

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas



PROYECTO INTEGRADOR

**“ESTUDIO ESTADÍSTICO DE LA PELIGROSIDAD SÍSMICA
DEL ECUADOR MEDIANTE LA TEORÍA DE VALORES
EXTREMOS”**

Previo la obtención del Título de:
INGENIERO EN ESTADÍSTICA

Presentado por:
Angel Jasmani Chancay Mesa

GUAYAQUIL - ECUADOR

Año: 2018

AGRADECIMIENTOS

A mi madre Magali Mesa por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor.

A mi padre Fernando Chancay por el valor mostrado para salir adelante, por los ejemplos de perseverancia y constancia que lo caracterizan y por su apoyo incondicional.

A mi tutor PhD. Sandra García por su ayuda, motivación y conocimientos brindados para la culminación del presente trabajo.

DEDICATORIA

El presente proyecto lo dedico a mis padres, mis hermanos y a mi sobrino por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo perfectamente mantenido a través del tiempo.

A mis amigos, compañeros y a todos aquellos que participaron directa o indirectamente en la elaboración de este proyecto.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Yo Angel Jasmani Chancay Mesa doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Angel Chancay Mesa

EVALUADORES



Ph.D. Sandra García

PROFESOR DE LA MATERIA Y TUTOR

RESUMEN

En el presente trabajo se estudia la ocurrencia sísmica en el Ecuador, describir el comportamiento de las máximas magnitudes sísmicas registradas conforma uno de los principales objetivos. Se utilizaron algunas técnicas estadísticas como lo son los principios del Análisis de Valores Extremos, a fin de encontrar una distribución de probabilidad que se ajuste a los datos de las máximas magnitudes, la técnica de clustering denominada K-Medias, permite agrupar las observaciones de las ocurrencias sísmicas de acuerdo con sus características similares. El modelo Gumbel resulto ser el más adecuado para ajustar las máximas magnitudes a nivel general de los sismos ocurridos en el Ecuador y dentro las cinco zonas resultantes del análisis clúster. Se concluyó que un sismo de $7.8 M_w$. o mayor tendría un periodo de retorno de 50 años en el Ecuador.

Palabras Clave: Peligrosidad sísmica, clustering, valores extremos.

ABSTRACT

The seismic occurrence in Ecuador is studied in the present work. Describe the behavior of the maximum registered seismic magnitudes, is one of the main objectives. Some statistical techniques were used such as the principles of Extreme Value Analysis, to find a probability distribution that fits the data of the maximum magnitudes, the clustering technique K-Medias, allows to group the observations of the seismic occurrences according to their similar characteristics. The Gumbel model was the most suitable to fit the maximum magnitudes of the earthquakes occurred in Ecuador and within the five zones resulting from the cluster analysis. An earthquake of 7.8 M_w . or greater would have a return period of 50 years in Ecuador.

Keywords: *Seismic hazard, clustering, extreme values.*

ÍNDICE GENERAL

ÍNDICE GENERAL.....	III
ABREVIATURAS	V
SIMBOLOGÍA	VI
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS	IX
CAPÍTULO 1	10
1. INTRODUCCIÓN	10
1.1 Descripción del problema	10
1.2 Justificación del problema.....	10
1.3 Alcance del Proyecto	11
1.4 Objetivos.....	11
1.4.1 Objetivo General	11
1.4.2 Objetivos Específicos	11
1.5 Marco Teórico.....	12
1.6 Marco Conceptual.....	14
1.6.1 Descripción Territorial del Ecuador	14
1.6.2 Teoría de Sismos	14
CAPÍTULO 2.....	18
2. METODOLOGÍA:	18
2.1 Enfoque de la investigación	18
2.1.1 Tipo de investigación.....	18
2.1.2 Diseño de investigación	18
2.2 Herramientas Estadísticas.....	19
2.2.1 Estadísticos de Orden	19

2.2.2	Análisis de valores extremos.....	20
2.2.3	Máximo por Bloques (BM).....	20
2.2.4	Picos por encima del umbral (POT)	26
2.2.5	Niveles y periodos de retorno.....	28
2.2.6	Criterios de Información	28
2.2.7	Prueba de Razón de Verosimilitud.....	28
2.2.8	Método de Imputación de datos: Hot-Deck	29
2.2.9	Métodos de Clustering	29
CAPÍTULO 3.....		32
3.	RESULTADOS Y ANÁLISIS:	32
3.1	Análisis descriptivo	32
3.2	Modelos de distribución de valores extremos.....	40
3.2.1	Modelos de Máximos por Bloques	41
3.2.2	Modelo de Picos Sobre un Umbral.....	44
3.3	Zonificación y comparación de modelos en el Ecuador.....	46
3.3.1	Teoría de Valores Extremos aplicada a la zonificación del Ecuador.....	51
CAPÍTULO 4.....		58
4.	CONCLUSIONES Y RECOMENDACIONES	58
	Conclusiones	58
	Recomendaciones.....	60
5.	BIBLIOGRAFÍA	60

ABREVIATURAS

AIC	=	Criterio de Información de Akaike (<i>Akaike Information Criterion</i>)
BIC	=	Criterio de Información Bayesiano (<i>Bayesian Information Criterion</i>)
BM	=	Máximo por Bloques (<i>Block Maxima</i>)
GEV	=	Distribución de Valores Extremos Generalizada (<i>Generalized Extreme Value</i>)
GP	=	Distribución de Pareto Generalizada (<i>Generalized Pareto</i>)
IC	=	Intervalo de Confianza (<i>Maximum Likelihood Estimation</i>)
IRIS	=	Incorporated Research Institutions for Seismology
MLE	=	Estimación por Máxima Verosimilitud
POT	=	Picos Sobre un Umbral (<i>Peaks Over Threshold</i>)
QQPLOT	=	Gráfico Cuantil-Cuantil (Quantil-Quantil)
SCE	=	Sumatoria Cuadrática del Error
UTM	=	Sistema de Coordenadas Universal Transversal de Mercator (<i>Universal Transverse Mercator</i>)
TVE	=	Teoría de Valores Extremos

SIMBOLOGÍA

μ	Parámetro de localización
σ	Parámetro de escala
ξ	Parámetro de forma
M_w	Magnitud momento
Z_p	Nivel de retorno

ÍNDICE DE FIGURAS

Figura 2.1 Ejemplo gráfico de selección de datos (Metodología de Máximo por Bloque).	21
Figura 2.2 Ejemplo gráfico de selección de datos (Metodología de Picos por encima del Umbral).	26
Figura 3.1 Histograma: Magnitud de los sismos ocurridos en el Ecuador (1906-2017).	32
Figura 3.2 Diagrama de caja: Magnitud de sismos ocurridos en el Ecuador (1906-2017).	33
Figura 3.3 Mapa: Epicentros de los sismos ocurridos en el Ecuador (1906-2017).	34
Figura 3.4 Gráfico de curvas de nivel: Concentración de sismos ocurridos en el Ecuador (1906-2017).	35
Figura 3.5 Gráfico de curvas de nivel: Concentración de sismos por intensidad en el Ecuador (1906-2017).	36
Figura 3.6 Gráfico de barras: Recuento de sismos por provincias e intensidad (1906-2017).	37
Figura 3.7 Serie temporal: Máximas magnitudes sísmicas por década en el Ecuador.	37
Figura 3.8 Mapa: Actividad sísmica en el Ecuador de las tres últimas décadas.	38
Figura 3.9 Gráfico de barras: Media de la magnitud de sismos por provincia, +/- 2 error estándar.	39
Figura 3.10 Gráfico de barras: Máximas magnitudes de sismos por provincias (1906-2017).	39
Figura 3.11 Diagrama de Caja: Magnitud de sismos por provincias (1906-2017).	40
Figura 3.12 Serie temporal: Magnitud máxima reportada por año (con imputación por método Hot-Deck).	41
Figura 3.13 Gráficos del modelo por distribución de Gumbel.	43
Figura 3.14 Vida media residual de la magnitud de los sismos ocurridos en el Ecuador.	44

Figura 3.15 Variación del parámetro de forma del modelo POT aplicado a la magnitud de los sismos ocurridos en el Ecuador.	45
Figura 3.16 Gráficos del modelo por distribución de Pareto Generalizada.	46
Figura 3.17 Suma de Cuadrado entre Clústers (K-Medias).	47
Figura 3.18 Índice de Silhouette (K-Medias).	48
Figura 3.19 Índice de Calinski (K-Medias).	48
Figura 3.20 Gráfico de los clústers por K-Medias de los sismos ocurridos en Ecuador.	49
Figura 3.21 Gráfico de los clústers por Ward de los sismos ocurridos en Ecuador. .	49
Figura 3.22: Gráfico de los epicentros de sismos ocurridos en el Ecuador, clasificados por las 5 zonas.....	51

ÍNDICE DE TABLAS

Tabla 1.1: Clasificación de los sismos según su magnitud.	15
Tabla 1.2: Clasificación resumen de los sismos según su magnitud.	16
Tabla 2.1 Características de la distribución de Gumbel.	22
Tabla 2.2 Características de la distribución de Fréchet.	23
Tabla 2.3 Características de la distribución de Weibull.	24
Tabla 2.4 Características de la distribución de Valores Extremos Generalizada.	25
Tabla 2.5 Características de la distribución de Pareto Generalizada.	27
Tabla 3.1 Modelo GEV aplicado a las máximas magnitudes en el Ecuador.	42
Tabla 3.2 Modelo Gumbel aplicado a las máximas magnitudes en el Ecuador.	42
Tabla 3.3 Test de razón de verosimilitud (Gumbel vs. GEV)	42
Tabla 3.4 Periodos de retorno modelo GEV.	43
Tabla 3.5 Modelo GP aplicado a las magnitudes de sismos en el Ecuador.	45
Tabla 3.6 Periodos de retorno del modelo GP.	46
Tabla 3.7 Características de las variables estandarizadas de los clústeres.	50
Tabla 3.8 Características de las variables no estandarizadas.	51
Tabla 3.9 Modelo Gumbel aplicado a la Zona 1.	52
Tabla 3.10 Test de razón de verosimilitud (Gumbel vs. GEV)	52
Tabla 3.11 Zona1-Periodos de retorno de las magnitudes estimadas.	53
Tabla 3.12 Modelo Gumbel aplicado a la Zona 2.	53
Tabla 3.13 Test de razón de verosimilitud (Gumbel vs. GEV)	53
Tabla 3.14 Zona 2-Periodos de retorno de las magnitudes estimadas.	54
Tabla 3.15 Modelo Gumbel aplicado a la Zona 3.	54
Tabla 3.16 Test de razón de verosimilitud (Gumbel vs. GEV)	54
Tabla 3.17 Zona 3-Periodos de retorno de las magnitudes estimadas.	55
Tabla 3.18 Modelo Gumbel aplicado a la Zona 4.	55
Tabla 3.19 Test de razón de verosimilitud	55
Tabla 3.20 Zona 4-Periodos de retorno de las magnitudes estimadas.	56
Tabla 3.21 Modelo Gumbel Aplicado a la Zona 5.	56
Tabla 3.22 Test de razón de verosimilitud (Gumbel vs. GEV)	56
Tabla 3.23 Zona 5-Periodos de retorno de las magnitudes estimadas.	57

CAPÍTULO 1

1. INTRODUCCIÓN

Ecuador es un país con elevada actividad sísmica, debido a que posee una de las zonas de subducción más relevantes del mundo y además forma parte del Cinturón de Fuego del Pacífico (Parra, Benito, & Gaspar, 2016). A continuación se explica detalles para conocer cuan importante es culturizarse en medidas de prevención cuando se vive en un sector de riesgo.

1.1 Descripción del problema

En el territorio ecuatoriano se han registrado sismos de elevada magnitud entre ellos destaca el ocurrido en 1906, cuya magnitud fue de $8.8 M_w$. y el reciente sismo catastrófico del 16 de abril de 2016 tuvo una magnitud de $7.8 M_w$. (IGEPN, 2018). Este último produjo notables daños en las provincias de Esmeraldas y Manabí.

Una de las razones que más ha incrementado el riesgo ante esta situación es la carente preparación para enfrentar eventos de gran magnitud y, por lo tanto, las secuelas que más han afectado al país se denotan en el ámbito económico y social. Por ello es necesario que los habitantes tomen conciencia de que viven en una zona de elevada actividad sísmica, y estimar la peligrosidad a la que debe desafiar. Ante lo dicho, es imperioso la preparación y prevención ante desastres, tras identificar y caracterizar las zonas de mayor actividad sísmica.

No obstante, el proyecto no consiste en predecir con exactitud en dónde, cuándo y qué magnitud tendrán futuros terremotos, sino más bien, hacer un análisis de la ocurrencia sísmica del Ecuador, y de esta manera dar una brecha de tiempo de los sucesos atípicos (sismos de elevadas magnitudes) que pueden suceder en una zona específica.

1.2 Justificación del problema

Este trabajo tiene una gran importancia, puesto que está orientado a dar una estimación del rango de tiempo en el que pudiese ocurrir un sismo igual o de mayor magnitud que el ocurrido el pasado 16 de abril del 2016 y así con este fundamento estadístico las

autoridades correspondientes vean la necesidad de realizar planes de contingencia y crear una cultura de prevención en nuestro país. Para este estudio se abarcará la zonificación del Ecuador de acuerdo con las características de los terremotos que se registren en estas, con la ayuda de técnicas de clustering.

De este modo, el trabajo propone un método para la agrupación de datos sísmicos que puede ayudar en la visualización e interpretación de estructuras sísmicas que, mediante el uso de los métodos de Máximo por Bloques y Picos Sobre el Umbral, se elaborarán los modelos que mejor se ajusten a una distribución de probabilidad en particular. Esto permitirá estimar la probabilidad a un nivel de confianza específico de que se produzca un sismo de cierta magnitud en un intervalo de tiempo definido, estos modelos serán el resultado de la aplicación de la teoría de valores extremos, los análisis a realizarse tendrán un enfoque estadístico.

1.3 Alcance del Proyecto

- **Espacio:** Ecuador (Zona continental y marítima comprendidas entre las coordenadas (Longitud: [-83; -75] y Latitud: [-5;2])
- **Tiempo:** Periodo comprendido entre los años 1906-2017
- **Población:** Epicentros de sismos registrados en el territorio ecuatoriano.

1.4 Objetivos

1.4.1 Objetivo General

Analizar la ocurrencia sísmica del Ecuador durante el periodo 1906-2017, con el uso de la Teoría de Valores Extremos y técnicas de clustering para que los habitantes conozcan cómo se comportan estos eventos en las zonas que residen.

1.4.2 Objetivos Específicos

- Describir el comportamiento de los sismos registrados en el Ecuador en el periodo de 1906-2017 para la obtención de características que permitan conocer en primera instancia cuando y donde sucedieron.

- Modelar una distribución que represente las máximas magnitudes de los sismos mediante el uso de la Teoría de Valores Extremos para la estimación de probabilidades de ocurrencia de sismos de cierta magnitud.
- Considerar las principales características de los terremotos registrados en las zonas identificadas del Ecuador para la observación de posibles periodos de retorno de 10, 50 y 100 años.

1.5 Marco Teórico

Parra, Benito, y Gaspar (2016) propusieron un mapa de peligrosidad sísmica del Ecuador Continental a través de un modelo zonal para evaluar la probabilidad de excedencia de varios niveles de aceleración PGA (Peak Ground Acceleration), considerando datos de terremotos en el periodo de 1584 a 2014. El Código Ecuatoriano de Construcción (CEC, 2001) presenta el primer mapa de riesgo sísmico para el Ecuador, considerando cuatro zonas sísmicas clasificadas por los valores del PGA, que van de 0.15 a 0.4g. La versión actualizada, publicada en la Norma Ecuatoriana de Construcción (NEC-11, 2015) considera que el mapa de peligrosidad sísmica está dividido en seis zonas con un rango de 0.15 a 0.5g.

Eventos inusuales como grandes magnitudes de sismos, no se los puede analizar de manera convencional, ante la necesidad de analizar eventos extremos, surge la Teoría de Valores Extremos (TVE). La raíz de este tipo de análisis es incierta, muchos autores atribuyen a diferentes personas la base para el estudio de los valores extremos, sin embargo, no tuvo un gran fundamento matemático hasta que Fisher y Tippett (1928) propusieron el Teorema de la Teoría de Valores Extremos, que describe la distribución asintótica de los valores máximos o mínimos de una variable aleatoria, y esta teoría obtuvo una metodología formal para modelar este tipo de sucesos a partir del 1950 (Coles S. , 2001).

Existen muchas aplicaciones relacionadas a la TVE, por ejemplo, según Kotz y Nadarajah (2000) se puede aplicar a la medición y análisis de contaminación en el aire y

de corrosión. János G. (1978) aplica esta teoría en el cálculo de probabilidades de inundaciones atípicas y sequías. Coles (2001) explica cómo la teoría de valores extremos puede ayudar a la construcción de un rompeolas para salvaguardar a la población de diversos niveles del mar que se esperan durante los próximos 100 años en Países Bajos. Reiss y Thomas (2007) mencionan la forma en que se puede analizar la resistencia de materiales, a través de la TVE, puesto que la fuerza que compone un metal es el mínimo de las fuerzas de las fibras que las componen. Así mismo, Albeverio, Jentsch y Kantz (2005) detallan que la epilepsia, se considera como un “Valor extremo dentro del cerebro humano”, por lo que se pueden estimar periodos de retorno. Otros ejemplos se evidencian en el cálculo de grandes pérdidas de una aseguradora, con la finalidad de que no sobrepase cierto umbral; otro campo es el de riesgos financieros, siendo más preciso, se puede calcular el VAR (Value at Risk) en los requerimientos de capital del riesgo de mercado y operacional que puede tener un banco, a través de la TVE (Alcalde, 2005), entre otros campos de aplicación.

En el contexto del estudio de la vulnerabilidad sísmica, el modelo de Gumbel fue utilizado por Burton y Makropoulos (1985) para modelar el riesgo sísmico de los terremotos del Cinturón del Pacífico. Así mismo, Zimbidis, Frangos y Pantelous (2007) analizaron los datos históricos de los terremotos ocurridos en la zona fronteriza de Grecia, para producir un modelo fiable del riesgo de la magnitud de los terremotos, utilizando técnicas avanzadas de la Teoría de Valores Extremos. Chunga, Michetti y Pazmiño (2009) integraron datos geológicos, sismológicos y sismo-tectónicos, a fin de estimar los máximos niveles de sismicidad en el litoral ecuatoriano. Aguiar, Castro, Garzón, Yanchatuña, Cumbal y La Fave (2009) presentaron las magnitudes máximas esperadas en 20 zonas del Ecuador, las que fueron obtenidas considerando la máxima magnitud registrada instrumentalmente en el periodo comprendido entre 1540 y 2009, y en base a una relación entre la longitud de ruptura y la magnitud.

En lo que respecta al análisis clúster, Kamath Rajani (2017) realizó una clasificación no supervisada de los últimos diez años de eventos sísmicos ocurridos en la India, en función de su ubicación espacial y su magnitud mediante el uso de la técnica de

agrupación de K-Medias, este trabajo muestra el rendimiento de la agrupación con K-Medias ajustando el número de clústeres e iteraciones.

1.6 Marco Conceptual

1.6.1 Descripción Territorial del Ecuador

El Ecuador es una unidad geográfica que comprende el espacio continental y marítimo, las islas adyacentes, el mar territorial, el Archipiélago de Galápagos, el suelo, la plataforma submarina, el subsuelo y el espacio suprayacente continental, insular y marítimo; además el Estado ecuatoriano se organiza territorialmente en regiones, provincias, cantones y parroquias rurales. (Constitución de la República del Ecuador, 2008).

1.6.2 Teoría de Sismos

1.6.2.1 Definición de Sismo

De acuerdo con Zúñiga (2011), un sismo, temblor, terremoto o movimiento telúrico es un fenómeno de sacudida o vibración brusca y pasajera de la corteza terrestre, estos se producen por una liberación rápida y espontánea de energía acumulada, la cual se propaga en todas las direcciones desde su origen en forma de ondas sísmicas.

1.6.2.2 Clasificación de los sismos según su origen

Según Zúñiga (2011) a los sismos se los puede clasificar en 2 principales clases según la forma cómo se originan:

- **Sismos Naturales:** Son aquellos en los que no interviene el ser humano o algún otro agente externo. Entre ellos se encuentran:
 - Sismos Volcánicos: producidos por la actividad volcánica del planeta. Suelen ser de magnitud leve y son imperceptibles para los pueblos aledaños al volcán.

- Sismos Tectónicos: son producidos por la actividad entre placas o intraplaca. Estos son los más concurrentes y que pueden generar mayor peligro.
- **Sismos Artificiales**: Se producen de manera voluntaria o involuntaria por los humanos, de tal manera que se altera el equilibrio de fuerzas en la corteza terrestre. Su profundidad no suele ser mayor a 2 kilómetros y su origen es muy puntual (Zúñiga, 2011). Entre estos tenemos aquellos ocasionados por la industria minera, pruebas nucleares o la demolición de edificios.

1.6.2.3 Clasificación de sismos según su magnitud.

Se puede representar la energía liberada por un sismo mediante la siguiente clasificación según el UPSEIS de la MTU (Michigan Tech University, 2016).

Clase	Magnitud (M_w)
Gran Intensidad	>8
Mayor Intensidad	7 – 7.9
Fuerte Intensidad	6 – 6.9
Moderada Intensidad	5 – 5.9
Leve Intensidad	4 – 4.9
Menor Intensidad	3 – 3.9

Tabla 1.1: Clasificación de los sismos según su magnitud.

Fuente: UPSEIS de la MTU (Michigan Tech University)

En el presente proyecto, con el propósito de realizar comparaciones entre sismos de diferentes clasificaciones, se resumirán las diferentes clases en 3 categorías.

Clase	Magnitud (M_w)
Leve Intensidad	< 4.5
Moderada Intensidad	4.5 – 5.5
Grave Intensidad	> 5.5

**Tabla 1.2: Clasificación resumen de los sismos
según su magnitud.**

Fuente: Elaboración propia.

1.6.2.4 Escala de Magnitud de Momento (M_W)

En 1971 los científicos Hanks y Kanamori (1979) crearon una escala basada en la medición de la energía total que se libera en un sismo. Debido a su precisión se ha convertido en la más usada por sismólogos.

$$M_W = \frac{2}{3}(\log_{10} M_0(\text{ergs}) - 16.0) \quad (1.3)$$

Siendo:

- M_0 es el momento Sísmico, la energía medida en ergs.

$$M_0 = \mu Au \quad (1.4)$$

- μ es el módulo de deformación de las rocas que fueron afectadas por el sismo, usualmente es de 30 gigapascales.
- A es el área donde se produce la ruptura, a lo largo de la falla donde ocurrió el sismo o terremoto.
- u es el desplazamiento promedio del área donde se produjo la ruptura.

1.6.2.5 Ley de Gutenberg – Richter

Gutenberg y Richter (1944) postularon, en base a la estadística de leyes de recurrencia, que existe una relación entre **Log N** y **M**, y esta debía ser lineal. La ley de Gutenberg-Richter establece que la relación de estas variables obedece a la siguiente ecuación:

$$\log_{10} N = a - b * M \quad (1.5)$$

Dónde:

- **M** es un umbral de la magnitud.
- **N** es la recurrencia sísmica anual de magnitud $\geq \mathbf{M}$.
- **a** y **b** son constantes determinadas por la naturaleza sísmica de la región.

Las constantes **a** y **b** pueden calcularse con diferentes métodos, debido a que la naturaleza es cambiante, es necesario actualizar estas constantes frecuentemente. Esta relación no varía significativamente en el tiempo o entre regiones y es sorprendentemente robusta.

CAPÍTULO 2

2. METODOLOGÍA:

En el presente proyecto se desarrolla un análisis descriptivo desde un punto de vista general de los sismos ocurridos en el Ecuador, con la finalidad de encontrar ubicaciones de mayor concentración de epicentros de diferentes intensidades, y así comparar con el análisis de los modelos explicados en el capítulo 3, mediante la implementación de las distribuciones de valores extremos.

2.1 Enfoque de la investigación

2.1.1 Tipo de investigación

La presente investigación es de tipo cuantitativo, documental, exploratorio e inferencial.

2.1.2 Diseño de investigación

Para la consecución de los objetivos planteados, se extrajo información de las bases de datos proporcionadas por el sitio web del Incorporated Research Institutions for Seismology de Washington DC. (IRIS, 2016) , y contrastada junto con la base de sismos proporcionada por el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN, 2016). La data contiene registros de sismos ocurridos durante el periodo comprendido entre 1906 y 2017.

Se procedió a calcular estadísticos de orden, posición y dispersión, los cuales permitirán describir las principales características de los sismos, después de este análisis descriptivo se estimaron curvas de nivel en base a la ocurrencia de los sismos, estos dos análisis sirvieron de preámbulo para el desarrollo del análisis clúster.

El análisis clúster se realizó con el método de K-Medias y método de Ward, a fin de hacer una comparación entre las agrupaciones resultantes, para la selección del número de clúster se calculó la variación dentro y entre clúster, y los índices de Silhouette y Calinski, el método escogido fue aquel cuyo resultado se ajustó visualmente al mapa de densidades de ocurrencia de sismos.

Posteriormente se diseñaron los modelos mediante los dos enfoques de Máximo por Bloques y Picos Sobre el Umbral, que fueron aplicados sobre los datos de los sismos a nivel general y de las zonas resultantes del análisis clúster, esto a fin de predecir los periodos de retornos de los sismos de elevada magnitud, y también para estimar la probabilidad de ocurrencia. En los modelos BM se escogió la distribución con mejor ajuste a los datos de las máximas magnitudes teniendo en cuenta el parámetro de forma; la selección del modelo más adecuado fue realizada mediante una prueba de razón de verosimilitud, además que se tuvo en consideración los criterios de AIC y BIC de los modelos. Se elaboraron las gráficas de exceso medio y de variación del parámetro forma que sirven como criterio para fijar un umbral adecuado en el diseño del modelo POT.

2.2 Herramientas Estadísticas

2.2.1 Estadísticos de Orden

Definición: Sea (X_1, X_2, \dots, X_n) una muestra procedente de una población. Si los valores de la secuencia X_1, X_2, \dots, X_n se ordenan en orden creciente, dados como $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, entonces el valor r -ésimo de esta nueva secuencia se denomina estadístico de orden r de la muestra dada (Evans & Rosenthal, 2005).

Para este trabajo con estadísticos de orden, se hace el supuesto de que X_1, X_2, \dots, X_n , son independientes e idénticamente distribuidas, con función de densidad $F(x)$. Sean $X_{(1)} = \text{Min}(X_1, X_2, \dots, X_n)$ y $X_{(n)} = \text{Max}(X_1, X_2, \dots, X_n)$, su función de densidad y distribución viene dada de la siguiente manera:

Para el mínimo

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \quad (2.1)$$

$$f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} \cdot f_X(x) \quad (2.2)$$

Y para el máximo

$$F_{X_{(n)}}(x) = (F_X(x))^n \quad (2.3)$$

$$f_{X_{(n)}}(x) = n(F_X(x))^{n-1} \cdot f_X(x) \quad (2.4)$$

2.2.2 Análisis de valores extremos

2.2.2.1 Teorema de Fisher-Tippet-Gnedenko

Con la colaboración de Fisher, Tippet y Gnedenko en sus trabajos (1928), y Gnedenko (1943) se elaboró el siguiente teorema de la Teoría de Valores Extremos:

Sean X_1, X_2, \dots, X_n variables independientes e idénticamente distribuidas, $M_n = \max(X_1, X_2, \dots, X_n)$ y las sucesiones de constantes $a_n > 0$ y b_n , tal que:

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z), \text{ cuando } n \rightarrow \infty \quad (2.5)$$

Esta teoría es una analogía al Teorema de Límite Central (Grinstead & Snell, 1997), que se resta a la variable de interés (x), un parámetro de tendencia y se lo divide por uno de dispersión, determinando dicha función de distribución, se ajusta a $G(z)$ cuando n tiende al infinito.

2.2.2.2 Dos Enfoques

Existen dos metodologías para interpretar y modelar los valores extremos de una variable aleatoria (Edd & Khairi, 2016), que son Máximo por Bloques (*Block Maxima*, “*BM*”) y Picos Sobre el Umbral (*Peaks Over Threshold*, “*POT*”), estas poseen interpretaciones y métodos de cálculo diferentes, pero ambas parten de la Teoría de Valores Extremos.

2.2.3 Máximo por Bloques (BM)

La variable aleatoria de los máximos se las toma por periodos de tiempo, y lo ajusta a una distribución de probabilidad $G(z)$ que pueden tender a tres distribuciones las cuales se describen a continuación:

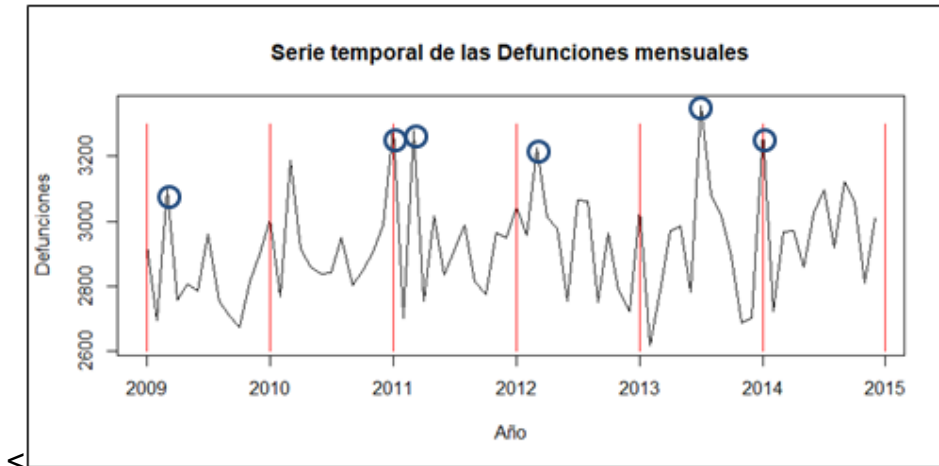


Figura 2.1 Ejemplo gráfico de selección de datos (Metodología de Máximo por Bloque).

Fuente: Elaborado por autor.

2.2.3.1 Distribución de Gumbel

Descubierta por el matemático alemán Emil Julius Gumbel, la Distribución de Gumbel permite encontrar un modelo para representar los valores extremos (sean estos máximo o mínimos) de un evento en particular (1934) y (1935), la cual viene dada de la siguiente manera:

- Distribución acumulada de Gumbel

$$F(x) = e^{-e^{\frac{(\mu-x)}{\sigma}}} \quad (2.6)$$

- Distribución de probabilidad de Gumbel

$$f(x) = \frac{e^{\frac{(\mu-x)}{\sigma}} e^{-e^{\frac{(\mu-x)}{\sigma}}}}{\sigma}, \quad x \in (-\infty, +\infty) \quad (2.7)$$

La distribución de Gumbel es un caso particular de la distribución de valores extremos generalizada, que se menciona más adelante, que corresponde cuando los parámetros poseen los valores de $\mu = 0$ y $\beta = 1$. Por ello la distribución acumulada y la distribución de probabilidad de Gumbel quedan de la siguiente manera:

- Distribución acumulada de Gumbel

$$F(x) = e^{-e^x} \quad (2.8)$$

- Distribución de probabilidad de Gumbel

$$f(x) = e^{-e^x}, \quad x \in (-\infty, +\infty) \quad (2.9)$$

Características	
Media	$\mu + \sigma\gamma$, dónde γ es la constante de Euler.
Varianza	$\frac{\pi^2}{6}\sigma^2$
Moda	μ
Mediana	$\mu - \sigma\ln(\ln(2))$

Tabla 2.1 Características de la distribución de Gumbel.
Fuente: Elaboración propia.

2.2.3.2 Distribución Fréchet

Descubierta por el matemático francés Maurice Fréchet (1927), la distribución Fréchet tiene como principal aplicación en el campo de la hidrología, viene dada de la siguiente manera:

- Distribución acumulada de Fréchet

$$F(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}}, \text{ si } x > \mu \quad (2.10)$$

- Distribución de probabilidad de Fréchet

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{-1-\xi} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}}, \text{ si } x > \mu \quad (2.11)$$

Donde $\xi \in (0, \infty)$ es el parámetro de forma, $\sigma \in (0, \infty)$ el de escala y $\mu \in (-\infty, +\infty)$ el de localización. Se puede dar el caso en que $\mu = 0$, consecuencia de esto la función de distribución pasa a ser:

- Distribución acumulada de Fréchet

$$F(x) = e^{-\left(\frac{x}{\sigma}\right)^{-\xi}}, \text{ si } x > 0 \quad (2.12)$$

- Distribución de probabilidad de Fréchet

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x}{\sigma}\right)^{-1-\xi} * e^{-\left(\frac{x}{\sigma}\right)^{-\xi}}, \text{ si } x > 0 \quad (2.13)$$

Características	
Media	$\mu + \sigma \Gamma\left(1 - \frac{1}{\xi}\right)$, siempre que $\xi > 1$
Varianza	$\sigma^2 \left[\Gamma\left(1 - \frac{2}{\xi}\right) - \left(\Gamma\left(1 - \frac{1}{\xi}\right)\right)^2 \right]$, siempre que $\xi > 2$
Moda	$\mu + \sigma \left(\frac{\xi}{1 + \xi}\right)^{\frac{1}{\xi}}$
Mediana	$\mu + \frac{\sigma}{\sqrt[\xi]{\ln(2)}}$

Tabla 2.2 Características de la distribución de Fréchet.

Fuente: Elaboración propia.

2.2.3.3 Distribución Weibull

La distribución de Weibull que fue descubierta en el año 1927 por Maurice Fréchet (1927) y definida por el matemático sueco Waloddi Weibull (1951), y viene dada de la siguiente manera:

- Distribución acumulada de Weibull

$$F(x) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^{\xi}}, \text{ si } x \geq 0 \quad (2.14)$$

- Distribución de probabilidad de Weibull

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{\xi-1} e^{-\left(\frac{x-\mu}{\sigma}\right)^\xi}, \text{ si } x \geq 0 \quad (2.15)$$

Siendo ξ el parámetro de forma y $\sigma > 0$ el parámetro de escala de la distribución.

Características	
Media	$\mu + \sigma \Gamma\left(1 + \frac{1}{\xi}\right)$
Varianza	$\sigma^2 \left[\Gamma\left(1 + \frac{2}{\xi}\right) - \left(\Gamma\left(1 + \frac{1}{\xi}\right)\right)^2 \right]$
Moda	$\mu + \sigma \left(\frac{\xi - 1}{\xi}\right)^{\frac{1}{\xi}}, \text{ si } \xi > 1$
Mediana	$\mu + \sigma (\ln(2))^{\frac{1}{\xi}}$

Tabla 2.3 Características de la distribución de Weibull.

Fuente: Elaboración propia.

2.2.3.4 Distribución de Valores Extremos Generalizada (GEV)

Mises (1954) y Jenkinson (1955) propusieron una reformulación del teorema de Fisher y Tippet, en el cual se generaliza las familias de distribuciones de Gumbel, Fréchet y Weibull en una única distribución general que viene dada por la siguiente función:

$$G(z) = e^{-\left(1 + \xi \left(\frac{z-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}}, \text{ dado } 1 + \xi \left(\frac{z-\mu}{\sigma}\right) > 0 \quad (2.16)$$

Parámetros:

- Posición / Localización: rango y tendencia central (μ)
- Escala: tendencia central y dispersión (σ)
- Forma: dispersión y momentos de alto orden (ξ)

Dicha función generalizada posee el parámetro de forma (ξ) y según su valor dependerá a qué familia va a aproximar su función de distribución (Gumbel, Fréchet y Weibull).

Distribución:

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

Características	
Media	$\begin{cases} \mu + \sigma \frac{\Gamma(1 - \xi) - 1}{\xi} & , si \xi \neq 0, \xi < 0 \\ \mu + \sigma \gamma & , si \xi = 0 \\ No existe & , si \xi \geq 1 \end{cases}$ <p>Donde $\Gamma()$ es la función Gamma, y γ es la constante de Euler.</p>
Varianza	$\begin{cases} \sigma^2 \frac{(g_2 - g_1^2)}{\xi^2} & , si \xi \neq 0, \xi < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6} & , si \xi = 0 \\ No existe & , si \xi \geq \frac{1}{2} \end{cases}$ <p>Donde $g_k = \Gamma(1 - k\xi)$.</p>
Mediana	$\begin{cases} \mu + \sigma \frac{\ln(2)^{-\xi} - 1}{\xi} & , si \xi \neq 0 \\ \mu - \sigma \ln(\ln(2)) & , si \xi = 0 \end{cases}$
Cuantil p	$\begin{cases} \mu + \sigma \frac{\ln(\frac{1}{p})^{-\xi} - 1}{\xi} & , si \xi \neq 0 \\ \mu - \sigma \ln(\ln(\frac{1}{p})) & , si \xi = 0 \end{cases}$

Tabla 2.4 Características de la distribución de Valores Extremos Generalizada.

Fuente: Elaborado por autor.

2.2.4 Picos por encima del umbral (POT)

Definición: Sea X una variable aleatoria unidimensional, y sea U un umbral fijado, se dice que el suceso $\{X = x\}$ es una excedencia del umbral U , si se cumple que $x > U$.

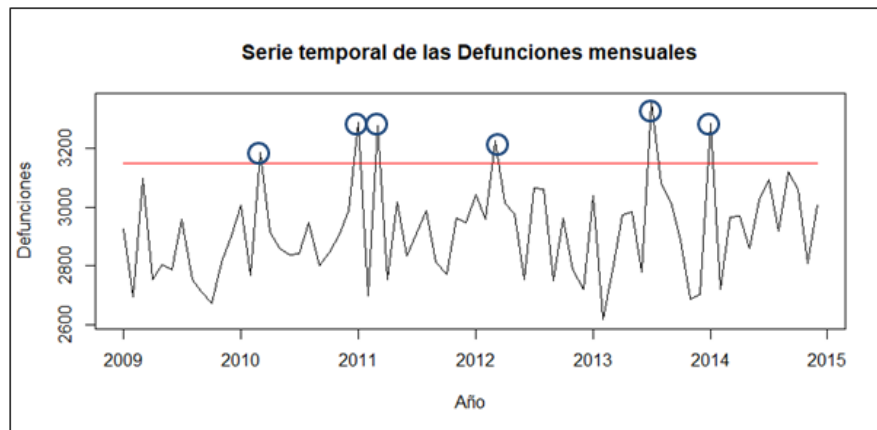


Figura 2.2 Ejemplo gráfico de selección de datos (Metodología de Picos por encima del Umbral).

Fuente: Elaborado por autor.

2.2.4.1 Distribución de Pareto Generalizada (GP)

La distribución de Pareto Generalizada tiene un uso específico que consiste en acoplar los datos de los modelos que usan un umbral. En otras palabras, es la distribución de los valores que son excedencias del umbral, la cual usa los mismos tres parámetros que se vieron en el método de Máximo por Bloques, los parámetros de localización (μ), de escala (σ) y de forma (ξ). Esta distribución viene dada por la siguiente forma:

- Distribución acumulada de Pareto Generalizada

$$F(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}, & \text{si } \xi \neq 0 \\ 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)}, & \text{si } \xi = 0 \end{cases} \quad (2.17)$$

Todo esto para $x \geq \mu$, cuando $\xi \geq 0$, y $x \leq \mu - \frac{\sigma}{\xi}$ cuando $\xi < 0$.

- Distribución de probabilidad de Pareto Generalizada

$$f(x) = \frac{1}{\sigma} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} \quad (2.18)$$

O

$$f(x) = \frac{\sigma^{1/\xi}}{(\sigma + \xi(x-\mu))^{\left(\frac{1}{\xi}+1\right)}} \quad (2.19)$$

Todo esto para $x \geq \mu$, cuando $\xi \geq 0$, y $\mu \leq x \leq \mu - \frac{\sigma}{\xi}$ cuando $\xi < 0$, donde $\mu \in \mathbb{R}$.

Características	
Media	$\mu + \frac{\sigma}{1-\xi}$, siempre que $\xi < 1$
Varianza	$\frac{\sigma^2}{(1-\xi)^2(1-2\xi)}$, siempre que $\xi < \frac{1}{2}$
Mediana	$\mu + \frac{\sigma(2^\xi - 1)}{\xi}$

Tabla 2.5 Características de la distribución de Pareto Generalizada.

Fuente: Elaboración propia.

Adicional, cuando la distribución de Pareto toma el valor de $\mu = 0$, es decir cuando tiene dos parámetros (σ, ξ) , viene dada por:

$$F(x) = \begin{cases} 1 - \left(1 - \frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}}, & \text{si } \left(1 - \frac{\xi x}{\sigma}\right) \geq 0, \xi \neq 0, \sigma > 0 \\ 1 - e^{-\left(\frac{x}{\sigma}\right)}, & \text{si } x \geq 0, \xi = 0, \sigma > 0 \end{cases} \quad (2.20)$$

Y su función de densidad es:

$$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 - \frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}-1}, & \text{si } \left(1 - \frac{\xi x}{\sigma}\right) \geq 0, \xi \neq 0, \sigma > 0 \\ \frac{1}{\sigma} e^{-\left(\frac{x}{\sigma}\right)}, & \text{si } x \geq 0, \xi = 0, \sigma > 0 \end{cases} \quad (2.21)$$

2.2.5 Niveles y periodos de retorno

Una de las razones por la que se concibe, que el análisis de valores extremos tenga tantos aplicativos en diferentes ramas de la ciencia es la estimación de los niveles y periodos de retorno, este último se lo considera como los cuantiles de los valores extremos.

Dado que, si Z_p o nivel de retorno, es el cuantil de orden p de una variable con distribución GEV, entonces p es la probabilidad de que haya superado Z_p una vez al año. El periodo de retorno $\frac{1}{p}$ es el número de unidades de tiempo, años en este caso, que transcurrirán en medida en que la variable supere por segunda ocasión el valor de Z_p .

2.2.6 Criterios de Información

Los criterios de información de Akaike (AIC) y Bayesiano (BIC) (Akaike, 1977) y (Schwarz, 1978) son medidas de la calidad de los modelos estadísticos que ayudan a seleccionar un modelo entre un conjunto finito de modelos. El BIC es más sensible al número de parámetros utilizados por tamaño de la muestra.

$$AIC = 2k - 2 * \ln(L) \quad (2.22)$$

$$BIC = -2 * \ln(L) + k \ln(n) \quad (2.23)$$

Siendo k el número de parámetros, L el máximo valor de la función de verosimilitud del modelo (Hald, 1999) y n el tamaño de la muestra. Por regla general, el modelo que mantenga lo menor posible los valores de AIC y BIC será el óptimo.

2.2.7 Prueba de Razón de Verosimilitud

Es una prueba de hipótesis que permite contrastar dos modelos, uno con p y otro con $p+1$ parámetros, siendo la hipótesis nula:

$H_0 =$ El parámetro adicional del segundo modelo es igual a cero

2.2.8 Método de Imputación de datos: Hot-Deck

Cuando sea necesario realizar alguna imputación de datos y se desea mantener la distribución de probabilidad de las variables con datos incompletos, se puede recurrir al algoritmo de imputación no paramétrico denominado Hot-Deck (Madow, Nisselson y Olkin, 1983). Este método fue desarrollado por estadísticos de encuestas y tiene como objetivo llenar los registros vacíos con información de campos de registros completos, y los datos faltantes se reemplazan a partir de una selección aleatoria de valores observados, lo que evita que se establezcan sesgos en la varianza del estimador.

2.2.9 Métodos de Clustering

El análisis clúster agrupa las observaciones de un conjunto de datos, basado solo en la información que describen las observaciones provistas por los mismos datos, con el objetivo de que las observaciones pertenecientes a un grupo sean similares (estén relacionadas) entre sí y sean diferentes a las observaciones pertenecientes a otro grupo.

2.2.9.1 Método de clustering: K-Means

Según Hartigan (1974) el método de clustering llamado K-Means es el más utilizado y conocido debido a su simplicidad de aplicación y su eficiencia; creado por MacQueen en el año 1967. Su procedimiento consiste en la clasificación de un conjunto de objetos en un determinado número K de clústeres y su nombre se debe a que cada uno de los clústeres viene representado por la media (o media ponderada) de sus puntos, es decir por su centroide. La ventaja de usar centroide es que su representación tiene un significado gráfico y estadístico. El algoritmo se resume en 4 etapas:

- Etapa 1: Se elige aleatoriamente K objetos que forman así los K clústeres iniciales. Para cada clúster k, el valor inicial del centro es igual a x_i . Con los x_i únicos de D_n pertenecientes al clúster.

$$\hat{s} = \operatorname{argmin} \|u_k - x\|^2 \quad (2.29)$$

- Etapa 2: Se reasignan los objetos del clúster. Para cada objeto x , el prototipo que se le asigna es el que es más próximo al objeto, según una medida de distancia.
- Etapa 3: Una vez que todos los objetos son colocados, se recalculan los centros de los K clústeres o también llamados Baricentros.
- Etapa 4: Repetir las etapas 2 y 3 hasta que no se hagan más reasignaciones.

2.2.9.2 Método de clustering: Ward

El método de Ward es un procedimiento jerárquico en el cual, en cada etapa, se unen dos clústers los cuales se tengan el menor incremento en el valor de la suma de los cuadrados de las diferencias, dentro de cada clúster, de cada individuo al centroide del clúster. Notemos por:

- x_{ij}^k : al valor de la j -ésima variable sobre el i -ésimo individuo del k -ésimo clúster, suponiendo que dicho clúster posee n_k individuos.
- m^k : al centroide del clúster k , con componentes m_j^k .
- E_k : a la suma de cuadrados de los errores del clúster k , o sea la distancia euclídea al cuadrado entre cada individuo del clúster k a su centroide.

$$E_k = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k - m_j^k)^2$$

- E : a la suma de cuadrado de los errores para todos los clústers, o sea, si suponemos que hay h clústers.

$$E = \sum_{k=1}^h E_k$$

El proceso comienza con m clústers, cada uno de los cuales está compuesto por un solo individuo, por lo que cada individuo coincide con el centro del clúster y por lo tanto en este primer paso se tendrá $E_k = 0$ para cada clúster y con ello, $E = 0$. El objetivo del método de Ward es encontrar en cada etapa aquellos dos clústers cuya unión proporcione el menor incremento en la suma total de errores, E . Supongamos ahora que los clústers C_p y C_q se unen resultando un nuevo clúster C_t . Entonces el incremento de E será:

$$\Delta E_{pq} = E_t - E_p - E_q = \frac{n_p n_q}{n_t} \sum_{j=1}^n (m_j^p - m_j^q)^2$$

Así el menor incremento de los errores cuadráticos es proporcional a la distancia euclídea al cuadrado de los centroides de los clústers unidos.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS:

En la presente sección se realiza un análisis descriptivo de los sismos registrados en el Ecuador. Para ello se tomó en cuenta la frecuencia de los sismos en las diferentes provincias, y variables como la magnitud (medida en la escala de magnitud de momento M_w) y profundidad (medida en kilómetros).

3.1 Análisis descriptivo

Las magnitudes de los sismos ocurrido en el Ecuador desde 1906 hasta 2017, tuvieron una fuerte acumulación de observaciones entre el rango 3 y $4M_w$. (figura 3.1), con media de $3.53M_w$. y desviación $0.56M_w$., teniendo una curva de distribución que se asemeja a la una función exponencial, cumpliendo con la ley de Gutenberg-Richter.

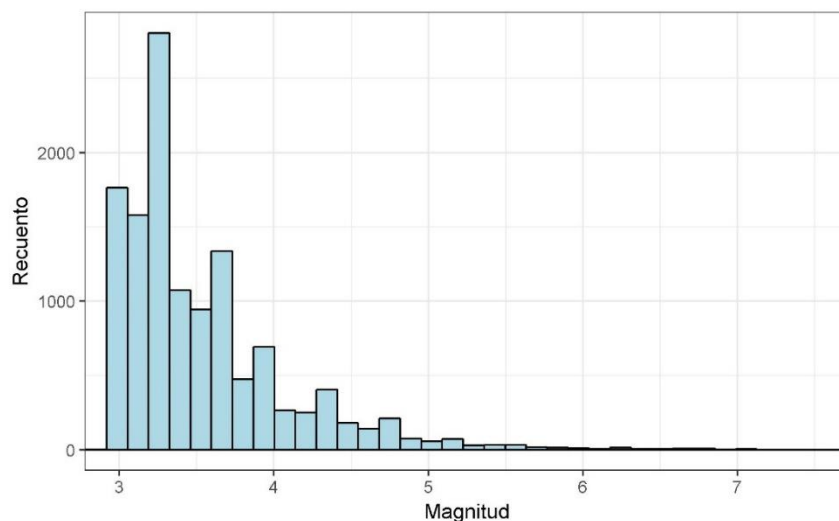


Figura 3.1 Histograma: Magnitud de los sismos ocurridos en el Ecuador (1906-2017).

Fuente: Elaboración propia.

El 75% de sismos registrados tuvieron una magnitud menor o igual a $4M_w$., además de que existe una proporción alta de datos atípicos que tienen una magnitud superior a los $5M_w$. (figura 3.2)

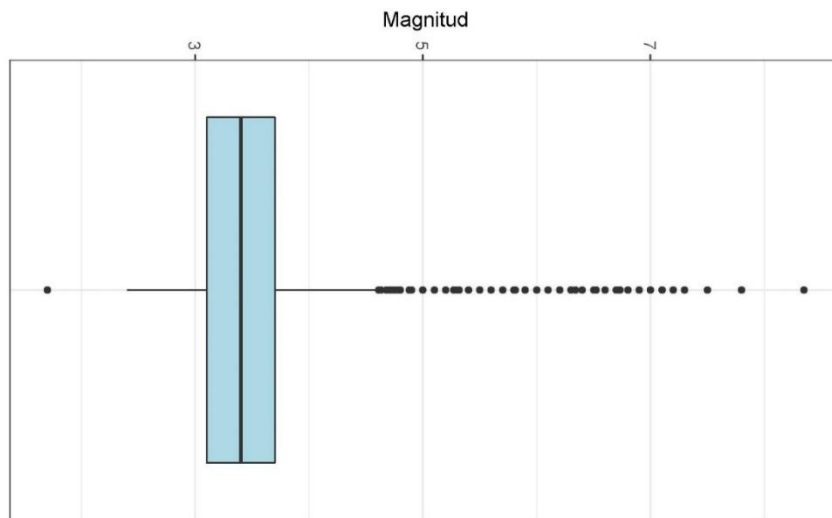


Figura 3.2 Diagrama de caja: Magnitud de sismos ocurridos en el Ecuador (1906-2017).

Fuente: Elaboración propia.

Las provincias de Esmeraldas, Manabí, Guayas, Los Ríos, Pichincha, Cotopaxi, Tungurahua y Morona Santiago presentaron acumulaciones de epicentros; en el resto del Ecuador también se registraron epicentros, aunque no tan concentrados como en las provincias antes mencionadas (figura 3.3). Además con respecto a su profundidad se evidenciaron ciertos comportamientos de los epicentros, entre los cuales se puede mencionar que la mayor parte de sismos ocurridos (80% aproximadamente) han tenido profundidades de 0 a 50Km. y estos se encuentra dispersos por todo el territorio ecuatoriano; los sismos ocurridos a la altura de la ciudad de Quevedo tuvieron la una profundidad de aproximadamente 100Km.; a la altura de las provincias de Pastaza y Morona Santiago se registraron sismos de profundidad de 200-250 Km. Por último, es válido recalcar que solo se han reportado 3 sismos con profundidades mayores a 300Km.

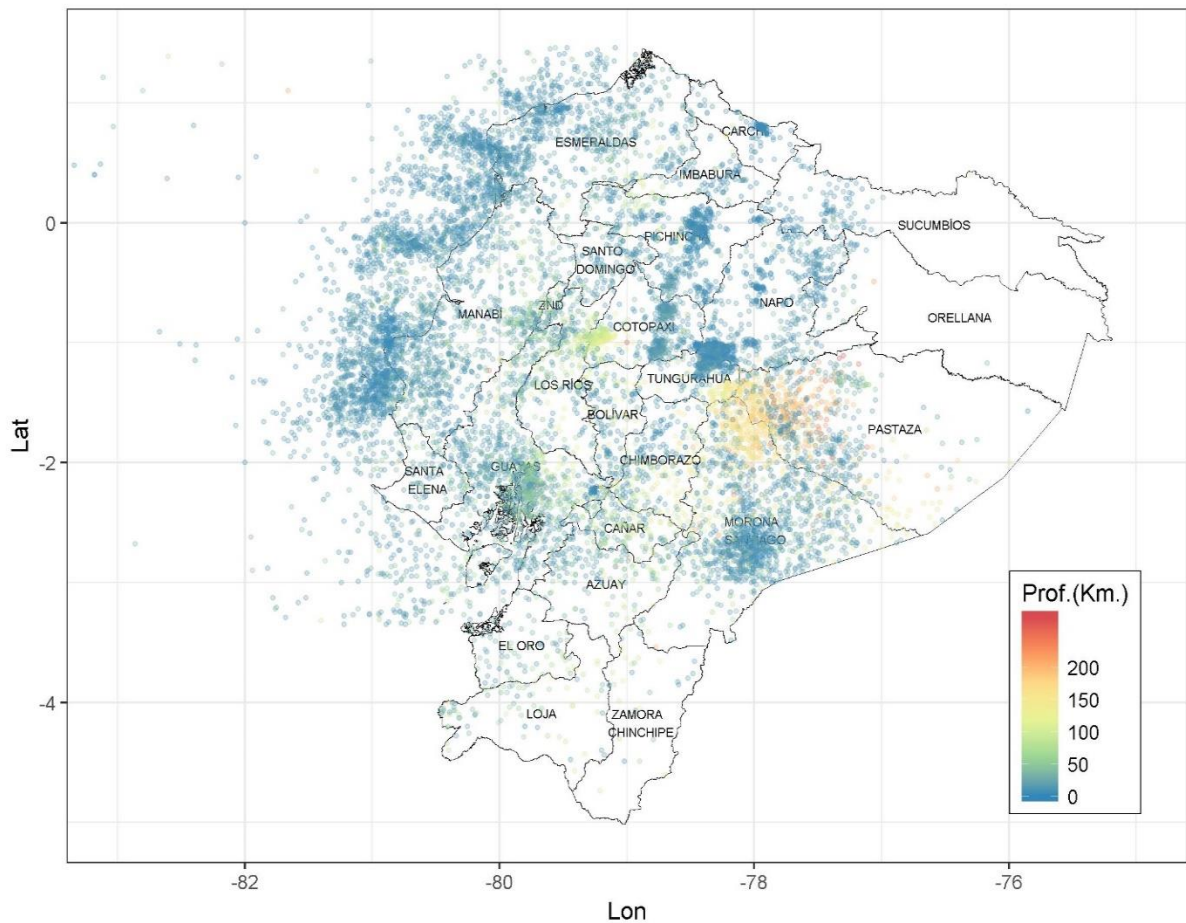


Figura 3.3 Mapa: Epicentros de los sismos ocurridos en el Ecuador (1906-2017).
Fuente: Elaboración propia.

En la costa ecuatoriana se evidenció que, existen dos puntos donde se acumulan los epicentros de sismos, uno está ubicado en la parte suroeste de Manabí a la altura de los cantones Manta, Jipijapa, Montecristi; mientras que el otro punto se encuentra en Esmeraldas, a la altura del cantón de Atacames, adicional a esto, hay un punto de mayor concentración de epicentros en la región Sierra a la altura de las fronteras de las provincias de Cotopaxi, Napo y Tungurahua, otro punto de acumulación en Morona Santiago y dos puntos de acumulación en las provincias de Los Ríos y Pichincha (figura 3.4).

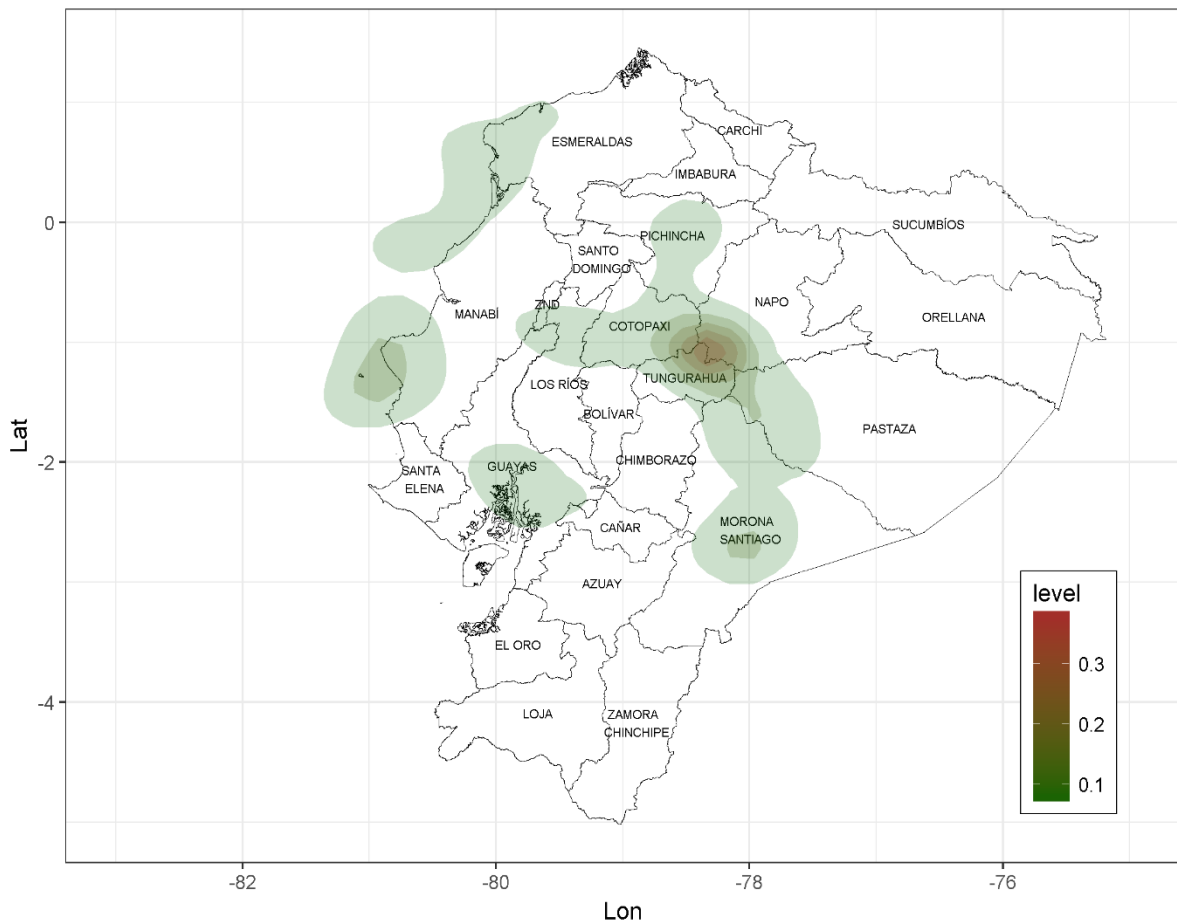


Figura 3.4 Gráfico de curvas de nivel: Concentración de sismos ocurridos en el Ecuador (1906-2017).
 Fuente: Elaboración propia.

Segmentando la figura anterior, según la intensidad de los sismos ocurridos, se obtiene la figura 3.5, la cual indica que los sismos ocurridos en el punto a la altura de Cotopaxi fueron de intensidad leve en su mayoría, puesto que en el gráfico de intensidad moderada e intensidad grave se disipa dicha concentración, debido a que estos sismos son de tipo volcánico. Adicionalmente, se puede descomponer en tres grupos los puntos de concentración de epicentros de la costa, puesto que el primer grupo se sitúa al suroeste de Manabí entre los cantones Manta y Montecristi, tiene una intensidad moderada, y al llegar a intensidad grave, dicho punto de concentración es más intenso que el segundo grupo el cual tiene como ubicación la provincia de Esmeraldas a la altura de los cantones Muisne y Atacames, este punto al llegar a una intensidad grave la concentración de epicentros incrementa. El tercer grupo está ubicado en la provincia del

Guayas a la altura del cantón Guayaquil, tiene una intensidad leve y al llegar a la intensidad moderada y grave este punto se disipa.

También se puede observar los puntos de acumulación de la provincia de Los Ríos y Pichincha los cuales tienen un comportamiento parecido al de la provincia del Guayas, puesto que estos al pasar a la intensidad moderada y grave se disipan. El punto de concentración más significativo se encuentra ubicado cerca de las fronteras de las provincias de Cotopaxi, Napo Tungurahua y Pastaza, este resulta de la alta actividad volcánica que ocurre en esa zona sin tomar en cuenta la magnitud, intensidad o profundidad del sismo.

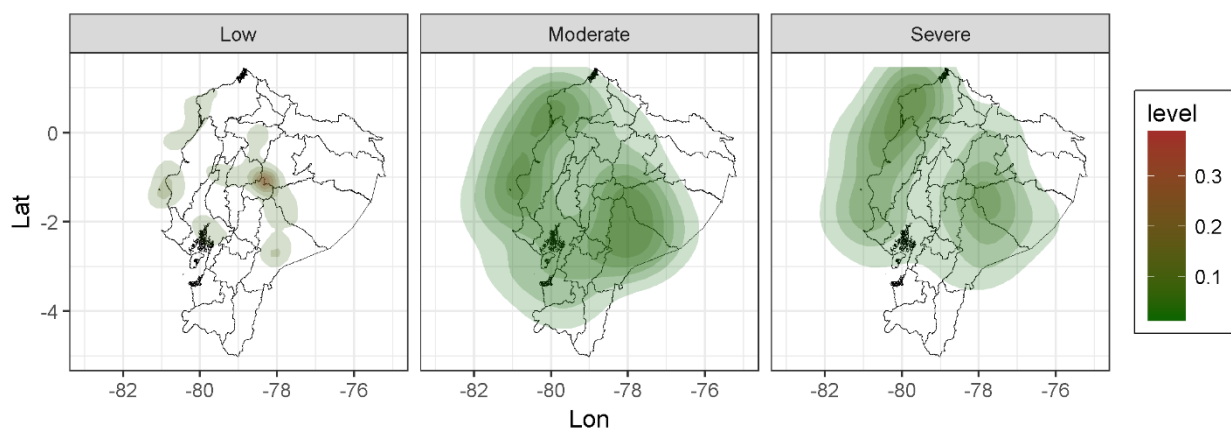


Figura 3.5 Gráfico de curvas de nivel: Concentración de sismos por intensidad en el Ecuador (1906-2017).

Fuente: Elaboración propia.

En la zona marítima del Ecuador se reportaron 2819 sismos, cantidad que es mayor teniendo solo en cuenta la zona continental de las provincias que componen la región Costa. De acuerdo con la figura 3.6, entre las provincias que han registrado la mayor cantidad de ocurrencias sísmicas de intensidad leve tenemos Morona Santiago y Napo, seguido de las provincias Guayas, Manabí, Pastaza, Esmeraldas, Pichincha, Cotopaxi, Tungurahua y Los Ríos. Las provincias de Esmeraldas, Morona Santiago y Pastaza presentaron la mayor concentración de sismos de intensidad moderada y grave.

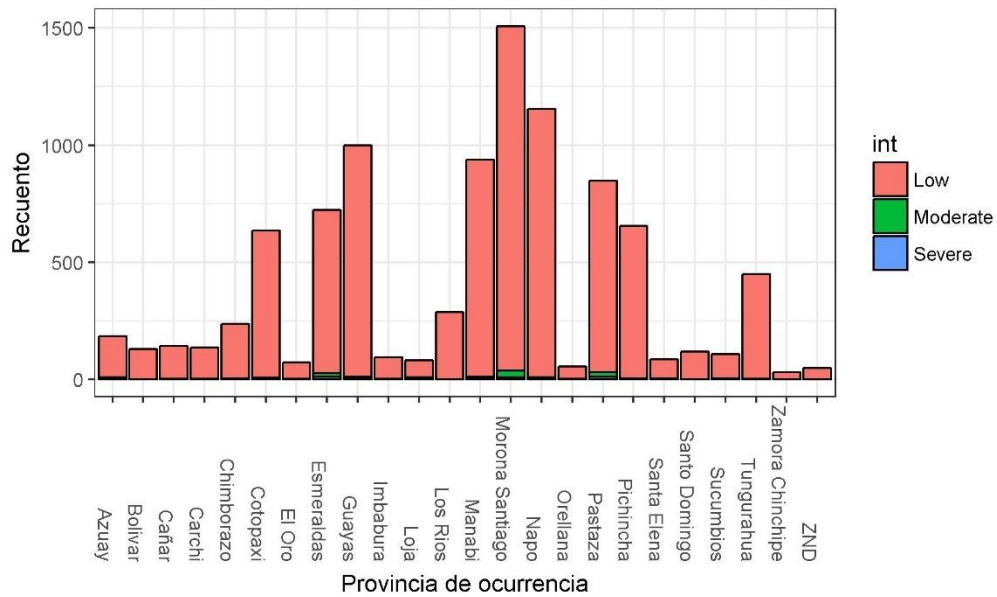


Figura 3.6 Gráfico de barras: Recuento de sismos por provincias e intensidad (1906-2017).
 Fuente: Elaboración propia.

En el Ecuador se han registrado epicentros con magnitudes que sobrepasan los 5 M_w . desde la década de 1950, desde la década de 1960 a 2000 las máximas magnitudes sísmicas se han mantenido por debajo de los 7 M_w . (figura 3.7).

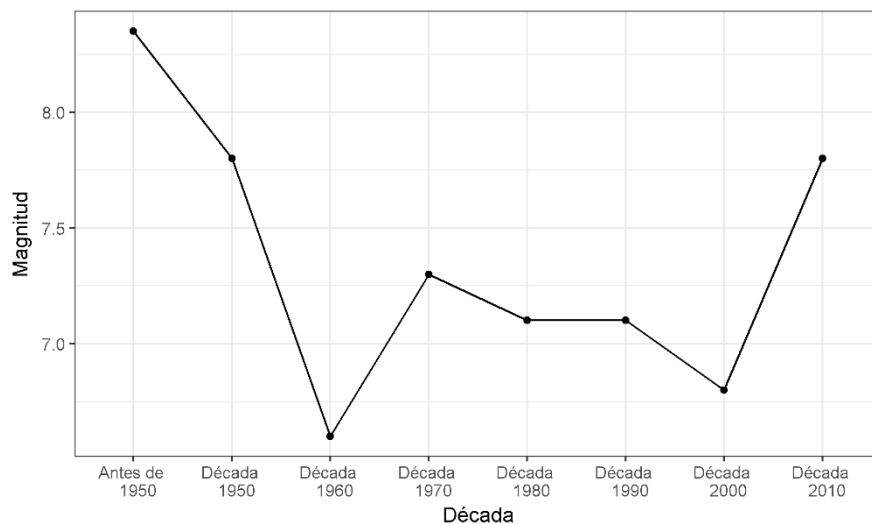


Figura 3.7 Serie temporal: Máximas magnitudes sísmicas por década en el Ecuador.
 Fuente: Elaboración propia.

En la década de 1990, las provincias con mayor actividad sísmica fueron Morona Santiago y Napo, la década con mayor actividad sísmica ha sido la década del 2000, en la cual se concentró la actividad sísmica en provincias como Morona Santiago y Napo, seguido de una actividad sísmica moderada en las provincias Cotopaxi, Guayas, Pichincha, Manabí, Pastaza, Tungurahua y Esmeraldas. En la década del 2010 la mayor cantidad de ocurrencias sísmicas han sido reportadas en la Provincias de Morona Santiago, Guayas, Manabí, Esmeraldas, Pastaza, Pichincha, Carchi y Napo (figura 3.8).

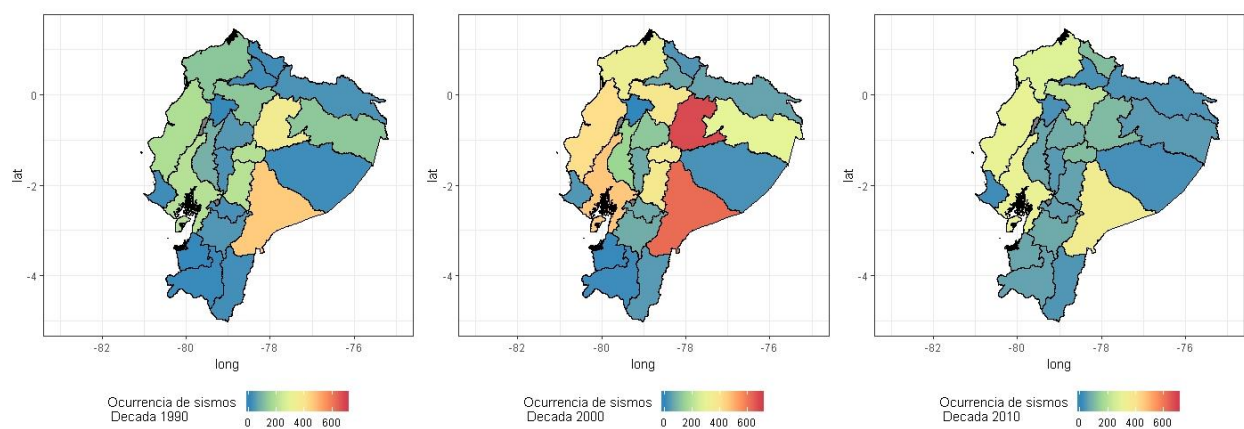


Figura 3.8 Mapa: Actividad sísmica en el Ecuador de las tres últimas décadas.
Fuente: Elaboración propia.

De acuerdo con la figura 3.9, la provincia de Loja posee el territorio con mayor media en magnitud, seguido de las provincias de El Oro y Zamora Chinchipe, estas junto con la provincia de Orellana son las que presentan mayor variación en las magnitudes de los sismos, sin embargo, también se observa que la media de magnitud por provincia está entre 3 y $4M_w$.

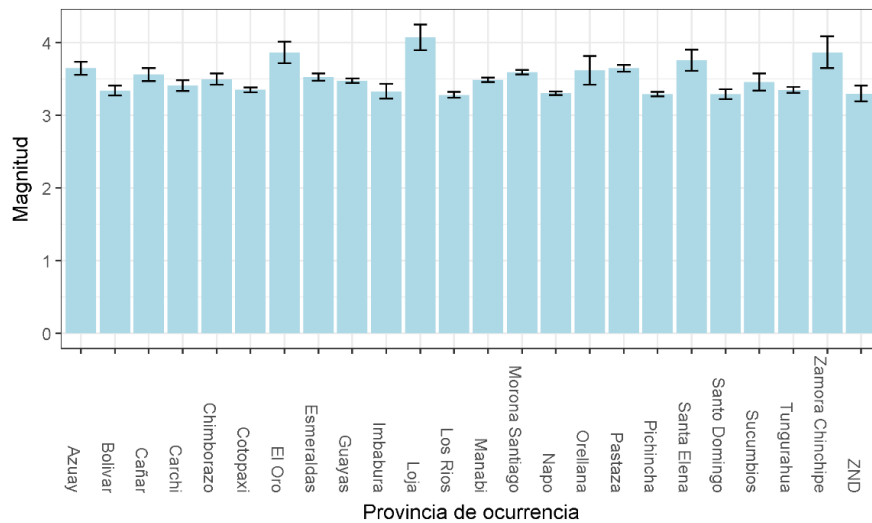


Figura 3.9 Gráfico de barras: Media de la magnitud de sismos por provincia, +/- 2 error estándar.
Fuente: Elaboración propia.

Se pueden destacar provincias como Esmeraldas, Chimborazo, Morona Santiago, Napo, Pastaza y Manabí, las cuales disponen de las máximas magnitudes de sismos ocurridas en el territorio ecuatoriano, además cabe mencionar provincias como Los Ríos, Bolívar, Santo Domingo, donde la máxima magnitud de sismo registrada es de intensidad leve (figura 3.10).

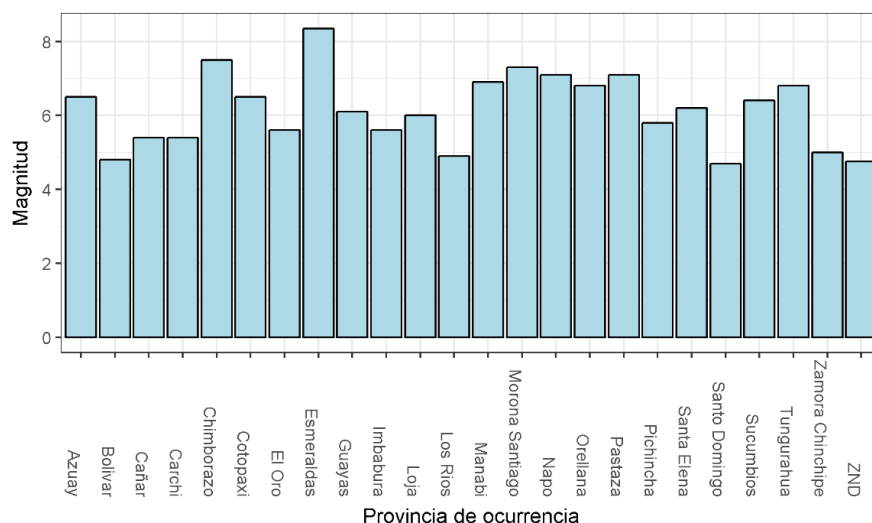


Figura 3.10 Gráfico de barras: Máximas magnitudes de sismos por provincias (1906-2017).
Fuente: Elaboración propia.

En la figura 3.11 se presenta el diagrama de caja de la magnitud de los sismos por cada una de las provincias, el cual indica que la mediana de las magnitudes ocurridas en las diferentes provincias del Ecuador ronda entre los 3 y 4 grados de magnitud M_w ; además que las provincias con mayor variación son Loja, Orellana, Pastaza, Esmeraldas, Santa Elena y el Oro, y además de estas existen provincias donde valdría la pena analizar sus datos atípicos como Azuay, Chimborazo, Cotopaxi, Guayas, Sucumbíos y Tungurahua, puesto que hay observaciones muy alejadas de sus respectivas cajas, y dan indicio de terremotos de intensidad moderada y grave a lo largo de los años.

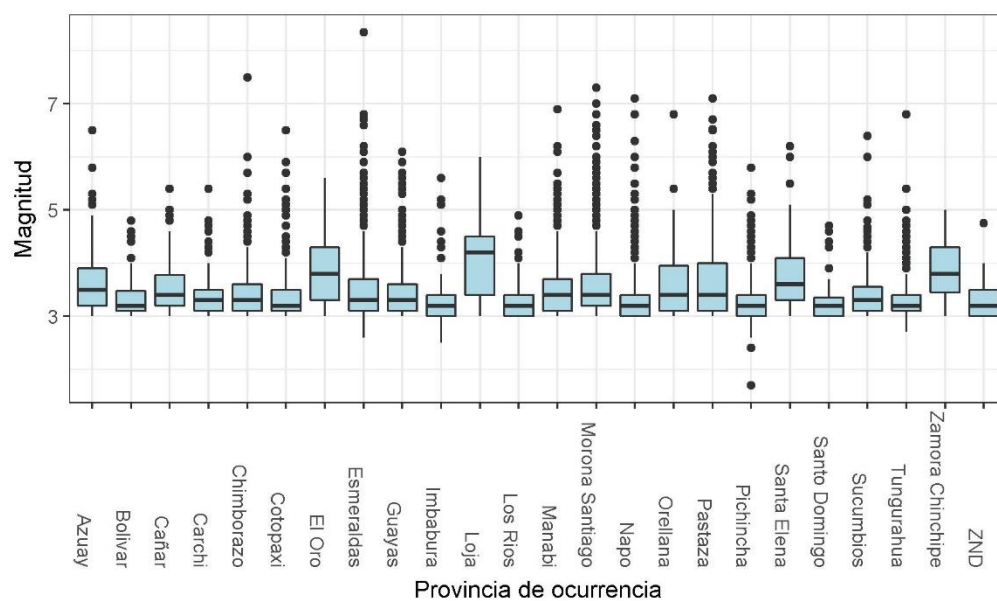


Figura 3.11 Diagrama de Caja: Magnitud de sismos por provincias (1906-2017).
Fuente: Elaboración propia.

3.2 Modelos de distribución de valores extremos

En esta sección se presentan las estimaciones de los modelos teóricos de Análisis de Valores Extremos de los sismos ocurridos en el Ecuador desde el año 1938, desde los dos enfoques vistos en capítulo 1, Máximos por Bloque y Picos Sobre el Umbral. Primero se han realizado los modelos sobre los sismos con epicentros en el Ecuador tanto en la zona marítima y continental de manera general, posteriormente se ha usado un método de clustering para particionar al Ecuador en diferentes zonas, según el nivel de

peligrosidad sísmica y luego se ha obtenido el mejor método del Análisis de Valores Extremos.

3.2.1 Modelos de Máximos por Bloques

En este análisis se ha fijado como variable constante de bloque, al año en el cual se produjo el sismo, es decir se trabajó con la magnitud máxima por año en el Ecuador; la falta de registros de sismos en ciertos años fue uno de los problemas más importantes que ha afectado el análisis, esto debido a que se dificulta el aplicar un modelo de mediante el enfoque BM; para solucionar dicho problema se optó por reemplazar los máximos con el método de imputación de Hot-Deck; para este análisis se ha imputado el 20% de los datos desde el año 1938.

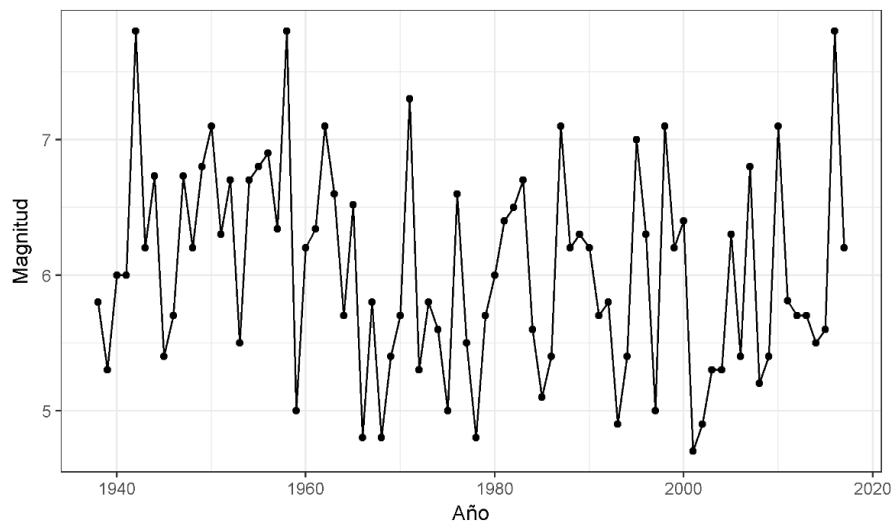


Figura 3.12 Serie temporal: Magnitud máxima reportada por año (con imputación por método Hot-Deck).

Fuente: Elaboración propia.

Una vez solucionado el problema de los años sin observaciones, se procedió a aplicar la primera técnica de análisis de valores extremos de Máximos por Bloque a todo el Ecuador, mediante el uso de la distribución GEV, estimando los parámetros por el método de máxima verosimilitud, teniendo como resultado que la distribución a la que se ajustan las máximas magnitudes es una distribución de Weibull, puesto que, el parámetro de forma es menor que 0 (Tabla 3.1), sin embargo, debido a que el parámetro de forma de este modelo es cercano cero pudo no ser significativo, se diseñó también un modelo a partir de la distribución de Gumbel (tabla 3.2).

	Parámetros		
	Localización	Escala	Forma
Estimación	5.6618794	0.6850781	- 0.1489366
Error estándar	0.09465122	0.07023308	0.11048562
AIC	165.5192	BIC	172.3072

Tabla 3.1 Modelo GEV aplicado a las máximas magnitudes en el Ecuador.

Fuente: Elaboración propia.

	Parámetros		
	Localización	Escala	
Estimación	5.6086780	0.6505541	
Error estándar	0.08157150	0.06036888	
AIC	165.0404	BIC	169.5658

Tabla 3.2 Modelo Gumbel aplicado a las máximas magnitudes en el Ecuador.

Fuente: Elaboración propia.

El modelo Gumbel tuvo menor AIC y BIC con respecto al modelo Weibull, por lo que se procedió a contrastar estos dos modelos para verificar si ξ es igual a 0.

Test de Razón de Verosimilitud	
Razón de verosimilitud	1.5213
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.0915

Tabla 3.3 Test de razón de verosimilitud (Gumbel vs. GEV)

Fuente: Elaboración propia.

Se concluyó que a un nivel de confianza del 95%, no existe suficiente evidencia estadística para rechazar la hipótesis de que el parámetro de forma sea igual a cero, por lo que se usará la distribución de Gumbel para modelar las máximas magnitudes de los sismos en el Ecuador. También se elaboraron los gráficos de densidad empírica vs teórica, además del gráfico qqplot, y la gráfica de niveles de retorno del modelo escogido (figura 3.13).

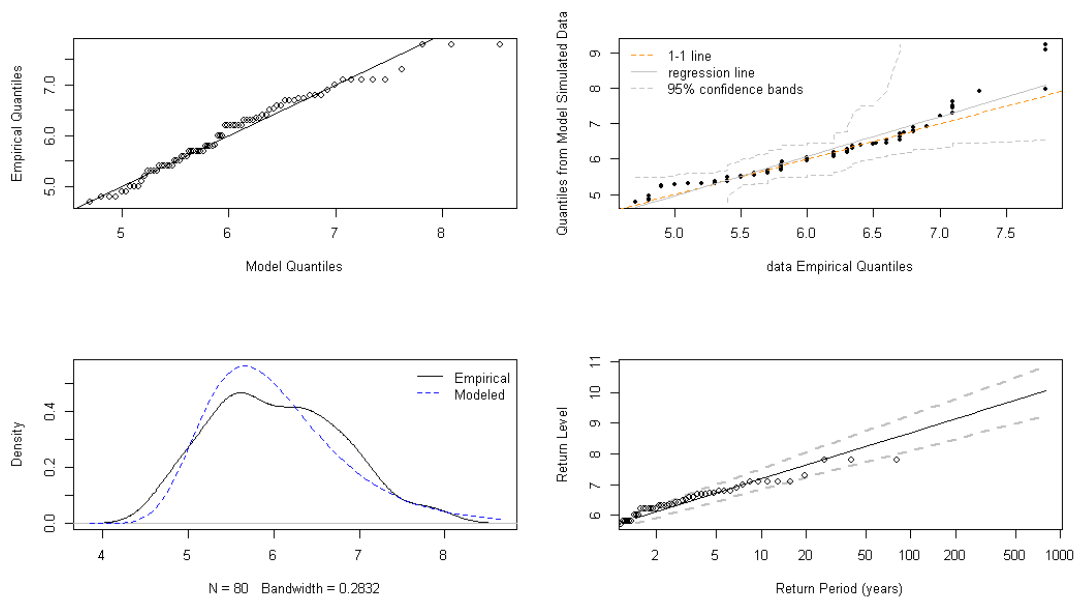


Figura 3.13 Gráficos del modelo por distribución de Gumbel.
Fuente: Elaboración propia.

Al analizar el gráfico de niveles de retorno, se pueden observar los periodos de retorno versus su respectivo nivel, además del intervalo de confianza resultante. Se puede apreciar que, para 50 años de retorno, habría un sismo de al menos 7.83 grados en escala M_w y que, para un periodo de retorno de 100 años, tendríamos un sismo de una magnitud máxima de al menos 8.08 M_w . Como era de esperar, el intervalo de confianza cada vez va aumentando con el pasar del tiempo, debido a que, la incertidumbre aumenta. Una de las desventajas de este enfoque es que no posee una cota que limite el grado de magnitud con el pasar el tiempo, es decir, si se habla de periodos de retorno de 1000 años o más, se esperarán resultados de magnitudes mayores a 10 M_w , lo cual es poco probable que suceda.

Periodo de Retorno	de	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años		5.893077	6.047028	6.200980
5 años		6.553322	6.736285	6.919248
10 años		6.909030	7.127081	7.345132
50 años		7.435998	7.835298	8.234598
100 años		7.572085	8.081677	8.591269

Tabla 3.4 Periodos de retorno modelo GEV.
Fuente: Elaborado por autor.

3.2.2 Modelo de Picos Sobre un Umbral

También se procedió a analizar un modelo basado en el enfoque POT utilizando la distribución de Pareto Generalizada; se elaboró la gráfica de Vida Media Residual que detalla la media de residuos tomando en cuenta varios umbrales de la misma muestra, de los cuales se escoge aquel umbral que represente la mayor diferencia con respecto al anterior o anteriores, es decir un cambio drástico; también se diseñó el gráfico de Variación del Parámetro de Forma, el cual permite ver cómo varía la estimación del parámetro de forma con diferentes umbrales.

La grafica de Vida Media Residual presentó un cambio de crecimiento a partir de $4M_w$ y de decrecimiento a partir de $6M_w$. (figura 3.14), el gráfico de Variación del Parámetro Forma (figura 3.15) ha evaluado el modelo POT con diferentes umbrales, en esta se puede observar que hasta un umbral de $4.88M_w$ el parámetro de forma es mayor a cero y no presenta mucha variación; mientras que con un umbral a partir de $4.9 M_w$ el parámetro de forma será negativo. El umbral escogido para trabajar fue de $5 M_w$. debido a que se encuentra en un intervalo que no presenta mucha variación y el tamaño de muestra resultante de la selección de este umbral es considerado adecuado diseñar el modelo.

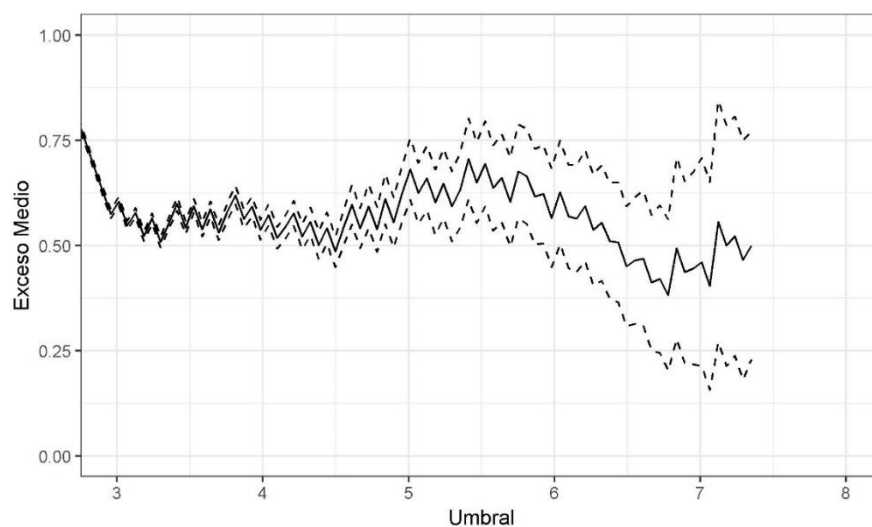


Figura 3.14 Vida media residual de la magnitud de los sismos ocurridos en el Ecuador.

Fuente: Elaboración propia.

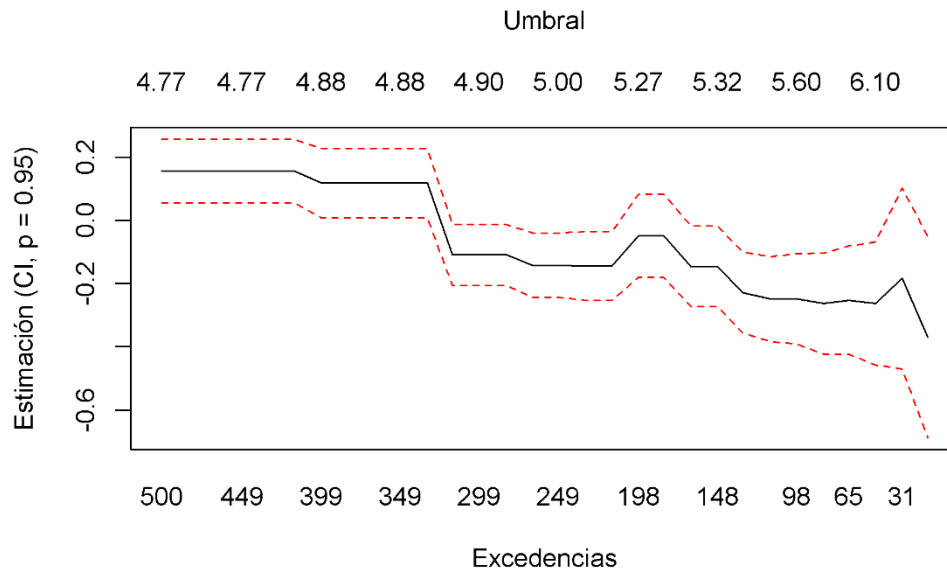


Figura 3.15 Variación del parámetro de forma del modelo POT aplicado a la magnitud de los sismos ocurridos en el Ecuador.

Fuente: Elaboración propia.

	Parámetros	
	Escala	Forma
Estimación	0.7903158	-0.1422843
Error estándar	0.06403416	0.05467631
<hr/>		
AIC	345.0551	BIC 352.2814

Tabla 3.5 Modelo GP aplicado a las magnitudes de los sismos ocurridos en el Ecuador.

Fuente: Elaboración propia.

A su vez se obtiene la distribución empírica versus la teórica de Pareto Generalizada, junto con los Gráficos quantil-quantil del respectivo modelo (figura 3.16).

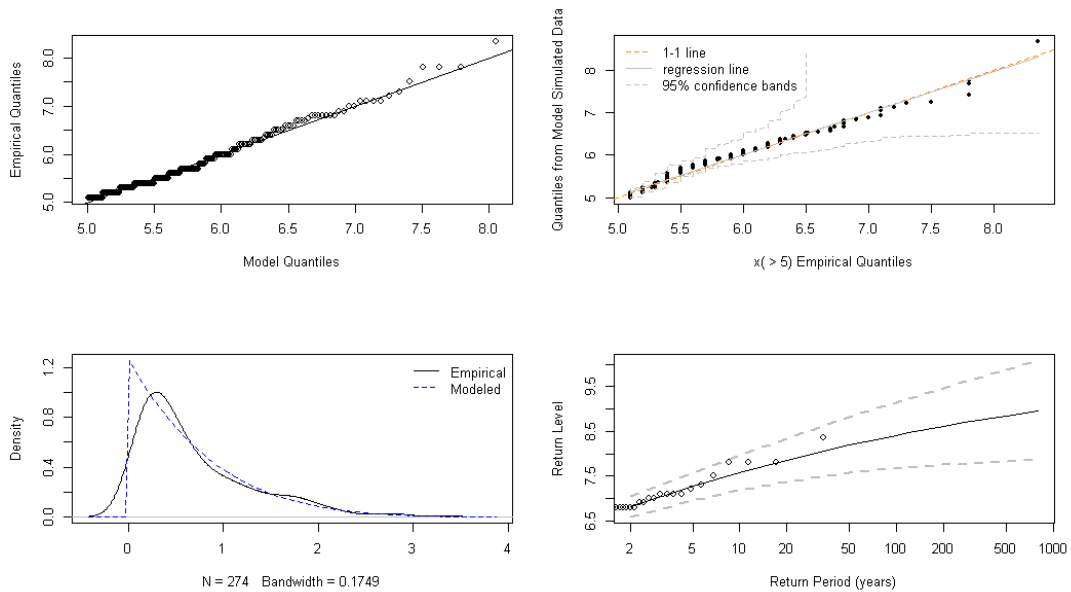


Figura 3.16 Gráficos del modelo por distribución de Pareto Generalizada.

Fuente: Elaboración propia.

Al analizar el gráfico de niveles de retorno se puede observar los periodos de retorno versus su respectivo nivel, además del intervalo de confianza resultante; un sismo como el ocurrido el pasado 16 de abril de 2016, de aproximadamente 7.8 grados M_w o mayor tiene un periodo de retorno de 25 años, con una probabilidad de ocurrencia en un año cualquiera de 1/25.

Periodo de Retorno	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años	6.580419	6.809750	7.039082
5 años	6.962463	7.267479	7.572496
10 años	7.192501	7.576184	7.959867
25 años	7.428844	7.940230	8.451615
50 años	7.564389	8.185752	8.807115
100 años	7.668870	8.408215	9.147560

Tabla 3.6 Periodos de retorno del modelo GP.

Fuente: Elaboración propia.

3.3 Zonificación y comparación de modelos en el Ecuador

Se procedió a dividir por zonas características al Ecuador, mediante los métodos de K-Medias y Ward, el cálculo de los índices de Silhouette y Calinski para las agrupaciones

resultantes con el método de Ward presentaron problemas para ser calculados debido a limitaciones computacionales, por lo cual no se pudo obtener un número óptimo de clústers mediante estos criterios, sin embargo, no se presentó esta complicación al evaluar las agrupación resultante de la aplicación del algoritmo de K-Medias.

Pese a que visualmente al proyectar los epicentros de los sismos pertenecientes a cada clúster sobre las coordenadas de longitud y latitud se obtuvieron agrupaciones similares, esto se puede apreciar en la figura 3.20 y 3.21; se escogió la clasificación por K-Medias debido a que el tamaño de los clústeres presentaba mayor variabilidad con respecto al método de Ward.

El gráfico de la figura 3.17 permite observar el comportamiento de la suma de cuadrados entre clúster obtenidos a medida que se aumenta el número de clúster (K). El criterio para seleccionar el número de clúster mediante esta gráfica es escoger el punto en donde exista un “codo”, por lo cual los posibles valores van de K=4 a K=6.

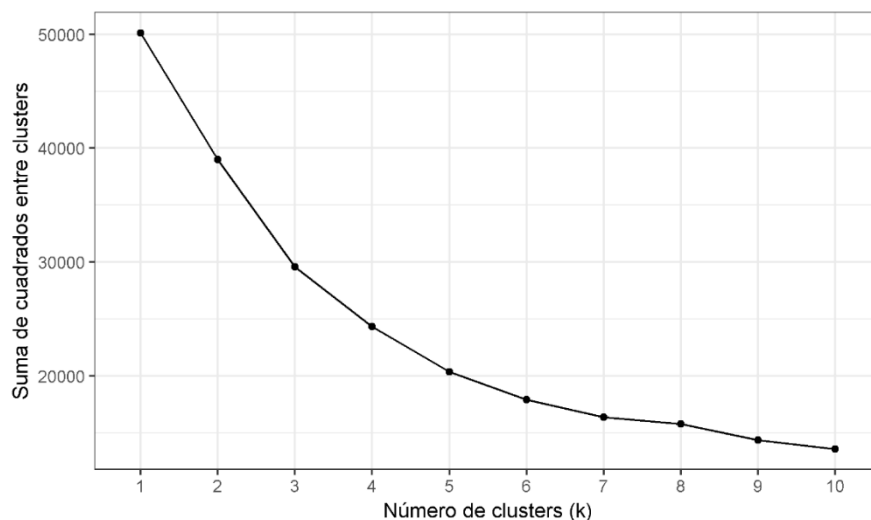


Figura 3.17 Suma de Cuadrado entre Clústers (K-Medias).

Fuente: Elaborado por autor.

El índice de Silhouette (figura 3.18), el cual permite interpretar y validar la consistencia de los datos dentro del clúster ya que este valor es una medida de cuan similares son los objetos que pertenecen a un mismo clúster; para el método de K-Medias este índice

alcanzó su máximo con una cantidad de clústeres $K=7$, sin embargo, los valores de este índice con $K=5$ y $K=6$ se encuentran muy cercanos al óptimo.

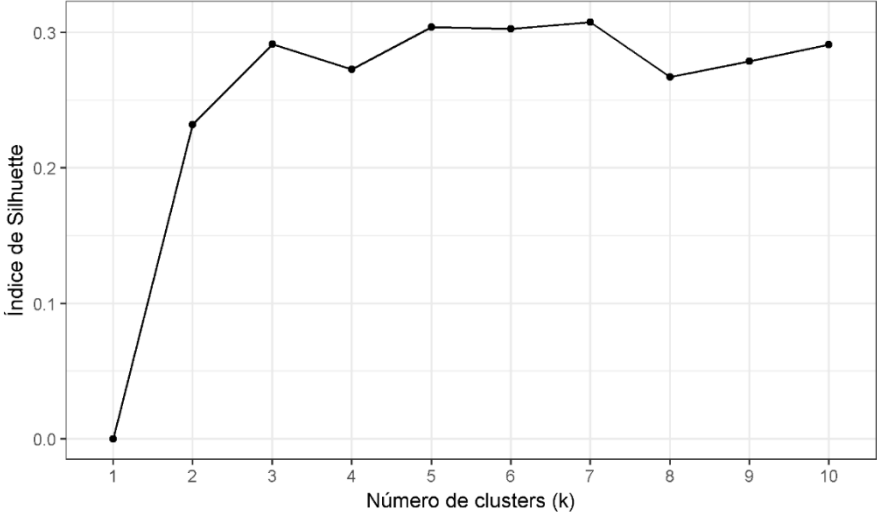


Figura 3.18 Índice de Silhouette (K-Medias).
Fuente: Elaborado por autor.

Además, también se calculó el índice de Calinski (figura 3.19), cuyo máximo se obtuvo con una cantidad de clústeres $K=5$.

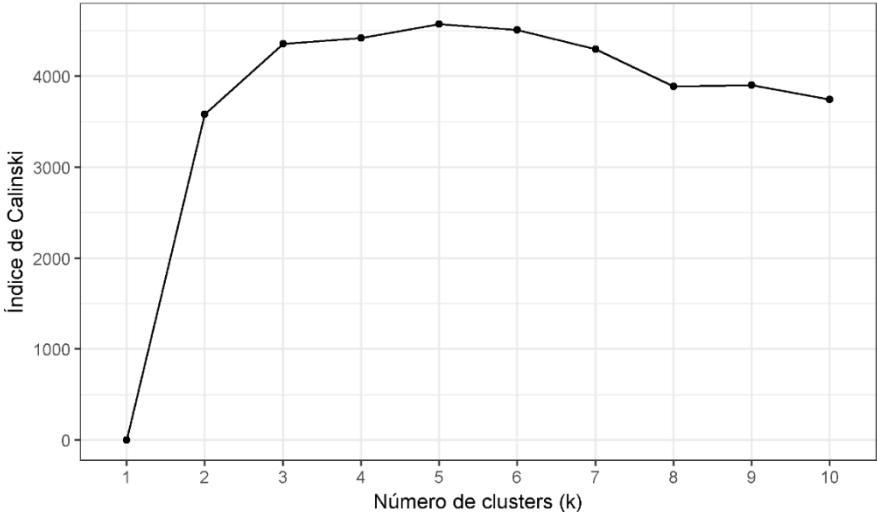


Figura 3.19 Índice de Calinski (K-Medias).
Fuente: Elaborado por autor.

Al tomar en cuenta el objetivo principal de elaborar una zonificación del territorio ecuatoriano de acuerdo con las características de los sismos que se generan allí, inicialmente se analizaron las agrupaciones resultantes de la aplicación del método de K-Medias y Ward con $K=7$.

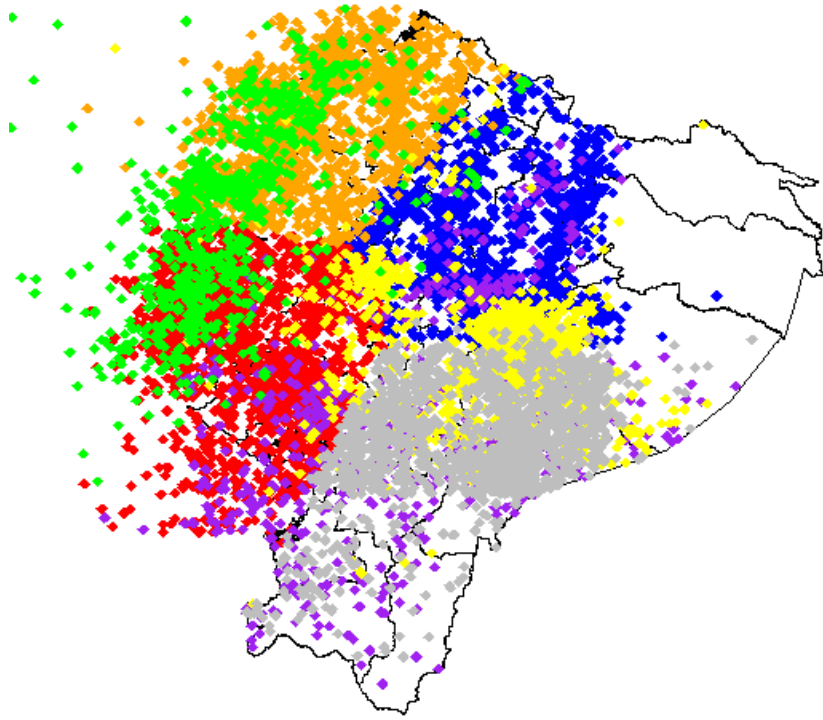


Figura 3.20 Gráfico de los clústers por K-Medias de los sismos ocurridos en Ecuador.
Fuente: Elaborado por autor.

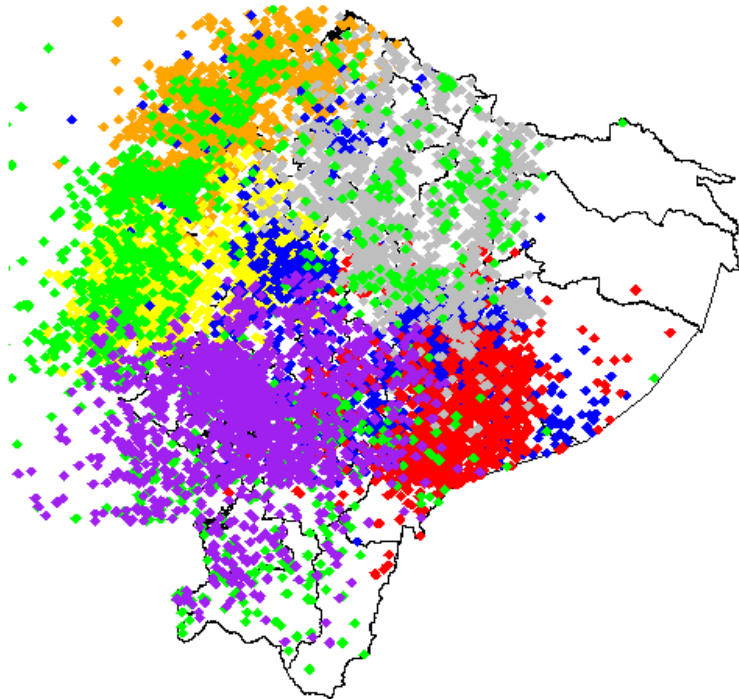


Figura 3.21 Gráfico de los clústers por Ward de los sismos ocurridos en Ecuador.
Fuente: Elaborado por autor.

Grupo	Color	Mag.	Prof.	Utm-X	Utm-Y
1	Verde	1,7410	-0,3380	-1,3530	0,6709
2	Azul	-0,5106	-0,4419	0,7351	0,3897
3	Rojo	-0,1674	-0,2864	-0,9667	-0,3901
4	Anaranjado	2,1895	0,4701	0,2692	-1,1188
5	Morado	-0,2720	-0,3340	0,5122	-1,3810
6	Amarillo	0,0169	2,4698	0,6989	-0,4189
7	Gris	-0,2866	-0,3119	-0,7015	-1,2254

Tabla 3.7 Características de las variables estandarizadas de los clústeres por K-Medias.

Fuente: Elaborado por autor.

En la Tabla 3.7, la variable profundidad del grupo #6 presentó una media que se encuentra está más distante del origen con respecto a la de los otros grupos, esto quiere decir que, en su mayoría, este grupo está formado por sismos que registraron profundidades muy altas comparadas con su media, así también el grupo #4 su variable de magnitud se encuentra más distante del origen, esto quiere decir que, en su mayoría, este grupo está formado por sismos que registraron magnitudes muy altas comparadas a su media.

Los grupos antes mencionados no fueron válidos para crear una zonificación, en cambio, los grupos restantes (grupo #1, grupo #2, grupo #3, grupo #5 y grupo #7), tuvieron medias cercanas al origen, y se encuentran dispersas en el mapa de manera regular y cubriendo áreas extensas, ideales para crear una zonificación de esta. A continuación, en la figura 3.22 se presenta la zonificación por 5 clústers.

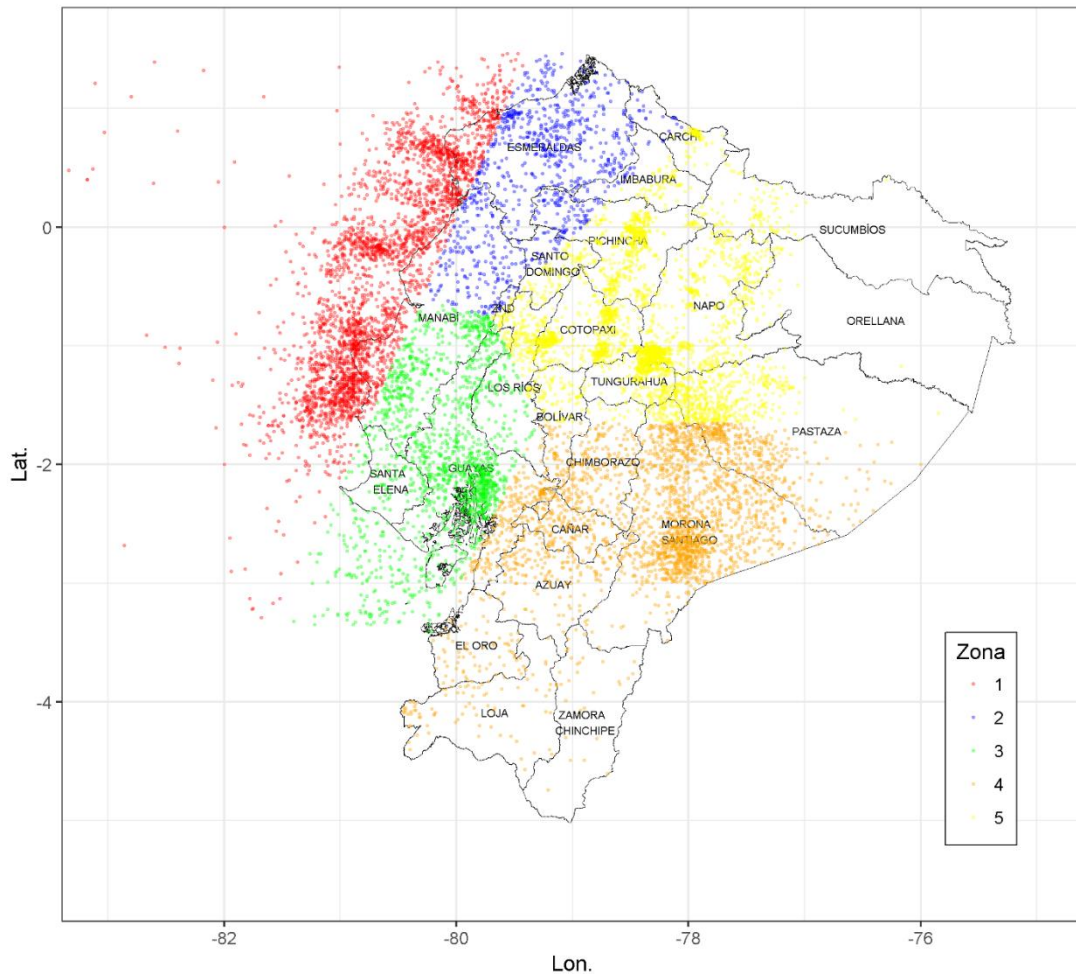


Figura 3.22: Gráfico de los epicentros de sismos ocurridos en el Ecuador, clasificados por las 5 zonas.
Fuente: Elaborado por autor.

Grupo	Color	Mag.	Prof.	Utm-X	Utm-Y
1	Rojo	3,7128	14,0534	528370.4	9955005.7
2	Azul	3,3887	25.0216	690794,6	10047886.2
3	Verde	3,5363	30.0572	794232.5	9790216.2
4	Anaranjado	3.6094	53.0987	790853,9	9727706.0
5	Amarillo	3.3626	36.9872	795122.0	9909860.0

Tabla 3.8 Características de las variables no estandarizadas.
Fuente: Elaborado por autor.

3.3.1 Teoría de Valores Extremos aplicada a la zonificación del Ecuador.

Una vez ya establecida la zonificación del territorio ecuatoriano, se procedió a aplicar los modelos MB y POT para cada una de las zonas.

La zona 1 está ubicada al noroeste del territorio ecuatoriano, abarcando porciones de las provincias de Esmeraldas y Manabí con su respectiva zona marítima, representada con el color rojo en el mapa de zonificación (figura 3.22). Una curiosidad importante de esta zona es que en ella se produjo el terremoto de 7.8 M_w , del pasado 16 de abril del 2016. El modelo que mejor se ajustó a los datos fue el de máximo por bloque, siguiendo una distribución de probabilidad de Gumbel.

	Parámetros		
	Localización	Escala	
Estimación	4.8951	0.5542	
Error estándar	0.0689	0.0543	
AIC	150.30	BIC	154.83

Tabla 3.9 Modelo Gumbel aplicado a la Zona 1.

Fuente: Elaborado por autor.

Test de Razón de Verosimilitud	
Razón de verosimilitud	3.6389
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.0564

Tabla 3.10 Test de razón de verosimilitud (Gumbel vs. GEV)

Fuente: Elaborado por autor.

Los periodos de retorno de esta zona tienen probabilidades elevadas para sismos de alta magnitud. Por ejemplo, para un sismo de magnitud 7 en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 50 años, también se lo puede leer como si la probabilidad de que en un año cualquiera ocurra un sismo de magnitud 7 M_w o mayor sería de 1/50. Para un sismo o terremoto como el que ocurrió el pasado 16 de abril de 2016 de 7.8 grados en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 200 años, leyéndolo de otra manera, la probabilidad de que en esta zona ocurra un sismo de 7.8 grados en escala M_w sería de 1/200. Se considera que esta zona es altamente sísmica dadas las probabilidades considerables de sismos de magnitud elevada.

Periodo de Retorno	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años	4.954090	5.104014	5.253938
5 años	5.507406	5.742836	5.978267
10 años	5.861075	6.165792	6.470510
50 años	6.628652	7.096651	7.564651
100 años	6.951217	7.490177	8.029137

Tabla 3.11 Zona1-Periodos de retorno de las magnitudes estimadas.

Fuente: Elaborado por autor.

La zona 2 también se ubica al noroeste del territorio ecuatoriano, abarcando en gran parte a la provincia de Esmeraldas y parcialmente a Manabí, Pichincha, Santo Domingo de los Tsáchilas, Carchi e Imbabura, representada con el color azul en el mapa de zonificación (figura 3.22). El modelo que mejor se ajustó a los datos fue el de máximo por bloque, siguiendo una distribución de probabilidad Gumbel.

	Parámetros		
	Localización	Escala	
Estimación	4.4180675	0.5222119	
Error estándar	0.05970413	0.04601093	
AIC	165.73	BIC	170.59

Tabla 3.12 Modelo Gumbel aplicado a la Zona 2.

Fuente: Elaborado por autor.

Test de Razón de Verosimilitud	
Razón de verosimilitud	2.491
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.1145

Tabla 3.13 Test de razón de verosimilitud (Gumbel vs. GEV)

Fuente: Elaborado por autor.

Esta zona tiene probabilidades poco considerables para sismos de alta magnitud. Por ejemplo, para un sismo de magnitud 7 en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 200 años, también se lo puede leer como si la probabilidad de que en un año cualquiera ocurra un sismo de magnitud 7 M_w o mayor sería de 1/200. Para un sismo o terremoto como el que ocurrió el pasado 16 de abril de 2016 de 7.8 grados en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 550 años, leyéndolo de otra manera, la probabilidad de que en esta zona ocurra un sismo de

7.8 grados M_w . sería de 1/550 Se considera que esta zona, a pesar de ser altamente concurrida por sismos, tiene magnitudes que no se consideran altamente peligrosas.

Periodo de Retorno	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años	4.478712	4.609465	4.740218
5 años	4.997770	5.201354	5.404938
10 años	5.330434	5.593236	5.856038
50 años	6.053077	6.455706	6.858335
100 años	6.356867	6.820320	7.283773

Tabla 3.14 Zona 2-Periodos de retorno de las magnitudes estimadas.

Fuente: Elaborado por autor.

Para la zona 3, la cual se ubica al suroeste del territorio ecuatoriano, abarcando parcialmente a las provincias de Manabí, Los Ríos, Guayas y Santa Elena, éstas 2 últimas con su zona marítima; está representada por el color verde en el mapa de zonificación (figura 3.22). El modelo que mejor se ajustó a los datos fue el máximo por bloques, siguiendo una distribución de probabilidad Gumbel.

	Parámetros		
	Localización	Escala	
Estimación	4.74842	0.44289	
Error estándar	0.05094213	0.03759742	
AIC	131.1477	BIC	136.0093

Tabla 3.15 Modelo Gumbel aplicado a la Zona 3.

Fuente: Elaborado por autor.

Test de Razón de Verosimilitud	
Razón de verosimilitud	0.2636
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.6077

Tabla 3.16 Test de razón de verosimilitud (Gumbel vs. GEV)

Fuente: Elaborado por autor.

Esta zona tiene probabilidades considerables bajas para sismos de alta magnitud. Por ejemplo, para un sismo de magnitud 7 en la escala M_w , tiene asociado un periodo de 200 años, también se lo puede leer como si la probabilidad de que en un año cualquiera ocurra un sismo de magnitud 7 M_w . o mayor sería de 1/200. Para un sismo o terremoto como el que ocurrió el pasado 16 de abril de 2016 de 7.8 grados en la escala M_w , tiene

asociado un periodo de retorno mucho mayor a 1000 años, asumiendo la improbabilidad. Se considera que esta zona, a pesar de ser altamente concurrida por sismos, tiene magnitudes que no se consideran peligrosas.

Periodo de Retorno	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años	4.799375	4.910753	5.022131
5 años	5.241954	5.412747	5.583541
10 años	5.526143	5.745111	5.964079
50 años	6.143705	6.476591	6.809477
100 años	6.403338	6.785828	7.168319

Tabla 3.17 Zona 3-Periodos de retorno de las magnitudes estimadas.
Fuente: Elaborado por autor.

Para la zona 4 ubicada al sur del territorio ecuatoriano, abarcando las provincias de Morona Santiago, Zamora Chinchipe, El Oro, Loja, Azuay, Cañar, Chimborazo y parcialmente a Guayas, Los Ríos, Pastaza y Bolívar; representada por el color Naranja en el mapa de Zonificación (figura 3.22). El modelo que mejor se ajustó a los datos fue el de máximo por bloque, siguiendo una distribución de probabilidad Gumbel.

	Parámetros		
	Localización	Escala	
Estimación	5.0254467	0.4563548	
Error estándar	0.05294226	0.03975774	
AIC	136.43	BIC	141.24

Tabla 3.18 Modelo Gumbel aplicado a la Zona 4.
Fuente: Elaborado por autor.

Test de Razón de Verosimilitud	
Razón de verosimilitud	0.34382
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.5576

Tabla 3.19 Test de razón de verosimilitud (Gumbel vs. GEV)
Fuente: Elaborado por autor.

Esta zona tiene probabilidades considerables (Mediana-Alta) para sismos de alta magnitud. Por ejemplo, para un sismo de magnitud 7 en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 50 años, también se lo puede leer como si la

probabilidad de que en un año cualquiera ocurra un sismo de magnitud 7 M_w o mayor sería de 1/50. Para un sismo o terremoto como el que ocurrió el pasado 16 de abril de 2016 de 7.8 grados en la escala M_w , tiene asociado un periodo de retorno mayor a 200 años, asumiendo la improbabilidad. Se considera que esta zona es altamente sísmica dadas las probabilidades considerables de sismos de magnitud elevada.

Periodo de Retorno	IC_95% Límite Inferior	Magnitud Estimada	IC_95% Límite Superior
2 años	5.200673	5.330290	5.459906
5 años	5.707940	5.907405	6.106870
10 años	6.033421	6.289505	6.545590
50 años	6.740550	7.130448	7.520347
100 años	7.037813	7.485961	7.934109

Tabla 3.20 Zona 4-Periodos de retorno de las magnitudes estimadas.
Fuente: Elaborado por autor.

La zona 5 también se ubica en el centro-norte del territorio ecuatoriano, abarcando las provincias de Napo, Cotopaxi, Tungurahua y parcialmente a Pichincha, Santo Domingo de los Tsáchilas, Chimborazo, Bolívar, Los Ríos, Imbabura, Carchi, Sucumbíos, Orellana y Pastaza; representada con el color amarillo en el mapa de zonificación (figura 3.22). El modelo que mejor se ajustó a los datos fue el de máximo por bloque, siguiendo una distribución de probabilidad Gumbel.

	Parámetros		
	Localización	Escala	
Estimación	4.9228281	0.5126105	
Error estándar	0.05590259	0.04202462	
AIC	175.09	BIC	180.16

Tabla 3.21 Modelo Gumbel Aplicado a la Zona 5.
Fuente: Elaborado por autor.

Test de Razón de Verosimilitud	
Razón de verosimilitud	0.09974
Valor crítico Chi-cuadrado	3.8415
Alpha	0.05
Grados de libertad	1
Valor-p	0.7521

Tabla 3.22 Test de razón de verosimilitud (Gumbel vs. GEV)
Fuente: Elaborado por autor.

Esta zona tiene probabilidades poco considerables (Media alta) para sismos de alta magnitud. Por ejemplo, para un sismo de magnitud 7 en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 50 años, también se lo puede leer como si la probabilidad de que en un año cualquiera ocurra un sismo de magnitud 7 M_w o mayor sería de 1/50. Para un sismo o terremoto como el que ocurrió el pasado 16 de abril de 2016 de 7.8 grados en la escala M_w , tiene asociado un periodo de retorno de aproximadamente 200 años, leyéndolo de otra manera, la probabilidad de que en esta zona ocurra un sismo de 7.8 grados M_w sería de 1/200.

Periodo de Retorno	IC 95% Límite Inferior	Magnitud Estimada	IC 95% Límite Superior
2 años	5.060917	5.193149	5.325380
5 años	5.616902	5.820897	6.024891
10 años	5.974305	6.236521	6.498736
50 años	6.751450	7.151242	7.551035
100 años	7.078275	7.537946	7.997617

Tabla 3.23 Zona 5-Periodos de retorno de las magnitudes estimadas.
Fuente: Elaborado por autor.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

Conclusiones

La actividad sísmica del Ecuador está conformada en su gran parte por sismos de intensidad leve (97.8% de sismos registrados), los principales puntos de acumulación de sismos de esta intensidad se encuentran ubicados en las provincias de Esmeraldas, Manabí Guayas, Pichincha, Los Ríos, Morona Santiago y Napo; en cuanto a los sismos de intensidad moderada y grave (1.6 y 0.6% de sismos registrados), estos se concentran principalmente en la zona marítima del Ecuador a la altura de las provincias de Esmeraldas y Manabí, además de otro punto de concentración de estas intensidades entre las provincias de Pastaza y Morona Santiago.

El modelo que mejor ajusta la distribución empírica de Valores Extremos de las magnitudes de sismos por el enfoque de Máximos por Bloques es la Distribución de Gumbel, que además indica que un sismo de magnitud $7.12 M_w$ o mayor tiene un periodo de retorno de 10 años y que un sismos de $7.8 M_w$ o mayor tendría un periodo de retorno de 50 años en el Ecuador; mientras que el modelo que ajusta a la distribución empírica de Valores Extremos de las magnitudes de sismos por el enfoque de Picos Sobre Un Umbral, indica que un sismo de magnitud $7.26 M_w$ o mayor tiene un periodo de retorno de apenas 5 años, y que un sismos de $7.8 M_w$ o mayor tendría un periodo de retorno de 25 años.

En la zonificación mediante el uso del algoritmo de K-medias, se obtuvieron 5 grupos diferentes; al aplicar los modelos de valores extremos sobre estas zonas, la distribución que mejor se ajusta a estas es la de Gumbel.

La zona 1 del Ecuador es conformada por la zona costera de las provincias de Esmeraldas y Manabí. La magnitud promedio de los sismos de esta zona son de $3.71M_w$.

y con una profundidad promedio de 14Km. Para un sismo de $7.8M_w$, como el ocurrido 16 de abril del 2016, esta zona tiene un periodo de retorno de 200 años.

La zona 2 del Ecuador es conformada por gran parte del territorio continental de las provincias de Esmeraldas y Manabí y parte de las provincias Carchi, Imbabura, Pichincha y Santo Domingo. La magnitud promedio de los sismos de esta zona son de $3.38M_w$, y con una profundidad promedio de 25Km. Para un sismo de $7.8M_w$, esta zona tiene un periodo de retorno que sobrepasa los 500 años.

La zona 3 del Ecuador es conformada por las provincias de Guayas, Santa Elena, y parte sur de las provincias de Manabí y Los Ríos. La magnitud promedio de los sismos de esta zona son de $3.5M_w$, y con una profundidad promedio de 30Km. Un sismo de $7.8M_w$, tiene una probabilidad de ocurrencia de 1/1000.

La zona 4 del Ecuador es conformada por las provincias de Chimborazo, Cañar, Azuay, El Oro, Loja, Zamora Chinchipe, Morona Santiago y la parte sur de las provincias del Guayas, Bolívar y Pastaza. La magnitud promedio de los sismos de esta zona son de $3.6M_w$, y con una profundidad promedio de 53Km. Para un sismo de $7.8M_w$, esta zona tiene un periodo de retorno de 200 años.

La zona 5 del Ecuador es conformada por las provincias de Carchi, Imbabura, Pichincha, Cotopaxi, Santo Domingo, Sucumbíos, Napo, Tungurahua, y la parte norte de las provincias de Los Ríos, Bolívar y Pastaza. La magnitud promedio de los sismos de esta zona son de $3.36M_w$, y con una profundidad promedio de 36Km. Para un sismo de $7.8M_w$, esta zona tiene un periodo de retorno de 200 años.

Recomendaciones

Se recomienda utilizar otros métodos de imputación de datos a fin de comparar los resultados obtenidos con el método de Hot_Deck, y evaluar con que método se obtendría modelos más adecuados.

Probar diferentes métodos de clustering, y formar otras agrupaciones, que permitan reconocer las características más relevantes, que dé como resultado una mejor zonificación, con el fin de diseñar y comparar modelos de valores extremos en dichas zonas.

También se recomienda usar variables que no se han tomado en cuenta en el presente estudio, como la distancia del epicentro del sismo a la falla más cercana, la distancia al límite convergente de las placas de Nazca con la de Sudamérica o la distancia al volcán más cercano, a fin de poder hacer agrupaciones que permitan describir origen de los sismos.

5. BIBLIOGRAFÍA

- Aguiar, R. C. (2009). *Magnitud máxima en zonas fuentes para estudios de peligrosidad sísmica del Ecuador*. Rev Cienc ESPE Univ Fuerzas Armadas.
- Aguiar, R. C. (2009). *Magnitud máxima en zonas fuentes para estudios de peligrosidad sísmica del Ecuador*. Rev Cienc ESPE Univ Fuerzas Armadas. 12(2):109–121.
- Akaike, H. (1977). "On entropy maximization principle". In: Krishnaiah, P.R. (Editor). *Applications of Statistics*, North-Holland, Amsterdam, pp. 27–41.
- Albeverio, S., Jentsch, V., & Kantz, H. (2005). *Extreme Events in Nature and Society*. Springer. ISBN: 978-3540286103.
- Albeverio, S., Jentsch, V., & Kantz, H. (2005). *Extreme Events in Nature and Society*. Springer. ISBN: 978-3540286103.
- Alcalde, F. (2005). *La Teoría de los Eventos Extremos, aplicación para evaluación de riesgos*. Centro de Investigación en Métodos Cuantitativos Aplicados a la Economía y la Gestión.. Buenos Aires: Centro de Investigación en Métodos

Cuantitativos Aplicados a la Economía y la Gestión-Facultad de Ciencias Económicas-Universidad de Buenos Aires.

- Ameer, A., Sharma, M., Wason, H., & Alsinawi, A. (2004). *Seismic Hazard Characterization and Risk Evaluation using Gumbel's Method of Extremes (G1 and G3) and G-R Formula for Iraq. Paper presented at the 13th World Conference on Earthquake Engineering. Vancouver, B.C., Canada.*
- Asamblea Nacional Constituyente. (2008). *Constitución de la República del Ecuador.*
- Balkema, A., & Haan, L. (1978a). *Limit distributions for order statistics, I. Theory Probab. Appl. 23: 77-92.*
- Balkema, A., & Haan, L. (1978b). *Limit distributions for order statistics, II. Theory Probab. Appl. 23: 341-358.*
- Burtun, A., & Makropoulus, K. (1985). *Seismic Risk of Circum-Pacific Earthquakes: II. Extreme Values Using Gumbel's Third Distribution and the Relationship with Strain Energy Release. PAGEOPH: 123.*
- Castillo, E. (1988). *estadística de Valores Extremos: Distribuciones Asintóticas. Estadística Española: 116:5-34.*
- CEC. (2 de 08 de 2001). *Requisitos generales de diseño: peligro sísmico, espectros de diseño y requisitos mínimos de cálculos para diseño sismo resistente. Código Ecuatoriano de la Construcción. Registro Oficial No. 382 . Obtenido de <http://www.normacon>*
- Chunga, K., Michetti, A., Pazmiño, N., Martillo, C., Romero, A., & Quiñonez, M. (2009). *Estimación de máximos niveles de sismicidad para el litoral ecuatoriano a través de la integración de datos geológicos, sismológicos y sismotectónicos. Revista Científica y Tecnológica UPSE, 73-86.*
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer. ISBN: 978-1852334598.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer. ISBN: 978-1852334598.
- Dzwinel, W., A.Yuen, D., Boryczko, K., Ben-Zion, Y., Yoshioka, S., & Ito, T. (2003). *Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space.* Earth and Planetary Sci. Letters.

- Edd, S., & Khairi, A. (2016). *Evaluation of Peaks-Over-Threshold Method*. *Ocean Sci. Discuss.*, doi: 10.5194/os-2016-47, 2016.
- Evans, M. J., & Rosenthal, J. S. (2005). *Probabilidad y estadística*. Editorial Reverté.
- Fisher, R., & Tippett, L. (1928). *Limiting forms of the frequency distribution of the largest and smallest member of a sample*. *Proc . Cambridge Phil. Soc.* 24: 180-190, doi:10.1017/s0305004100015681.
- Fréchet, M. (1927). *Sur la loi de probabilité de l'écart maximum*. *Ann. .Soc. Polon. Math, Cracow.* 6, 93.
- Gnedenko, B. V. (1943). *Sur la distribution limite du terme maximum d'une série aléatoire*. *Ann. Math.* 4a, 423-453.
- Green, H. W. (1994). *Solving the paradox of deep earthquakes*. *Scientific American*, 271(3), 64-71.
- Grinstead, C. M., & Snell, j. (1997). *Central Limit Theorem. Introduction to Probability*. AMS Bookstore. pp. 325-360. ISBN 0821807498.
- Gumbel, E. J. (1934). *Leê moments des distributions finales de la première et de la dernière valeur*. *Comptes Rendus de l'Académie des Sciences, Paris*, 198, I 41-143.
- Gumbel, E. J. (1935). *Les valeurs extremes des distributions statistiques*. *Annales de l'institut Henri Poincaré*, 5, 115-158.
- Gutenberg, B., & Richter, C. F. (1944). *Frecuency of earthquakes in California*, *BSSA*. 34: 185-188.
- Guzmán, J., & Vela, Y. (2015). *Determinación de la Amenaza Sísmica en Santander mediante el análisis de Valores Extremos (Tesis de Grado)*. Universidad Distrital Francisco José de Caldas. Bogotá DC, Colombia.
- Hald, A. (1999). *On the history of maximum likelihood in relation to inverse probability and least squares*. *Statistical Science* 14 (2): 214-222.
- Hanks, T. C., & Kanamori, H. (1979). *"A moment magnitude scale"*. *Journal of Geophysical Research* 84 (B5): 2348-50.
- Hartigan, J. (1974). *The K-Means Algorithm*. En R. Bradley, D. Kendall, J. Hunter, G. Watson (Eds.), *Wiley Series in Probability and Mathematical Statistics* (págs. 84-112). New York: John Wiley & Sons.

- IGEPN. (2016). *Instituto Geofísico Escuela Politécnica Nacional. Quito, Ecuador.* Obtenido de <http://www.igepn.edu.ec/solicitud-de-datos>.
- IGEPN. (2018). *Instituto Geofísico Escuela Politecnica Nacional.* Obtenido de <https://www.igepn.edu.ec/ultimos-sismos>
- INEC (Instituto Nacional de Estadísticas y Censos). (2017). *Clasificador Geográfico Estadístico 2017.*
- IRIS. (2016). *Incorporated Research Institutions for Seismology. Interactive Earthquake Browser.* Obtenido de <http://ds.iris.edu/ieb/>
- János, G. (1978). *The asymptotic theory of extreme order statistics.* New York: Wiley ISBN-10: 0471021482.
- Jenkinson, A. F. (1955). *The frequency distribution of the annual maximum (or minimum) values of meteorological elements.* *Quart. J. Roy. Meteor. Soc.* 1(81): 158-171.
- Kamat, R., & Kamath, R. (2017). *Earthquake Cluster Analysis: K-Means Approach.* Journal of Chemical and Pharmaceutical Sciences.
- Kotz, S., & Nadarajah, S. (2000). *Extreme Value Distribution: theory and applications.* Imperial College Press; ISBN 10: 1860942245.
- Kotz, S., & Nadarajah, S. (2001). Some extremal type elliptical distributions. *Statistics & Probability Letters*, 171-182.
- Lei, L. (2010). Identify earthquake hot spots with 3-dimensional density-based clustering analysis. *IEEE International Geoscience and Remote Sensing Symposium.* Honolulu, HI, USA: ISBN: 978-1-4244-9564-1.
- Marchioni Fleck, E., & De Araújo Santos, R. (2007). Clustering seismic data: A multivariable approach. En B. Hardage, *SEG Technical Program Expanded Abstracts 2007* (págs. 2075-2079.). Society of Exploration Geophysicists; ISSN (online): 1949-4645.
- Medina, A., & Galván, M. (2007). *Estudios estadísticos y prospectivos, Imputación de datos: Teoría y práctica. División Estadística y proyecciones Económicas Naciones Unidas, CEPAL.*
- Michigan Tech University. (2016). *UPSEIS an educational site for budding seismologists. Earthquake Magnitude Classes Recovered from. Michigan, USA.* Obtenido de <http://www.geo.mtu.edu/UPSeis/magnitude.html>.

- NEC-11. (10 de 01 de 2015). *NEC-11 Norma Ecuatoriana de la Construcción. Registro Oficial No. 413*. Obtenido de <http://www.normaconstruccion.ec/>.
- Ortiz, O. (2013). *Sismo tectónica y Peligrosidad Sísmica en Ecuador. Universidad Complutense de Madrid, Madrid, España*.
- Parra, H., Benito, M., & Gaspar, J. (2016). *Seismic hazard assessment in continental Ecuador. Bull Earthquake Eng. DOI 10.1007/s10518-016-9906-7*.
- Pickands, J. (1975). *Statistical inference using extreme order statistics. Ann. Statist. 3, 1 19-131*.
- Pisarenko, V., Sornette, D., & Rodkin, M. (2010). *Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007). Earth, Planets and Space: 62(7): 567–578*.
- Presidencia de la República del Ecuador. (2010). *Código Orgánico de Organización Territorial*. Quito.
- Reiss, R.-D., & Thomas, M. (2007). *Statistical Analysis of Extreme Values With Applications to Insurance, Finance, Hydrology and Other Fields. Third Edition*. Springer. ISBN: 978-3764372309.
- Richter, C. F. (1935). *An instrumental earthquake magnitude scale. Bulletin of the Seismological Society of America, 25(1), 1-32*.
- Saenz, F., Hidalgo, F., & Santa, A. (1991). Estimación de Precipitaciones máximas mediante el método regional del índice de avenida. *Revista de Obras Públicas*, 9-22.
- Schwarz, E. (1978). «*Estimating the dimension of a model*». *Annals of Statistics 6 (2): 461–464. MR 468014. doi:10.1214/aos/1176344136*.
- UPSEIS. (2016). *UPSeis an educational site for budding seismologists. Earthquake Magnitude Classes Recovered from. Michigan, USA*. Obtenido de <http://www.geo.mtu.edu/UPSeis/magnitude.html>
- Von, M. R. (1936). La distribución de la plus grande de n valeurs. Reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271-294.
- Weibull, W. (1951). *A statistical distribution function of wide applicability. J. Appl. Mech.-Trans. ASME 18 (3): 293-297*.

Wood, H. O., & Neumann, F. (1931). *Modified Mercalli intensity scale of 1931. Bulletin of the Seismological Society of America*, 21(4), 277-283.

Zimbidis, A., Frangos, N., & Pantelous, A. (2007). *Modeling Earthquake Risk via Extreme Value Theory and Pricing the Respective Catastrophe Bonds. Astin Bulletin* 37(1), 163-183. doi: 10.2143/AST.37.1.2020804.

Zúñiga, F. R. (2011). *“Notas Introductorias “Sismos”, Posgrado en Ciencias de la tierra, Centro de Geociencias, UNAM – Campus Juriquilla. Querétaro, México.*