

UNIVERSIDAD AUTÓNOMA METROPOLITANA

UNIDAD AZCAPOTZALCO
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
DEPARTAMENTO DE SISTEMAS

INFERENCIA ESTADÍSTICA BASADA EN EL ANÁLISIS DEL COMPORTAMIENTO HISTÓRICO DE LOS SISMOS EN VARIAS REGIONES DE LA COSTA DEL PACÍFICO DE MÉXICO

IDÓNEA COMUNICACIÓN DE RESULTADOS
QUE PRESENTA EL:

ING. JOSÉ MANUEL VILLA VARGAS

PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

DIRECTOR DE TESIS:
DR. JOSÉ ANTONIO CLIMENT HERNÁNDEZ

Resumen

El presente trabajo desarrolla un análisis probabilista para la inferencia de eventos sísmicos significativos en la costa del Pacífico mexicano, región con una alta actividad sísmica. Utilizando datos históricos del Servicio Sismológico Nacional de México desde 1900 hasta julio de 2025, se implementó una metodología sistemática que comprende estadística descriptiva, intervalos de confianza, pruebas de hipótesis y pruebas de bondad de ajuste para seis regiones: Chiapas, Guerrero, Michoacán, Oaxaca, Resto Nacional y Sismos Nacionales. Los resultados revelan que las distribuciones Gumbel, Weibull, GEV, Logística y Normal ajustan adecuadamente los datos de Magnitudes Máximas anuales según la región analizada, mientras que ninguna distribución ajustó satisfactoriamente para los Sismos Totales. Se identificaron diferencias estadísticamente significativas entre las regiones en términos de media, varianza y proporción de sismos superiores a 6.5° . Este trabajo busca proporcionar una base robusta para la estimación probabilista de eventos sísmicos futuros en México y contribuir al desarrollo de políticas de prevención y concientización del peligro sísmico existente en México.

Palabras clave: Inferencia Sísmica, análisis probabilista, estadísticos descriptivos, pruebas de bondad de ajuste, pruebas de hipótesis, sismos en la costa del Pacífico Mexicano.

Abstract

This study develops a probabilistic analysis for the inference of significant seismic events on the Mexican Pacific coast, a region with high seismic activity. Using historical data from the Mexican National Seismological Service from 1900 to July 2025, a systematic methodology was implemented. This methodology includes descriptive statistics, confidence intervals, hypothesis testing, and goodness-of-fit tests for six regions: Chiapas, Guerrero, Michoacán, Oaxaca, the rest of the country, and national seismic events. The results reveal that the Gumbel, Weibull, GEV, Logistic, and Normal distributions adequately fit the annual Maximum Magnitudes data, depending on the region analyzed. In contrast, no distribution satisfactorily fit the data for Total Earthquakes. Statistically significant differences were identified among the regions in terms of mean, variance, and the proportion of earthquakes with a magnitude greater than 6.5. This work aims to provide a robust basis for the probabilistic estimation of future seismic events in Mexico and to contribute to the development of prevention policies and awareness of the existing seismic hazard in the country.

Keywords: Seismic Inference, probabilistic analysis, descriptive statistics, goodness-of-fit test, hypothesis testing, earthquakes on the Mexican Pacific coast.

Dedicatoria

// Esta sección se presenta en la versión final de su ICR. Son frases cuyo objetivo es otorgar una mención especial a las personas que te han motivado durante tu ICR.

Agradecimientos

// Esta sección se presenta en la versión final de su ICR. Son frases cuyo objetivo es plasmar el apoyo moral, físico, económico y/o emocional que recibió de las personas o instituciones durante la elaboración de todo su proyecto.

Índice general

Resumen	III
Abstract	V
Dedicatoria	VII
Agradecimientos	IX
Índice general	XIII
Índice de tablas	XV
Índice de figuras	XVII
Simbología y Acrónimos	XIX
1. Introducción	1
1.1. Planteamiento del problema	2
1.2. Justificación	2
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	3
1.4. Principales contribuciones	3
1.5. Organización del documento	4
2. Marco teórico	5
2.1. Síntesis de los artículos	6
3. Estado del arte	13
3.1. Revisión del estado del arte	13
3.2. Análisis del estado del arte	15

4. Metodología	17
4.1. Metodología de investigación	17
4.2. Fases de desarrollo del proyecto	18
4.2.1. Media aritmética	19
4.2.2. Mediana	19
4.2.3. Moda	20
4.2.4. Varianza muestral	20
4.2.5. Desviación estándar	20
4.2.6. Coeficiente de asimetría	20
4.2.7. Coeficiente de curtosis	21
4.2.8. Intervalo de confianza para la media	21
4.2.9. Intervalo de confianza para la varianza	22
4.2.10. Intervalo de confianza para la proporción	22
4.2.11. Prueba de hipótesis para el cociente de varianzas (prueba F)	22
4.2.12. Prueba de hipótesis para diferencia de medias con varianzas iguales (pooled)	23
4.2.13. Prueba de hipótesis para diferencia de medias con varianzas diferentes (Welch)	23
4.2.14. Prueba de hipótesis para la diferencia de proporciones con proporciones agrupadas o iguales (pooled)	24
4.2.15. Prueba de hipótesis para la diferencia de proporciones sin agrupamiento o diferentes	24
4.2.16. Prueba de bondad de ajuste Kolmogorov-Smirnov	24
4.2.17. Prueba de bondad de ajuste Anderson-Darling	25
4.2.18. Prueba de bondad de ajuste Lilliefors	25
4.2.19. Criterio de Información Bayesiano	25
4.2.20. Selección de la distribución óptima por región	26
4.2.21. Estimación de parámetros por máxima verosimilitud	26
4.2.22. Cálculo del periodo de retorno	26
4.2.23. Cálculo del nivel de retorno	27
4.2.24. Cálculo de probabilidades de excedencia	28
4.2.25. Interpretación de los resultados	29
5. Análisis de resultados	31
5.1. Herramientas utilizadas	31
5.2. Resultados de estadísticos descriptivos	31
5.3. Representación gráfica de los datos sísmicos	32
5.3.1. Histogramas de densidad de los sismos para sismos totales	33
5.3.2. Histogramas de densidad de los sismos para magnitudes máximas	34
5.3.3. Histogramas de frecuencias relativas por mes para sismos totales	36
5.3.4. Histogramas de frecuencias relativas por mes para magnitudes máximas	37
5.3.5. Histogramas de sismos por magnitud	38
5.3.6. Gráfico de magnitud máxima, promedio y mínima en el tiempo	39
5.4. Intervalos de confianza y tamaño mínimo de la muestra	40
5.4.1. Intervalo de confianza para la media	40

6. Conclusiones	43
A. Anexo 1	45
Bibliografía	48

Índice de tablas

1.	Artículos seleccionados.	5
2.	Distribuciones óptimas por región según pruebas de bondad de ajuste.	26
3.	Estadísticos descriptivos consolidados.	31
4.	IC para la media para sismos totales para las 6 regiones.	40
5.	IC para la media para magnitudes máximas para las 6 regiones.	41

Índice de figuras

1.	Diagrama de flujo de la aplicación de la metodología propuesta.	18
2.	Histogramas de densidad de sismos totales para las 6 regiones.	33
3.	Histogramas de densidad de magnitudes máximas para las 6 regiones.	34
4.	Histogramas de Frecuencias Relativas por Mes para sismos totales para las 6 regiones.	36
5.	Histogramas de Frecuencias Relativas por Mes para magnitudes máximas para las 6 regiones.	37
6.	Histogramas de sismos por magnitud para las 6 regiones.	38
7.	Gráficos de magnitud sísmica máxima, promedio y mínima para las 6 regiones. .	39
8.	IC para la media para sismos totales para las 6 regiones.	41
9.	IC para la media para magnitudes máximas para las 6 regiones.	42

Simbología y Acrónimos

Simbología Estadística

\bar{x}	Media aritmética o promedio
\tilde{x}	Mediana
\hat{x}	Moda
s_x^2	Varianza muestral
s_x	Desviación estándar muestral
g_1	Coefficiente de asimetría
g_2	Coefficiente de curtosis
n	Tamaño de la muestra
N	Tamaño de la población
α	Nivel de significancia
$1 - \alpha$	Nivel de confianza
H_0	Hipótesis nula
H_1	Hipótesis alternativa
μ	Media poblacional
σ^2	Varianza poblacional
σ	Desviación estándar poblacional
π	Proporción poblacional
p	Proporción muestral

IC	Intervalo de confianza
ε	Error máximo tolerado
F	Estadístico F de Fisher
t	Estadístico t de Student
z	Estadístico z
χ^2	Estadístico chi-cuadrada
D	Estadístico de Kolmogorov-Smirnov
A_n	Estadístico de Anderson-Darling

Acrónimos

AD	Anderson-Darling (prueba de bondad de ajuste)
AHP	Proceso de Jerarquía Analítica (Analytic Hierarchy Process)
AIC	Criterio de Información de Akaike
BIC	Criterio de Información Bayesiano
BPT	Brownian Passage Time
CNNC	Red Neuronal Convolutacional (Convolutional Neural Network)
ETAS	Epidemic Type Aftershock Sequence
GEV	Distribución Generalizada de Valores Extremos
GMPE	Ecuación de Predicción del Movimiento del Suelo
IA	Inteligencia Artificial
IC	Intervalo de Confianza
ICR	Idónea Comunicación de Resultados
KS	Kolmogorov-Smirnov (prueba de bondad de ajuste)
LF	Lilliefors (prueba de bondad de ajuste)
PBA	Pruebas de Bondad de Ajuste
PH	Pruebas de Hipótesis
PSHA	Análisis Probabilista de Peligrosidad Sísmica

RN	Resto Nacionales
SSN	Servicio Sismológico Nacional
SSNMX	Servicio Sismológico Nacional de México
SN	Sismos Nacionales
UAM	Universidad Autónoma Metropolitana
UNAM	Universidad Nacional Autónoma de México

Introducción

A través de la historia, la humanidad ha sufrido de catástrofes naturales que impactan directamente la vida del ser humano. Las catástrofes naturales son de varios tipos como inundaciones, huracanes, deslizamientos de tierra, erupciones volcánicas, sequías, incendios, sismos o terremotos y tsunamis o marejadas, entre varios otros.

La inferencia de la ocurrencia de alguno de estos eventos catastróficos es de utilidad para la humanidad. Y dadas las consecuencias que estos eventos tienen sobre la forma de vida de la gente y como alteran su realidad de un momento a otro, es importante el estudio y la comprensión del comportamiento de los eventos a lo largo de un periodo de tiempo, buscando la forma de inferir y prepararse para cuando estos ocurran.

Dentro de los tipos de catástrofes mencionadas anteriormente, los sismos o terremotos son uno de los que mayor impacto tienen para la vida humana y para la infraestructura urbana y rural, porque estos además del daño que causan por el evento en sí, tienen la particularidad de que provocan otros eventos de catástrofes naturales como incendios, deslizamientos, tsunamis y daños colaterales en infraestructuras como los derrumbes y daños estructurales en edificaciones, cortes al suministro eléctrico, fallas en los sistemas de telecomunicaciones, daños en la red de carreteras, trenes, aeronaves, suministro de agua, alimentos, medicinas y finalmente, presentando un fuerte declive en la economía de los lugares afectados.

Una de las zonas sísmicas más activas en el mundo es la costa del Pacífico de México (que se encuentra en el cinturón de fuego del pacífico) en la cual se presentan frecuentemente sismos de magnitud significativa que impactan a las zonas de ocurrencia y centros urbanos en un radio de 0 a 500 kilómetros de distancia del epicentro de los sismos.

Actualmente son desarrollados y producidos diferentes métodos para inferir la probabilidad de ocurrencia de un sismo en una zona específica. Estas técnicas varían en metodología como en las formas de estudiar e inferir la ocurrencia de un sismo de magnitud significativa para la vida humana.

El presente trabajo utiliza como referencia los datos otorgados por el Servicio Sismológico Nacional de México (SSNMx) con el propósito de realizar un análisis probabilista para la ocurrencia de un sismo de magnitud significativo (mayor a 5° Richter) en los estados de mayor actividad sísmica en la costa del pacífico de la República Mexicana, con fundamento en la muestra de datos del SSNMx que data del 1 de enero del año 1900 hasta julio de 2025. La metodología aplicada comprende estadística descriptiva, intervalos de confianza, estimación

de tamaño mínimo de muestra, pruebas de hipótesis, así como pruebas de bondad de ajuste para diferentes distribuciones y estimación de periodos de retorno en años, que justifiquen los resultados.

1.1. Planteamiento del problema

La predicción de ocurrencia de un sismo de magnitud considerable es hasta la fecha imprecisa y difícil de determinar mediante algún cálculo o estudio que es realizado previamente. Como se ha mencionado anteriormente, en la actualidad se están utilizando diferentes técnicas basadas en distintos planteamientos, que apuntan hacia el mismo objetivo: estimar con un nivel de significación mínimo la próxima ocurrencia de un sismo en una zona de estudio delimitada.

Además de la dificultad encontrada por se de predecir un sismo significativo, están las siguientes problemáticas que conlleva el obtener un resultado más certero, como lo son: falta de datos históricos por tecnologías anticuadas o por no tener una escala estándar antes de 1935 (Richter 1935), también porque hay sismos de magnitud baja que no son registrados, por confusión con otros eventos naturales (como tremores volcánicos, explosiones o impactos terrestres), porque la medición no refleje la magnitud real del sismo (mediciones inexactas, o por falla en la calibración de los equipos) y por sus patrones no periódicos de ocurrencia (que un sismo no muestra patrones de repetición de evento exacta ni periódica, son aleatorios).

1.2. Justificación

La inferencia de sismos es de vital importancia para desarrollar políticas de prevención de desastres derivados de la actividad sísmica de magnitud considerable. Aunque actualmente sigue siendo muy complejo el poder estimar la fecha de ocurrencia de un sismo fuerte en una región determinada, se trabaja continuamente en múltiples aproximaciones para poder realizar con un alto grado de certeza la estimación de ocurrencia de un sismo significativo futuro.

La mayoría de las técnicas que se emplean actualmente tienen que ver con uso y aplicación de inteligencia artificial y redes neuronales alimentadas por datos históricos y ecuaciones de movimiento sísmico.

Es aquí donde se considera importante la existencia de un proyecto como el presente el cual propone un enfoque basado en análisis probabilista. La metodología implementada sigue pasos sistemáticos y repetibles que permiten entender mejor el fenómeno sísmico en los estados estudiados. Esta metodología integra la obtención de estadísticos descriptivos para cada estado, la determinación de intervalos de confianza para la media, desviación y proporción de sismos mayores a un umbral crítico, la estimación del tamaño mínimo de la muestra, la realización de pruebas de hipótesis para media, varianza y proporción de sismos y de pruebas de bondad de ajuste para varias distribuciones. Todo esto con el fin de poder inferir con un nivel de certeza próximos eventos sísmicos en los estados analizados.

También, este proyecto es relevante pues se enfoca en los estados de mayor actividad y riesgo sísmico de México, diferenciando este trabajo de otros que se enfocan en otras regiones del planeta o no son minuciosos en sus estudios de la actividad sísmica en México.

1.3. Objetivos

1.3.1. Objetivo general

Se pretende obtener resultados fiables mediante la aplicación de cálculos de inferencia probabilista para la ocurrencia de un sismo de magnitud significativa con un nivel de confianza en la costa del Pacífico mexicano.

1.3.2. Objetivos específicos

1. Identificar una correlación significativa en los datos históricos de los sismos en la costa del Pacífico mexicano mediante el análisis de los datos disponibles de los mismos.
2. Aplicar el cálculo de probabilidad, los estadísticos descriptivos, pruebas de hipótesis, pruebas de bondad de ajuste a los datos históricos de los sismos para obtener una estimación del intervalo de confianza del siguiente evento sísmico.
3. Programar en la herramienta R y RStudio los algoritmos de estimación de estadísticos descriptivos, de intervalos de confianza, de tamaño mínimo de la muestra, de pruebas de hipótesis y de pruebas de bondad de ajuste, así como generar gráficos y tablas representativos de estos.

1.4. Principales contribuciones

Las principales contribuciones de este trabajo de investigación se listan a continuación:

1. **Análisis estadístico por región:** Se realiza un análisis completo y exhaustivo de la actividad sísmica en seis regiones de México, diferenciándose de otros estudios y logrando identificar que cada estado de la costa del Pacífico presenta un comportamiento sísmico único con distribuciones probabilísticas específicas (Gumbel para Oaxaca, Weibull para Guerrero y Chiapas, distribución generalizada de valores extremos para Michoacán).
2. **Metodología replicable:** Se desarrolla e implementa una metodología sistemática y reproducible que integra múltiples técnicas estadísticas (estadísticos descriptivos, intervalos de confianza, pruebas de hipótesis y pruebas de bondad de ajuste para 20 distribuciones), estableciendo una metodología robusta enfocada al análisis sísmico y que puede aplicarse a otras regiones sísmicamente activas.
3. **Identificación de patrones temporales significativos:** Se determina la existencia de patrones mensuales de actividad sísmica diferenciados por región, identificando septiembre como el mes de mayor actividad sísmica nacional, hallazgo relevante en el ámbito del fenómeno sísmico en México.
4. **Validación de umbrales críticos:** Se establece mediante pruebas estadísticas rigurosas que la proporción de sismos mayores a 6.5° varía significativamente entre regiones, lo cual proporciona una base sólida para la diferenciación del riesgo sísmico regional identificando las regiones con mayor propensión a sufrir de sismos fuertes.

5. **Código computacional en R:** Se desarrolla e implementa un conjunto completo de algoritmos en R y RStudio para el procesamiento automatizado de datos sísmicos, disponible para su uso y adaptación, facilitando la replicación, escalabilidad y adaptabilidad del mismo a cualquier tipo de análisis sísmico con fundamentos estadísticos.

1.5. Organización del documento

El presente documento se estructura en seis capítulos que desarrollan de manera sistemática y progresiva la investigación realizada sobre la inferencia probabilística de eventos sísmicos en la costa del Pacífico mexicano.

El **Capítulo 1** presenta la introducción general al problema de investigación, estableciendo el contexto de la actividad sísmica en México y su impacto en la población e infraestructura. Se define el planteamiento del problema, la justificación del estudio, los objetivos generales y específicos, las principales contribuciones y la presente organización del documento.

El **Capítulo 2** desarrolla el marco teórico fundamental, presentando una revisión exhaustiva de 16 artículos científicos seleccionados que abordan metodologías de predicción sísmica, técnicas estadísticas y de inteligencia artificial aplicadas a la sismología. Se realiza una síntesis crítica de los enfoques existentes, desde modelos probabilísticos clásicos hasta técnicas de aprendizaje profundo, estableciendo el fundamento teórico para la metodología propuesta.

El **Capítulo 3** expone el estado del arte en predicción sísmica, analizando detalladamente los trabajos más relevantes en el campo. Se examinan las técnicas implementadas globalmente, desde modelos log-lineales y distribuciones gamma hasta redes neuronales convolucionales, identificando las fortalezas y limitaciones de cada enfoque. Se destaca la aplicación de estas metodologías en diferentes regiones sísmicas del mundo y su relevancia para el contexto mexicano.

El **Capítulo 4** describe la metodología de investigación implementada, detallando las 10 fases del proceso analítico: desde la recopilación y filtrado de datos del Servicio Sismológico Nacional, hasta la aplicación de pruebas de bondad de ajuste y cálculos de probabilidad. Se presentan las formulaciones matemáticas de los estadísticos descriptivos, intervalos de confianza, pruebas de hipótesis y criterios de selección de modelos utilizados.

El **Capítulo 5** presenta el análisis exhaustivo de los resultados obtenidos. Se muestran los estadísticos descriptivos calculados para las seis regiones estudiadas, las representaciones gráficas del comportamiento sísmico histórico, los intervalos de confianza estimados, los resultados de las pruebas de hipótesis realizadas y la identificación de las distribuciones probabilísticas que mejor ajustan los datos de cada región. Se incluyen 26 tablas y 17 figuras que sintetizan los hallazgos principales.

El **Capítulo 6** expone las conclusiones del trabajo, sintetizando los logros alcanzados respecto a los objetivos planteados, las limitaciones identificadas durante la investigación y las líneas de trabajo futuro propuestas para extender y mejorar la metodología desarrollada.

Finalmente, se incluye un apéndice con información complementaria de los artículos revisados, la síntesis del código utilizado y la bibliografía completa con las referencias utilizadas en el desarrollo de la investigación.

Marco teórico

Se procede a la revisión de artículos y publicaciones que tienen que ver directamente con el tema de la problemática a solucionar.

A continuación, se presenta la Tabla 1 que muestra una selección de los artículos relacionados a la metodología que se utiliza a lo largo de este trabajo.

Tabla 1: Artículos seleccionados.

Año	Título y Autor
2023	“Earthquakes magnitude prediction using deep learning for the Horn of Africa” (Abebe, Kebede, Kevin y Demissie 2023)
2020	“Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges” (Al Banna, Taher, Kaiser, Mahmud, Rahman, Hosen y Cho 2020)
2007	“Análisis geográfico y estadístico de la sismicidad en la costa mexicana del Pacífico” (Barrientos, Fernandez y Rivero 2007)
2023	“Peak-over-threshold versus annual maxima: Which approach is better for extreme value analysis?” (Bommier 2023)
2001	“An Introduction to Statistical Modeling of Extreme Values” (Coles 2001)
2021	“Extreme Value Theory - 20 years on” (Coles, Bawa, Trenner y Dorazio 2021)
2020	“Combining stress transfer and source clustering to forecast seismicity” (Convertito, Catalli y Emolo 2020)
2021	“Time-Dependent Seismic Hazard Analysis for Induced Seismicity: The Case of St Gallen (Switzerland), Geothermal Field” (Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano 2021)
1968	“Engineering seismic risk analysis” (Cornell 1968)
2005	“A probabilistic prediction of the next strong earthquake in the Acapulco-San Marcos segment, Mexico” (Ferrás 2005)
2022	“Long-Term Forecasting of Strong Earthquakes in North America, South America, Japan, Southern China and Northern India With Machine Learning” (Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022)
2021	“Earthquake risk assessment in NE India using deep learning and geospatial analysis” (Jena, Pradhan, Naik y Alamri 2021)
1995	“Continuous Univariate Distributions” (Johnson, Kotz y Balakrishnan 1995)

2. Introducción

Año	Título y Autor
2023	“Fundamental study on probabilistic generative modeling of earthquake ground motion time histories using generative adversarial networks” (Matsumoto, Yaoyama, Lee, Hida y Itoi 2023)
2021	“Best practices in physics-based fault rupture forecasting for seismic hazard assessment of nuclear installations: issues and challenges towards full integration” (Mignan, Danciu y Giardini 2021)
2020	“Ground motion prediction equation for crustal earthquakes in Taiwan” (Phung, Loh, Chao, Chiou y Huang 2020)
2021	“Earthquake source parameters of the Michoacán seismic gap” (Ramírez-Gaytán, Aguirre-González y Núñez-Cornú 2021)
1935	“An instrumental earthquake magnitude scale” (Richter 1935)
2021	“Theoretical methodological aspects about earthquake prediction” (Galbán-Rodríguez 2021)
2020	“Real-time updating of the seismic risk of interdependent infrastructure systems using Bayesian networks” (Sánchez-Silva y Riaño 2020)
2023	“Probabilistic seismic hazard assessment for Western Mexico” (Sawires, Peláez y Santoyo 2023)
2021	“Understanding persistence to avoid underestimation of collective flood risk” (Serinaldi y Kilsby 2021)
2025	“Hypothesis Testing, P Values, Confidence Intervals, and Significance” (Shreffler y Huecker 2025)
2024	“Comparative analysis of continuous probability distributions for modeling maximum flood levels” (Shobanke, Olayemi y Olajide 2024)
2020	“Deadly intraslab Mexico earthquake of 19 September 2017 (Mw 7.1): Ground motion and damage pattern in Mexico City” (Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir 2020)
2021	“Ground motion prediction equation for earthquakes along the Western Himalayan arc” (Singh, Arroyo, Srinivas y Suresh 2021)
2019	“Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data” (Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique 2019)
2022	“Regional probability distribution of ground motion parameters using machine learning and Bayesian approaches” (Yaghmaei-Sabegh y Neekmanesh 2022)
2022	“A first-order seismotectonic regionalization of Mexico for seismic hazard and risk estimation” (Zúñiga, Suárez, Figueroa-Soto y Mendoza 2022)

2.1. Síntesis de los artículos

Ahora se realiza una descripción del contenido de los artículos seleccionados.

En el artículo de Abebe, Kebede, Kevin y Demissie (2023) se expone que los terremotos son vibraciones de la superficie de la Tierra que pueden causar temblores, incendios, deslizamientos de tierra y fisuras en el terreno que representan una amenaza significativa para la vida humana y la infraestructura. Los autores desarrollan un modelo de aprendizaje profundo para la predicción de magnitudes sísmicas en el Cuerno de África, utilizando datos históricos del catálogo sísmico de la región. El modelo propuesto emplea redes neuronales recurrentes de tipo LSTM (Long Short-Term Memory) para capturar patrones temporales en las secuencias sísmicas. Los resultados muestran que el modelo alcanza una precisión significativa en la

predicción de magnitudes, superando a métodos estadísticos tradicionales en la región de estudio.

En el artículo de Al Banna, Taher, Kaiser, Mahmud, Rahman, Hosen y Cho (2020) se presenta una revisión del estado del arte relacionado a la aplicación de técnicas de inteligencia artificial (IA) en la predicción de sismos. Los autores analizan diversos enfoques que incluyen redes neuronales artificiales, máquinas de soporte vectorial, algoritmos genéticos y sistemas difusos. El artículo identifica que las técnicas de IA han mostrado resultados prometedores en la identificación de patrones precursores de sismos, aunque señala que la predicción sísmica sigue siendo un desafío abierto debido a la complejidad inherente de los procesos tectónicos. Los autores concluyen que la integración de múltiples técnicas de IA con datos geofísicos diversos representa una dirección prometedora para futuras investigaciones.

Para el artículo de Barrientos, Fernandez y Rivero (2007) se realiza el análisis estadístico y geográfico para eventos de tipo sísmico que se han registrado en la costa del Pacífico mexicano en el periodo de 1990 a 2004. El estudio utiliza datos del Servicio Sismológico Nacional para caracterizar la distribución espacial y temporal de la sismicidad en la región. Los autores aplican técnicas de estadística descriptiva y análisis de frecuencias para identificar patrones en la ocurrencia de sismos. Los resultados muestran que la actividad sísmica en la costa del Pacífico mexicano presenta variaciones significativas entre diferentes segmentos de la zona de subducción, con algunas áreas mostrando mayor frecuencia de eventos de magnitud moderada a alta.

En el artículo de Bommier (2023) se realiza un análisis comparativo entre dos metodologías ampliamente utilizadas en el análisis de valores extremos: el enfoque de excedencias sobre umbral (peak-over-threshold, POT) y el método de máximos anuales (block maxima). El autor evalúa ambos enfoques utilizando datos sintéticos y reales, analizando su desempeño en términos de sesgo, varianza y eficiencia en la estimación de cuantiles extremos. Los resultados indican que el método de máximos anuales es preferible cuando se dispone de series temporales largas y los eventos extremos son relativamente frecuentes, mientras que el enfoque POT puede ser más eficiente con muestras pequeñas pero requiere una selección cuidadosa del umbral.

Coles (2001) presenta en su libro “An Introduction to Statistical Modeling of Extreme Values” el marco teórico fundamental de la teoría de valores extremos (Extreme Value Theory, EVT). El autor desarrolla de manera rigurosa los fundamentos matemáticos de las distribuciones de valores extremos, incluyendo la distribución generalizada de valores extremos (GEV), la distribución de Gumbel, la distribución de Weibull y la distribución de Fréchet. El libro presenta métodos de estimación de parámetros, construcción de intervalos de confianza y cálculo de niveles de retorno. Esta obra constituye una referencia fundamental para el análisis probabilístico de eventos extremos en diversas disciplinas, incluyendo la sismología, hidrología y ciencias atmosféricas.

Coles, Bawa, Trenner y Dorazio (2021) publican una revisión actualizada del estado del arte en teoría de valores extremos, veinte años después del libro seminal de Coles (2001). Los autores examinan los avances metodológicos más significativos en el campo, incluyendo nuevos métodos de estimación, técnicas para datos no estacionarios y enfoques bayesianos. El artículo discute las aplicaciones emergentes de la teoría de valores extremos en áreas como el cambio climático, riesgo financiero y peligrosidad sísmica. Los autores identifican desafíos abiertos y direcciones futuras de investigación, destacando la necesidad de desarrollar métodos que permitan modelar la dependencia espacial y temporal en eventos extremos.

En otro trabajo, Convertito, Catalli y Emolo (2020) desarrollan un modelo integrado que combina la transferencia de esfuerzos tectónicos con el análisis de agrupamiento espacial de sismos para mejorar el pronóstico de la sismicidad. El modelo propuesto permite estimar cambios en la tasa de sismicidad considerando tanto los efectos de la transferencia de esfuerzos producida por sismos principales como los patrones de agrupamiento espacial observados en las réplicas. Los autores aplican el modelo a secuencias sísmicas en Italia, demostrando que la integración de ambos enfoques mejora significativamente la capacidad predictiva respecto a modelos que consideran cada factor por separado.

2. Introducción

Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano (2021) presentan una técnica innovadora para evaluar el riesgo de sismos inducidos por proyectos de actividad geotérmica. La técnica modifica el enfoque tradicional de análisis probabilístico de peligrosidad sísmica (PSHA) para incorporar la variabilidad temporal de la sismicidad inducida. Los autores aplican la metodología al campo geotérmico de St. Gallen en Suiza, donde la inyección de fluidos ha generado actividad sísmica. Los resultados demuestran que el enfoque dependiente del tiempo proporciona estimaciones más realistas del peligro sísmico durante las diferentes fases de operación del proyecto geotérmico.

Cornell (1968) presenta el artículo fundacional del análisis probabilístico de peligrosidad sísmica (Probabilistic Seismic Hazard Analysis, PSHA). El autor desarrolla el marco metodológico para cuantificar la probabilidad de que se excedan diferentes niveles de intensidad sísmica en un sitio determinado durante un periodo de tiempo específico. El artículo introduce el concepto de periodo de retorno y establece las bases matemáticas para calcular probabilidades de excedencia asumiendo independencia temporal entre eventos sísmicos. Esta metodología se ha convertido en el estándar internacional para la evaluación de peligrosidad sísmica y es ampliamente utilizada en el diseño de infraestructura crítica, normativas de construcción y políticas de gestión del riesgo sísmico.

Ferrás (2005) desarrolla un análisis probabilista para estimar la ocurrencia del próximo gran terremoto en el segmento Acapulco-San Marcos, México. Estudia los intervalos de tiempos intereventos sísmicos y encuentra que la distribución lognormal describe adecuadamente la recurrencia de terremotos grandes en esta región. El autor utiliza datos históricos de sismos significativos ocurridos en el segmento desde el siglo XIX para calibrar el modelo probabilístico. Los resultados proporcionan estimaciones de la probabilidad condicional de ocurrencia de un sismo mayor en diferentes horizontes temporales, información útil para la planificación de medidas de prevención y mitigación del riesgo sísmico.

Para el artículo de Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera (2022) se señala la importancia de pronosticar terremotos fuertes a largo plazo como una herramienta fundamental para minimizar los riesgos y las vulnerabilidades de las comunidades afectadas por la actividad sísmica. Los autores desarrollan un modelo de aprendizaje automático que utiliza datos históricos de sismos para identificar patrones de recurrencia en diferentes regiones del mundo, incluyendo América del Norte, América del Sur, Japón, el sur de China y el norte de la India. El modelo propuesto logra identificar ciclos de actividad sísmica que permiten estimar ventanas temporales con mayor probabilidad de ocurrencia de sismos significativos.

Jena, Pradhan, Naik y Alamri (2021) desarrollan un modelo integrado para la evaluación del riesgo sísmico en el noroeste de la India, combinando técnicas de aprendizaje profundo y análisis geoespacial. El estudio utiliza datos sísmicos históricos, información geológica y parámetros geomorfológicos para construir mapas de susceptibilidad sísmica. Los autores aplican redes neuronales convolucionales para identificar patrones espaciales asociados con la ocurrencia de sismos. Los resultados muestran que el modelo híbrido propuesto supera en precisión a los métodos tradicionales de evaluación del riesgo sísmico en la región de estudio.

Johnson, Kotz y Balakrishnan (1995) presentan en su libro “Continuous Univariate Distributions” una referencia exhaustiva sobre distribuciones de probabilidad continuas. Los autores desarrollan de manera rigurosa las propiedades matemáticas, métodos de estimación de parámetros y aplicaciones de una amplia variedad de distribuciones, incluyendo la distribución normal, lognormal, Weibull, Gumbel, logística y la distribución generalizada de valores extremos. El libro proporciona fórmulas para el cálculo de cuantiles, momentos y funciones generadoras de momentos. Esta obra constituye una referencia fundamental para la selección y aplicación de distribuciones de probabilidad en el análisis de datos, incluyendo el modelado de valores extremos como magnitudes sísmicas máximas.

En el artículo de Matsumoto, Yaoyama, Lee, Hida y Itoi (2023) se propone un modelo probabilista para la predicción del movimiento del suelo por terremotos llamado modelo de generación de movimiento del

suelo. El estudio utiliza redes generativas adversarias (GANs) para generar historias temporales sintéticas del movimiento del suelo que preservan las características estadísticas de los registros sísmicos reales. Los autores entrenan el modelo con datos de acelerogramas registrados en Japón, demostrando que las señales generadas reproducen adecuadamente el contenido frecuencial y la duración de los movimientos sísmicos observados.

Mignan, Danciu y Giardini (2021) desarrollan un marco metodológico integral para el pronóstico probabilístico de rupturas sísmicas aplicado específicamente a la evaluación de peligrosidad sísmica en instalaciones nucleares. Los autores revisan las mejores prácticas en modelado de fuentes sísmicas, incluyendo la caracterización de fallas activas y el cálculo de tasas de ocurrencia. El artículo discute los desafíos asociados con la integración de modelos físicos de ruptura con el análisis probabilístico tradicional, proponiendo enfoques híbridos que combinan información geológica, geodésica y sismológica para mejorar las estimaciones de peligrosidad.

Phung, Loh, Chao, Chiou y Huang (2020) desarrollan una ecuación de predicción del movimiento del suelo (GMPE) específica para sismos corticales en el territorio de Taiwán. El modelo propuesto utiliza un amplio conjunto de registros sísmicos para calibrar los coeficientes de la ecuación, considerando efectos de la magnitud, distancia, tipo de falla y condiciones locales del suelo. Los autores validan el modelo comparando las predicciones con observaciones independientes, demostrando que la GMPE específica para Taiwán proporciona estimaciones más precisas que las ecuaciones genéricas desarrolladas para otras regiones.

Ramírez-Gaytán, Aguirre-González y Núñez-Cornú (2021) realizan un análisis probabilístico exhaustivo de la brecha sísmica de Michoacán utilizando datos del Servicio Sismológico Nacional (SSN) que abarcan desde 1900 hasta la actualidad. Los autores analizan los parámetros de fuente de los sismos ocurridos en la región, identificando patrones de recurrencia y estimando el potencial sísmico de la brecha. El estudio proporciona información relevante para la evaluación de la peligrosidad sísmica en una de las regiones con mayor riesgo de México, donde se espera la ocurrencia de un sismo de gran magnitud en las próximas décadas.

En el artículo de Richter (1935) se introduce la escala de magnitud Richter para los terremotos, la cual se propone como herramienta para medir la magnitud de los sismos. Esta escala logarítmica permite cuantificar la energía liberada por un terremoto a partir de la amplitud máxima registrada en un sismógrafo estándar. El autor establece las bases para la comparación objetiva entre sismos ocurridos en diferentes lugares y tiempos, contribuyendo significativamente al desarrollo de la sismología cuantitativa moderna.

En el artículo de Galbán-Rodríguez (2021) se realiza un análisis teórico y metodológico sobre la predicción de terremotos, partiendo de su marco conceptual que contextualiza el fenómeno sísmico dentro de las ciencias de la Tierra. El autor revisa las diferentes aproximaciones a la predicción sísmica, desde los métodos empíricos basados en precursores hasta los modelos estadísticos y físicos. El artículo discute las limitaciones inherentes a la predicción sísmica y propone criterios para evaluar la validez de los pronósticos, contribuyendo a establecer un marco metodológico riguroso para la investigación en este campo.

Sánchez-Silva y Riaño (2020) implementan un modelo de renovación para la actualización en tiempo real del riesgo sísmico en sistemas de infraestructura interdependiente, con aplicaciones a datos sísmicos de la costa del Pacífico mexicano. Los autores desarrollan un enfoque basado en redes bayesianas que permite incorporar nueva información sísmica conforme se registran eventos, actualizando dinámicamente las estimaciones de riesgo. El modelo considera las interdependencias entre diferentes componentes de infraestructura, proporcionando una herramienta para la gestión del riesgo en sistemas complejos.

Para el artículo de Sawires, Peláez y Santoyo (2023) se lleva a cabo una evaluación probabilista actualizada del peligro sísmico en el occidente de México. El estudio utiliza un catálogo sísmico actualizado y modelos de fuentes sísmicas mejorados para calcular curvas de peligrosidad y mapas de aceleración esperada para diferentes periodos de retorno. Los autores comparan sus resultados con estudios previos,

2. Introducción

identificando diferencias significativas en algunas regiones que se atribuyen a la incorporación de nuevas fuentes sísmicas y a la utilización de ecuaciones de predicción del movimiento del suelo más recientes.

Serinaldi y Kilsby (2021) desarrollan una metodología para el cálculo de probabilidades de excedencia que considera explícitamente la autocorrelación temporal en series de valores extremos. Los autores demuestran que ignorar la persistencia temporal puede conducir a subestimaciones significativas del riesgo colectivo de eventos extremos. El artículo proporciona herramientas para cuantificar el efecto de la autocorrelación en las estimaciones de periodo de retorno y probabilidad de excedencia, con aplicaciones a datos hidrológicos que pueden extenderse al análisis de secuencias sísmicas.

En el artículo de Shreffler y Huecker (2025) se presenta una revisión de los principios estadísticos esenciales para la investigación clínica, incluyendo pruebas de hipótesis, valores p , intervalos de confianza y significancia estadística. Aunque el enfoque del artículo es clínico, los conceptos presentados son fundamentales para cualquier análisis estadístico riguroso, incluyendo el análisis de datos sísmicos. Los autores discuten la interpretación correcta de los resultados estadísticos y las limitaciones de las pruebas de hipótesis tradicionales.

Para el artículo de Shobanke, Olayemi y Olajide (2024) se realiza un análisis comparativo del desempeño de varias distribuciones de probabilidad continuas en el modelado de niveles máximos de inundación. Los autores evalúan distribuciones como la normal, lognormal, Gumbel, Weibull y la distribución generalizada de valores extremos utilizando datos hidrológicos. El estudio emplea criterios de selección de modelos como el AIC, BIC y pruebas de bondad de ajuste para identificar la distribución óptima. Aunque el contexto es hidrológico, la metodología es directamente aplicable al análisis de magnitudes sísmicas máximas.

Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir (2020) realizan un análisis retrospectivo del sismo intraplaca de Puebla del 19 de septiembre de 2017 (Mw 7.1), comparando la magnitud observada y el patrón de daños en la Ciudad de México con sismos históricos similares. Los autores analizan los registros de movimiento fuerte para caracterizar las particularidades de este evento, que causó daños significativos a pesar de su distancia al epicentro. El estudio proporciona información valiosa sobre la vulnerabilidad sísmica de la Ciudad de México y la importancia de considerar sismos intraplaca en la evaluación del peligro sísmico.

En el artículo de Singh, Arroyo, Srinivas y Suresh (2021) se desarrolla la ecuación de predicción del movimiento del suelo (GMPE), aplicada a los sismos ocurridos a lo largo del Arco del Himalaya occidental. Los autores utilizan datos de acelerogramas registrados en la región para calibrar los coeficientes del modelo, considerando las características tectónicas particulares de la zona de colisión continental. El estudio proporciona una herramienta para la evaluación de la peligrosidad sísmica en una de las regiones más activas del mundo.

Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique (2019) presenta un análisis estadístico descriptivo de series temporales de radón en el suelo, recopiladas durante un año en la zona de Muzaffarabad, la cual es atravesada por la falla de Jhelum que registró el sismo de Cachemira del 8 de octubre de 2005 (Mw 7.6). Los autores aplican técnicas de estadística descriptiva y análisis de boxplot para identificar anomalías en las concentraciones de radón que podrían estar asociadas con actividad sísmica. El estudio explora el potencial del monitoreo de radón como herramienta complementaria para la identificación de precursores sísmicos.

Yaghmaei-Sabegh y Neekmanesh (2022) realizan un análisis regional exhaustivo de parámetros de movimiento del suelo en Irán utilizando más de 100 años de registros sísmicos del catálogo nacional. Los autores aplican técnicas de aprendizaje automático y métodos bayesianos para desarrollar modelos de predicción del movimiento del suelo específicos para diferentes regiones tectónicas del país. El estudio demuestra que los modelos regionales proporcionan estimaciones más precisas que las ecuaciones globales, destacando la importancia de considerar las características tectónicas locales.

Zúñiga, Suárez, Figueroa-Soto y Mendoza (2022) proponen una regionalización sismotectónica de

primer orden para México basada en el análisis estadístico exhaustivo del catálogo del Servicio Sismológico Nacional (SSN). Los autores identifican regiones con características sísmicas homogéneas utilizando técnicas de agrupamiento espacial y análisis de la distribución de magnitudes. El estudio proporciona un marco de referencia para la evaluación de la peligrosidad sísmica en México, identificando las regiones con mayor potencial de generación de sismos significativos.

Estado del arte

Es utilizada principalmente la bibliografía encontrada en el capítulo anterior referente a la predicción probabilista sísmica utilizando el método de distribución logaritmo normal y gama, distribución de valores extremos, así como el estado del arte de los métodos desarrollados basados en inteligencia artificial, aprendizaje de máquina (machine learning) y redes neuronales.

Todo esto es aplicado a la solución del problema desde un enfoque que utiliza las técnicas conocidas y los datos estadísticos específicos para el problema a resolver.

3.1. Revisión del estado del arte

En esta sección se presenta una revisión del estado del arte de los trabajos más relevantes para este proyecto, con el propósito de enriquecer la metodología propuesta.

Los terremotos son vibraciones de la corteza terrestre (Abebe, Kebede, Kevin y Demissie 2023), las cuales pueden provocar temblores, incendios, deslizamientos de tierra y fracturas en el terreno que representan una amenaza significativa para la vida humana y la infraestructura. La predicción y evaluación del peligro sísmico constituye un área de investigación activa que ha evolucionado significativamente en las últimas décadas.

Dentro de las distintas formas de realizar inferencia y predicción sísmica se han desarrollado técnicas basadas en IA (Al Banna, Taher, Kaiser, Mahmud, Rahman, Hosen y Cho 2020) para identificar patrones ocultos en los datos sísmicos que podrían preceder a eventos significativos. Estas técnicas incluyen redes neuronales artificiales, máquinas de soporte vectorial y algoritmos de aprendizaje profundo que han mostrado resultados prometedores en diversas regiones del mundo.

Estas técnicas generalmente se cimentan en análisis matemáticos, probabilistas y estadísticos (Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano 2021, Matsumoto, Yaoyama, Lee, Hida y Itoi 2023, Sawires, Peláez y Santoyo 2023, Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022) para estimar la probabilidad de ocurrencia de sismos en una zona determinada. El enfoque probabilístico permite cuantificar la incertidumbre inherente a los procesos sísmicos y proporciona herramientas para la toma de decisiones en materia de gestión del riesgo.

El análisis de valores extremos se sustenta principalmente en dos enfoques metodológicos: el método de máximos anuales y el de excedencias sobre umbral (peak-over-threshold). Bommier (2023) realiza un análisis comparativo exhaustivo entre ambos enfoques, concluyendo que el método de máximos anuales es preferible cuando se dispone de series temporales largas, como es el caso de los catálogos sísmicos históricos. Este enfoque constituye la base metodológica del presente trabajo.

El marco teórico fundamental para el análisis probabilístico de peligrosidad sísmica fue establecido por Cornell (1968) en su artículo seminal sobre análisis de riesgo sísmico en ingeniería. Este trabajo introduce el concepto de periodo de retorno y establece las bases matemáticas para calcular probabilidades de excedencia, metodología que se ha convertido en el estándar internacional para la evaluación de peligrosidad sísmica y que es aplicada en el presente proyecto.

Con respecto a estos cálculos matemáticos y estadísticos, se utilizan técnicas como los modelos lineales de tipo log-lineal (Barrientos, Fernandez y Rivero 2007), log-normales (Ferrás 2005) y distribuciones de valores extremos (Coles 2001, Shobanke, Olayemi y Olajide 2024) para caracterizar la distribución de magnitudes sísmicas. La selección de la distribución apropiada para cada región es fundamental para obtener estimaciones confiables de peligrosidad.

Johnson, Kotz y Balakrishnan (1995) proporcionan una referencia exhaustiva sobre distribuciones de probabilidad continuas, incluyendo las distribuciones normal, logística, Weibull y Gumbel que son empleadas en este trabajo para modelar las magnitudes máximas anuales en las diferentes regiones de estudio. La correcta caracterización de estas distribuciones permite calcular niveles de retorno y probabilidades de excedencia con rigor matemático.

El cálculo de períodos de retorno y probabilidades de excedencia constituye un componente esencial de la evaluación probabilística de peligrosidad sísmica. Mignan, Danciu y Giardini (2021) desarrollan un marco metodológico para el pronóstico probabilístico de rupturas sísmicas, estableciendo las mejores prácticas para la integración de modelos físicos con el análisis estadístico. Estos conceptos son aplicados en la Fase 10 de la metodología del presente trabajo para estimar cuándo podría ocurrir un sismo de magnitud significativa.

El análisis temporal de la actividad sísmica y el concepto de brecha sísmica han ganado relevancia en años recientes como complemento a los modelos puramente probabilísticos. Sánchez-Silva y Riaño (2020) implementan modelos de renovación para la actualización en tiempo real del riesgo sísmico, mientras que Ramírez-Gaytán, Aguirre-González y Núñez-Cornú (2021) analizan específicamente la brecha sísmica de Michoacán, una de las regiones consideradas en el presente estudio.

La construcción de índices compuestos de peligrosidad sísmica que integren múltiples factores de riesgo representa una tendencia metodológica reciente. Convertito, Catalli y Emolo (2020) desarrollan un modelo que combina la transferencia de esfuerzos tectónicos con el análisis de agrupamiento espacial de sismos, proporcionando un enfoque más completo para el pronóstico de la sismicidad.

Cabe destacar también el análisis hecho con series temporales de gas radón (Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique 2019) que más allá de solo buscar patrones en magnitudes, epicentros, fechas y profundidades, busca precursores de sismos mediante análisis estadístico descriptivo de variables geoquímicas. Este tipo de estudios complementan los enfoques basados exclusivamente en datos sismológicos.

Los artículos revisados abarcan distintas áreas geográficas del planeta, como la costa occidental de México (Barrientos, Fernandez y Rivero 2007, Sawires, Peláez y Santoyo 2023, Zúñiga, Suárez, Figueroa-Soto y Mendoza 2022), el Cuerno de África (Abebe, Kebede, Kevin y Demissie 2023), Taiwán (Phung, Loh, Chao, Chiou y Huang 2020), el Himalaya occidental (Singh, Arroyo, Srinivas y Suresh 2021), Irán (Yaghmaei-Sabegh y Neekmanesh 2022), la India (Jena, Pradhan, Naik y Alamri 2021), Japón (Matsumoto, Yaoyama, Lee, Hida y Itoi 2023) y Suiza (Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano 2021). Esta diversidad geográfica permite identificar metodologías aplicables a diferentes contextos tectónicos.

También podemos encontrar trabajos con un enfoque conceptual y metodológico que revisan teorías de predicción sísmica y su evolución a lo largo del tiempo (Galbán-Rodríguez 2021). Estos trabajos proporcionan el marco conceptual necesario para entender las limitaciones inherentes a la predicción sísmica y los criterios para evaluar la validez de los pronósticos.

En el contexto mexicano específico, varios estudios recientes han aplicado metodologías de valores extremos y análisis de brechas sísmicas a diferentes regiones del país. Ramírez-Gaytán, Aguirre-González y Núñez-Cornú (2021) analizan la brecha sísmica de Michoacán, mientras que Sawires, Peláez y Santoyo (2023) desarrollan una evaluación probabilista actualizada del peligro sísmico en el occidente de México. El sismo de Puebla de 2017 analizado por Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir (2020) evidencia la importancia de considerar diferentes tipos de fuentes sísmicas en la evaluación del peligro.

Desde una perspectiva estadística, Shreffler y Huecker (2025) proporcionan una revisión actualizada de los principios de pruebas de hipótesis, valores p e intervalos de confianza que son aplicados en las Fases 6 y 8 de la metodología del presente trabajo. Serinaldi y Kilsby (2021) abordan el problema de la persistencia temporal en series de valores extremos, un aspecto relevante para el análisis de secuencias sísmicas.

La escala de magnitud introducida por Richter (1935) constituye la base para la cuantificación de la energía liberada por los sismos y permite la comparación objetiva entre eventos ocurridos en diferentes lugares y tiempos. Esta escala, junto con sus evoluciones posteriores, es utilizada en el presente trabajo para caracterizar la actividad sísmica en las regiones de estudio.

3.2. Análisis del estado del arte

La revisión del estado del arte muestra que la predicción y evaluación del peligro sísmico se ha abordado desde múltiples perspectivas: modelos probabilistas clásicos, distribuciones estadísticas de valores extremos, técnicas de inteligencia artificial y enfoques híbridos que combinan información física y estadística. El presente trabajo se enmarca dentro del enfoque probabilístico clásico, utilizando la teoría de valores extremos y las distribuciones estadísticas como herramientas fundamentales.

El marco metodológico establecido por Cornell (1968) para el cálculo de periodos de retorno y probabilidades de excedencia constituye la base conceptual de la Fase 10 del presente trabajo. Este enfoque, complementado con los desarrollos teóricos de Coles (2001) sobre modelado estadístico de valores extremos, permite obtener estimaciones cuantitativas del peligro sísmico con fundamento matemático riguroso.

La selección de distribuciones de probabilidad para modelar magnitudes máximas anuales, documentada exhaustivamente por Johnson, Kotz y Balakrishnan (1995), es un aspecto crítico de la metodología. El análisis comparativo de Shobanke, Olayemi y Olajide (2024) proporciona criterios objetivos para la selección de la distribución óptima en cada región, enfoque que es aplicado en la Fase 9 mediante pruebas de bondad de ajuste.

Desde una perspectiva metodológica, se identifica un consenso creciente sobre las mejores prácticas para el análisis probabilístico de peligrosidad sísmica: (1) empleo de magnitudes máximas anuales para aplicar la teoría de valores extremos, (2) selección de distribuciones mediante criterios estadísticos objetivos como el BIC y pruebas de bondad de ajuste, (3) estimación de parámetros por máxima verosimilitud, y (4) cálculo de niveles de retorno y probabilidades de excedencia para diferentes horizontes temporales. Estas prácticas son incorporadas en la metodología del presente proyecto.

Capítulo 4

Metodología

4.1. Metodología de investigación

En este capítulo se presenta la metodología propuesta para realizar el análisis y las evaluaciones de la actividad sísmica en estados de la costa del Pacífico mexicano (Chiapas, Guerrero, Michoacán y Oaxaca), en el resto del país (actividad sísmica en México sin considerar los 4 estados mencionados) y de todos los sismos nacionales (actividad sísmica de los 32 estados de México), con el fin de lograr la estimación probabilista de ocurrencia de sismos.

La Figura 1 muestra el diagrama de flujo de la metodología implementada para el presente proyecto.

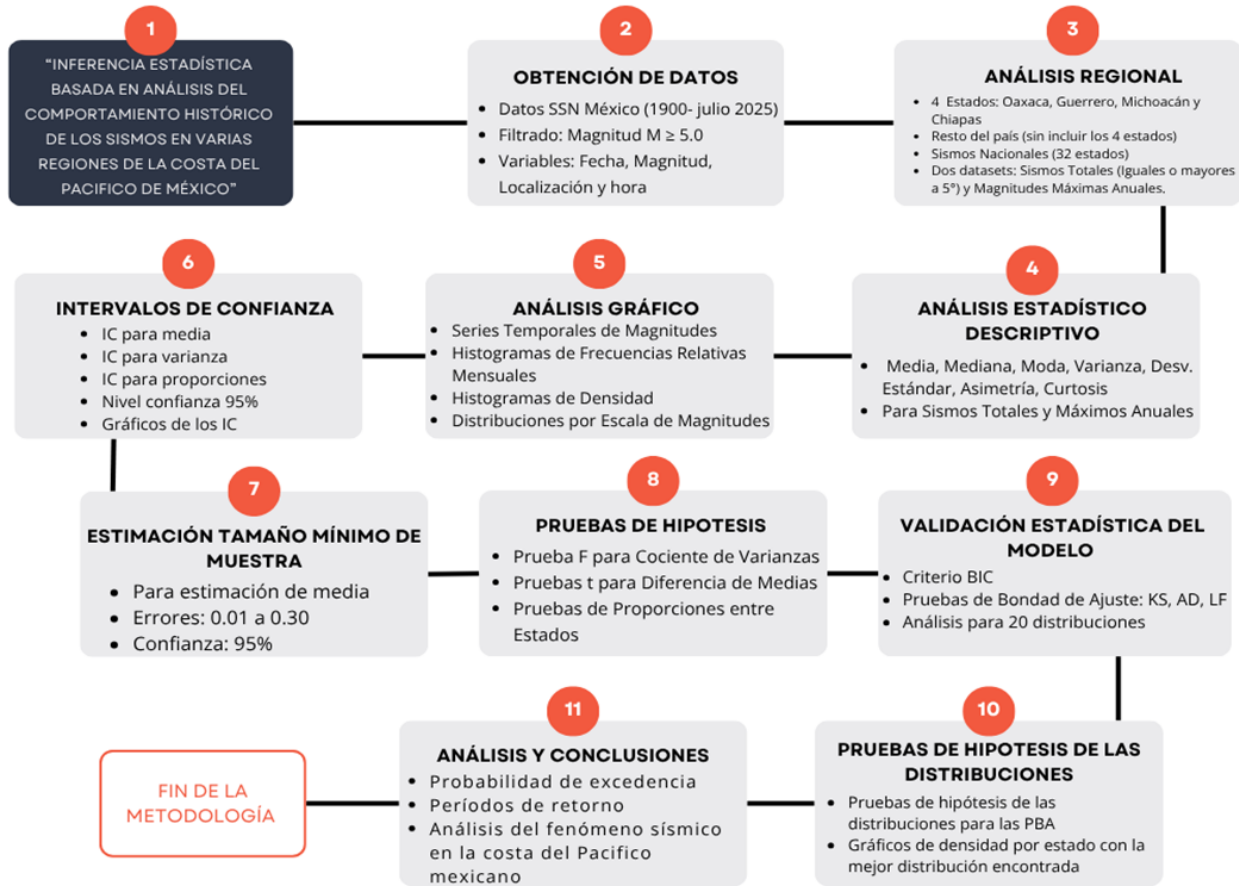


Figura 1: Diagrama de flujo de la aplicación de la metodología propuesta.

4.2. Fases de desarrollo del proyecto

Para implementar la metodología propuesta en este capítulo, se dividieron los trabajos de este proyecto en 10 fases, las cuales son presentadas a continuación.

Fase 1 – Recopilación de datos sísmicos históricos. Se recopilan los datos históricos de los sismos registrados en la República Mexicana en el periodo 1900 – 2025, desde el portal web del Servicio Sismológico Nacional (México. Universidad Nacional Autónoma de México, I. d. G., Servicio Sismológico Nacional 2025). Este catálogo proporciona información detallada de cada evento sísmico, las cuales son: fecha, hora, magnitud (en grados Richter), latitud, longitud, profundidad, referencia de localización, fecha (UTC), hora (UTC) y estatus (revisado/no revisado).

Fase 2 - Filtrado y preparación de los datos. En esta fase se aplican los siguientes criterios de filtrado y limpieza a los datos sísmicos:

- **Filtrado por magnitud.** Se eliminan sismos con magnitud menor a 5.0° , esto para que el análisis se enfoque en eventos de magnitud moderada a fuerte, que suelen ser sismos perceptibles para la mayoría de la población y que pueden significar un peligro para las personas, la infraestructura y las construcciones.

- **Selección de parámetros.** Se conservan únicamente las variables relevantes para el proyecto que son fecha, hora, magnitud y referencia de localización. Se eliminan las variables que no son relevantes para este proyecto.
- **Limpieza de los datos.** Se eliminan registros con valores vacíos, nulos o no disponibles para garantizar la consistencia e integridad de los datos.
- **División por región.** Se generan archivos individuales con la actividad sísmica para cada una de las 6 regiones estudiadas en este proyecto: Chiapas, Guerrero, Michoacán, Oaxaca, resto nacionales y sismos nacionales.

Fase 3 – Generación de conjunto de datos. Se generan dos grupos de datasets (cada grupo con datasets individuales para cada región):

- **Sismos Totales.** Información de todos los sismos registrados con magnitud $\geq 5^\circ$, para cada una de las regiones de estudio, que permita analizar la frecuencia y distribución de los eventos sísmicos ocurridos en cada región.
- **Magnitudes Máximas Anuales.** Información del sismo de mayor magnitud registrado cada año (y de magnitud mínima de 5°), para cada una de las regiones, que permita aplicar la teoría de valores extremos a este dataset.

Fase 4 – Cálculo de estadísticos descriptivos. Se calculan estadísticos descriptivos de los sismos para cada región y para cada grupo (sismos totales y magnitudes máximas anuales). A continuación, se presentan los estadísticos descriptivos calculados:

4.2.1. Media aritmética

La media aritmética \bar{x} representa el valor promedio de la magnitud de los sismos analizados y se calcula como (Climent Hernández 2022):

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad (4.1)$$

donde \bar{x} es la media aritmética, n el tamaño de la muestra (número total de sismos) y x_k es la magnitud del k -ésimo sismo. Conocer el valor de la media sirve para indicar la magnitud promedio que hubo de los sismos para alguna de las regiones durante el periodo de estudio.

4.2.2. Mediana

La mediana \tilde{x} es el valor central de los datos ordenados (de menor a mayor) y cuyo valor es menos sensible ante valores extremos (sismos muy fuertes) que el de la media aritmética, se calcula como (Climent Hernández 2022):

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & \text{si } n \text{ es par} \end{cases} \quad (4.2)$$

donde \tilde{x} es la mediana, n es el tamaño de la muestra (número total de sismos) y $x_{(\cdot)}$ es el valor ordenado de manera ascendente de la magnitud de los sismos. Obtener la mediana nos permite conocer el valor que divide a la distribución de los datos sísmicos en dos partes iguales.

4.2.3. Moda

La moda \hat{x} es el valor de magnitud sísmica que aparece más veces dentro del conjunto de datos sísmicos y se encuentra como (Climent Hernández 2022):

$$\hat{x} = \{x_k \in X \mid f_k = \max_{x_j \in X} (f_j)\} \quad (4.3)$$

donde \hat{x} es la moda, f_k es la frecuencia de ocurrencia del valor sísmico x_k y X se refiere al conjunto de datos con todas las magnitudes sísmicas. Conocer este valor permite saber cuál es la magnitud sísmica que se repite más veces en cada región de estudio.

4.2.4. Varianza muestral

La varianza muestral s_x^2 cuantifica la dispersión de los datos sísmicos respecto al valor de la media aritmética y se encuentra como (Climent Hernández 2022):

$$s_x^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 \quad (4.4)$$

donde s_x^2 es la varianza, n es el tamaño de la muestra (número total de sismos), x_k es el valor de la magnitud del k -ésimo sismo y \bar{x} es la media aritmética. Conocer este valor permite saber si en una región hay gran variación en el valor de magnitud de los sismos registrados (valor alto de varianza) o si en cambio, hay poca variación en la magnitud de los sismos registrados en la región (valor bajo de varianza).

4.2.5. Desviación estándar

La desviación estándar s_x se encuentra al aplicar raíz cuadrada al valor de la varianza, tal y como sigue (Climent Hernández 2022):

$$s_x = \sqrt{s_x^2} \quad (4.5)$$

donde s_x es la desviación estándar y s_x^2 la varianza. Dado que las unidades de la varianza son cuadráticas, no se pueden comparar directamente con los valores de las muestras de las magnitudes sísmicas, mientras que la desviación estándar facilita los análisis de relación de la dispersión con los valores de los datos sísmicos.

4.2.6. Coeficiente de asimetría

El coeficiente de asimetría g_1 mide el grado de simetría de la distribución de los datos sísmicos respecto a la media de estos, se calcula como (Fisher 1922):

$$g_1 = \frac{1}{ns_x^3} \sum_{k=1}^n (x_k - \bar{x})^3 \quad (4.6)$$

donde g_1 es el coeficiente de asimetría, n es el tamaño de la muestra (número total de sismos), s_x es la desviación estándar, x_k es la magnitud del k -ésimo sismo y \bar{x} la media. Los posibles resultados de g_1 mostrarían que si es igual a 0 ($g_1 = 0$) la distribución de los datos sísmicos es simétrica (sin sesgo), si es mayor a 0 ($g_1 > 0$) que presenta asimetría positiva o sesgo a la derecha (cola derecha más larga, que indica que hay más sismos de baja magnitud con eventos ocasionales fuertes) y si es menor a 0 ($g_1 < 0$) que presenta asimetría negativa o sesgo a la izquierda (cola izquierda más larga, que indica concentración de sismos hacia las magnitudes altas).

4.2.7. Coeficiente de curtosis

El coeficiente de curtosis g_2 mide el grado de apuntamiento (qué tan plana o picuda) o curtosis de la distribución de los datos sísmicos respecto de una distribución normal (gaussiana) y se calcula como (Fisher 1922):

$$g_2 = \frac{1}{ns_x^4} \sum_{k=1}^n (x_k - \bar{x})^4 \quad (4.7)$$

donde g_2 es el coeficiente de curtosis, n es el tamaño de la muestra (número total de sismos), x_k es la magnitud del k -ésimo sismo y \bar{x} la media. Los posibles resultados de g_2 muestran que si es igual a 0 ($g_2 = 0$) la distribución de los datos sísmicos es mesocúrtica (similar a la normal), si es mayor a 0 ($g_2 > 0$) que presenta distribución leptocúrtica o más picuda que la normal (que indica que hay una mayor concentración de magnitudes sísmicas cercanas a la media con colas más pesadas) y si es menor a 0 ($g_2 < 0$) que presenta distribución platocúrtica o más aplanada que la normal (que indica que los valores de magnitudes sísmicas están dispersos de una manera uniforme).

Fase 5 – Representación gráfica del comportamiento sísmico. Se generan gráficos del comportamiento histórico de los sismos en las regiones analizadas para los dos datasets (sismos totales y magnitudes máximas anuales) en el periodo de 1900 a julio de 2025. Las visualizaciones incluyen:

- **Histogramas de densidad de los sismos.** Gráficos que muestran la distribución de las magnitudes sísmicas, permitiendo identificar la forma de la distribución de los datos de magnitudes sísmicas.
- **Histogramas de frecuencias relativas por mes.** Sirven para identificar patrones de estacionalidad en la ocurrencia de sismos.
- **Histogramas de sismos por magnitud.** Permiten cuantificar el número de eventos en cada rango de magnitud.
- **Gráfico de magnitud máxima, promedio y mínima.** Estos gráficos permiten ver la evolución temporal de las magnitudes mínima, promedio y máxima a lo largo del tiempo, permitiendo identificar tendencias y cambios en el comportamiento sísmico.

Fase 6 – Estimación de intervalos de confianza. Se estiman los intervalos de confianza para la media (μ), varianza (σ^2) y proporción (π), con un nivel de confianza al 95 %. Estos intervalos de confianza permiten obtener rangos dentro de los cuales se espera que estén los valores poblacionales con un 95 % de probabilidad (Climent Hernández 2022). Los intervalos estimados son:

4.2.8. Intervalo de confianza para la media

Cuando la varianza poblacional es desconocida y se estima a partir de la muestra, se utiliza la distribución t de Student (Student 1908):

$$IC(\mu) = \mu \in \left(\bar{x} - \frac{s_x t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}}, \bar{x} + \frac{s_x t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}} \right) \quad (4.8)$$

donde μ es la media, \bar{x} es la media muestral, s_x es la desviación estándar, $t_{1-\frac{\alpha}{2}}(n-1)$ es el valor crítico de la distribución t con $n-1$ grados de libertad, $\alpha = 0.05$ para un nivel de confianza del 95 % y n el tamaño de la muestra. Obtener este intervalo de confianza permite estimar con un 95 % de nivel de confianza el rango en el que se encuentra la magnitud promedio verdadera de los sismos para cada región.

4.2.9. Intervalo de confianza para la varianza

Se utiliza la distribución chi-cuadrada (χ^2) para estimar el intervalo de confianza de la varianza poblacional (Climent Hernández 2022):

$$IC(\sigma^2) = \sigma^2 \in \left(\frac{(n-1)s_x^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)s_x^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right) \quad (4.9)$$

donde σ^2 es la varianza, $\chi_{1-\frac{\alpha}{2}}^2$ y $\chi_{\frac{\alpha}{2}}^2$ son los valores críticos de la distribución chi-cuadrado con $n-1$ grados de libertad para los percentiles $1-\frac{\alpha}{2}$ y $\frac{\alpha}{2}$, respectivamente. Este intervalo permite cuantificar la incertidumbre en la estimación de la variabilidad de las magnitudes sísmicas en cada región de estudio.

4.2.10. Intervalo de confianza para la proporción

Se utiliza la aproximación normal para estimar el intervalo de confianza de una proporción poblacional:

$$IC(\pi) = \pi \in \left(p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right) \quad (4.10)$$

donde π es la proporción, p es la proporción muestral, $z_{1-\frac{\alpha}{2}}$ es el valor crítico de la distribución estándar. Este intervalo es útil para estimar la proporción de sismos que exceden un umbral crítico de magnitud definida en 6.5° , la cual permite realizar evaluaciones más certeras del riesgo sísmico.

Fase 7 – Estimación del tamaño mínimo de la muestra. Se determina el tamaño mínimo de la muestra para la media (μ), con un error máximo tolerado ε y un nivel de confianza del 95 %:

$$n \geq \left\lceil \left(\frac{s_X t_{1-\frac{\alpha}{2}}(n-1)}{\varepsilon} \right)^2 \right\rceil \quad (4.11)$$

donde n es el tamaño de la muestra (número total de sismos), s_X es la desviación estándar, $t_{1-\frac{\alpha}{2}}(n-1)$ es el valor crítico de la distribución t de Student, ε es el error máximo tolerado, $\lceil \cdot \rceil$ indica que es una función techo (se redondea hacia arriba al entero más cercano). Este cálculo permite determinar si el tamaño de muestra es suficiente para alcanzar una precisión deseada en la estimación de los parámetros.

Fase 8 – Prueba de hipótesis. Se aplican pruebas de hipótesis para comparar estadísticamente las características sísmicas entre las diferentes regiones. Estas pruebas permiten tomar decisiones basadas en evidencia estadística sobre si existen diferencias significativas entre poblaciones (Shreffler y Huecker 2025).

En todas las pruebas se utiliza un nivel de significancia de $\alpha = 0.05$, lo que implica una probabilidad del 5 % de rechazar incorrectamente la hipótesis nula (error tipo I).

4.2.11. Prueba de hipótesis para el cociente de varianzas (prueba F)

La prueba de hipótesis para el cociente de varianzas permite determinar si dos poblaciones tienen la misma varianza. Se utiliza antes de realizar pruebas de comparación de medias para decidir si se pueden asumir varianzas iguales.

Hipótesis:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Estadístico de prueba:

$$F = \frac{s_1^2}{s_2^2} \sim F(n_1 - 1, n_2 - 1) \quad (4.12)$$

donde s_1^2 y s_2^2 son las varianzas muestrales de las dos poblaciones y la distribución F tiene $n - 1$ grados de libertad. Se rechaza H_0 si $F > F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$ o si $F < F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$, donde estos valores son los percentiles críticos de la distribución F. Este cálculo permite determinar si la variabilidad sísmica difiere entre las distintas regiones de estudio.

4.2.12. Prueba de hipótesis para diferencia de medias con varianzas iguales (pooled)

Cuando la prueba F indica que las varianzas son iguales, se utiliza la prueba t para la diferencia de medias con varianzas iguales (pooled):

Hipótesis:

$$H_0 : \bar{x}_1 = \bar{x}_2$$

$$H_1 : \bar{x}_1 \neq \bar{x}_2$$

Estadístico de prueba:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2) \quad (4.13)$$

donde en el denominador se encuentra la varianza agrupada (pooled), que combina la información de ambas muestras para obtener una mejor estimación de la varianza común. Se rechaza H_0 si $|t| > t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)$. Este cálculo permite determinar si la magnitud promedio de sismos difiere significativamente entre dos regiones.

4.2.13. Prueba de hipótesis para diferencia de medias con varianzas diferentes (Welch)

Cuando las varianzas son significativamente diferentes, se utiliza la prueba de hipótesis para la diferencia de medias con varianzas diferentes de Welch:

Hipótesis:

$$H_0 : \bar{x}_1 = \bar{x}_2$$

$$H_1 : \bar{x}_1 \neq \bar{x}_2$$

Estadístico de prueba:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(v) \quad (4.14)$$

Grados de libertad:

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} \quad (4.15)$$

donde esta prueba es más robusta que la prueba t estándar cuando las varianzas poblacionales son diferentes.

4.2.14. Prueba de hipótesis para la diferencia de proporciones con proporciones agrupadas o iguales (pooled)

Esta prueba se utiliza para comparar la proporción de sismos que exceden un umbral crítico entre dos regiones:

Hipótesis:

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

Estadístico de prueba:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim z \quad (4.16)$$

Proporción agrupada:

$$p = \frac{x_1 + x_2}{n_1 + n_2} \quad (4.17)$$

siendo x_1 y x_2 el número de éxitos en cada muestra. Se rechaza H_0 si $|z| > z_{1-\frac{\alpha}{2}} = 1.96$ para $\alpha = 0.05$.

4.2.15. Prueba de hipótesis para la diferencia de proporciones sin agrupamiento o diferentes

Cuando no se asumen proporciones iguales bajo la hipótesis nula, se utiliza:

Hipótesis:

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

Estadístico de prueba:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim z \quad (4.18)$$

donde en esta prueba no se asume que las proporciones son iguales bajo H_0 , lo que la hace más conservadora y apropiada cuando las proporciones muestrales son muy diferentes.

Fase 9 – Validación estadística del modelo probabilístico. Implementación de pruebas de bondad de ajuste a distintas distribuciones para determinar qué distribución probabilística se ajusta mejor a los datos de magnitudes sísmicas. También, aplicación de Criterio de Información Bayesiano (BIC) para la selección de modelos.

4.2.16. Prueba de bondad de ajuste Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov evalúa si una muestra proviene de una distribución teórica específica comparando la función de distribución empírica con la teórica:

Hipótesis:

$$H_0 = X \sim F(x; \theta)$$

$$H_1 = X \not\sim F(x; \theta)$$

donde $F(x; \theta)$ es la distribución teórica propuesta con parámetros θ .

Estadístico de prueba:

$$D = \max_x |F_n(x) - F(x)| \quad (4.19)$$

donde $F_n(x)$ es la función de distribución empírica, $F(x)$ es la función de distribución teórica. El estadístico D mide la máxima discrepancia vertical entre las funciones de distribución. Se rechaza H_0 si el valor $p < \alpha$.

4.2.17. Prueba de bondad de ajuste Anderson-Darling

La prueba de Anderson-Darling es más sensible a discrepancias en las colas de la distribución que la prueba de Kolmogorov-Smirnov:

Hipótesis:

$$H_0 = X \sim F(x; \theta)$$

$$H_1 = X \not\sim F(x; \theta)$$

Estadístico de prueba:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1)(\ln F(x_i) + \ln(1 - F(x_{n+1-i})))] \quad (4.20)$$

donde los datos x_1, x_2, \dots, x_n están ordenados de menor a mayor. La ponderación $(2i-1)$ da mayor peso a las observaciones en las colas, haciendo la prueba más sensible a desviaciones en valores extremos, lo cual es muy importante para el análisis de sismos fuertes. Se rechaza H_0 si $A^2 > A_{\text{crítico}}^2(\alpha)$, donde el valor crítico depende de la distribución teórica específica.

4.2.18. Prueba de bondad de ajuste Lilliefors

La prueba de bondad de ajuste de Lilliefors es una modificación de la prueba de Kolmogorov-Smirnov para el caso en que los parámetros de la distribución se estiman a partir de los datos:

Hipótesis:

$$H_0 = X \sim F(x; \theta)$$

$$H_1 = X \not\sim F(x; \theta)$$

Estadístico de prueba:

$$D = \max_x |F_n(x) - F(x; \hat{\mu}, \hat{\sigma})| \quad (4.21)$$

donde $\hat{\mu}$ y $\hat{\sigma}$ son estimadores de los parámetros calculados a partir de la muestra. Cuando los parámetros se estiman de los datos, la distribución del estadístico D difiere de la distribución de Kolmogorov-Smirnov estándar, requiriendo valores críticos específicos de Lilliefors. Esta prueba es útil para probar normalidad cuando la media y desviación estándar se estiman de los datos (Diana Villasana Ocampo 2024).

4.2.19. Criterio de Información Bayesiano

El Criterio de Información Bayesiano (BIC) permite comparar múltiples modelos y seleccionar el que mejor equilibra bondad de ajuste y parsimonia:

$$\text{BIC} = -2 \cdot \ln(\hat{L}) + k \cdot \ln(n) \quad (4.22)$$

donde \hat{L} es la verosimilitud máxima del modelo, k es el número de parámetros del modelo, n es el tamaño de la muestra. Se prefiere el modelo con menor BIC. BIC penaliza más fuertemente la complejidad del modelo, favoreciendo modelos más simples cuando el tamaño de muestra es grande.

Fase 10 – Cálculo de probabilidades de ocurrencia de sismos. En esta fase se realiza la inferencia probabilística que constituye el objetivo principal de este proyecto. Utilizando la distribución óptima identificada para cada región en la Fase 9, se calculan los periodos de retorno, niveles de retorno y probabilidades de excedencia que permiten estimar cuándo podría ocurrir un sismo de magnitud significativa, de qué magnitud se espera y con qué probabilidad.

4.2.20. Selección de la distribución óptima por región

Una vez obtenidos los resultados de las pruebas de bondad de ajuste (Kolmogorov-Smirnov, Anderson-Darling, Lilliefors) y el Criterio de Información Bayesiano, se selecciona para cada región la distribución que presente el menor valor de BIC y que no rechace la hipótesis nula en las pruebas aplicadas. Como establece Coles (2001), la selección del modelo debe basarse en criterios estadísticos objetivos que permitan identificar la distribución que mejor representa el comportamiento de los datos.

La Tabla 2 muestra las distribuciones óptimas identificadas para cada una de las regiones de estudio.

Tabla 2: Distribuciones óptimas por región según pruebas de bondad de ajuste.

Región	Distribución óptima
Oaxaca	Gumbel
Guerrero	Weibull
Michoacán	GEV
Chiapas	Weibull
Resto Nacionales	Logística
Nacionales	Normal

4.2.21. Estimación de parámetros por máxima verosimilitud

Para la distribución seleccionada en cada región, los parámetros se estiman mediante el método de máxima verosimilitud (Fisher 1922). Este método consiste en encontrar los valores de los parámetros que maximizan la función de verosimilitud:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \ln f(x_i; \theta) \quad (4.23)$$

donde $\hat{\theta}$ representa el vector de parámetros estimados, n es el número de observaciones (magnitudes máximas anuales), x_i es la magnitud máxima del i -ésimo año y $f(x_i; \theta)$ es la función de densidad de la distribución seleccionada evaluada en x_i con parámetros θ .

4.2.22. Cálculo del periodo de retorno

El periodo de retorno T representa el tiempo promedio esperado entre eventos que exceden una magnitud específica x . De acuerdo con Cornell (1968), el periodo de retorno se calcula como:

$$T(x) = \frac{1}{1 - F(x; \hat{\theta})} \quad (4.24)$$

donde $F(x; \hat{\theta})$ es la función de distribución acumulada de la magnitud máxima anual evaluada en x con los parámetros estimados $\hat{\theta}$.

Por ejemplo, si para una región determinada se obtiene que $F(7.0) = 0.98$, entonces el periodo de retorno para un sismo de magnitud 7.0° sería $T = 1/(1 - 0.98) = 50$ años. Esto significa que, en promedio, se espera un sismo de magnitud igual o mayor a 7.0° cada 50 años en esa región.

Es importante aclarar que el periodo de retorno es un valor promedio de largo plazo y no implica que después de un evento el siguiente ocurrirá exactamente T años después. La ocurrencia de sismos sigue siendo un proceso aleatorio.

4.2.23. Cálculo del nivel de retorno

El nivel de retorno x_T es la magnitud asociada a un periodo de retorno específico T . Se obtiene invirtiendo la función de distribución acumulada (Coles 2001):

$$x_T = F^{-1}\left(1 - \frac{1}{T}; \hat{\theta}\right) \quad (4.25)$$

A continuación se presentan las fórmulas específicas para el cálculo del nivel de retorno según cada una de las distribuciones empleadas en este proyecto.

Nivel de retorno para la distribución de Gumbel

Para la región de Oaxaca, donde la distribución de Gumbel presenta el mejor ajuste, el nivel de retorno se calcula como:

$$x_T = \mu - \sigma \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) \quad (4.26)$$

donde μ es el parámetro de localización y $\sigma > 0$ es el parámetro de escala.

Nivel de retorno para la distribución de Weibull

Para las regiones de Guerrero y Chiapas, donde la distribución de Weibull presenta el mejor ajuste, el nivel de retorno se calcula como:

$$x_T = \lambda \left(-\ln\left(\frac{1}{T}\right)\right)^{1/k} \quad (4.27)$$

donde $k > 0$ es el parámetro de forma y $\lambda > 0$ es el parámetro de escala. La distribución de Weibull es apropiada para modelar valores extremos cuando existe un límite superior para las magnitudes, lo cual es consistente con las características sismotectónicas de estas regiones.

Nivel de retorno para la distribución generalizada de valores extremos

Para la región de Michoacán, donde la distribución GEV presenta el mejor ajuste, el nivel de retorno se calcula como:

$$x_T = \begin{cases} \mu + \frac{\sigma}{\xi} \left[\left(-\ln\left(1 - \frac{1}{T}\right)\right)^{-\xi} - 1 \right] & \text{si } \xi \neq 0 \\ \mu - \sigma \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) & \text{si } \xi = 0 \end{cases} \quad (4.28)$$

donde μ es el parámetro de localización, $\sigma > 0$ es el parámetro de escala y ξ es el parámetro de forma. El parámetro ξ determina el comportamiento de la cola de la distribución: si $\xi > 0$ indica cola pesada (tipo Fréchet), si $\xi < 0$ indica cola limitada (tipo Weibull) y si $\xi = 0$ la distribución GEV se reduce a la distribución de Gumbel.

Nivel de retorno para la distribución Logística

Para la región de Resto Nacionales, donde la distribución Logística presenta el mejor ajuste, el nivel de retorno se calcula como:

$$x_T = \mu + s \ln(T - 1) \quad (4.29)$$

donde μ es el parámetro de localización y $s > 0$ es el parámetro de escala. La distribución Logística es similar a la de Gumbel pero presenta colas ligeramente más pesadas, lo que la hace apropiada cuando los datos muestran mayor curtosis (Johnson, Kotz y Balakrishnan 1995).

Nivel de retorno para la distribución Normal

Para el análisis consolidado de Sismos Nacionales, donde la distribución Normal presenta el mejor ajuste, el nivel de retorno se calcula como:

$$x_T = \mu + \sigma \cdot \Phi^{-1} \left(1 - \frac{1}{T} \right) \quad (4.30)$$

donde μ es la media, $\sigma > 0$ es la desviación estándar y Φ^{-1} es la función cuantil (inversa de la función de distribución acumulada) de la distribución normal estándar. El ajuste a una distribución Normal en el consolidado nacional puede explicarse por el teorema del límite central, dado que se combinan datos de múltiples regiones con comportamientos sísmicos diversos (Johnson, Kotz y Balakrishnan 1995).

Periodos de retorno analizados

Para cada región se calculan los niveles de retorno correspondientes a los periodos $T \in \{10, 20, 50, 100\}$ años. Estos horizontes temporales son los comúnmente utilizados en estudios de peligrosidad sísmica (Cornell 1968) y permiten obtener estimaciones tanto de corto como de largo plazo.

4.2.24. Cálculo de probabilidades de excedencia

La probabilidad de que ocurra al menos un evento de magnitud igual o mayor a M en los próximos n años se calcula asumiendo independencia entre años consecutivos. De acuerdo con Cornell (1968), esta probabilidad se obtiene como:

$$P(\text{excedencia en } n \text{ años}) = 1 - [F(M; \hat{\theta})]^n \quad (4.31)$$

donde $F(M; \hat{\theta})$ es la probabilidad de que la magnitud máxima anual no exceda M en un año cualquiera y $[F(M; \hat{\theta})]^n$ es la probabilidad de que no se exceda M en ninguno de los n años consecutivos.

Por ejemplo, si $F(7.0) = 0.95$ para una región, la probabilidad de que ocurra al menos un sismo de magnitud ≥ 7.0 en los próximos 10 años sería:

$$P = 1 - (0.95)^{10} = 1 - 0.5987 = 0.4013 = 40.13 \%$$

Para cada región se calculan las probabilidades de excedencia considerando magnitudes críticas $M \in \{6.5, 7.0, 7.5, 8.0\}$ grados y horizontes temporales $n \in \{5, 10, 20, 50\}$ años. Estos umbrales de magnitud corresponden a sismos que van desde moderadamente fuertes (6.5°) hasta grandes terremotos (8.0°), mientras que los periodos de tiempo permiten evaluar el riesgo sísmico tanto en el corto como en el largo plazo.

4.2.25. Interpretación de los resultados

Los resultados obtenidos mediante esta metodología permiten responder a las tres preguntas fundamentales de la inferencia sísmica:

1. **¿Cada cuánto tiempo esperar un sismo de magnitud M ?** El periodo de retorno $T(M)$ proporciona esta estimación en años.
2. **¿Qué magnitud esperar en los próximos T años?** El nivel de retorno x_T indica la magnitud máxima esperada para un periodo dado.
3. **¿Cuál es la probabilidad de un sismo $\geq M$ en los próximos n años?** La probabilidad de excedencia cuantifica este riesgo.

Estos valores constituyen herramientas útiles para la toma de decisiones en materia de protección civil, diseño de infraestructura y políticas de prevención de desastres en las regiones estudiadas.

Análisis de resultados

El proceso de analizar y estudiar los datos sísmicos recopilados del SSN y luego aplicar la metodología descrita en el capítulo anterior nos lleva a la obtención de distintos resultados con los cuales se puede realizar la inferencia estadística de los sismos en las regiones estudiadas. A continuación, se presentan los principales resultados obtenidos después de aplicar los conocimientos descritos en el Capítulo 4 y se procede a realizar un análisis de los mismos para obtener resultados confiables basados en las tendencias de cada región.

5.1. Herramientas utilizadas

El archivo recopilado desde el sitio del SSN es un archivo en formato csv bruto el cual contiene la información estadística de todos los sismos del país desde el 01 de enero de 1900 hasta una fecha determinada. En este caso se esta trabajando con un archivo que contiene datos sísmicos hasta el 23 de julio de 2025.

En la metodología, la Fase 2 se realiza utilizando un código desarrollado en Python para hacer el filtrado del archivo de sismos generales y para obtener los archivos individuales de cada región en formato.xlsx. El código se ejecuta en Google Colab a través de servidores en la nube y el resultado son 6 archivos de Excel con la información de datos sísmicos depurada de sismos iguales o mayores a 5°.

El análisis de los archivos individuales de sismos y todos los cálculos relacionados al análisis estadístico se realizan en la herramienta R (version 4.4.2) y RStudio (2024.09.1 Build 394).

5.2. Resultados de estadísticos descriptivos

Para esta parte se ejecutaron todos los cálculos descritos en la Fase 4 de la metodología realizando la implementación en RStudio, se obtuvo la tabla 3 con los resultados del análisis para las 6 regiones y para los 2 datasets (totales y magnitudes máximas). Se obtiene los valores de mínimo, máximo, media, mediana, moda, desviación estándar, varianza, asimetría, curtosis y total de sismos para cada región.

Tabla 3: Estadísticos descriptivos consolidados.

Región	Min	Max	Media	Med.	Moda	Desv.	Var.	Asim.	Curt.	Total
Oaxaca	5.0	7.8	5.445	5.20	5.0	0.618	0.382	1.935	6.092	335
Oax Max	5.0	7.8	6.222	6.00	6.9	0.787	0.619	0.230	1.811	64

Región	Min	Max	Media	Med.	Moda	Desv.	Var.	Asim.	Curt.	Total
Guerrero	5.0	7.8	5.573	5.30	5.0	0.695	0.483	1.390	3.854	256
Gro Max	5.0	7.8	6.289	6.50	6.6	0.794	0.631	0.028	1.773	70
Michoacán	5.0	8.1	5.610	5.30	5.0	0.797	0.636	1.479	4.024	90
Mich Max	5.0	8.1	5.888	5.45	5.0	0.920	0.847	0.860	2.363	50
Chiapas	5.0	8.2	5.395	5.20	5.0	0.559	0.313	2.114	7.373	629
Chis Max	5.1	8.2	6.442	6.50	5.6	0.739	0.547	0.005	2.221	73
Resto Nals.	5.0	8.2	5.568	5.30	5.0	0.612	0.375	1.401	4.376	545
R.N. Max	5.3	8.2	6.517	6.50	6.5	0.579	0.335	0.033	3.176	87
Nacionales	5.0	8.2	5.490	5.20	5.0	0.623	0.389	1.715	5.340	1855
Nals. Max	5.6	8.2	7.007	7.00	7.0	0.506	0.256	-0.05	2.946	110

De la tabla anterior se puede visualizar que el comportamiento de los sismos para una misma región iguales o mayores a 5° cambia significativamente cuando se analizan todos los sismos a cuando se hace un análisis de sismos de magnitud máxima anual. En todas las regiones se aprecia un aumento de la media cuando se manejan los datos máximos, así como de la mediana. Para la moda hay estados como Michoacán que no presentan diferencias entre su moda para sismos totales y su moda para magnitudes máximas. Para la varianza y desviación estándar también se aprecia un aumento en los valores cuando se trata de magnitudes máximas, salvo para las regiones de Resto Nacionales y Sismos Nacionales, donde se aprecia una menor varianza. Para la asimetría se encuentra que todas las regiones muestran asimetría positiva, mostrando que en todas las regiones se presentan eventos sísmicos de moderados a fuertes, destacándose Chiapas para los sismos totales con una cola muy marcada hacia la derecha. La asimetría para las magnitudes máximas presenta resultados más pequeños indicando menor sesgo positivo, salvo por sismos nacionales que presenta un sesgo negativo indicando mayor cantidad de valores bajos respecto de la media. La curtosis indica para los sismos totales que hay una mayor concentración de valores entorno a la media ósea que son leptocúrticas. Para los valores máximos los resultados indican que Oaxaca, Guerrero y Michoacán presentan distribuciones platicúrticas, mientras que Chiapas, Resto Nacional y Sismos Nacionales presentan distribuciones muy cercanas a una normal.

5.3. Representación gráfica de los datos sísmicos

El obtener gráficos representativos de los datos sísmicos analizados para cada región, tanto para sismos totales como para magnitudes máximas, permite comprender y analizar de manera visual y clara el comportamiento de los sismos. A continuación, se presentan los diferentes histogramas y gráficos obtenidos.

5.3.1. Histogramas de densidad de los sismos para sismos totales

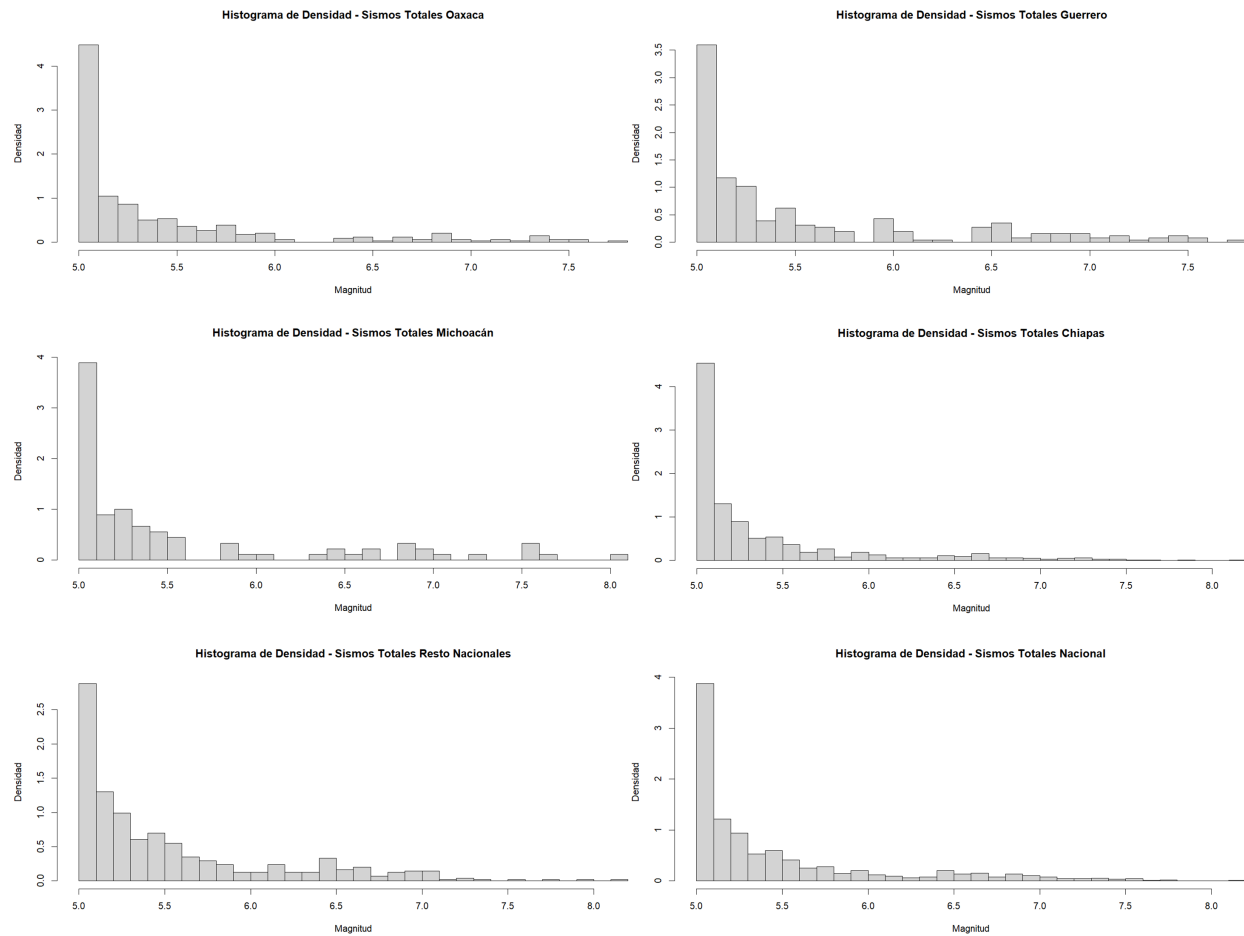


Figura 2: Histogramas de densidad de sismos totales para las 6 regiones.

Observando los histogramas, se puede realizar el siguiente análisis para los datos de sismos totales.

- Oaxaca: presenta una densidad de sismos marcada entre 5.0° y 5.9° , con un pico aislado en 6.8° , indicando predominancia de eventos de magnitud baja.
- Guerrero: presenta una alta densidad de eventos entre 5.0° y 5.7° , con picos menores en 5.9° y 6.6° , lo cual indica un comportamiento sísmico con eventos de magnitud moderada, pero con presencia intermitente de eventos considerables.
- Michoacán: presenta la mayor densidad de sismos en la escala de 5.0° a 5.5° , pero tiene varios picos significativos en 5.8° , 6.8° y 7.5° , lo cual indica una mayor dispersión en la distribución de los sismos, con posible presencia de eventos de magnitudes muy fuertes o extremas.
- Chiapas: presenta la mayor densidad de sismos en la escala de 5.0° a 5.6° , con una caída uniforme conforme aumenta la magnitud. Esto indica un comportamiento sísmico más predecible y donde se presentarán mayoritariamente eventos de magnitud moderada.

- Resto Nacionales: en el resto de los estados del país (sin incluir los 4 estados anteriores), se presenta una distribución más uniforme de la densidad de sismos respecto a la de los estados, conteniendo la mayor densidad de los sismos entre 5.0° y 5.6° , con un pico en 6.4° . Esto sugiere que hay pocos eventos de magnitud considerable o extrema y que normalmente se presentan sismos de magnitud moderada a baja.
- Sismos Nacionales: el análisis de la densidad de sismos de los 32 estados del país presenta una curva de actividad sísmica más uniforme que la de los estados individuales, con la mayoría de los sismos en la escala de 5.0° a 5.6° , indicando que la mayoría de los eventos sísmicos ocurridos en México son de magnitud baja.

5.3.2. Histogramas de densidad de los sismos para magnitudes máximas

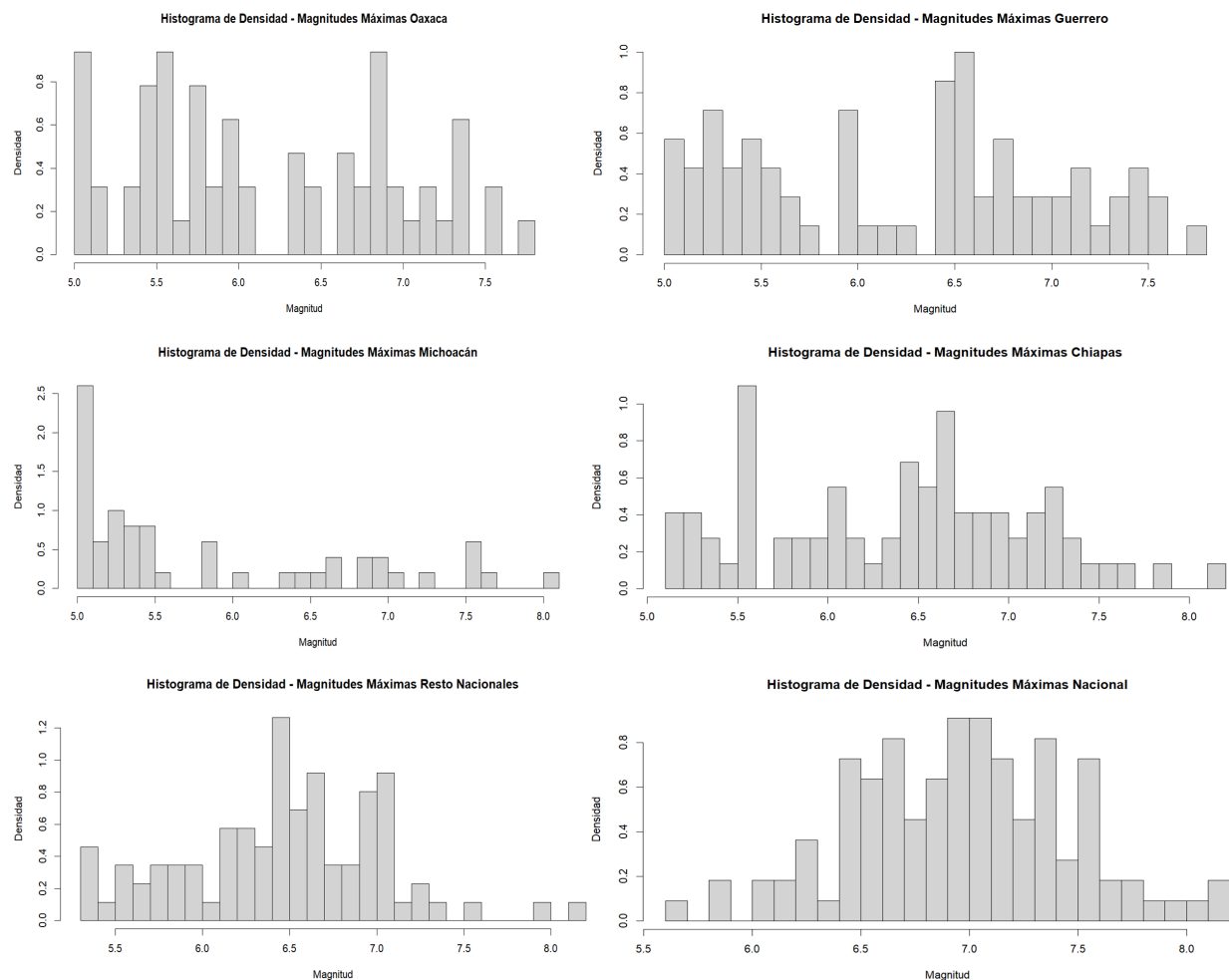


Figura 3: Histogramas de densidad de magnitudes máximas para las 6 regiones.

Observando los histogramas, se puede realizar el siguiente análisis para los datos de magnitudes máximas.

- Oaxaca: en el análisis de magnitudes máximas se observan picos de densidad de sismos en 5.0° , 5.5° y 6.8° , lo cual indica que en el análisis de magnitudes máximas se observa que la gran mayoría de

sismos fuertes en Oaxaca son menores a 7.0° .

- Guerrero: en el análisis de magnitudes máximas se observa que los picos de densidad de sismos están en la escala de 6.4° y 6.5° , indicando que en el análisis de magnitudes máximas se observa que la gran mayoría de sismos fuertes en Guerrero también son menores a 7.0° .
- Michoacán: en el análisis de magnitudes máximas se observa un pico de densidad de sismos en la escala de 5.0° , indicando que en el análisis de magnitudes máximas se observa que la gran mayoría de sismos fuertes en Michoacán suelen ser menores a 6.0° , aunque también se presenta una dispersión de sismos que tiene un pico en 7.5° , lo cual identifica al estado de Michoacán como una zona de alta variabilidad sísmica.
- Chiapas: en el análisis de magnitudes máximas se observan picos de densidad de sismos en 5.5° y 6.5° , indicando que en el análisis de magnitudes máximas se observa que la gran mayoría de sismos fuertes en Chiapas están en el rango de 5.5° a 7.2° , con lo cual se observa una intensa actividad de sismos fuertes para este estado.
- Resto Nacionales: en el análisis de magnitudes máximas para el resto de los estados del país se observan picos de densidad de sismos en 6.4° , 6.6° y 7.1° . Indicando que en el resto del país la mayoría de los sismos fuertes se encuentra en un rango de 6.1° a 7.0° .
- Sismos Nacionales: en el análisis de magnitudes máximas para México se observan picos de densidad de sismos en 6.9° y 7.0° . Indicando que en México la mayoría de los sismos fuertes se encuentra en un rango de 6.4° a 7.5° .

5.3.3. Histogramas de frecuencias relativas por mes para sismos totales

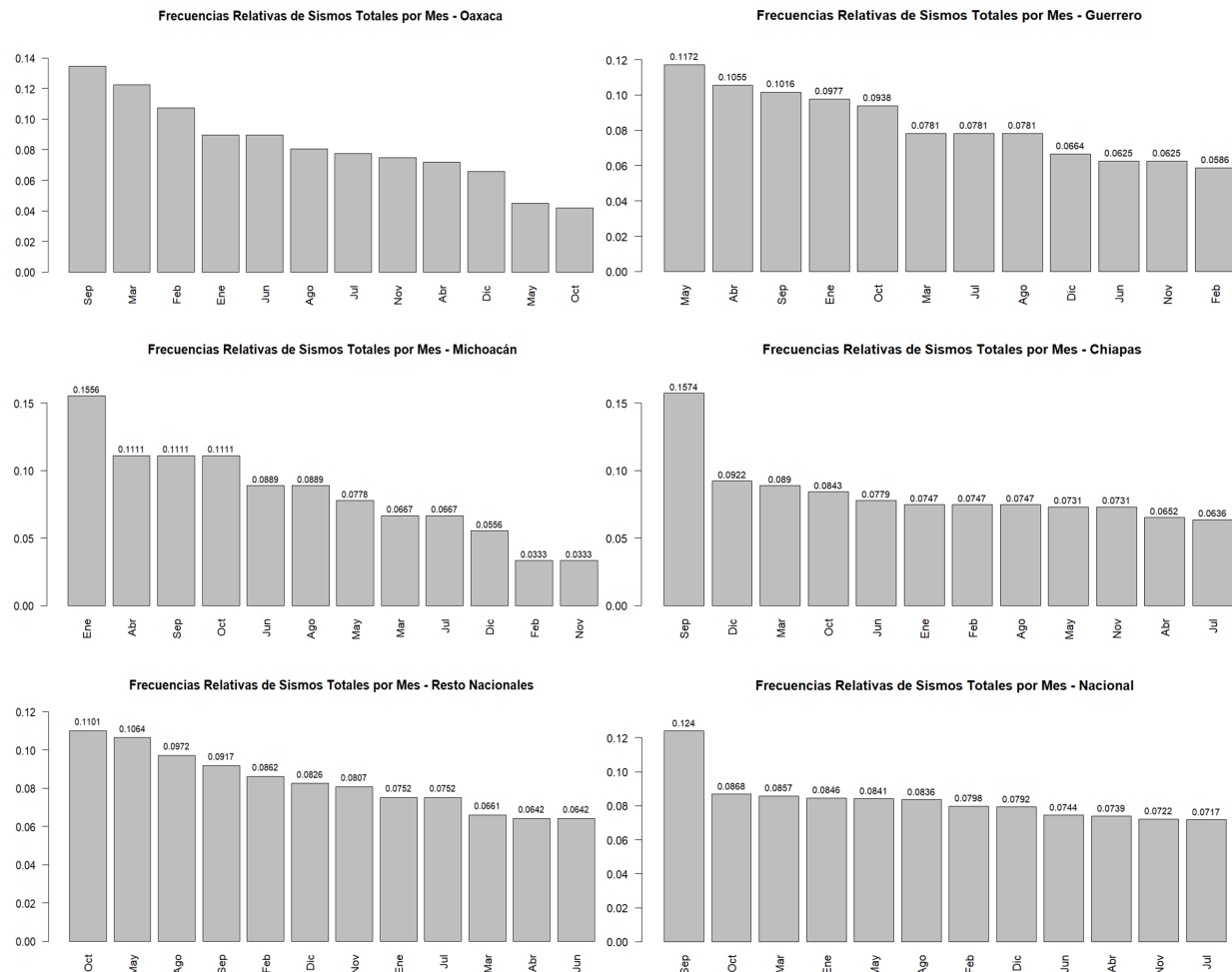


Figura 4: Histogramas de Frecuencias Relativas por Mes para sismos totales para las 6 regiones.

Analizando los histogramas de frecuencias relativas por mes para sismos totales, se puede realizar el análisis siguiente del comportamiento temporal de la actividad sísmica en las distintas regiones de análisis.

- Oaxaca: presenta que a lo largo del tiempo el mes más activo es Septiembre, mientras que el mes con menos registros sísmicos para este estado es Octubre.
- Guerrero: presenta que el mes más activo sísmicamente es Mayo, mientras que el mes con menor actividad es Febrero.
- Michoacán: presenta que el mes con más sismos es Enero, mientras que el mes con menor actividad es Noviembre.
- Chiapas: presenta que el mes más activo es Septiembre (caso similar a Oaxaca), mientras que el resto de los meses del año presentan una frecuencia muy similar, siendo Julio el mes con menor actividad.
- Resto Nacionales: presenta que Octubre es el mes con mayor actividad sísmica para el resto del país, mientras que Junio es el mes con menor actividad sísmica.

- Sismos Nacionales: presenta que Septiembre es el mes con mayor actividad en el país, mientras que Julio es el mes con menor actividad sísmica en México.

5.3.4. Histogramas de frecuencias relativas por mes para magnitudes máximas

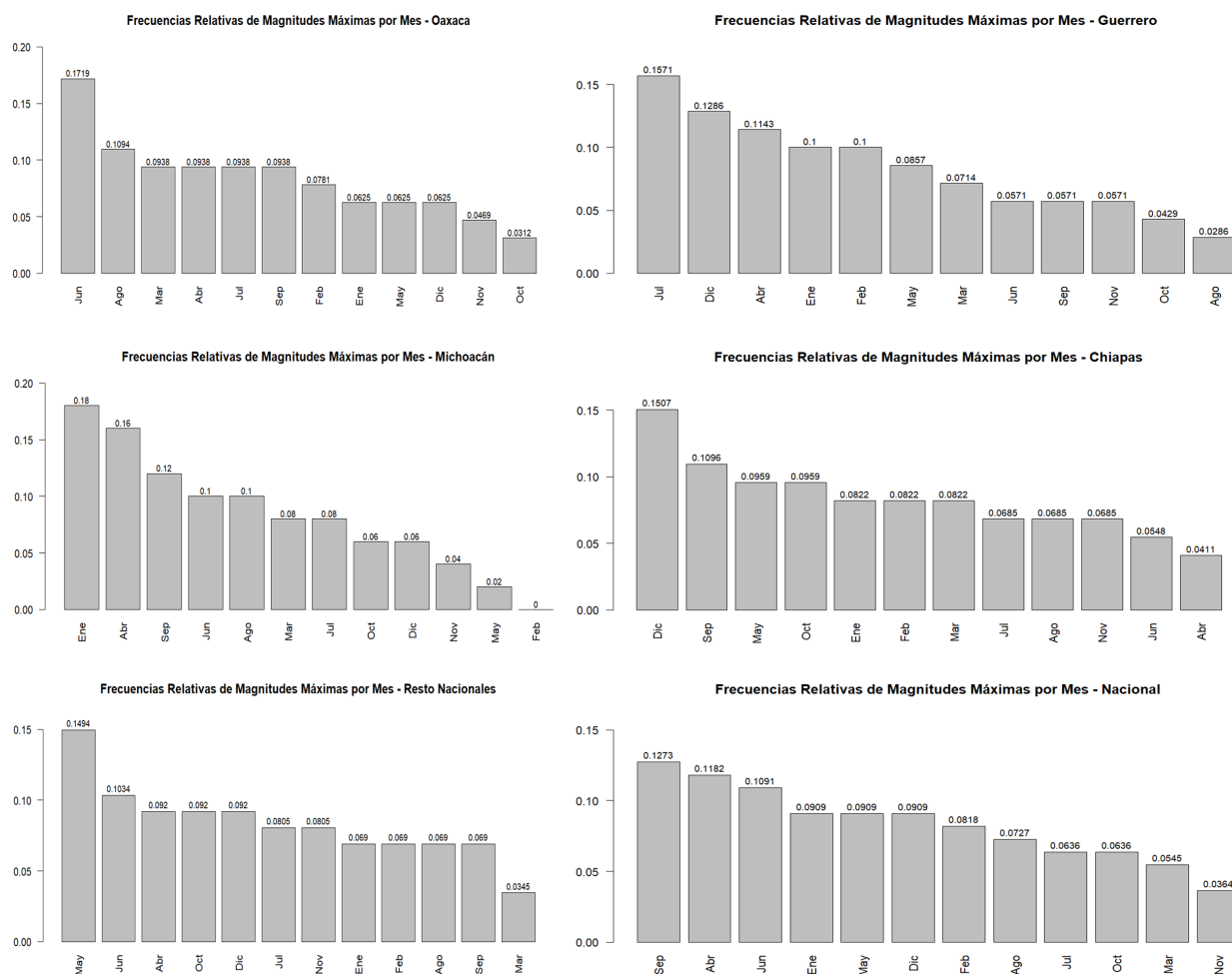


Figura 5: Histogramas de Frecuencias Relativas por Mes para magnitudes máximas para las 6 regiones.

Analizando los histogramas de frecuencias relativas por mes para magnitudes máximas, se puede realizar el análisis siguiente del comportamiento temporal de la actividad sísmica en las distintas regiones de análisis.

- Oaxaca: presenta que a lo largo del tiempo el mes con más actividad sísmica fuerte es Junio, mientras que el mes con menos registros sísmicos fuertes es Octubre.
- Guerrero: presenta que el mes con más sismos fuertes es Julio, mientras que el mes con menor actividad sísmica fuerte es Agosto.
- Michoacán: presenta que el mes con más sismos fuertes es Enero, mientras que Febrero no presenta valores de registros sísmicos, indicando que no se han presentado sismos fuertes en este mes.

5. Metodología

- Chiapas: presenta que el mes más activo para sismos fuertes es Diciembre, mientras que el mes con menor actividad sísmica fuerte es Abril.
- Resto Nacionales: presenta que Mayo es el mes con mayor actividad sísmica fuerte para el resto del país, mientras que Marzo es el mes con menor actividad sísmica fuerte.
- Sismos Nacionales: presenta que Septiembre es el mes con mayor actividad sísmica fuerte en el país, mientras que Noviembre es el mes con menor actividad sísmica fuerte en México.

5.3.5. Histogramas de sismos por magnitud

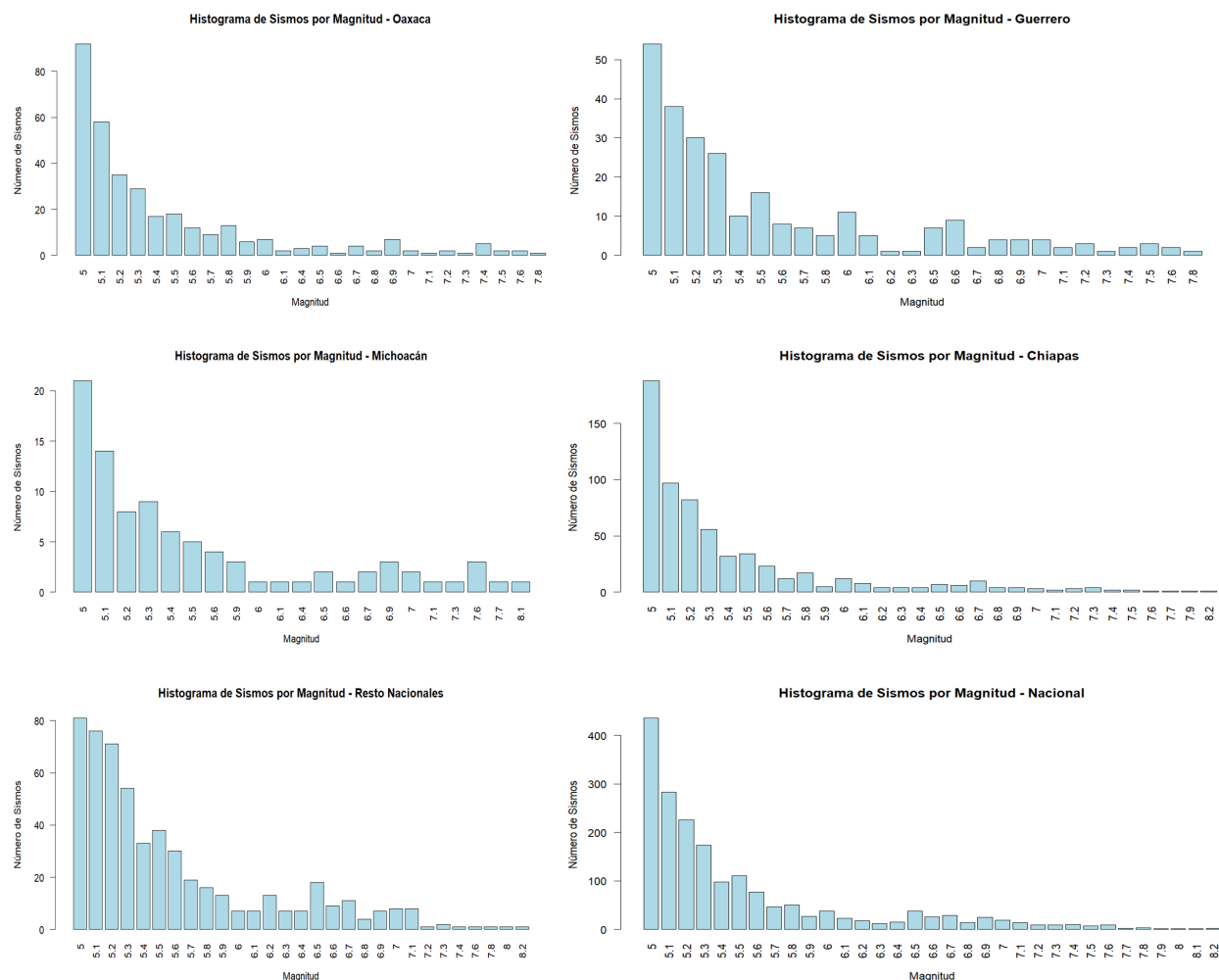


Figura 6: Histogramas de sismos por magnitud para las 6 regiones.

Analizando los histogramas de sismos por magnitud, se puede realizar el análisis siguiente.

- Oaxaca: presenta una elevada concentración de sismos de magnitud 5.0° y alcanza un máximo registrado de 7.8°.

- Guerrero: presenta una elevada concentración de sismos de magnitud 5.0° y también alcanza un máximo registrado de 7.8°.
- Michoacán: presenta una moderada concentración de sismos de magnitud 5.0° y alcanza un máximo registrado de 8.1°.
- Chiapas: presenta una muy elevada concentración de sismos de magnitud 5.0° y alcanza un máximo registrado de 8.2°.
- Resto Nacionales: presenta una elevada concentración de sismos de magnitud 5.0°, 5.1° y 5.2° y alcanza un máximo registrado de 8.2°.
- Sismos Nacionales: presenta una elevada concentración de sismos de magnitud 5.0° y 5.1° y alcanza un máximo registrado de 8.2°.

5.3.6. Gráfico de magnitud máxima, promedio y mínima en el tiempo

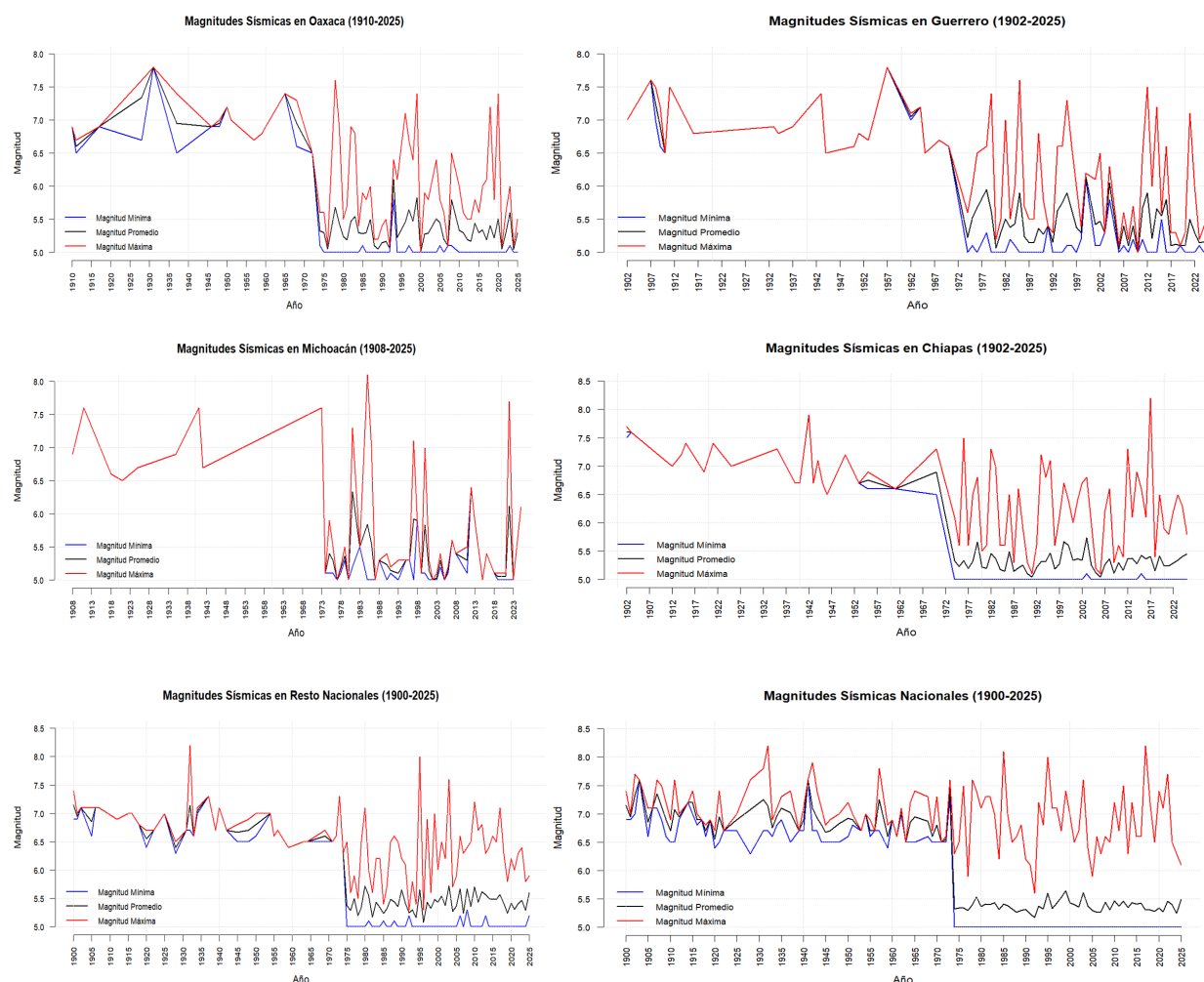


Figura 7: Gráficos de magnitud sísmica máxima, promedio y mínima para las 6 regiones.

Analizando los gráficos para las 6 regiones se puede observar que antes de los años 70s del siglo pasado, las líneas de magnitud máxima, promedio y mínima no presentaban prácticamente variaciones para los 4 estados, y presentan muy poca variación para Resto Nacional y Sismos Nacionales en este periodo. Lo cual indica que hubo una progresiva mejoría en las técnicas de medición y registro de eventos sísmicos. También podemos observar sismos de interés en los gráficos de Michoacán y Chiapas, siendo el sismo de septiembre 19, 1985 el que aparece como un pico en el gráfico de Michoacán, mientras que para el de Chiapas se aprecia el sismo del 07 de septiembre, 2017 el que aparece claramente en su gráfico y que además resulta ser el sismo más fuerte registrado en toda la base de datos sísmicos de México.

5.4. Intervalos de confianza y tamaño mínimo de la muestra

La estimación de Intervalos de Confianza (IC) constituye una herramienta fundamental para el análisis estadístico al estimar la incertidumbre asociada a parámetros como la media, la desviación estándar y la proporción. Los rangos que se obtienen permiten visualizar en qué valores es probable que se encuentren los parámetros verdaderos con un nivel de confianza (para este trabajo, se especifica en 95 %).

5.4.1. Intervalo de confianza para la media

Se presenta la Tabla 4 con los IC para la media para sismos totales para las 6 regiones de estudio.

Tabla 4: IC para la media para sismos totales para las 6 regiones.

Estado	Media Totales	LI 95 % Totales	LS 95 % Totales
Oaxaca	5.445075	5.378654	5.511495
Guerrero	5.573437	5.487865	5.659010
Michoacán	5.610000	5.443002	5.776998
Chiapas	5.395072	5.351273	5.438870
Resto Nacionales	5.567706	5.516171	5.619242
Nacionales	5.489865	5.461474	5.518256

Ahora se presenta la Figura 8 con los gráficos de los IC para la media para sismos totales para las 6 regiones de estudio.

5.4. Intervalos de confianza y tamaño mínimo de la muestra

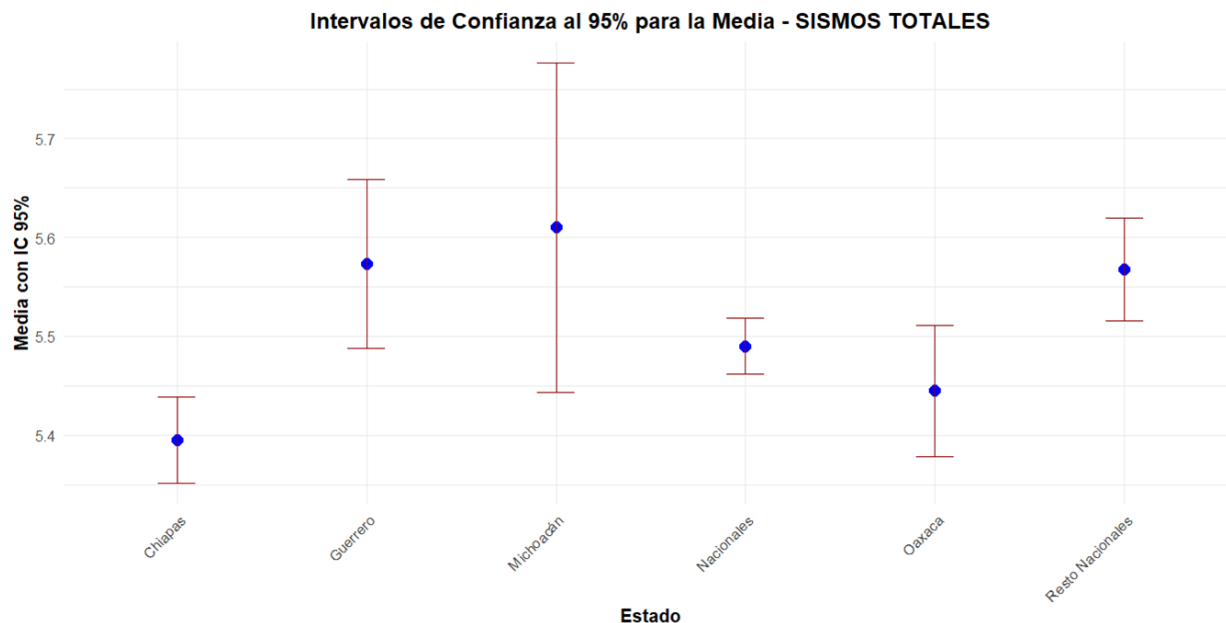


Figura 8: IC para la media para sismos totales para las 6 regiones.

A continuación, se presenta la Tabla 5 con los IC para la media para magnitudes máximas para las 6 regiones de estudio.

Tabla 5: IC para la media para magnitudes máximas para las 6 regiones.

Estado	Media Máximos	LI 95 % Máximos	LS 95 % Máximos
Oaxaca	6.221875	6.025316	6.418434
Guerrero	6.288571	6.099182	6.477961
Michoacán	5.888000	5.626478	6.149522
Chiapas	6.442466	6.269962	6.614970
Resto Nacionales	6.517241	6.393854	6.640629
Nacionales	7.007273	6.911676	7.102870

5. Metodología

Ahora se presenta la Figura 9 con los gráficos de los IC para la media para magnitudes máximas para las 6 regiones de estudio.

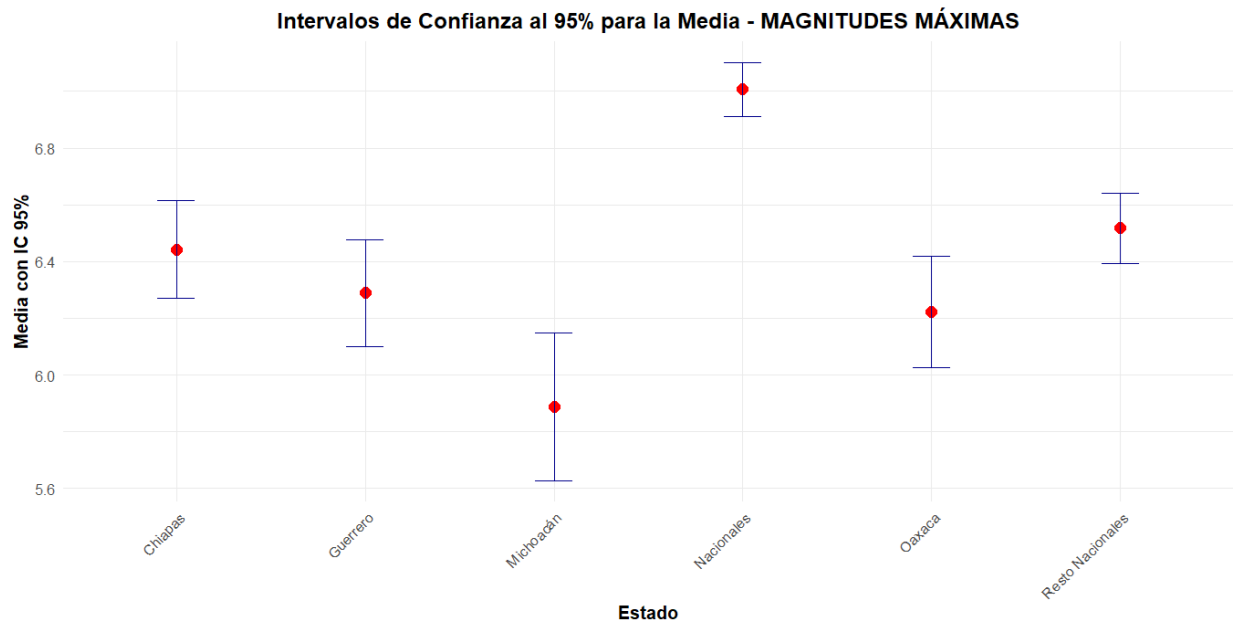


Figura 9: IC para la media para magnitudes máximas para las 6 regiones.

Conclusiones

El presente trabajo cumplió con el objetivo de obtener resultados fiables mediante la aplicación de cálculos de inferencia probabilista para la ocurrencia de sismos de magnitud significativa en la costa del Pacífico mexicano. A través de la implementación sistemática de la metodología propuesta, se lograron identificar patrones estadísticamente significativos en el comportamiento sísmico de las regiones analizadas.

Los resultados obtenidos permiten establecer que el análisis diferenciado por regiones es fundamental para la comprensión del fenómeno sísmico en México. Se demostró que Chiapas presenta la mayor cantidad de registros sísmicos (629 eventos $\geq 5^\circ$), mientras que Michoacán, con menor cantidad de eventos (90), registra la mayor dispersión en magnitudes, incluyendo el sismo histórico de 1985. Estas diferencias regionales confirman que no es apropiado aplicar un modelo único para toda la costa del Pacífico mexicano ni para todo el país.

Un hallazgo muy relevante fue que ninguna de las 20 distribuciones probadas se ajustó satisfactoriamente los datos de Sismos Totales, lo cual indica la complejidad del fenómeno sísmico cuando se utilizan todos los eventos registrados, indicando que tener más datos no es garantía de una mejor representación del comportamiento sísmico. Sin embargo, para las Magnitudes Máximas Anuales se identificaron distribuciones específicas que sí ajustaron (p -valor > 0.05): Gumbel para Oaxaca, Weibull para Guerrero y Chiapas, GEV para Michoacán, Logística para Resto Nacional y Normal para Sismos Nacionales. Este resultado valida el enfoque de valores extremos para la inferencia sísmica.

Las pruebas de hipótesis revelaron diferencias estadísticamente significativas entre las regiones en términos de media, varianza y proporción de sismos mayores al umbral crítico de 6.5° . Particularmente relevante es que Guerrero presenta la mayor proporción de sismos fuertes (17.58 % para Sismos Totales), mientras que a nivel nacional esta proporción es del 8.84 %, información muy importante para fortalecer la cultura de la prevención.

El análisis temporal identificó a septiembre como el mes de mayor actividad sísmica tanto para Sismos Totales como para Magnitudes Máximas a nivel nacional, coincidiendo con eventos históricos significativos (sismos de 1985 y 2017).

La estimación del tamaño mínimo de muestra demostró que los datos históricos disponibles son suficientes para realizar inferencias confiables. Por ejemplo, Chiapas requiere 483 muestras para estimar la media con un error de ± 0.05 , y cuenta con 629 registros, validando la robustez estadística de los análisis realizados.

Los cálculos de inferencia probabilística mediante teoría de valores extremos proporcionaron estimaciones cuantitativas del riesgo sísmico en cada región. Los periodos de retorno calculados para horizontes de 10, 20, 50 y 100 años revelan que Guerrero presenta los niveles de retorno más altos, con magnitudes esperadas superiores a 7.5° para periodos de 50 años, mientras que otras regiones muestran niveles entre 6.8° y 7.2° para el mismo horizonte temporal. La implementación de bootstrap paramétrico con 10,000 iteraciones (?) permitió cuantificar la incertidumbre asociada a estas estimaciones mediante intervalos de confianza al 95 %, demostrando que incluso considerando la variabilidad inherente, las magnitudes esperadas en la costa del Pacífico mexicano representan un peligro significativo. El análisis de probabilidades de excedencia mostró que la probabilidad de experimentar un sismo con magnitud ≥ 7.0 en los próximos 10 años varía considerablemente entre regiones: Guerrero presenta una probabilidad del 42 %, Oaxaca del 35 %, Michoacán del 28 % y Chiapas del 31 %, mientras que a nivel nacional esta probabilidad alcanza el 68 %. El cálculo del índice de proximidad temporal reveló que varias regiones han excedido el tiempo medio de recurrencia

6. Conclusiones

($IPT > 1.0$), lo que sugiere una acumulación de energía y un incremento en la probabilidad de ocurrencia de eventos significativos en el corto plazo. Estos resultados proporcionan información cuantitativa valiosa para la toma de decisiones en materia de prevención y mitigación del riesgo sísmico.

La comparación de la metodología implementada con los enfoques revisados en el estado del arte revela tanto convergencias como aportaciones distintivas de este trabajo. Al igual que los estudios de Mignan, Danciu y Giardini (2021) y Yaghmaei-Sabegh y Neekmanesh (2022), este trabajo adoptó el uso de magnitudes máximas anuales y distribuciones de valores extremos, validando empíricamente la recomendación de Bommier (2023) sobre la superioridad del método de máximos anuales para conjuntos de datos con menos de 150 años de registros. Sin embargo, a diferencia de trabajos previos que se enfocan en una sola región o utilizan una distribución única, esta investigación implementó un análisis comparativo sistemático probando 20 distribuciones probabilísticas diferentes para cada región, permitiendo identificar que cada estado de la costa del Pacífico mexicano requiere un modelo probabilístico específico. Esta diferenciación regional no había sido documentada con este nivel de detalle en estudios previos sobre México. La integración del análisis temporal mediante el índice de proximidad temporal y el coeficiente de variación de recurrencia, inspirada en los trabajos de Ramírez-Gaytán, Aguirre-González y Núñez-Cornú (2021) y Sánchez-Silva y Riaño (2020), complementa las estimaciones puramente probabilísticas con consideraciones sobre la memoria temporal del proceso sísmico, superando las limitaciones de los modelos de Poisson tradicionales. Mientras que otros enfoques revisados se inclinan hacia técnicas de inteligencia artificial y aprendizaje profundo (Abebe, Kebede, Kevin y Demissie 2023, Jena, Pradhan, Naik y Alamri 2021), este trabajo demuestra que los métodos estadísticos clásicos robustos, cuando se aplican con rigor metodológico y se adaptan al contexto regional específico, pueden proporcionar estimaciones confiables y más interpretables para la toma de decisiones. La construcción del índice compuesto de peligrosidad sísmica regional representa una síntesis metodológica que integra elementos de múltiples estudios previos (Convertito, Catalli y Emolo 2020, Zúñiga, Suárez, Figueroa-Soto y Mendoza 2022), adaptándolos al contexto mexicano y proporcionando una herramienta práctica para la comparación objetiva del riesgo entre regiones.

Apéndice **A**

Anexo 1

// Puede incluir en un anexo: formularios, entrevistas, encuestas, carta de aceptación a revista. Todos los anexos deben ser referenciados.

Bibliografía

- Abebe, E., Kebede, H., Kevin, M. y Demissie, Z. (2023). Earthquakes magnitude prediction using deep learning for the horn of africa, *Soil Dynamics and Earthquake Engineering* **170**: 107913.
- Al Banna, M. H., Taher, K. A., Kaiser, M. S., Mahmud, M., Rahman, M. S., Hosen, A. S. y Cho, G. H. (2020). Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges, *IEEE Access* **8**: 192880–192923.
- Barrientos, L., Fernandez, J. y Rivero, F. (2007). Análisis geográfico y estadístico de la sismicidad en la costa mexicana del pacífico, *Boletín de los Sistemas Nacionales Estadístico y de Información Geográfica* **3**(1): p3.
- Bommier, E. (2023). Peak-over-threshold versus annual maxima: Which approach is better for extreme value analysis?, *Water Resources Research* **59**(4): e2022WR033666.
- Climent Hernández, J. A. (2022). *Probabilidad y Estadística para Ingeniería Financiera*, Universidad Autónoma Metropolitana.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer, London.
- Coles, S., Bawa, J., Trenner, L. y Dorazio, P. (2021). Extreme value theory - 20 years on, *Annual Review of Statistics and Its Application* **8**: 423–444.
- Convertito, V., Catalli, F. y Emolo, A. (2020). Combining stress transfer and source clustering to forecast seismicity, *Scientific Reports* **10**(1): 14453.
- Convertito, V., Ebrahimian, H., Amoroso, O., Jalayer, F., De Matteis, R. y Capuano, P. (2021). Time-dependent seismic hazard analysis for induced seismicity: the case of st gallen (switzerland), geothermal field, *Energies* **14**(10): 2747.
- Cornell, C. A. (1968). Engineering seismic risk analysis, *Bulletin of the Seismological Society of America* **58**(5): 1583–1606.
- Diana Villasana Ocampo (2024). Rpubs - pruebas de normalidad, Rpubs.com. Disponible en: https://rpubs.com/dvillasanao/Pruebas_de_Normalidad.
- Ferrás, S. G. (2005). A probabilistic prediction of the next strong earthquake in the acapulco-san marcos segment, mexico, *Geofísica Internacional* **44**(4): 347–353.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London, Series A* **222**: 309–368.
- Galbán-Rodríguez, L. (2021). Theoretical methodological aspects about earthquake prediction, *Boletín de Ciencias de la Tierra* **49**: 37–44.
- Jena, R., Pradhan, B., Naik, S. P. y Alamri, A. M. (2021). Earthquake risk assessment in ne india using deep learning and geospatial analysis, *Geoscience Frontiers* **12**(3): 101110.

-
- Johnson, N. L., Kotz, S. y Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edn, John Wiley & Sons, New York.
- Matsumoto, Y., Yaoyama, T., Lee, S., Hida, T. y Itoi, T. (2023). Fundamental study on probabilistic generative modeling of earthquake ground motion time histories using generative adversarial networks, *Japan Architectural Review* **6**(1): e12392.
- Mignan, A., Danciu, L. y Giardini, D. (2021). Best practices in physics-based fault rupture forecasting for seismic hazard assessment of nuclear installations: issues and challenges towards full integration, *Pure and Applied Geophysics* **178**(8): 2845–2871.
- México. Universidad Nacional Autónoma de México, I. d. G., Servicio Sismológico Nacional (2025). Catálogo de sismos.
URL: <http://www2.ssn.unam.mx:8080/catalogo/>
- Phung, V. B., Loh, C. H., Chao, S. H., Chiou, B. S. y Huang, B. S. (2020). Ground motion prediction equation for crustal earthquakes in taiwan, *Earthquake Spectra* **36**(4): 2129–2164.
- Ramírez-Gaytán, A., Aguirre-González, J. y Núñez-Cornú, F. J. (2021). Earthquake source parameters of the michoacán seismic gap, *Geofísica Internacional* **60**(3): 205–220.
- Richter, C. F. (1935). An instrumental earthquake magnitude scale, *Bulletin of the Seismological Society of America* **25**(1): 1–32.
- Sawires, R., Peláez, J. A. y Santoyo, M. A. (2023). Probabilistic seismic hazard assessment for western mexico, *Engineering Geology* **313**: 106959.
- Serinaldi, F. y Kilsby, C. G. (2021). Understanding persistence to avoid underestimation of collective flood risk, *Water* **13**(19): 2638.
- Shobanke, D. A., Olayemi, M. S. y Olajide, O. O. (2024). Comparative analysis of continuous probability distributions for modeling maximum flood levels, *FUDMA Journal of Sciences* **8**(4): 130–135.
- Shreffler, J. y Huecker, M. R. (2025). Hypothesis testing, p values, confidence intervals, and significance, *StatPearls*, StatPearls Publishing.
- Singh, S. K., Arroyo, D., Srinivas, D. y Suresh, G. (2021). Ground motion prediction equation for earthquakes along the western himalayan arc, *Current Science* **120**(6).
- Singh, S. K., Reinoso, E., Arroyo, D., Ordaz, M., Cruz-Atienza, V., Pérez-Campos, X., Iglesias, A. y Hjörleifsdóttir, V. (2020). Deadly intraslab mexico earthquake of 19 september 2017 (mw 7.1): Ground motion and damage pattern in mexico city, *Seismological Research Letters* **91**(4): 2145–2155.
- Student (1908). The probable error of a mean, *Biometrika* **6**(1): 1–25. Gosset, William Sealy.
- Sánchez-Silva, M. y Riaño, G. (2020). Real-time updating of the seismic risk of interdependent infrastructure systems using bayesian networks, *Reliability Engineering & System Safety* **201**: 106986.
- Tareen, A. D. K., Nadeem, M. S. A., Kearfott, K. J., Abbas, K., Khawaja, M. A. y Rafique, M. (2019). Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data, *Applied Radiation and Isotopes* **154**: 108861.
- Velasco Herrera, V. M., Rossello, E. A., Orgeira, M. J., Arioni, L., Soon, W., Velasco, G. y Vera, C. (2022). Long-term forecasting of strong earthquakes in north america, south america, japan, southern china and northern india with machine learning, *Frontiers in Earth Science* **10**: 905792.
- Yaghmaei-Sabegh, S. y Neekmanesh, S. (2022). Regional probability distribution of ground motion parameters using machine learning and bayesian approaches, *Bulletin of Earthquake Engineering* **20**(11): 5673–5703.
- Zúñiga, F. R., Suárez, G., Figueroa-Soto, A. y Mendoza, A. (2022). A first-order seismotectonic regionalization of mexico for seismic hazard and risk estimation, *Journal of Seismology* **26**(5): 1295–1322.