forma parecida se podría establecer el rango adecuado de bloques entre 25 y 35.

**Figura 3.10. Estimación del Parámetro de Forma  $\xi$  con diferentes tamaños de Bloques de la GEVD usando L-Momentos**



En la figura 3.11 se representan las estimaciones para el parámetro de Localización  $\mu$ , donde la tendencia es creciente y también tiene una característica estabilidad entre tamaño de bloques de 25 y 30.

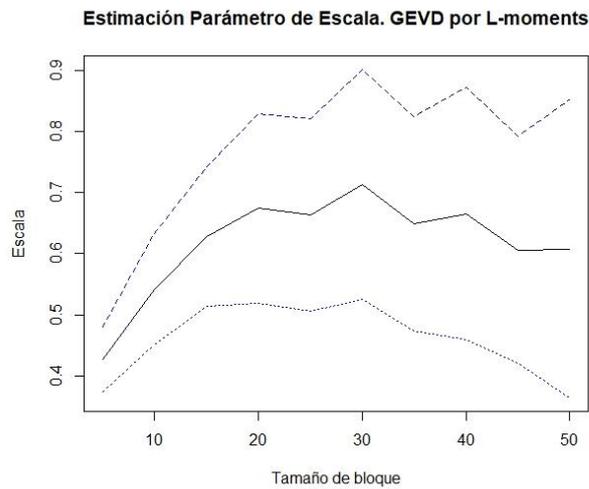
**Figura 3.11. Estimación del Parámetro de Localización  $\mu$  con diferentes tamaños de Bloques de la GEVD usando L-Momentos**



En la figura 3.12 se presenta la estimación del parámetro de Escala  $\sigma$  se nota mucha similitud con las estimaciones realizadas

por MLE. En el anexo B se puede ver el detalle de las estimaciones con diferentes números de bloques.

**Figura 3.12. Estimación del Parámetro de Escala  $\sigma$  con diferentes tamaños de Bloques de la GEVD usando L-Momentos**



Asimismo, se obtuvo la estimación usando bloques por cada año de registro de sismos, donde el número óptimo de bloques fue  $nb = 35$ , obteniéndose las estimaciones para los parámetros: Localización = 4.554, Forma = -0.201 y Escala = 0.423.

### 3.3. Estimación de parámetros de la Distribución Generalizada de Pareto

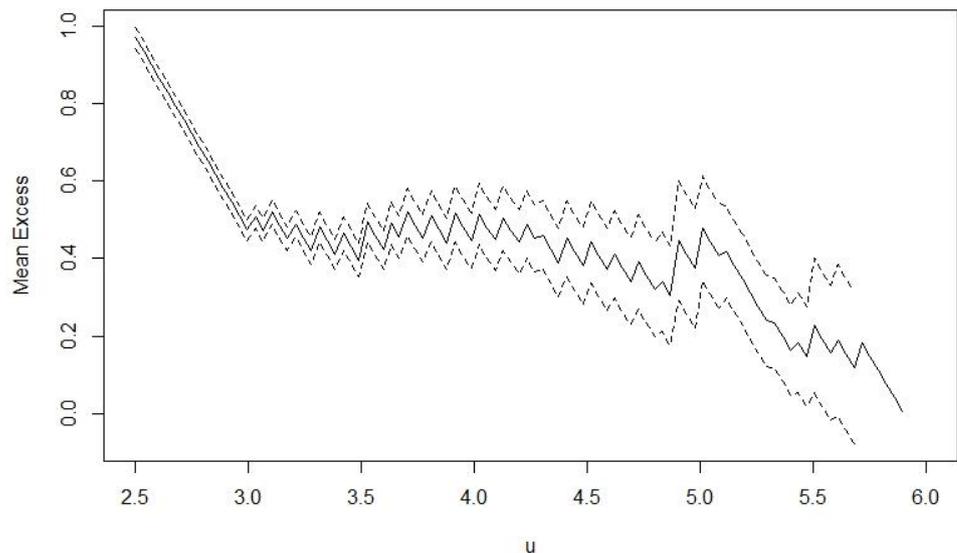
La selección del umbral implica un balance entre el sesgo y la varianza, en este caso, un umbral muy pequeño es muy probable que viole el principio asintótico del modelo, conduciendo a sesgos de estimación; por el contrario, un umbral demasiado alto va a generar pocas excedencias sobre el umbral con los cuales se estima el modelo, conduciendo a una varianza muy alta. Hay dos métodos para determinar el umbral óptimo. El primero es una técnica exploratoria previo a la estimación respectiva; el otro es una evaluación de la estabilidad de los parámetros estimados, basado en los modelos ajustados a lo largo de un rango de umbrales diferentes (Coles, 2001).

En la figura 3.13 se presenta el gráfico de vida residual media, donde a diferentes valores de umbrales  $u$  se calcula una aproximación válida de la distribución de las excedencias, la vida media residual debería ser aproximadamente lineal con respecto a  $u$  (Coles, 2001). Bajo este criterio, para umbrales menores a 3 claramente se aprecia linealidad, sin embargo, escoger valores de umbral muy bajos (respecto a la media) no es aconsejable. Las excedencias se calculan de acuerdo con la siguiente ecuación:

$$ME = \frac{1}{n_u} \sum_{i=1}^{n_u} (x(i) - u)$$

**Figura 3.13. Gráfico de Excedencia de medias, selección de umbral**

$u$ .



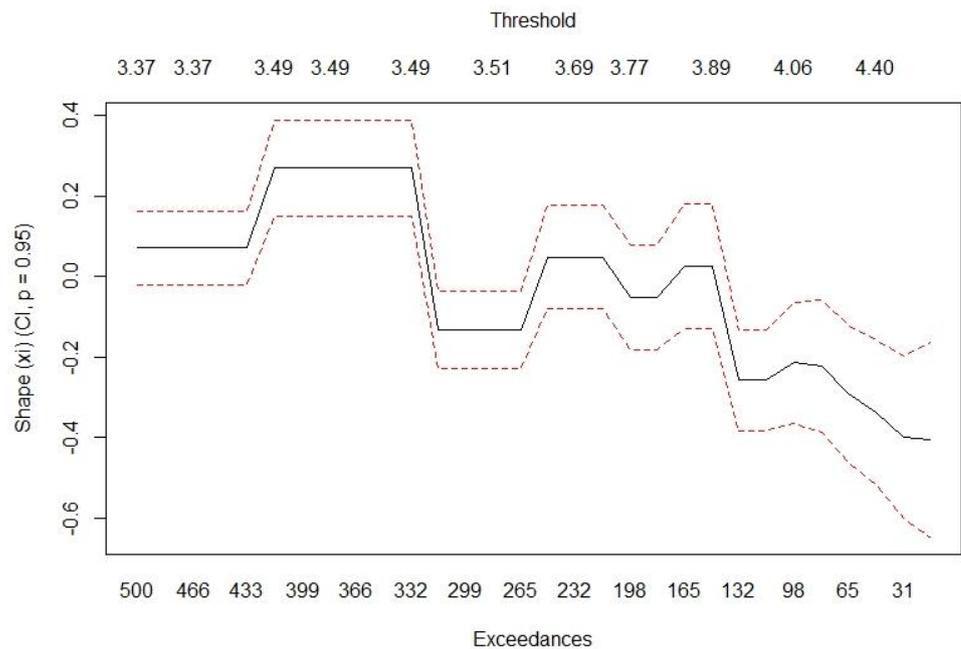
Umbrales entre 3.60 y 4.10 parecen ser bastante estables, manteniendo una tendencia lineal y cuasi constante en este caso, por lo que el umbral adecuado podría estar en este rango.

También se cuenta con la herramienta de selección mostrada en la Figura 3.14; en esta gráfica, además, se muestra el número de excedencias por cada umbral probado. Aquí el criterio adicional es que

haya un número suficientemente grande de excedencias sobre el umbral para que no haya sesgos de estimación.

Para la región entre 3.49 y 3.89 son aproximadamente estables (cuasi-lineales respecto al parámetro de Forma) y con un número de excedencias aceptable. Las gráficas se obtuvieron usando los paquetes R *ismev* y *evir* respectivamente.

**Figura 3.14. Gráfico de Excedencia de medias y parámetro de Forma  $\xi$  para la selección de umbral**

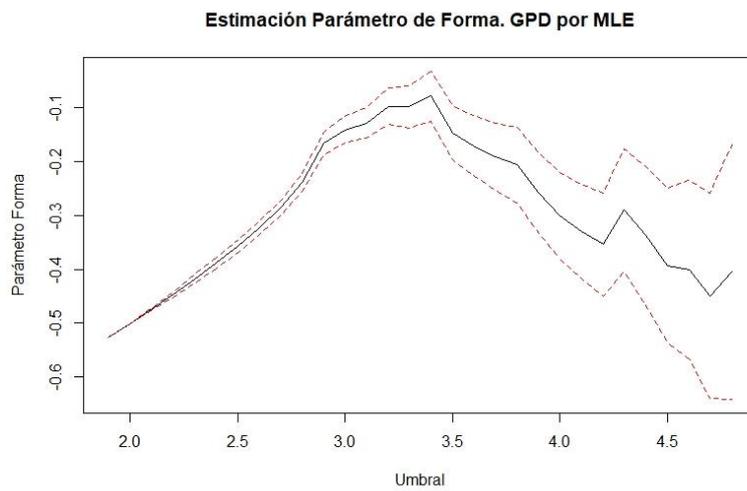


### 3.3.1. Estimación de parámetros de GPD usando el método de Máxima Verosimilitud

De acuerdo con Coles, sobre un umbral  $u_0$  en donde la condición asintótica para la distribución de Pareto es válida, las estimaciones del parámetro de forma  $\xi$  debe ser aproximadamente constantes, mientras que las estimaciones de escala  $\sigma$  deben ser lineales en  $u$  (Coles, 2001). En la figura 3.15 se presentan las estimaciones del parámetro de forma a diferentes umbrales.

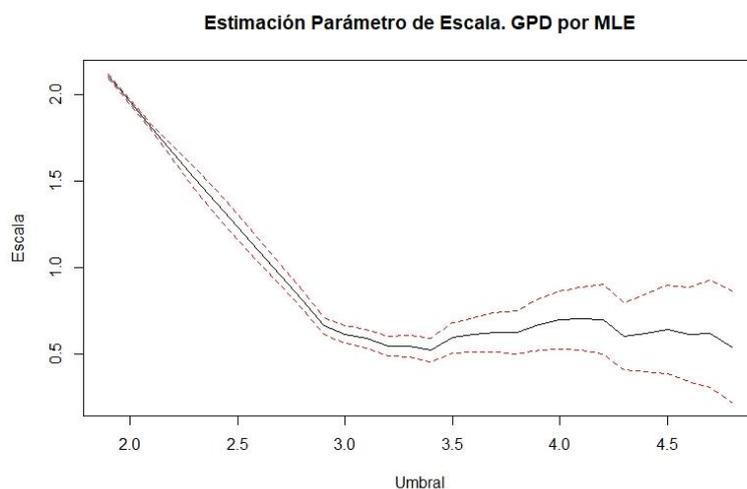
Se utilizó el paquete de R *extRemes* para las estimaciones de los parámetros. En el rango de interés de  $u$  (3.49 y 3.89) no se presentan valores aproximadamente constantes, pero, se nota que alrededor de 3.8 hay un punto de cambio del patrón.

**Figura 3.15. Estimación del Parámetro de Forma  $\xi$  con diferentes umbrales usando MLE**



En el anexo C se puede verificar que entre 3.6 y 3.8 el parámetro de forma tiene cierta estabilidad, aproximadamente constante.

**Figura 3.16. Estimación del Parámetro de Escala  $\sigma$  con diferentes umbrales usando MLE**



Las estimaciones del parámetro de escala se presentan en la figura 3.16, donde para valores de  $u$  de 3.60 y 3.80 se presenta una

tendencia casi constante y lineal, la misma que se comprueba en la figura 3.17 donde se presenta el parámetro de escala reparametrizado.

Con estas observaciones, se ha cerrado un poco más el intervalo adecuado para trabajar como umbral para el modelo de Pareto Generalizado. Para poder realizar afirmaciones respecto al parámetro de escala, se debe notar que el parámetro cambia con respecto a  $u$  durante la estimación.

Para remediar esta dificultad se puede reparametrizar la escala  $\sigma$  del Pareto generalizado de la siguiente manera:

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

Con este argumento, al graficar  $\sigma^*$  y  $\xi^*$  vs.  $u$ , se selecciona el umbral cuyas estimaciones permanecen aproximadamente constantes (Coles, 2001). Los intervalos de confianza para  $\xi^*$  se obtienen directamente de la matriz de varianzas y covarianzas  $V$ . Los intervalos de confianza para  $\sigma^*$  requieren el cálculo mediante el método Delta, usando:

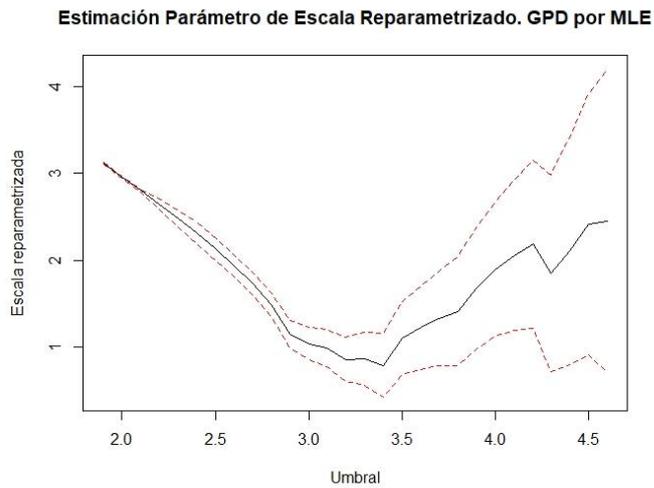
$$Var(\sigma^*) \approx \nabla \sigma^{*T} V \nabla \sigma^*,$$

donde

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

En la gráfica 3.17, se muestra el parámetro de escala reparametrizado. Realizando una inspección en la cual se estimó que el parámetro de forma se mantiene constante, nuevamente se presenta una condición cuasi-lineal alrededor de umbrales de 3.50 y 3.80. En el anexo C se recogen las estimaciones realizadas a diferentes niveles de umbrales.

**Figura 3.17. Parámetro de Escala reparametrizado para la GPD por MLE**



### 3.3.2. Estimación de los parámetros de la GPD usando el método de L-Momentos

Igual que con el método de MLE, la estimación para los parámetros de la GPD con L-momentos, se realizará con las ecuaciones 9, 10 y 11; luego de obtenidos los L-momentos, se emplean las siguientes relaciones para la solución analítica:

$$\xi = \frac{3\tau_3 - 1}{\tau_3 - \xi},$$

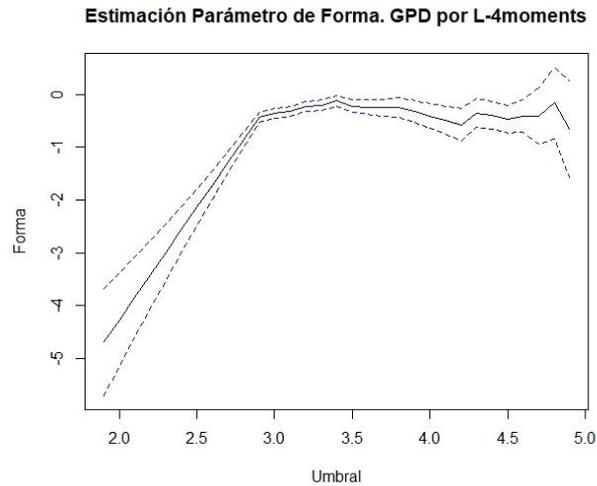
$$\sigma = \lambda_2(1 - \xi)(2 - \xi),$$

y

$$\mu = \lambda_2 - \frac{\sigma}{1 - \xi}.$$

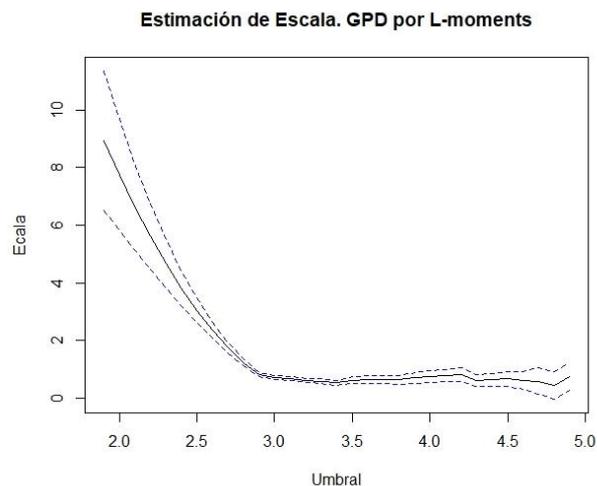
Todos estos cálculos se realizan con la ayuda el paquete R *Lmoments*, de manera alternativa, también está disponible el paquete R *extRemes*, este último siendo más fácil de usar debido que arroja directamente los valores de la estimación de parámetros, mientras que con el primero el cálculo es numérico, especialmente los gradientes para determinar la matriz de varianzas y covarianzas (Aras, 2019).

**Figura 3.18. Estimación del Parámetro de Forma  $\xi$  con diferentes umbrales usando L-momentos**



La búsqueda del umbral adecuado se realizó con un umbral inicial de 1.9 hasta 4.6. En la figura 3.18 se muestra la estimación del parámetro de Forma  $\xi$ , donde nuevamente vemos que se mantiene relativamente constante entre los valores de 3.50 y 3.70 (ver anexo D) mientras que en la figura 3.19 se presenta la estimación del parámetro de Escala  $\sigma$ , en el que se observa una gran estabilidad en un amplio rango de  $u$ , incluyendo el rango de interés 3.49 y 3.89.

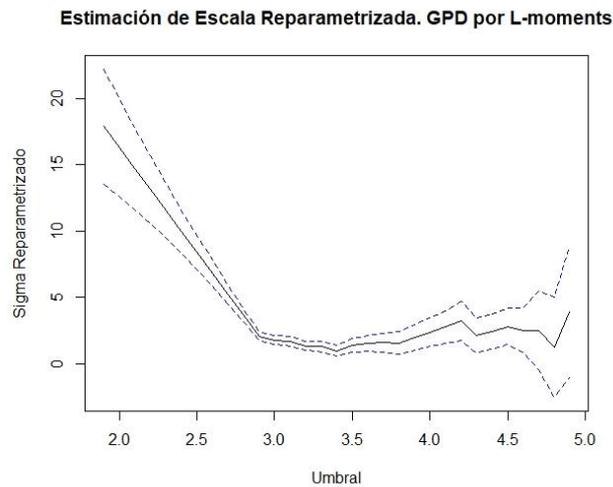
**Figura 3.19. Estimación del Parámetro de Escala  $\sigma$  con diferentes umbrales usando L-momentos**



Al igual que se realizó para el MLE en la sección anterior, también se realizó la reparametrización de  $\sigma$ , esto es, con la finalidad de

revisar el comportamiento de la estimación obtenida, llegándose a la misma conclusión que con el MLE (ver figura 3.20). En el anexo D se puede verificar que el comportamiento lineal se mantiene en valores de 3.5 a 3.7.

**Figura 3.20. Parámetro de Escala reparametrizado para la GPD usando L-momentos**



De los resultados obtenidos por los dos métodos, tanto el MLE como el método de L-Momentos permiten concluir que el umbral óptimo sería de 3.7, ya que en los dos casos el parámetro de Forma  $\xi$  es cuasi-constante en este valor y las reparametrizaciones de la Escala sus son lineales en este umbral.

### 3.4. Modelo Gutenberg-Richter: Regresión lineal y Regresión Poisson

La descripción más común de la ocurrencia de un sismo es la descrita por la ley de Gutenberg-Richter, la misma que permite obtener la relación entre la Frecuencia y la Magnitud de actividad sísmica de una zona específica. Los parámetros que se usan para la evaluación cuantitativa de la sismicidad son  $\alpha$  y  $\beta$ , de la relación entre magnitud y frecuencia:

$$\text{Log}N_m = \alpha - \beta M, \quad (14)$$

donde  $N$  es el número esperado de sismos que ocurren en una región durante un periodo de tiempo determinado en relación con sus magnitudes  $M$ , donde  $m$  es el número de años cubierto por el catálogo de datos de la sismicidad en estudio (Bayrak, 2008).

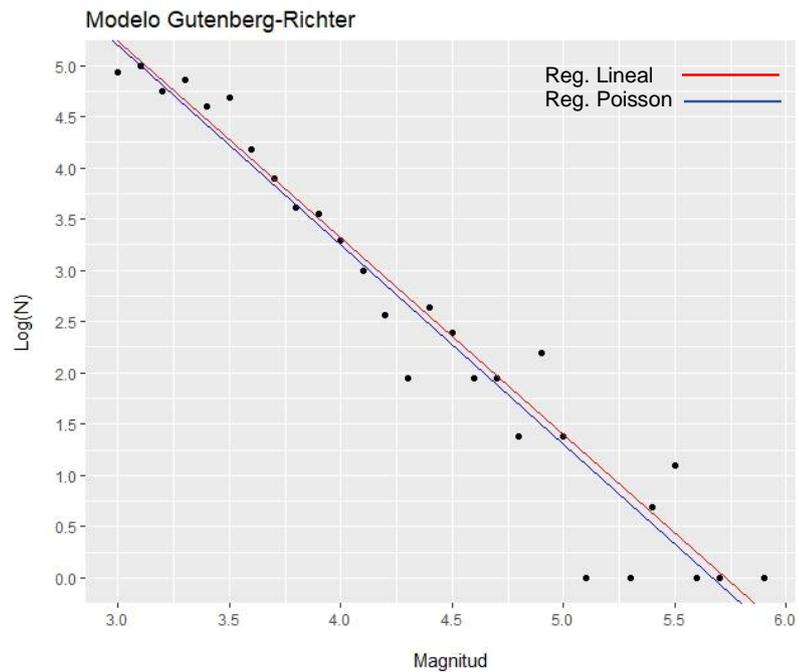
La ecuación (14) describe una relación lineal entre el logaritmo del número de observaciones y la magnitud asociada, que se puede modelar por medio de la regresión lineal clásica y también por la regresión de Poisson, esta última, a través de la estimación por máxima verosimilitud. Con la ayuda del paquete de *R stats*, se obtuvieron los modelos con sus respectivos parámetros estimados. En la tabla 3.1 se muestran las estimaciones de los parámetros, para los dos modelos, Lineal y Poisson. Con las estimaciones de los parámetros se pudo calcular el número de sismos o frecuencia de los mismos ligados a diferentes intervalos de magnitudes ( $M_w$ ); en general se puede decir que el modelo Poisson puede estimar tan bien o mejor que el modelo clásico lineal, esto reforzado por la raíz cuadrada del Error Cuadrático Medio o RMSE por sus siglas en inglés, siendo menor para el caso de la regresión de Poisson (12.51) que para la regresión Lineal (13.00), esta última comúnmente usada en el modelo de Gutenberg-Richter.

**Tabla 3.1. Estimaciones del número de sismos con los modelos Lineal y Poisson de los sismos de Guayas y Santa Elena**

Magnitud	Número de Sismos		
	Real	Regresión Lineal	Regresión Poisson
$M_w \leq 3.5$	738	704,13	718,11
$3.5 < M_w \leq 4.0$	213	197,62	224,72
$4.0 < M_w \leq 4.5$	65	74,63	110,40
$4.5 < M_w \leq 5.0$	31	28,18	68,85
$M_w > 5.0$	10	10,87	13,15
RMSE	-	13,00	12,51
Parámetros	$\hat{\alpha}$	11,04	10,81
	$\hat{\beta}$	-1,95	-1,87

También se puede comprobar que, para números grandes de sismos pequeños (e.g., frecuencia de sismos de magnitud  $\leq 4.0 M_w$ ), la regresión de Poisson predice mejor que la Lineal. Por otro lado, todo lo contrario, sucede con números más pequeños (e.g., frecuencia de sismos de magnitud  $> 4.0 M_w$ ). En la figura 3.21 se presentan las gráficas del número de sismos presentados por bandas de magnitudes disponibles en el catálogo, además de la estimación por regresión lineal clásica (en color rojo) y la línea de la estimación de la regresión de Poisson (en color azul). Es fácil ver el buen ajuste de ambos modelos de regresión.

**Figura 3.21. Modelo Gutenberg-Richter con las regresiones Lineal y Poisson**



# CAPÍTULO 4

## 4. RESULTADOS

Luego de haber realizado las estimaciones de los parámetros mediante los diferentes métodos propuestos en este trabajo, los resultados se muestran en la tabla 4.1. Aplicando el criterio de la distribución en bloques de igual tamaño de datos, se determinó que el número de bloques óptimo es de  $n_b = 30$  para la GEVD; las estimaciones son los resultados de las líneas 1 y 2 de la tabla 4.1, donde por el método de MLE se obtuvo además el AIC y BIC. Comparando con el modelo obtenido con el criterio de bloques por años  $n_b = 35$ , este último tiene un mejor ajuste de acuerdo con los índices AIC y BIC. Las estimaciones para los dos modelos presentan una diferencia notoria entre los 3 parámetros al ser calculados por los dos métodos de MLE y L-Momentos.

Llegado a este punto es importante someter a prueba el parámetro de Forma  $\xi$  para los modelos usando la GEVD tanto con  $n_b = 30$  como con  $n_b = 35$  para probar si este parámetro tiene un valor estadísticamente diferente de cero, es decir, que el modelo no tendría una distribución de Gumbel. Para esto, usando el paquete R *extRemes*, se obtuvo las estimaciones puntuales para los diferentes modelos, así como los resultados de la prueba de razón de verosimilitudes con los modelos GEVD ya obtenidos de acuerdo con la prueba de hipótesis del apartado 2.4. Las salidas obtenidas se muestran en la tabla 4.1.

El resultado del valor p ( $p\text{-value} > 0.05$ ) indica que es factible considerar para ambos casos un modelo de Gumbel ( $\xi = 0$ ), sin embargo, hay que ser muy criteriosos en la conclusión hasta que más adelante se revisen los resultados del cálculo de niveles de retorno con este modelo.

**Tabla 4.1. Resumen de estimación de parámetros**

	Localización $\mu$	Forma $\xi$	Escala $\sigma$	AIC	BIC	p-value	Likh. Ratio
GEVD MLE (30)	4,182	-0,257	0,688	83,10	87,85	-	-
GEVD L-mom (30)	4,154	-0,203	0,714	-	-	-	-
GEVD MLE - Gumbel (30)	4,093	-	0,623	82,59	85,75	0,223	1,483
GPD MLE (u=3,7)	-	-0,190	0,622	142,53	149,20	-	-
GPD L-mom (u=3,7)	-	-0,250	0,653	-	-	-	-
GEVD MLE (35)*	4,554	-0,229	0,414	45,92	50,59	-	-
GEVD L-mom (35)*	4,544	-0,201	0,423	-	-	-	-
GEVD MLE - Gumbel (35)*	4,505	-	0,393	46,35	49,46	0,119	2,430

\*Corresponde a los bloques realizados por año de medición: 35 bloques en el periodo 1980 - 2017

Al realizar las estimaciones para la GPD, se determinó que el umbral adecuado para este modelo es de  $u = 3.7$ , que fue aplicado con los dos métodos de estimación de MLE y L-Momentos. En ambos resultados, que se muestran en las filas 4 y 5 de la tabla 4.1, se nota una diferencia significativa, por simple observación, en el parámetro de Forma  $\xi$ . El AIC y BIC también se muestran para cuando la estimación se realizó por MLE.

Los tres últimos resultados de la tabla 4.1 corresponden al modelo de la GEVD, sosteniendo el criterio de realizar tantos bloques como años de observaciones, resultando  $nb = 35$  bloques en el periodo de 1980 hasta 2017 (1983 y 1992 no tienen observaciones registradas). Los tres parámetros estimados de este modelo son muy parecidos para los dos métodos, validando que, para este caso, la estimación podría ser realizada indistintamente por las dos vías, MLE o L-Momentos. De acuerdo con los índices de bondad de ajuste, sigue siendo mejor modelo el del criterio GEVD de bloques por años de observaciones  $nb = 35$  bloques.

#### 4.1. Niveles de retorno usando la GEVD

Una vez que se han estimado los parámetros mediante los métodos tratados en el apartado anterior, es posible calcular los niveles de retorno de sismos. Los niveles de retorno, o periodo de retorno, indican el tiempo en años en el que se espera que ocurra al menos un evento sísmico que supere el nivel de referencia. Para la GEVD, cuando  $\xi \neq 0$ , el nivel de retorno  $1/p$  se determina mediante (Coles, 2001):

$$x_m = \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}]. \quad (15)$$

Para obtener las estimaciones de los niveles de retorno, se reemplazan los parámetros de Localización, Forma y Escala en la ecuación (15). En la figura 4.1 se grafican los niveles de retorno (magnitud esperada) con los errores estándares para diferentes números de bloques usando la metodología de MLE para un periodo de retorno fijo de 1000 años. Se confirma que el tamaño adecuado para GEVD es 30 tomando en consideración que los errores estándares son mucho menores que los bloques iniciales.

Los errores de  $\hat{x}_m$  se obtienen mediante el método Delta:

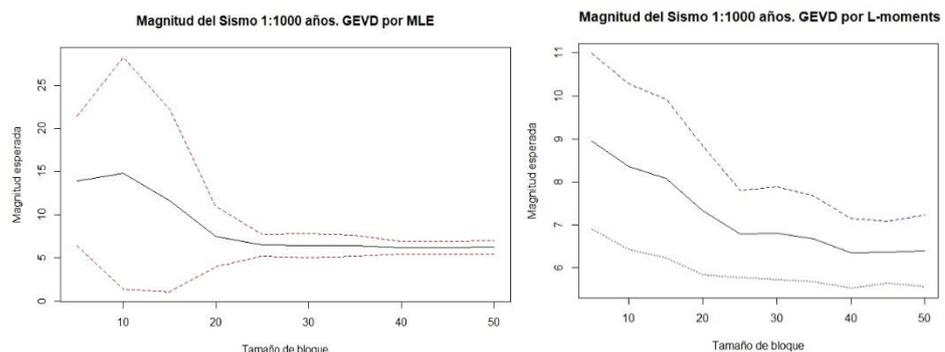
$$Var(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m, \quad (16)$$

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

$$\nabla x_m^T = \left[ \frac{\partial x_m}{\partial \mu}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \xi} \right] = [1, -\xi^{-1}(1 - y_p^{-\xi}), \sigma \xi^{-2}(1 - y_p^{-\xi}) - \sigma \xi^{-1} y_p^{-\xi} \log y_p], \quad (17)$$

donde  $y_p = -\log(1 - p)$  y evaluado en  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ . En la figura 4.1 también se presenta la gráfica de nivel de retorno calculado con diferentes números de bloques por el método de L-momentos.

**Figura 4.1. Niveles de retorno con tamaño de bloques para la GEVD por MLE y L-Momentos**



El procedimiento se puede aplicar realizando los cálculos con las ecuaciones (15) a (17), a través del paquete de R *extRemes*.

Con el método de L-Momentos también se puede verificar que, con un número de bloques de 30, el nivel de retorno tiende a estabilizarse. Por lo que es el número de bloques seleccionado atendiendo además al intervalo de confianza 95%. Si tomamos en cuenta las escalas de los gráficos de la figura 4.1, se puede verificar que el método de L-Momentos presenta una mejor estabilidad en cuanto al nivel de predicción, tomando también en consideración el error estándar representado en los intervalos de 95% de confianza graficados.

**Tabla 4.2. Resumen de Niveles de Retorno Sismos Guayas y Santa Elena: GEVD y GPD**

	<b>GEVD MLE nb = 30</b>	<b>GEVD Lmom nb = 30</b>	<b>Gumbel* MLE nb = 30</b>	<b>GPD MLE u = 3,7</b>
2-year return level	4,42	4,41	4,32	5,70
5-year return level	5,04	5,08	5,03	5,90
10-year return level	5,36	5,44	5,49	6,03
20-year return level	5,61	5,75	5,94	6,15
50-year return level	5,88	6,08	6,52	6,28
100-year return level	6,04	6,29	6,96	6,36
500-year return level	6,32	6,67	7,96	6,52
1000-year return level	6,41	6,80	8,40	6,58
	<b>GPD Lmom u = 3,7</b>	<b>GEVD MLE** nb = 35</b>	<b>GEVD Lmom** nb = 35</b>	<b>Gumbel MLE** nb = 35</b>
2-year return level	5,56	4,70	4,69	4,65
5-year return level	5,71	5,08	5,09	5,09
10-year return level	5,81	5,28	5,31	5,39
20-year return level	5,89	5,44	5,49	5,67
50-year return level	5,97	5,62	5,69	6,04
100-year return level	6,03	5,73	5,81	6,31
500-year return level	6,12	5,92	6,05	6,95
1000-year return level	6,15	5,99	6,12	7,22

\*Estimación de Nivel de retorno con el Modelo de Gumbel

\*\*Corresponde a los bloques realizados por año de medición:  
35 bloques en el periodo 1980 - 2017

Adicionalmente, se realizó el cálculo de nivel de retorno de las máximas magnitudes de sismos que se espera que ocurran por periodo de tiempo, asimismo con los métodos de MLE y L-Momentos. En la tabla 4.2 se resumen todos los niveles de retorno para los modelos y las metodologías causa de este estudio.

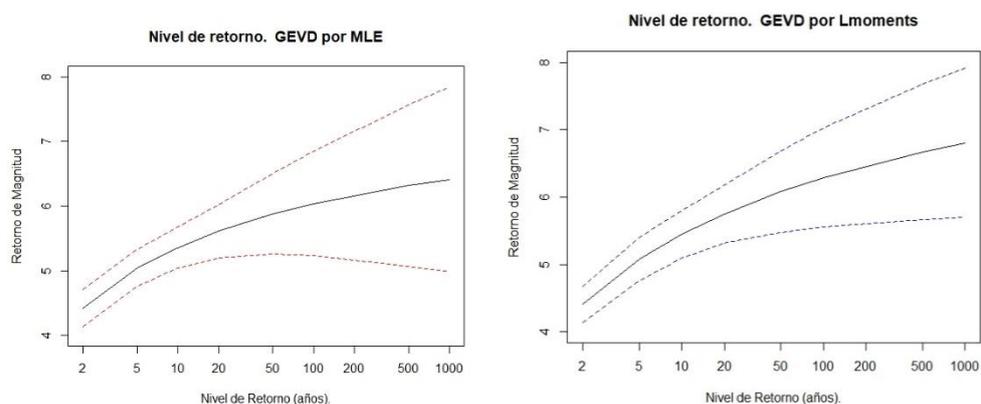
Las columnas 1 y 2 se presentan los niveles de retorno para diferentes intervalos de tiempo, para la GEVD por MLE y L-momentos. Dado que  $\hat{\xi} < 0$  es posible hacer inferencia del punto final (superior) de la distribución, que es el periodo de retorno infinito correspondiente a  $x_m$  con  $p = 0$ :  $x_m = \mu - \sigma/\xi$  a partir de (15), obteniéndose una distribución acotada (Coles, 2001).

La variación entre los valores de niveles de retorno calculados por los dos métodos se debe a las pequeñas diferencias en la estimación de los parámetros.

En el anexo E se puede ver la tabla 4.2 de niveles de retorno extendida, donde se incluyen los intervalos de confianza del 95%.

En la figura 4.2 se recogen las gráficas de nivel de retorno para los dos métodos. El eje de las abscisas está en escala logarítmica correspondiente a los periodos de retorno.

**Figura 4.2. Niveles de retorno para la GEVD por los métodos de MLE y L-Momentos usando  $nb = 30$**



Las estimaciones de niveles de retorno por ambos métodos fueron muy cercanas, especialmente en los periodos iniciales; para el periodo de 1000 años, las magnitudes fueron de 6.4 y 6.8, redondeando a un decimal, (véase tabla 4.2). Los errores estándares como se mencionó, se pudieron obtener con el método Delta para el caso MLE, sin embargo, por esta vía no se puede determinar para L-Momentos. En este último caso, se tuvo que usar el paquete R *extRemes* que aplica un método numérico (iterativo).

El periodo de retorno de 50 años suele ser de gran importancia para realizar muchas predicciones en el campo de la construcción, inversión de bienes raíces o también en decisiones de crecimiento urbanístico en muchas ciudades. Normativas sismos resistentes suelen adoptar un tiempo de exposición de 50 años; por ejemplo, en la española se considera una probabilidad de excedencia de un 10% en un periodo de exposición de 50 años y para estructuras de importancia pública como hospitales, escuelas, municipios, etc., una probabilidad de excedencia de 5% en 50 años de exposición (Rivas, 2014). En este caso, se espera que, en la ventana de tiempo de 50 años, se tenga al menos un sismo con magnitud entre 5.9 o 6.1 (ver tabla 4.2) en las provincias de Guayas y Santa Elena, según el método observado.

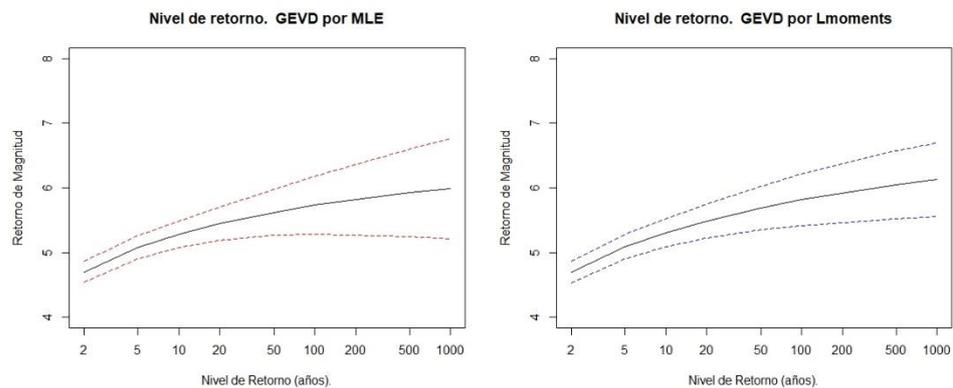
El modelo que mejor ajuste presentó fue el de bloques por año de observaciones ( $nb = 35$ ), basados en los AIC y BIC respectivos (ver en la tabla 4.1). También se construyó la gráfica de nivel de retorno con este modelo, tanto con el método de MLE como de L-Momentos (figura 4.3). En los primeros periodos de tiempo usados, los resultados obtenidos por MLE y L-Momentos son casi iguales. Para el periodo de 1000 años las magnitudes son de 6.0 y 6.1 (tabla 4.2).

Gracias al uso del paquete R *extRemes* se pudo realizar la determinación de los errores estándares y por tanto las gráficas de intervalos de confianza del 95%. Para el caso de L-Momentos se

obtuvieron intervalos ligeramente más estrechos que con el método de MLE, es decir presentaron menor varianza.

En un lapso de 50 años, se espera que ocurra al menos un sismo de 5.6 o 5.7 (tabla 4.2) según el método observado MLE o L-Momentos respectivamente, en comparación con los obtenidos con bloques  $nb = 30$ : 5.9 o 6.1, vemos que las predicciones con el modelo con número de bloques  $nb = 35$  son menores. Una segunda observación respecto a la comparación con el modelo de  $nb = 30$ , es que los intervalos de confianza son mucho más angostos, indicando menor error en la estimación.

**Figura 4.3. Niveles de retorno para la GEVD por MLE y L-momentos, usando bloques por año de observación de  $nb = 35$**

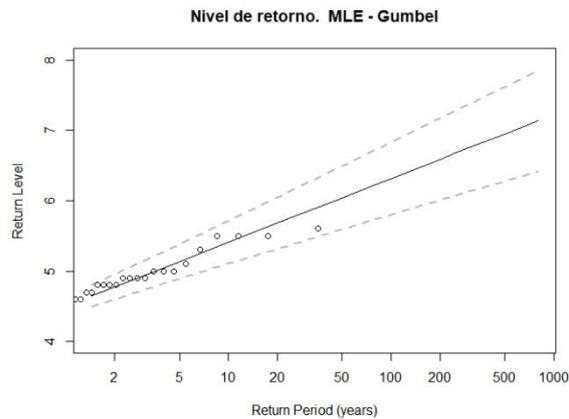


En la tabla 4.2 también se presentaron las predicciones de niveles de retorno para el modelo de Gumbel con  $nb = 30$ . En los periodos de tiempo iniciales, los niveles de retorno son muy similares a los de GEVD tanto con los métodos de MLE y L-Momentos, pero, a partir del periodo de 50 años, la diferencia se hace muy visible, teniendo valores más altos las predicciones del modelo Gumbel (6.5 vs 5.9 y 6.1). Si se siguen analizando los periodos incrementales, vemos que, para el intervalo de tiempo de 1000 años, la diferencia ya es significativa, con un registro de 8.4 vs 6.4 y 6.8. Se pueden hacer estimaciones con periodos más altos, por ejemplo, si el tiempo es  $1E16$  años, el nivel de retorno por da un valor de 27.0 vs 6.9 y 7.7 de los modelos de GEVD por MLE y L-

Momentos respectivamente, indicando una distorsión significativa de la predicción debido a errores de estimación que en este modelo serían mayores cuando el tiempo tiende al infinito.

El mismo ejercicio, con Gumbel, se realizó para el modelo con  $nb = 35$ , donde los resultados se pueden ver en la tabla 4.2. Para el periodo de retorno de 1000 años, los niveles son de 7.2 vs 6.0 y 6.1 mientras que, si el intervalo de tiempo se incrementa, por ejemplo, a 1E16 años, el nivel de retorno obtenido es de 19.0 vs 6.4 y 6.6 para los métodos de MLE y L-Momentos. En la figura 4.4 se presenta la gráfica de nivel de retorno para el Modelo de Gumbel.

**Figura 4.4. Niveles de retorno para la distribución de Gumbel por MLE usando bloques por año de observación,  $nb = 35$**



#### 4.2. Ajuste del modelo para la GEVD

Aunque es imposible realizar una validación de la extrapolación de un modelo GEVD, sí se puede realizar una evaluación del ajuste en base a los datos observados o existentes. Para este efecto se cuenta con los gráficos de probabilidad y de cuantiles. El gráfico de probabilidad es una comparación de la distribución empírica con la función de distribución ajustada. La distribución empírica ajustada evaluada en  $z_{(i)}$  está dada por:

$$\hat{F}(z_{(i)}) = \frac{i}{m+1}, \quad (18)$$

donde  $z_{(i)} \leq \dots \leq z_{(m)}$  son los valores máximos ordenados de cada bloque. Por sustitución de los parámetros estimados en la función del modelo GEV, se obtienen las estimaciones del modelo ajustado:

$$\hat{G}(z_{(i)}) = \exp \left\{ - \left[ 1 + \hat{\xi} \left( \frac{z_{(i)} - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}} \right\}. \quad (19)$$

Si el modelo GEV está funcionando bien en sus estimaciones,

$$\hat{F}(z_{(i)}) \approx \hat{G}(z_{(i)}), \quad (20)$$

para cada  $i$ , de tal manera que el gráfico de probabilidades consistirá en los siguientes puntos:

$$\left\{ \left( \hat{F}(z_{(i)}), \hat{G}(z_{(i)}) \right), i = 1, \dots, m \right\} \quad (21)$$

El gráfico de cuantiles evalúa la exactitud para valores grandes de  $z_{(i)}$  de mejor manera que el gráfico de probabilidades, debido a que ambos,  $\hat{F}(z_{(i)})$  y  $\hat{G}(z_{(i)})$  se aproximan a 1 para valores más altos de  $z_{(i)}$ . El gráfico de cuantiles es generado por los siguientes puntos:

$$\left\{ \left( \hat{G}^{-1}(i/(m+1)), z_{(i)} \right), i = 1, \dots, m \right\}, \quad (22)$$

donde tenemos,

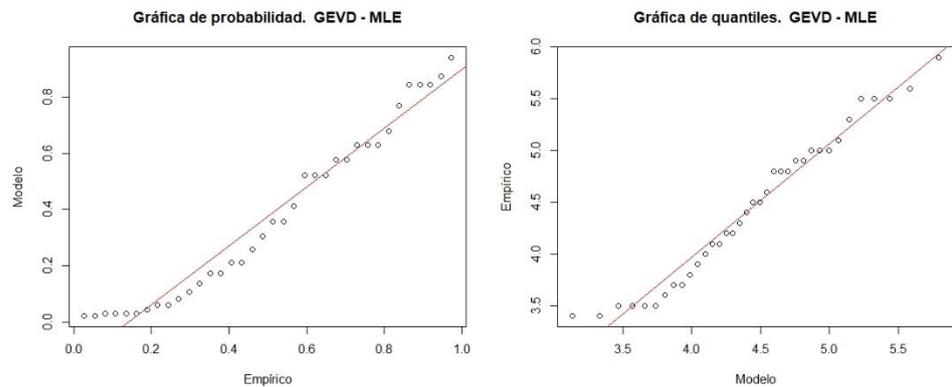
$$\hat{G}^{-1} \left( \frac{i}{m+1} \right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[ 1 - \left\{ -\log \left( \frac{i}{m+1} \right) \right\}^{-\hat{\xi}} \right]. \quad (23)$$

Los dos gráficos deben ser lineales con los puntos de probabilidad localizadas cerca de la diagonal de pendiente igual a 1 y el de cuantiles con comportamiento lineal (Coles, 2001).

Los gráficos correspondientes se presentan en la figura 4.5. Es evidente que la mayoría de los puntos se aproximan a la diagonal, con lo cual se

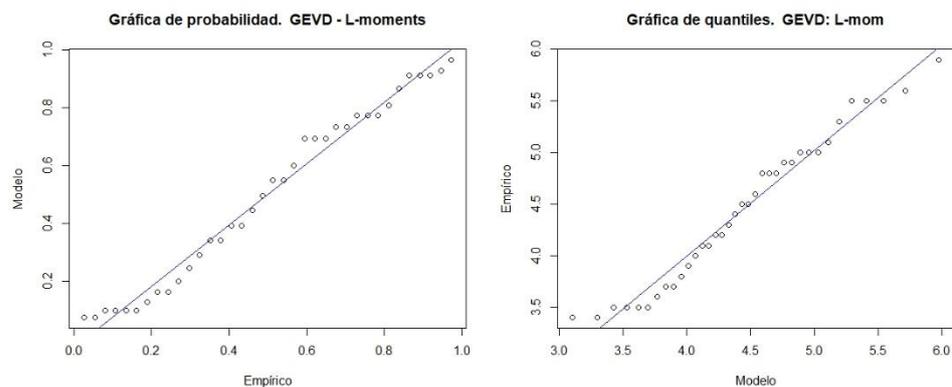
verifica el ajuste del modelo, en este caso GEVD por MLE, sin embargo, también se evidencia dispersión que vale tener presente en el análisis, especialmente en los primeros valores. En el caso del gráfico de cuantiles se tiene una correlación de  $R^2 = 0.9725$ , mostrando un buen ajuste.

**Figura 4.5. Gráficos de probabilidad y de cuantiles para la GEVD por MLE y  $nb = 30$**

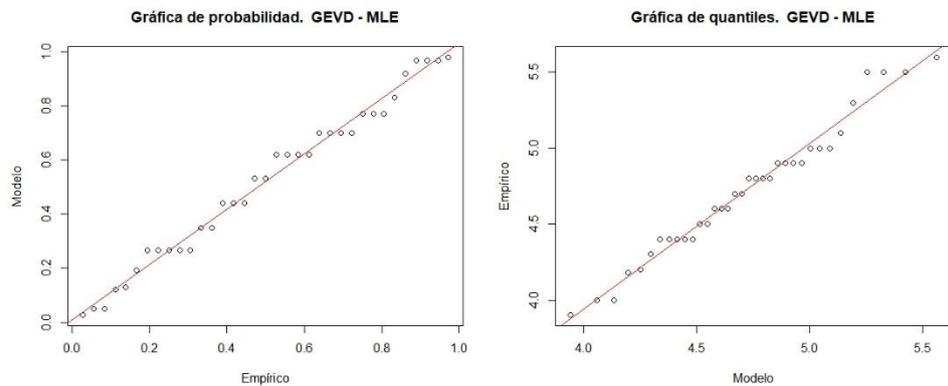


También se construyeron los gráficos de probabilidad y de cuantiles para las estimaciones realizadas por el método de L-Momentos. En la Figura 4.6, se evidencia un buen ajuste ya que la mayoría de los puntos están cerca de la diagonal, lo cual da validez también al modelo obtenido con este método con una correlación para cuantiles de  $R^2 = 0.9734$ . Se nota un mejor ajuste que en caso de MLE.

**Figura 4.6. Gráficos de probabilidad y de cuantiles para la GEVD por L-Momentos y  $nb = 30$**

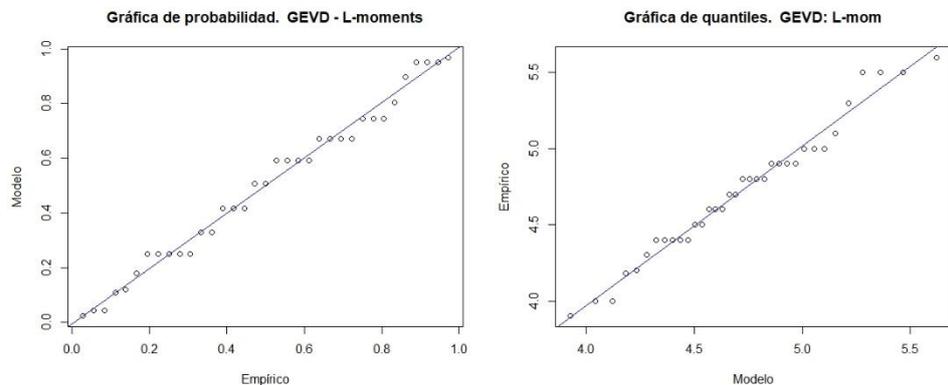


**Figura 4.7. Gráficos de probabilidad y de cuantiles para la GEVD por MLE y  $nb = 35$**



El modelo que mejor ajuste evidenció fue el que tiene número de bloques  $nb = 35$ , tomando cada año de observaciones como un bloque independiente; las gráficas de probabilidades y cuantiles se presenta en la figura 4.7. El ajuste del gráfico de probabilidades se acerca más al ideal de pendiente igual a 1. Por otro lado, el gráfico de cuantiles tiene un coeficiente de determinación  $R^2 = 0.9796$ . Es importante observar que, el ajuste es evidentemente mejor que el anterior modelo probado con  $nb = 30$ . Asimismo, se usaron las estimaciones de L-Momentos con el modelo de  $nb = 35$  generándose las gráficas de la figura 4.8.

**Figura 4.8. Gráficos de probabilidad y de cuantiles para la GEVD por L-Momentos y  $nb = 35$**



Comparando con el método de MLE, prácticamente no se ven diferencias. La mayoría de los puntos tienden a la diagonal en el caso del gráfico de probabilidad acercándose a recta identidad, por lo que se valida las estimaciones del método. No hay evidencias estadísticas de

mal ajuste del modelo; por otro lado, el ajuste del gráfico de cuantiles presenta un valor  $R^2$  de 0.9799.

Además, al poner en contraste con el modelo  $nb = 30$ , se llega a la misma conclusión. Se nota un mejor ajuste del modelo  $nb = 35$ , lo cual ya se evidenció al obtener menores valores de AIC y BIC, discutido al inicio de este capítulo. En los anexos F y G se muestra el gráfico completo de los modelos GEVD realizados con el paquete R *extRemes*, para los métodos de MLE y L-Momentos. Además, las gráficas de los modelos con el número de bloques  $nb = 35$  se pueden revisar en los anexos J y K.

#### 4.3. Niveles de retorno para la GPD

Tal como se realizó con las estimaciones para la GEVD, con modelo Pareto también es necesario verificar los niveles de retorno. Suele ser más conveniente conocer las probabilidades de ocurrencia de un sismo a una magnitud específica que simplemente obtener los valores individuales de los parámetros. Para el modelo de Pareto Generalizado, la literatura establece que para una variable  $X$ :

$$Pr\{X > x\} = \zeta_u \left[ 1 + \xi \left( \frac{x-u}{\sigma} \right) \right]^{-1/\xi}, \quad (24)$$

donde  $\zeta_u = Pr\{X > u\}$  y  $u$  es el umbral. La ecuación (24) se le puede asignar la probabilidad  $1/m$  de cuan frecuente puede ocurrir un evento sísmico con magnitud mayor a  $x_m$  y que excede en promedio al menos una vez cada  $m$  observaciones mediante:

$$\frac{1}{m} = \zeta_u \left[ 1 + \xi \left( \frac{x_m-u}{\sigma} \right) \right]^{-1/\xi}, \quad (25)$$

Rearreglando (25) para la máxima magnitud esperada  $x_m$  dado un cierto número de sismos, se obtiene

$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u)^\xi - 1]. \quad (26)$$

De la data disponible, es posible saber cuántos sismos ocurren en promedio, permitiendo que se pueda calcular la magnitud de un sismo en un determinado periodo de tiempo para algún  $x_m > u$ .

Los parámetros  $\xi$  y  $\sigma$  se estiman por los métodos ya revisados, mientras que  $\zeta_u$  tiene el siguiente estimador natural:

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

que es la proporción de la muestra de puntos que exceden al umbral  $u$ . Dado que el número de excedencias de  $u$  sigue una distribución Binomial  $(n, \zeta_u)$ ,  $\hat{\zeta}_u$  es también la estimación de máxima verosimilitud de  $\zeta_u$  (Coles, 2001). De las propiedades de la distribución binomial se tiene que  $Var(\hat{\zeta}_u) \approx \hat{\zeta}_u(1 - \hat{\zeta}_u)/n$ , y la matriz de varianzas y covarianzas para  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$  es aproximadamente:

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

Donde  $v_{i,j}$  denota los términos  $(i,j)$  de la matriz de varianzas y covarianzas de  $\hat{\sigma}$  y  $\hat{\xi}$ . Por lo tanto, por el método Delta:

$$Var(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m, \quad (27)$$

donde  $V$  es la matriz de varianzas y covarianzas de  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$  y

$$\nabla x_m^T = \left[ \frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \xi} \right] = \left[ \sigma m^\xi \zeta_u^{\xi-1}, \xi^{-1} \{ (m\zeta_u)^\xi - 1 \}, -\sigma \xi^{-1} \{ (m\zeta_u)^\xi - 1 \} + \sigma \xi^{-1} (m\zeta_u)^\xi \log(m\zeta_u) \right]. \quad (28)$$

Normalmente es más conveniente presentar los niveles de retorno en unidades de tiempo, en escala anual; la ecuación (26) se puede escribir en términos de  $m = N \times n_y$ , donde  $N$  es el número de años y  $n_y$  es el número de observaciones por año:

$$x_m = u + \frac{\sigma}{\xi} \left[ (N n_y \zeta_u)^\xi - 1 \right]. \quad (29)$$

Bommier difiere con Coles en el tema de la función de distribución de  $\zeta_u$ , pues sugiere que si se asume que las excedencias luego del umbral  $u$  son eventos poco frecuentes, se espera que  $\zeta_u$  siga la distribución de Poisson (Bommier, 2014). La distribución de Poisson es caracterizada por el parámetro  $\lambda$ , que corresponde a la media de las excedencias del umbral seleccionado por unidad de tiempo:

$$\hat{\zeta}_u = \frac{\hat{\lambda}}{n_y},$$

donde el estimador insesgado  $\lambda$  se determina mediante:

$$\hat{\lambda} = \frac{n_u}{M},$$

con  $n_u$  como el número de excedencias sobre el umbral seleccionado  $u$  y  $M$  el número de años de registro. Con estas consideraciones, y si  $\xi \neq 0$  la ecuación (29) se puede reescribir de la forma

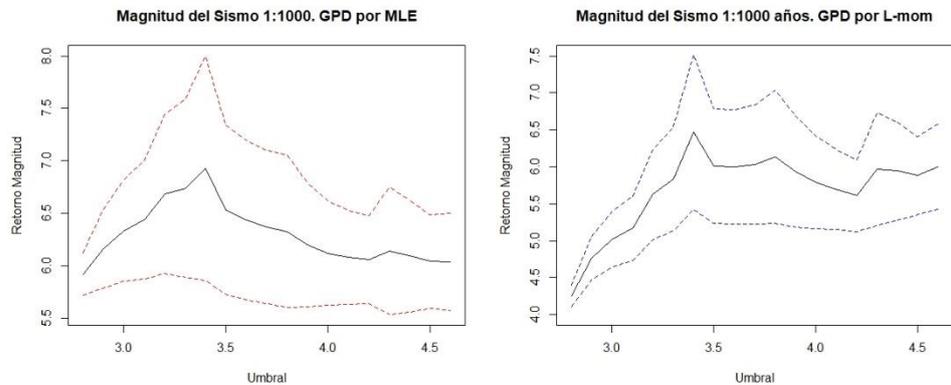
$$x_m = u + \frac{\sigma}{\xi} \left[ (\lambda N)^\xi - 1 \right]. \quad (30)$$

Con las estimaciones de parámetros realizadas al inicio del capítulo y mediante la aplicación de (29), se obtuvieron las gráficas de la Figura 4.9, correspondiente al nivel de retorno para diferentes umbrales con los respectivos intervalos de confianza del 95%, los cuales se obtuvieron

mediante el cálculo de la varianza de los niveles de retorno, por supuesto asumiendo una aproximación a la normalidad.

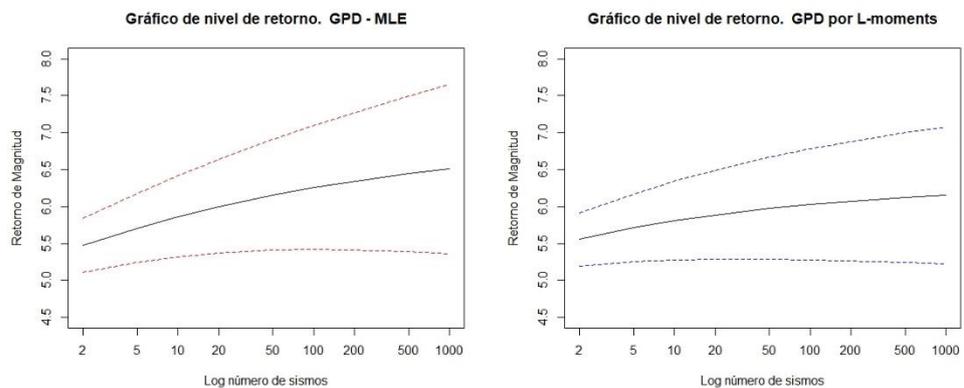
La varianza es calculada mediante la aplicación del método Delta a la matriz de varianzas y covarianzas de los parámetros estimados que se usaron para aplicar (29). También se usó el paquete R *extRemes* obteniéndose los mismos resultados. Las determinaciones también se hicieron por los métodos de MLE y L-Momentos.

**Figura 4.9. Niveles de retorno con diferentes umbrales para la GPD por MLE y L-momentos**



En las dos gráficas se puede observar que los niveles de retorno se estabilizan alrededor del umbral  $u = 3.7$ . Con el método de MLE no es tan evidente a primera vista como sí lo es con la gráfica de L-Momentos, por lo que se recomienda revisar los valores individuales en los anexos C y D.

**Figura 4.10. Niveles de retorno para la GPD por MLE y L-momentos**



Luego de fijar el umbral adecuado en  $u = 3.7$ , con el uso de (29) a diferentes periodos de tiempo, se obtuvieron las gráficas de nivel de retorno que son mostradas en la figura 4.10. Los valores individuales se presentaron en la tabla 4.2 con los dos métodos MLE y L-Momentos.

Para los primeros periodos de retorno, las magnitudes calculadas no tienen diferencia significativa, sin embargo, para los periodos finales de tiempo tenemos diferencias considerables, siendo el mayor valor para el método MLE con 6.6 y de 6.2 para el retorno por L-Momentos. Mediante la aplicación del método Delta se obtuvieron los errores estándares con los cuales se construyeron los intervalos del 95% de confianza, que también se presentan en las gráficas, observándose que con el método de L-Momentos se obtienen un menor error de predicción.

En un periodo de tiempo de 50 años se espera que ocurra al menos un sismo de 6.3 o 6.0 (tabla 4.2) según el método escogido, en comparación con los obtenidos con el modelo de número de bloques  $nb = 30$ : 5.9 o 6.1, se observa que las predicciones con el modelo GEVD son menores.

#### 4.4. Ajuste del modelo para la GPD

Los gráficos de probabilidad y de cuantiles, igual que en el apartado 4.2, también serán usados para comprobar la bondad de ajustes de los modelos Pareto obtenidos. Se espera que en ambos gráficos tiendan a la diagonal mostrando así que el modelo tiene buen ajuste. Asumiendo un umbral  $u$  y las excedencias  $y_{(1)} \leq \dots \leq y_{(k)}$  y un modelo de estimación  $\hat{H}$ , el gráfico de probabilidad consiste de los siguientes pares:

$$\{(i/(k+1), \hat{H}(y_{(i)})) ; i = 1, \dots, k\}, \quad (31)$$

donde

$$\hat{H}(y) = 1 - \left(1 + \frac{\xi y}{\hat{\sigma}}\right)^{-\frac{1}{\xi}}, \quad (32)$$

siendo  $\hat{H}$  es evaluado en  $(\sigma, \xi) = (\hat{\sigma}, \hat{\xi})$ . Asumiendo que  $\hat{\xi} \neq 0$ , el gráfico de cuantiles está compuesto por los siguientes pares:

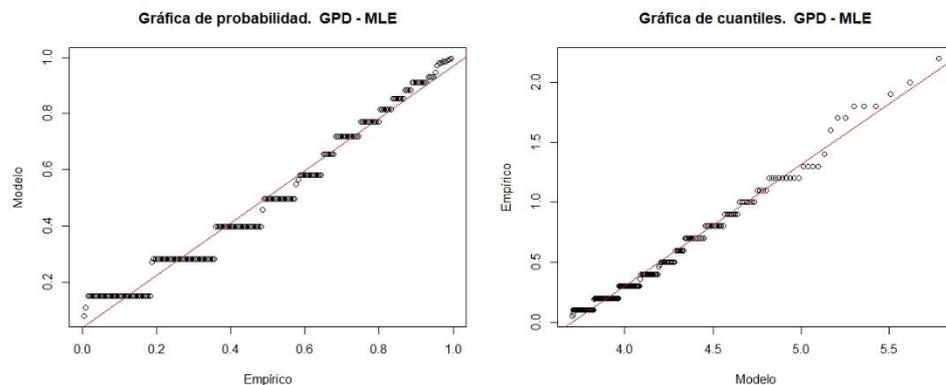
$$\{(\hat{H}^{-1}(i/(k+1)), y_{(i)}); i = 1, \dots, k\}, \quad (33)$$

donde

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}}(y^{-\hat{\xi}} - 1). \quad (34)$$

Los dos gráficos deben ser aproximadamente lineales si el modelo de Pareto generalizado es un modelo razonable para modelar las excedencias sobre el umbral  $u$ .

**Figura 4.11. Gráficos de probabilidad y de cuantiles para la GPD por MLE**

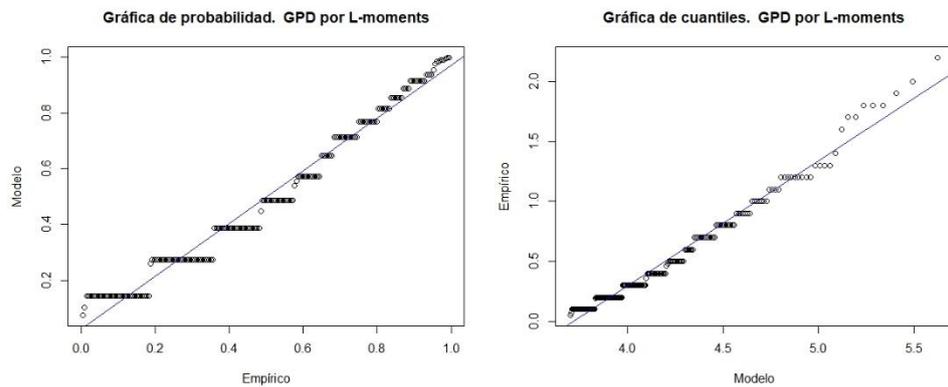


En la figura 4.11 se presentan los dos gráficos para el modelo GPD usando el método de máxima verosimilitud. En el de probabilidad, la tendencia de los puntos es acercarse a la línea con pendiente unitaria, mostrando por lo tanto un excelente ajuste. Asimismo, el gráfico de cuantiles muestra una tendencia con relación fuerte relación lineal ( $R^2=0.9871$ ). Los dos gráficos no muestran señales de falta de ajuste usando la GPD.

La misma rutina se realizó con las estimaciones obtenidas por el método de L-Momentos, con resultados mostrados en la Figura 4.12, muy

parecidos a los obtenidos con MLE. Los puntos de probabilidad en su mayoría se acercan a la diagonal unitaria y en el caso de los cuantiles, tenemos la mayor parte de los pares altamente correlacionados ( $R^2=0.9813$ ).

**Figura 4.12. Gráficos de probabilidad y de cuantiles para la GPD por el método de L-Momentos**



En los anexos H y I se muestra el gráfico completo de los modelos GPD realizados con el paquete R *extRemes*, usando los métodos de MLE y L-momentos.

# CAPÍTULO 5

## 5. CONCLUSIONES Y RECOMENDACIONES.

### *Conclusiones*

- Con el análisis descriptivo se determina que la sismicidad de las provincias de Guayas y Santa Elena está por debajo al promedio nacional en cuanto a magnitud.
- La distribución de los datos sísmicos del catálogo para las provincias del estudio presenta un marcado sesgo hacia la derecha, por lo tanto, para modelar su comportamiento se debe usar una distribución de probabilidades diferente a la normal.
- Mediante la prueba Aumentada de Raíz Unitaria de Dickey-Fuller o ADF se verifica la estacionariedad de los datos, por lo tanto, los modelos estadísticos empleados están validados. Esta conclusión se refuerza con la construcción de la gráfica de PACF, donde las excedencias en su mayoría caen dentro de los límites.
- En la selección del número de bloque máximo para la GEVD, se determina que el parámetro de forma se estabiliza con un número de bloques  $nb = 30$ , por lo que de acuerdo con el criterio de (Coles, 2001), este es el número óptimo de bloques con el cual se debe trabajar en la estimación de los parámetros.
- Los parámetros para la GEVD se obtienen mediante los métodos de Máxima Verosimilitud y L-Momentos mostrando gran similitud en los 3 valores, se concluye entonces que la metodología de L-Momentos es una de las vías alternativas para la estimación de los parámetros de este modelo.
- Al comparar los modelos realizados con la GEVD para los sistemas con número de bloques de 30 y 35, este último haciendo que cada bloque máximo represente cada año de observaciones, se demuestra que el modelo con número de bloques de 35 tiene mejor ajuste, esto gracias al cálculo de los índices de AIC y BIC.

- Con el modelo GPD se determina que el umbral óptimo es de 3.70 dado a que el parámetro de forma  $\xi$  se mantiene constante alrededor de este valor, mientras que el parámetro de escala  $\sigma$  se comporta lineal, por lo tanto, alineado con el criterio expuesto en (Coles, 2001), se justifica el valor del umbral seleccionado.
- Se demuestra además que el parámetro de forma es estadísticamente igual a 0 por lo que el modelo de Gumbel es el más apropiado que el GEVD para la representación de la sismicidad de las zonas en estudio, sin embargo, hay que analizar paralelamente los niveles de retorno calculados con este modelo.
- Los parámetros de forma y escala estimados con el modelo GPD, con dos métodos, máxima verosimilitud y L-Momentos, tienen mucha similitud, por lo que estos métodos se pueden usar indistintamente para la estimación, sin embargo, atendiendo los valores de AIC y BIC, el modelo GEVD por máxima verosimilitud es mejor desde el punto de vista de la bondad de ajuste.
- Según Coles (Coles, 2001), el parámetro de Forma, estimado tanto con GEVD como con GPD tienden a converger, este resultado se comprueba en la estimación del parámetro en mención con los modelos propuestos y por los dos métodos: máxima verosimilitud y L-Momentos, teniendo un valor estimado alrededor de -0.222.
- La regresión Poisson puede usarse con la misma validez que la regresión lineal clásica para modelar la relación entre la frecuencia de los sismos y su magnitud conocido como modelo Gutenberg-Richter, presentando una RMSE de 13.00 y 12.51 respectivamente.
- En el modelo GEVD con número de bloques de 30 se determina que las predicciones de niveles de retorno con los dos métodos, máxima verosimilitud y L-Momentos, dan resultados muy similares, con variaciones sutiles que corresponden a las diferencias de las estimaciones de los parámetros.
- Se comprueba que el modelo GEVD con número de bloques de 35 es el modelo de mejor ajuste, gracias a los valores obtenidos de AIC y BIC, y por lo tanto es el mejor modelo para realizar las predicciones.

Los niveles de retorno con MLE y L-Momentos, son prácticamente iguales con diferencias que corresponden a las variaciones de la estimación de los parámetros.

- Las predicciones del modelo Gumbel para periodos de retorno altos, tendientes al infinito, muestran una desviación significativa en contraste a los resultados del modelo GEVD, teniendo así predicciones de 27.0 vs 6.9 y 7.7 (máxima verosimilitud y L-Momentos) como resultado de probar los modelos con periodo de 1E16 años. El modelo de Gumbel, en este caso, solo es adecuado para periodos de retornos bajos, máximo 50 años.
- Los niveles de retorno calculados con la GPD, con los dos métodos máxima verosimilitud y L-Momentos, tiene resultados similares, las variaciones presentadas están asociadas a las diferencias en la estimación de los parámetros, por lo que, los dos métodos son válidos. Sin embargo, al comparar los resultados con el modelo GEVD con número de bloques de 35, se determina que el modelo GPD sobreestima los valores de niveles de retorno.
- Las predicciones de niveles de retorno con el método L-Momentos tienen menor error estándar comparadas con los resultados del método de máxima verosimilitud, en los dos modelos estudiados: GEVD y GPD.
- En cuanto a las evaluaciones de ajuste, con las gráficas de probabilidad y cuantiles, se comprueba que ambas herramientas son aplicables para determinar la bondad de ajuste, siendo el modelo con número de bloques de 35 el que tiene mejor ajuste -gráfico de cuantiles con un  $R^2 = 0.9799$ . Frente a la imposibilidad de calcular AIC y BIC para el método de L-Momentos, se evidencia que las gráficas de probabilidad y cuantiles son un método válido para la evaluación del ajuste.

## ***Recomendaciones***

- Extender el estudio para los datos a nivel nacional con la aplicación del método de L-Momentos.
- Es importante profundizar el estudio de las implicaciones de las predicciones de nivel de retorno del modelo Gumbel en las investigaciones de sismología dada las observaciones realizadas en este estudio.
- Ampliar la presente investigación con la información sismológica actualizada al 2021. Hasta el cierre de esta investigación se esperó contestación del Instituto Geográfico respecto a la actualización del catálogo sin obtener respuesta favorable.

## 6. Referencias.

- Aras, A. (2019). Predicting Catastrophes: an Analysis of the Earthquake Data in Groningen using Extreme Value Theory. *Faculty of Science and Engineering, University of Groningen*, 25-35.
- Bayrak, Y. Ö. (2008). Seismicity assessment for the different regions in and around Turkey based on instrumental data: Gumbel first asymptotic distribution and Gutenberg-Richter cumulative frequency law. *Natural Hazards and Earth System.*, 109-122.
- Benito, B. J. (1999). Peligrosidad Sísmica. En B. J. Benito, *Física de la Tierra* (págs. 13-47). Madrid, España: EUIT de Topografía (UPM).
- Bommier, E. (2014). *Peaks-Over-Threshold Modelling of Environmental Data*. Uppsala, Suecia: Departament of Mathematics, Uppsala University.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Bristol: Springer.
- García, S. N. (2018). Statistical analysis of the largest possible earthquake magnitudes on the Ecuadorian coast for selected return periods. *Georisk, Taylor&Francis Group*, 3.
- García, S. N. (2021). Zoning of Ecuador According to Maximum Magnitudes of Earthquakes and their Frequency of Occurrence using Statistical Models Estimated by Maximum Likelihood . *Journal of Science*.
- Hosking, J. R. (1990). *L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics*. Yorktown: Royal Statistical Society.
- INEC, I. N. (25 de Febrero de 2021). *INEC*. Obtenido de INEC: <https://www.ecuadorencifras.gob.ec/institucional/home/>
- Instituto Geofísico, E. P. (2020, 08 20). *Instituto Geofísico, Escuela Politécnica Nacional*. Retrieved from Instituto Geofísico, Escuela Politécnica Nacional: <http://www.igepn.edu.ec>
- Lev A. Maslov, V. M. (2012). Derivation of the Gutenberg-Richter empirical formula from the solution of the generalized logistic equation. *Natural Science*.

- Montesinos, A. (2011). Estudio de AIC y BIC en la Selección de Modelos de Vida con Datos Censurados. *Centro de investigación en Matemáticas, CIMAT*, 9-12.
- Muñoz, D. (1989). Conceptos básicos en riesgo sísmico. En D. Muñoz, *Física de la Tierra* (págs. 199-215). Madrid, España: Universidad Complutense.
- Rivas, A. (2014). *Contribución Metodológica Para Incorporar Fallas Activas en la Modelización de la Fuente Dirigida a Estimaciones de Peligrosidad Sísmica. Aplicación al Sur de España*. Madrid: Universidad Politécnica de Madrid.
- Seguí, P. (22 de 02 de 2021). *ovacen*. Obtenido de *ovacen*: <https://ovacen.com/evaluar-danos-casas-estructuras-sismo/>

## 7. Apéndices y anexos

### Anexo A. Tablas de modelo GEVD por MLE

**Tabla 1. Estimación de parámetros para diferentes números de bloques**

Tamaño de bloque	Localización $\mu$	Local. SE	Forma $\xi$	Forma SE	Escala $\sigma$	Escala SE	Nivel Ret. SE	AIC	BIC
5	3,418	0,030	0,342	0,079	0,374	3,836	2,862	339,68	350
10	3,551	0,055	0,317	0,140	0,449	6,877	5,112	208,69	217
15	3,695	0,091	0,198	0,177	0,535	5,427	2,653	157,73	165
20	3,868	0,106	-0,056	0,184	0,628	1,802	1,624	121,65	128
25	4,192	0,137	-0,225	0,150	0,644	0,676	0,412	93,73	99
30	4,182	0,143	-0,257	0,179	0,688	0,730	0,207	83,10	88
35	4,340	0,146	-0,253	0,175	0,621	0,663	0,272	66,06	70
40	4,393	0,146	-0,311	0,148	0,633	0,435	0,282	57,28	61
45	4,585	0,128	-0,329	0,171	0,577	0,417	0,145	46,81	50
50	4,538	0,151	-0,305	0,166	0,578	0,460	-	43,97	47
35*	4,554	0,079	-0,229	0,135	0,414	0,057	-	45,92	50,59

\*Corresponden a los bloques realizados por año de medición: 35 bloques en el periodo 1990 - 2017

## Anexo B. Tablas de modelo GEVD por L-Momentos

Tabla 2. Estimación de parámetros para diferentes números de bloques

Tamaño de bloque	Localización $\mu$	Local. SE	Forma $\xi$	Forma SE	Escala $\sigma$	Escala SE
5	3,452	0,029	0,165	0,063	0,427	0,027
10	3,611	0,056	0,067	0,084	0,541	0,045
15	3,750	0,076	0,000	0,097	0,628	0,060
20	3,872	0,108	-0,091	0,115	0,675	0,072
25	4,170	0,104	-0,181	0,116	0,664	0,088
30	4,154	0,119	-0,203	0,143	0,714	0,098
35	4,317	0,124	-0,211	0,164	0,649	0,095
40	4,378	0,125	-0,295	0,176	0,666	0,114
45	4,564	0,134	-0,291	0,154	0,606	0,100
50	4,519	0,133	-0,275	0,193	0,608	0,122
35*	4,544	-	-0,201	-	0,423	-

\*Corresponden a los bloques realizados por año de medición:  
35 bloques en el periodo 1990 - 2017

## Anexo C. Tablas de modelo GPD por MLE

Tabla 3. Estimación de parámetros para diferentes umbrales – MLE.

Umbral	Forma $\xi$	Escala $\sigma$	Forma SE	Escala SE	Escala Re-paramt.	Nueva Escala SE	VarCov sig/xi	AIC	BIC
1,90	-0,527	2,112	2,0E-08	5,9E-03	3,113	5,91E-03	-3,2E-15	2.585,49	2.595,42
2,00	-0,501	1,960	2,0E-08	6,4E-03	2,961	6,43E-03	-2,4E-15	2.481,60	2.491,53
2,10	-0,474	1,810	1,2E-03	9,2E-03	2,805	1,10E-02	-7,2E-06	2.369,36	2.379,28
2,20	-0,447	1,662	4,4E-03	0,020	2,645	2,97E-02	-8,4E-05	2.247,48	2.257,41
2,30	-0,418	1,517	8,1E-03	0,032	2,479	0,050	-2,5E-04	2.114,37	2.124,29
2,40	-0,389	1,374	0,011	0,038	2,307	0,063	-4,0E-04	1.967,98	1.977,91
2,50	-0,358	1,234	0,012	0,038	2,128	0,066	-4,2E-04	1.805,67	1.815,60
2,60	-0,324	1,095	0,012	0,034	1,938	0,065	-3,9E-04	1.623,86	1.633,79
2,70	-0,286	0,956	0,013	0,031	1,729	0,065	-3,6E-04	1.417,40	1.427,33
2,80	-0,238	0,814	0,016	0,028	1,481	0,069	-3,5E-04	1.178,13	1.188,05
2,90	-0,166	0,661	0,021	0,025	1,141	0,082	-3,9E-04	890,82	900,74
3,00	-0,141	0,612	0,025	0,025	1,034	0,095	-4,6E-04	683,96	693,61
3,10	-0,128	0,587	0,029	0,027	0,984	0,110	-5,6E-04	530,66	539,97
3,20	-0,097	0,544	0,034	0,028	0,853	0,131	-7,0E-04	391,79	400,77
3,30	-0,097	0,543	0,039	0,032	0,864	0,155	-9,3E-04	311,90	320,44
3,40	-0,078	0,521	0,047	0,035	0,786	0,188	-1,2E-03	237,56	245,69
3,50	-0,146	0,593	0,051	0,044	1,104	0,213	-1,7E-03	215,85	223,39
3,60	-0,170	0,611	0,056	0,051	1,221	0,244	-2,3E-03	176,48	183,57
3,70	-0,190	0,622	0,062	0,058	1,326	0,279	-2,9E-03	142,53	149,20
3,80	-0,206	0,625	0,070	0,064	1,407	0,320	-3,7E-03	113,46	119,72
3,90	-0,258	0,668	0,075	0,075	1,675	0,357	-4,7E-03	93,96	99,74
4,00	-0,300	0,695	0,080	0,085	1,897	0,397	-5,9E-03	75,96	81,31
4,10	-0,330	0,703	0,088	0,094	2,054	0,443	-7,2E-03	59,90	64,86
4,20	-0,353	0,700	0,096	0,102	2,184	0,495	-8,6E-03	46,30	50,88
4,30	-0,290	0,600	0,114	0,098	1,847	0,579	-9,7E-03	30,22	34,60
4,40	-0,338	0,620	0,129	0,114	2,105	0,669	-1,3E-02	23,16	27,06
4,50	-0,393	0,642	0,144	0,131	2,411	0,768	-1,7E-02	17,46	20,89
4,60	-0,400	0,612	0,166	0,140	2,453	0,895	-2,1E-02	11,36	14,41

## Anexo D. Tablas de modelo GPD por L-Momentos

Tabla 4. Estimación de parámetros para diferentes umbrales – L-Momentos.

Umbral	Forma $\xi$	Escala $\sigma$	Forma SE	Escala Re-paramt.	Escala SE	Nueva Escala SE	VarCov sig/xi
1,90	-5,069	9,384	0,591	17,876	1,405	3,205	-8,27E-01
2,00	-4,612	8,116	0,512	16,295	1,118	2,698	-5,70E-01
2,10	-4,154	6,939	0,439	14,714	0,873	2,227	-3,81E-01
2,20	-3,697	5,854	0,371	13,133	0,666	1,797	-2,46E-01
2,30	-3,240	4,860	0,309	11,552	0,495	1,412	-1,52E-01
2,40	-2,783	3,958	0,252	9,971	0,356	1,076	-8,89E-02
2,50	-2,326	3,147	0,201	8,390	0,247	0,790	-4,87E-02
2,60	-1,869	2,428	0,156	6,809	0,163	0,554	-2,46E-02
2,70	-1,411	1,800	0,116	5,228	0,101	0,368	-1,12E-02
2,80	-0,954	1,263	0,082	3,647	0,060	0,229	-4,40E-03
2,90	-0,497	0,818	0,054	2,066	0,034	0,136	-1,49E-03
3,00	-0,416	0,726	0,054	1,778	0,032	0,132	-1,35E-03
3,10	-0,389	0,685	0,057	1,681	0,034	0,141	-1,47E-03
3,20	-0,285	0,598	0,056	1,319	0,033	0,140	-1,33E-03
3,30	-0,262	0,580	0,060	1,267	0,037	0,160	-1,62E-03
3,40	-0,164	0,517	0,059	0,958	0,038	0,168	-1,63E-03
3,50	-0,250	0,589	0,068	1,396	0,047	0,206	-2,23E-03
3,60	-0,251	0,591	0,073	1,506	0,053	0,238	-2,74E-03
3,70	-0,236	0,580	0,080	1,577	0,063	0,286	-3,69E-03
3,80	-0,194	0,554	0,088	1,559	0,073	0,340	-5,07E-03
3,90	-0,236	0,592	0,102	1,954	0,089	0,422	-7,45E-03
4,00	-0,296	0,636	0,113	2,374	0,101	0,488	-9,36E-03
4,10	-0,393	0,695	0,123	2,763	0,115	0,561	-1,13E-02
4,20	-0,533	0,769	0,153	3,218	0,148	0,739	-1,88E-02
4,30	-0,327	0,599	0,148	2,125	0,135	0,699	-1,75E-02
4,40	-0,483	0,689	0,152	2,388	0,158	0,819	-2,11E-02
4,50	-0,527	0,687	0,154	2,807	0,179	0,936	-2,56E-02
4,60	-0,628	0,712	0,228	2,507	0,253	1,374	-5,80E-02

## Anexo E. Niveles de retorno

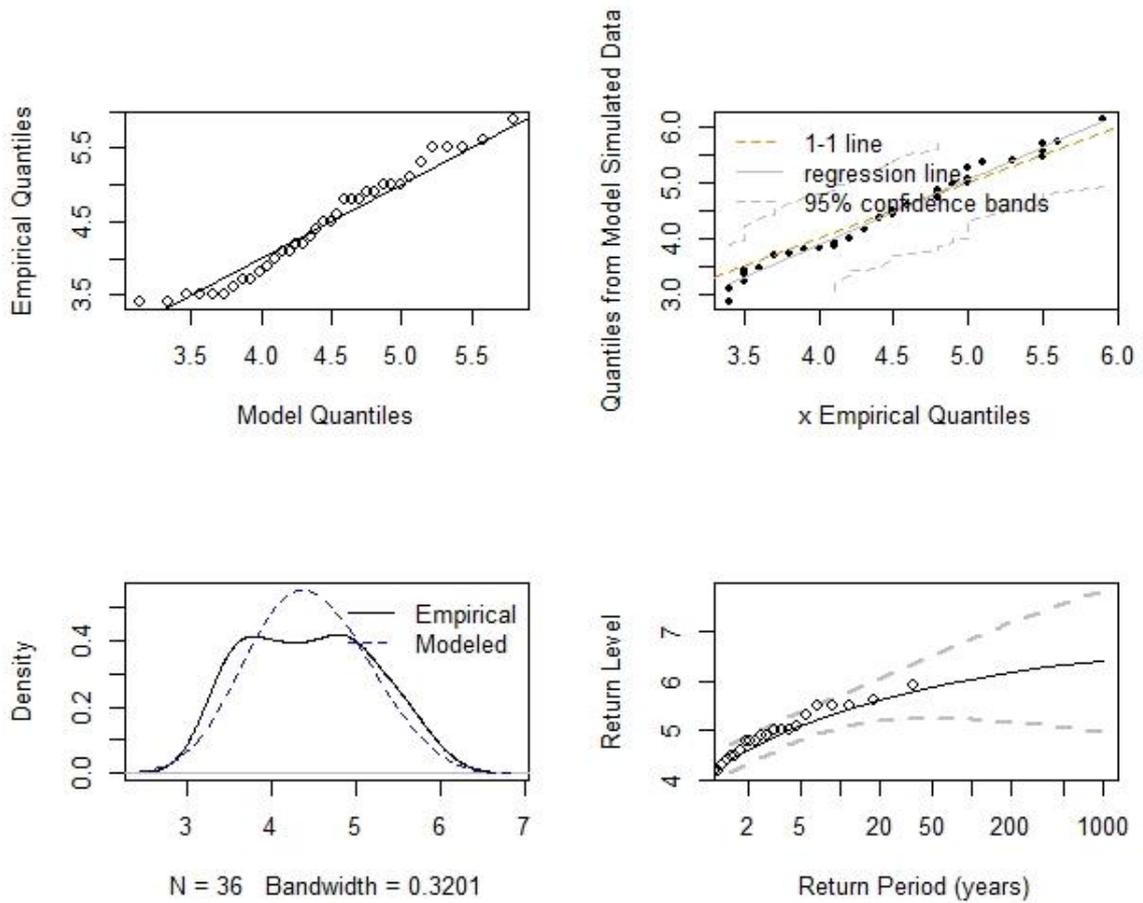
**Tabla 5. Niveles de retorno para los modelos estudiados con sus intervalos de confianza**

	GEVD MLE			GEVD Lmom			Gumbel MLE		
	95% lower	CI Estimate	95% upper	95% lower	CI Estimate	95% upper	95% lower	CI Estimate	95% upper
2-year return level	4,14	4,42	4,71	4,13	4,41	4,66	4,08	4,32	4,56
5-year return level	4,75	5,04	5,33	4,77	5,08	5,38	4,66	5,03	5,40
10-year return level	5,04	5,36	5,67	5,09	5,44	5,85	5,02	5,49	5,97
20-year return level	5,20	5,61	6,03	5,29	5,75	6,29	5,36	5,94	6,53
50-year return level	5,25	5,88	6,50	5,49	6,08	6,79	5,80	6,52	7,25
100-year return level	5,23	6,04	6,85	5,59	6,29	7,25	6,12	6,96	7,79
500-year return level	5,07	6,32	7,57	5,75	6,67	8,25	6,88	7,96	9,05
1000-year return level	4,98	6,41	7,84	5,81	6,80	8,69	7,20	8,40	9,59
	GPD MLE			GPD Lmom					
	95% lower	CI Estimate	95% upper	95% lower	CI Estimate	95% upper			
2-year return level	5,33	5,70	6,06	5,22	5,56	5,94			
5-year return level	5,43	5,90	6,37	5,29	5,71	6,22			
10-year return level	5,48	6,03	6,58	5,32	5,81	6,41			
20-year return level	5,52	6,15	6,78	5,35	5,89	6,62			
50-year return level	5,53	6,28	7,03	5,37	5,97	6,84			
100-year return level	5,53	6,36	7,20	5,38	6,03	6,99			
500-year return level	5,47	6,52	7,57	5,40	6,12	7,35			
1000-year return level	5,43	6,58	7,72	5,40	6,15	7,49			
	GEVD MLE*			GEVD Lmom*			Gumbel MLE*		
	95% lower	CI Estimate	95% upper	95% lower	CI Estimate	95% upper	95% lower	CI Estimate	95% upper
2-year return level	4,54	4,70	4,86	4,54	4,69	4,85	4,50	4,65	4,80
5-year return level	4,90	5,08	5,26	4,90	5,09	5,27	4,86	5,09	5,33
10-year return level	5,08	5,28	5,48	5,07	5,31	5,54	5,09	5,39	5,69
20-year return level	5,19	5,44	5,70	5,18	5,49	5,79	5,31	5,67	6,03
50-year return level	5,26	5,62	5,98	5,28	5,69	6,14	5,59	6,04	6,49
100-year return level	5,28	5,73	6,18	5,34	5,81	6,41	5,80	6,31	6,83
500-year return level	5,25	5,92	6,60	5,45	6,05	7,16	6,28	6,95	7,62
1000-year return level	5,21	5,99	6,76	5,47	6,12	7,46	6,49	7,22	7,95

\*Corresponde a los bloques realizados por año de medición: 35 bloques en el periodo 1980 - 2017

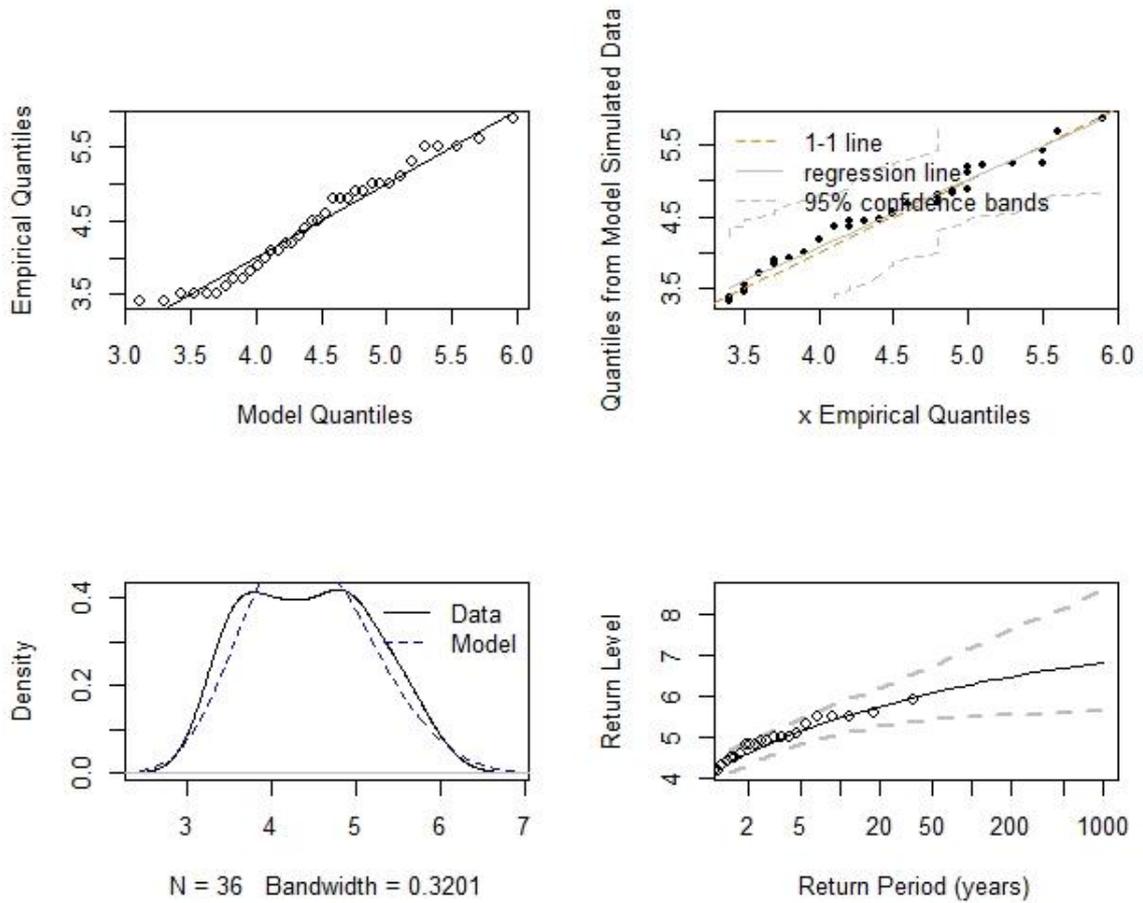
## Anexo F. Gráfica del modelo GEVD: MLE ( $nb = 30$ )

fevd(x = x, type = "GEV", method = "MLE")



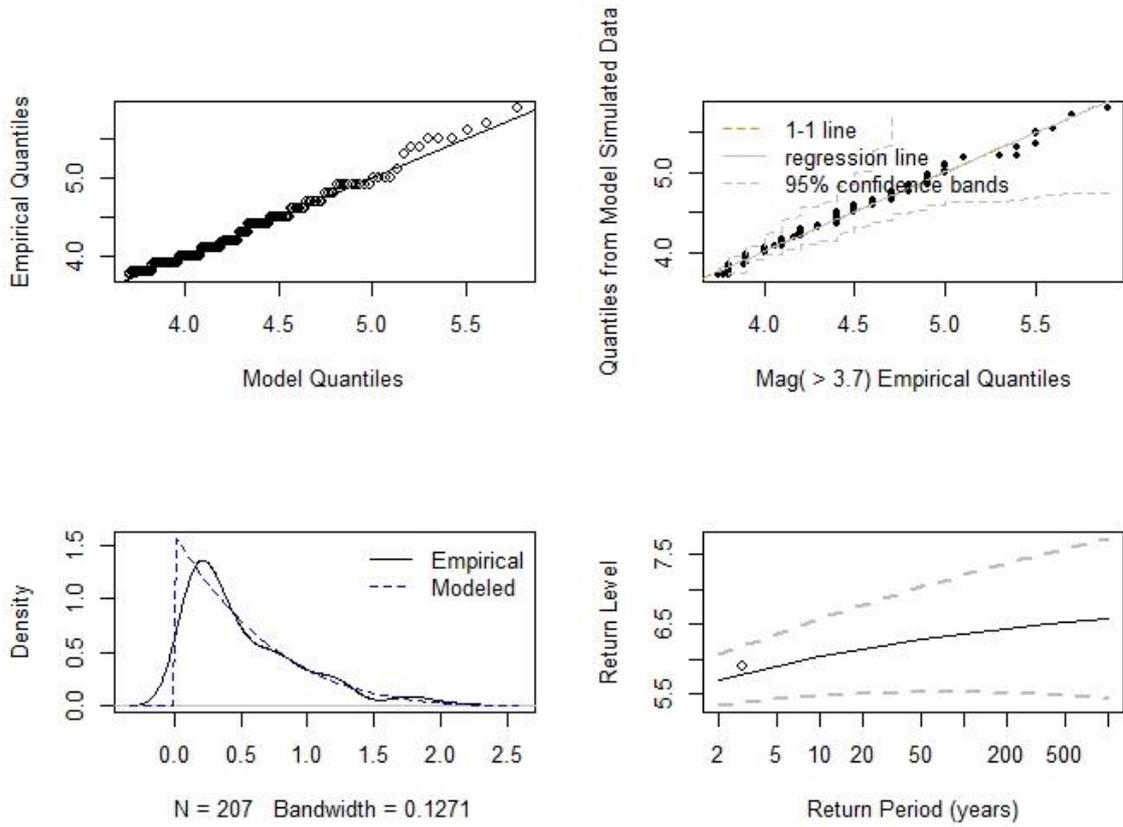
## Anexo G. Gráfica del modelo GEVD: L-Momentos ( $nb = 30$ )

fevd(x = x, type = "GEV", method = "Lmoments")



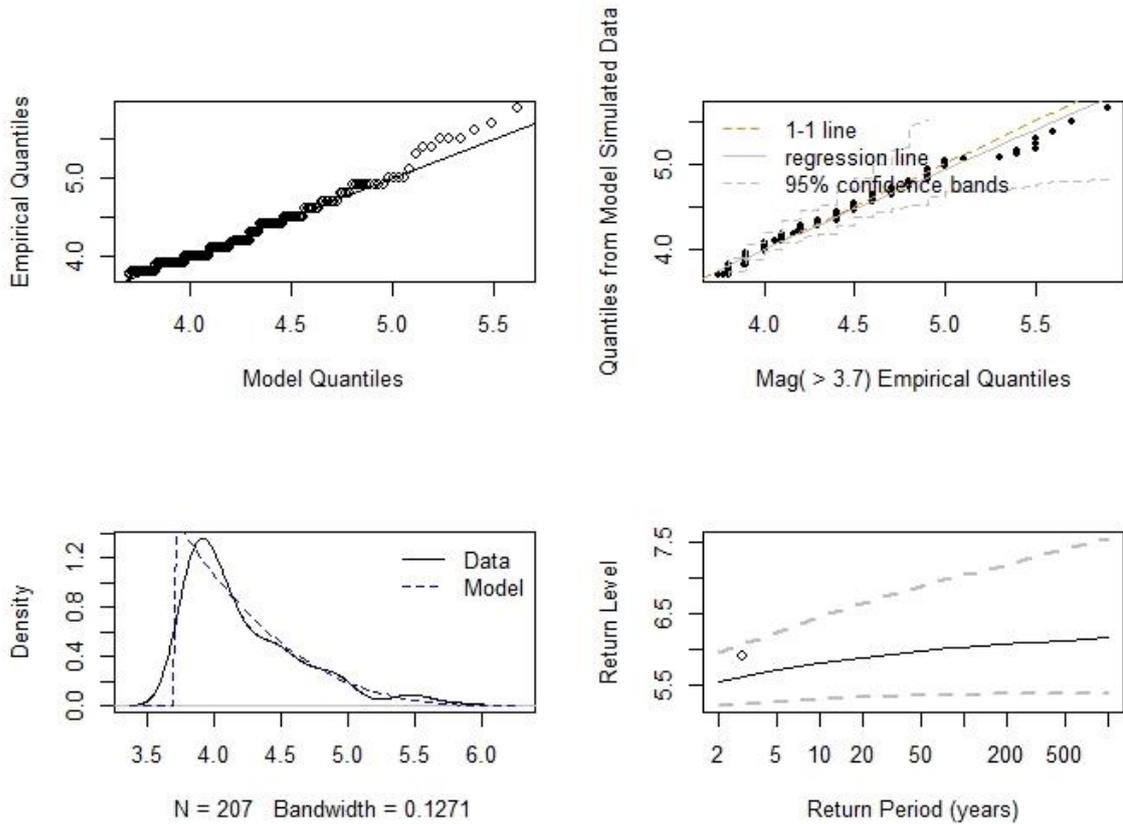
## Anexo H. Gráfica del modelo GPD: MLE

fevd(x = Mag, threshold = 3.7, type = "GP")



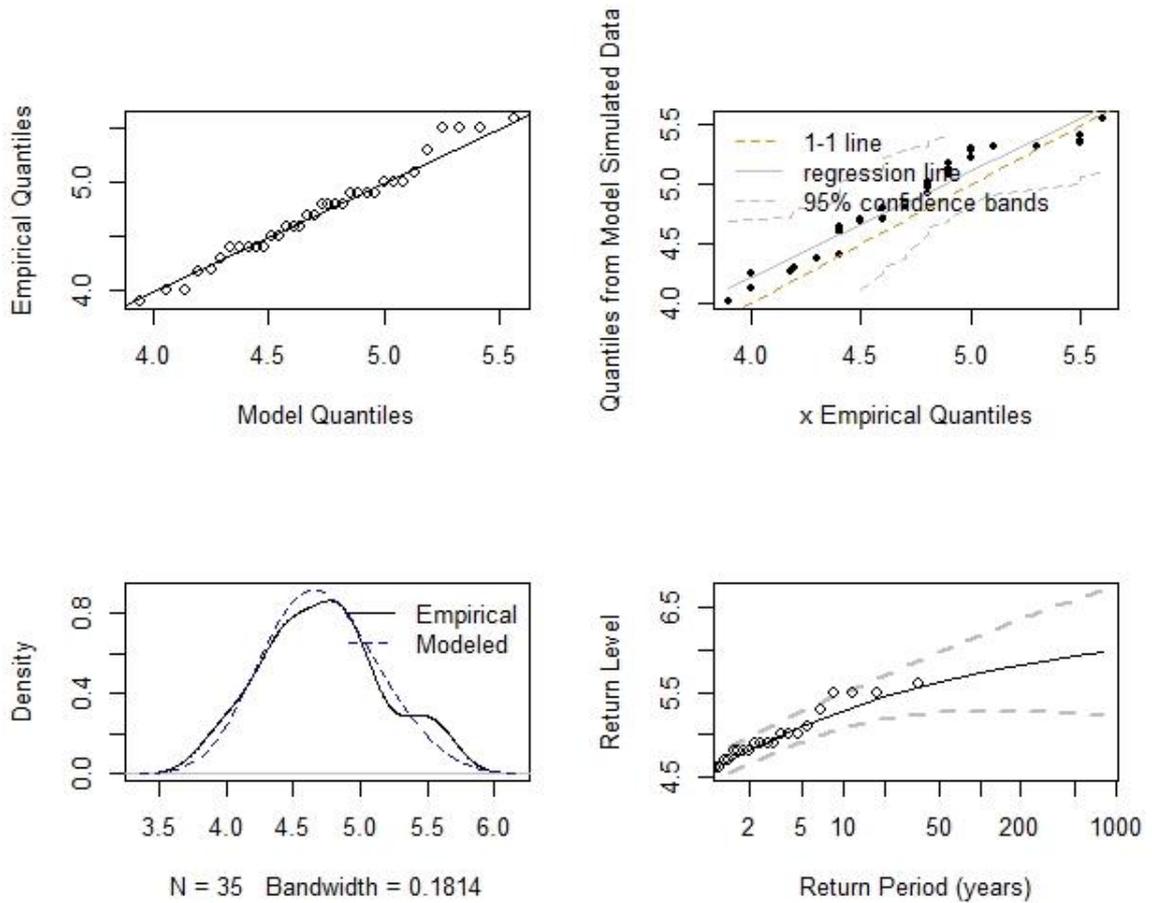
# Anexo I. Gráfica del modelo GPD: L-Momentos

fevd(x = Mag, threshold = 3.7, type = "GP", method = "Lmoments")



## Anexo J. Gráfica del modelo GEVD: MLE ( $nb = 35$ )

fevd(x = x, type = "GEV")



## Anexo K. Gráfica del modelo GEVD: L-Momentos ( $nb = 35$ )

fevd(x = x, type = "GEV", method = "Lmoments")

