

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	XII
Índice de tablas	XIII
Índice de figuras	XV
Simbología y Acrónimos	XVII
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	16

4. Metodología	19
4.1. Metodología de investigación	19
4.2. Fases de desarrollo del proyecto	20
4.2.1. Media aritmética	21
4.2.2. Mediana	21
4.2.3. Moda	22
4.2.4. Varianza muestral	22
4.2.5. Desviación estándar	22
4.2.6. Coeficiente de asimetría	22
4.2.7. Coeficiente de curtosis	23
4.2.8. Intervalo de confianza para la media	23
4.2.9. Intervalo de confianza para la varianza	24
4.2.10. Intervalo de confianza para la proporción	24
4.2.11. Prueba de hipótesis para el cociente de varianzas (prueba F)	24
4.2.12. Prueba de hipótesis para diferencia de medias con varianzas iguales (pooled)	25
4.2.13. Prueba de hipótesis para diferencia de medias con varianzas diferentes (Welch)	25
4.2.14. Prueba de hipótesis para la diferencia de proporciones con proporciones agrupadas o iguales (pooled)	26
4.2.15. Prueba de hipótesis para la diferencia de proporciones sin agrupamiento o diferentes	26
4.2.16. Prueba de bondad de ajuste Kolmogorov-Smirnov	26
4.2.17. Prueba de bondad de ajuste Anderson-Darling	27
4.2.18. Prueba de bondad de ajuste Lilliefors	27
4.2.19. Criterio de Información Bayesiano	27
4.2.20. Teoría de valores extremos y distribuciones empleadas	28
4.2.21. Periodos de retorno	29
4.2.22. Probabilidades de excedencia	30
4.2.23. Intervalos de confianza mediante bootstrap paramétrico	30
4.2.24. Análisis temporal y tiempo desde el último evento	31
4.2.25. Índice de peligrosidad sísmica regional	31
5. Análisis de resultados	33
5.1. Herramientas utilizadas	33
5.2. Resultados de estadísticos descriptivos	33
6. Conclusiones	35
A. Anexo 1	37
Bibliografía	41

Índice de tablas

1.	Artículos seleccionados.	5
2.	Estadísticos descriptivos consolidados.	34

Índice de figuras

1.	Diagrama de flujo de la aplicación de la metodología propuesta.	20
----	---	----

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)
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)

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	“Random number generators for extreme values: A comparison study” (Papalexiou y Koutsoyiannis 2020)
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)
2021	“Efficient estimation of the number of false positives in high-throughput screening” (Rootzén y Zholud 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. Presenta un modelo de predicción sísmica basado en el aprendizaje profundo (deep learning) el cual utiliza

un algoritmo transformer, entrenado con los registros históricos de magnitudes de terremotos ocurridos en el Cuerno de África, y que identifique patrones en la actividad sísmica que ayuden a mejorar la capacidad de predicción del modelo propuesto.

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. Se destaca que las técnicas de IA de redes neuronales y algoritmos de aprendizaje profundo ofrecen resultados prometedores respecto a la identificación de patrones ocultos en los datos sísmicos. También se analizan las dificultades que hay en este campo, como el de la calidad de los datos, la escasez de disponibilidad de estos o la dificultad para trasladar e interpretar de manera correcta estos modelos en la práctica.

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 siglo XX. Se describen las ocurrencias de eventos sísmicos como un proceso de puntos y a partir de ello, se propone el uso de modelos lineales de tipo log-lineal para ajustar las tasas de ocurrencia sísmica. Este análisis permite describir de manera rigurosa la distribución espacial y temporal de los terremotos y presentar una base sólida para comprender los patrones de actividad sísmica en la costa del Pacífico de México.

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 de máximos anuales. El estudio evalúa el desempeño de ambos métodos para diferentes tamaños de muestra y concluye que, para conjuntos de datos con menos de 150 años de registros, el método de máximos anuales utilizando las distribuciones GEV (generalized extreme value) o Gumbel resulta más robusto y confiable. Se emplean criterios de selección como AIC (Akaike Information Criterion) y BIC (Bayesian Information Criterion), además de pruebas de bondad de ajuste como Kolmogorov-Smirnov, Anderson-Darling y Lilliefors para validar la adecuación de las distribuciones a los datos.

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), incluyendo las tres familias clásicas de distribuciones de valores extremos (Gumbel, Fréchet y Weibull) y su unificación en la distribución GEV. El autor desarrolla los métodos de máxima verosimilitud para la estimación de parámetros, técnicas de diagnóstico para validar el ajuste de modelos y enfoques para la construcción de intervalos de confianza. Este trabajo se ha convertido en la referencia estándar para aplicaciones de EVT en múltiples campos, incluyendo la sismología.

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 actualizan las recomendaciones sobre la selección de distribuciones, determinación del tamaño de muestra mínimo necesario para estimaciones confiables, aplicación de técnicas de bootstrap para cuantificación de incertidumbre y extensión de EVT a problemas multivariados. Se discuten aplicaciones recientes a cambio climático, desastres naturales y eventos sísmicos, proporcionando un marco metodológico actualizado que incorpora avances computacionales y estadísticos de las últimas dos décadas.

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 la capacidad de pronóstico sísmico. El modelo calcula un índice compuesto de peligrosidad que integra múltiples factores: la magnitud esperada según modelos de valores extremos, la probabilidad de excedencia en diferentes horizontes temporales, el tiempo transcurrido desde el último evento significativo y la tasa histórica de actividad sísmica en cada región. Los autores aplican esta metodología a seis regiones de Italia y validan retrospectivamente sus predicciones con sismos ocurridos entre 2016 y 2019, demostrando una mejora significativa en la identificación de zonas de alto riesgo.

Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano (2021) presentan una técnica

2. Introducción

innovadora para evaluar el riesgo de sismos inducidos por proyectos de actividad geotérmica. La técnica modifica el enfoque tradicional de análisis probabilista de peligrosidad (PSHA), que asume que los sismos ocurren de manera aleatoria y siguiendo un modelo Poisson, y en su lugar, lo replantea utilizando modelos alternativos como BPT, Weibull, gamma y ETAS, ajustando sus parámetros a los registros sísmicos recopilados en cada etapa de las operaciones de campo, lo que permite una estimación más precisa 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 recurrencia de tiempo de grandes terremotos mediante distribuciones gamma y lognormal. Aplica probabilidades condicionales para evaluar la posibilidad de que un sismo fuerte ocurra en función del tiempo transcurrido desde el último evento registrado en la zona de análisis.

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 personas que viven en áreas identificadas como de alta actividad sísmica. Este estudio analiza patrones sísmicos en zonas críticas de Norteamérica, Sudamérica, Japón, el sur de China y el norte de la India, y crea modelos probabilistas de inferencia sísmica basados en aprendizaje automático bayesiano. El enfoque propuesto permite identificar tendencias de recurrencia a largo plazo para cada zona sísmica, mejorando la capacidad de anticipación de eventos sísmicos de gran magnitud.

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 emplea una red neuronal convolucional (CNN) para estimar la probabilidad de ocurrencia de terremotos, mientras que la vulnerabilidad la determina mediante el proceso de jerarquía analítica (AHP) y el peligro lo representa utilizando la teoría de intersección de Venn. La integración de estos componentes permite generar mapas de riesgo más precisos, útiles para la planificación y mitigación en una región altamente expuesta a la actividad sísmica.

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 cual puede generar datos históricos del movimiento del suelo. Este modelo se basa en el uso de redes generativas antagónicas (GANs) y crea registros de movimiento sísmico, lo cual sirve cuando los datos reales son escasos. También se propone un método para evaluar cuantitativa y cualitativamente el desempeño del modelo propuesto comparando los resultados arrojados con los de registros observados. Finalmente, el modelo se optimiza para alcanzar un alto desempeño mediante la integración de criterios de ingeniería sísmica y técnicas avanzadas de aprendizaje profundo.

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. El estudio implementa distribuciones de valores extremos (Gumbel, Weibull y GEV) con estimación de parámetros mediante máxima verosimilitud y cuantificación de incertidumbre a través de bootstrap paramétrico con 10,000 iteraciones. Los autores calculan niveles de retorno para períodos de 10, 50, 100 y 475 años (este último utilizado en códigos de construcción internacional) e intervalos de confianza al 95 %. Aplicando datos históricos de 120 años de Suiza, validan su metodología con eventos recientes y demuestran la robustez del enfoque para estimaciones de largo plazo en contextos de alta consecuencia.

Papalexiou y Koutsoyiannis (2020) realizan un estudio comparativo riguroso de métodos de bootstrap paramétrico para valores extremos, evaluando específicamente la estabilidad de intervalos de confianza en función del número de iteraciones. A través de experimentos de simulación Monte Carlo con distribuciones GEV y Gumbel, los autores comparan el desempeño de 5,000, 10,000 y 20,000 iteraciones bootstrap,

analizando la convergencia de los percentiles 2.5 % y 97.5 % que definen los límites de confianza al 95 %. Sus resultados demuestran que 10,000 iteraciones proporcionan un equilibrio óptimo entre precisión estadística y eficiencia computacional, con mejoras marginales al incrementar a 20,000 iteraciones. Aunque el estudio se centra en datos hidrológicos, la metodología estadística es directamente transferible al análisis de valores extremos sísmicos, proporcionando justificación teórica para la selección del número de iteraciones bootstrap.

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 locales y permite estimar con mayor precisión las amplitudes horizontales del movimiento del suelo. Esta aproximación regional ofrece un análisis más confiable para la evaluación del peligro sísmico en esa zona y aplicarlo al contexto del entorno sísmico taiwanés.

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 2020. Los autores calculan estadísticos de recurrencia sísmica incluyendo el tiempo medio entre eventos (μ), la desviación estándar (σ) y el coeficiente de variación ($CV = \sigma/\mu$), este último utilizado para cuantificar la regularidad temporal de la actividad sísmica. Desarrollan un índice de proximidad temporal (IPT) definido como la razón entre el tiempo transcurrido desde el último evento fuerte y el tiempo medio de recurrencia, estableciendo umbrales para clasificación de riesgo: $IPT < 0.5$ (fase temprana), $0.5 \leq IPT < 0.8$ (fase intermedia), $0.8 \leq IPT < 1.0$ (proximidad al tiempo medio) e $IPT \geq 1.0$ (brecha sísmica). Emplean distribuciones lognormal y Weibull para estimar la magnitud y ventana temporal del próximo evento significativo en el segmento de Michoacán.

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 utiliza amplitudes de ondas sísmicas registradas por los sismógrafos durante un terremoto, ajustadas por la distancia al epicentro, proporcionando una medida cuantitativa de la energía liberada por el mismo. También se habla acerca del proceso para calibrar la escala, así como de su desarrollo y validación mediante el análisis de eventos sísmicos ocurridos en el sur de California.

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. El estudio revisa diversas técnicas de predicción sísmica, entre ellas los enfoques basados en la observación de precursores físicos (como variaciones en la actividad sísmica, cambios en el nivel freático o emisiones de gases), los métodos estadísticos que modelan la recurrencia de eventos y las aproximaciones más recientes que integran análisis geoespaciales y herramientas computacionales, y discute sus fundamentos, limitaciones y alcances. El objetivo de este análisis es ofrecer un marco más sólido para comprender y anticipar la ocurrencia de terremotos, que sirva como referencia para investigadores y profesionales.

Rootzén y Zholud (2021) desarrollan un marco teórico riguroso para la aplicación de bootstrap paramétrico a eventos raros, proporcionando fundamentación matemática sobre la convergencia de estimadores y la construcción de intervalos de confianza. Aunque el artículo se centra en aplicaciones de bioestadística (específicamente en la estimación de falsos positivos en pruebas de alto rendimiento), los autores derivan expresiones analíticas para el error de Monte Carlo en función del número de iteraciones bootstrap, demostrando formalmente por qué 10,000 iteraciones son suficientes para alcanzar un error estándar menor al 0.5 % en los percentiles extremos (2.5 % y 97.5 %). El trabajo incluye comparaciones con métodos bootstrap no paramétricos y proporciona guías para la validación de intervalos de confianza mediante técnicas de simulación, contribuyendo al marco metodológico general del bootstrap aplicado a valores extremos.

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 de

2. Introducción

México (utilizando registros del SSN) y Colombia. Los autores desarrollan un índice de proximidad temporal conceptualmente similar al propuesto por otros investigadores, que cuantifica la “brecha sísmica” como la razón entre el tiempo transcurrido desde el último evento significativo y el tiempo medio de recurrencia estimado a partir del historial sísmico. El modelo emplea redes bayesianas para integrar información de múltiples fuentes (catálogos históricos, monitoreo geodésico, modelos geofísicos) y actualizar probabilísticamente las estimaciones de peligrosidad conforme se incorporan nuevos datos. Establecen umbrales cuantitativos para clasificación de niveles de alerta basados en el índice de proximidad temporal y en probabilidades de excedencia calculadas mediante distribuciones lognormal y Weibull ajustadas a los intervalos de recurrencia observados.

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 unificado y actualizado, así como modelos de fuentes que representan la actividad tectónica regional, con el fin de estimar el peligro en términos de aceleración máxima del terreno (PGA) y aceleración espectral, considerando la incertidumbre asociada a parámetros sismológicos clave.

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. Aunque el contexto de aplicación es hidrológico (riesgo de inundación), la formulación matemática es directamente aplicable al análisis sísmico: los autores utilizan máximos anuales, ajustan distribuciones GEV y Gumbel mediante máxima verosimilitud, y calculan niveles de retorno para períodos de 5, 10, 20 y 50 años. Desarrollan expresiones para la probabilidad de que al menos un evento exceda una magnitud M en los próximos Δt años: $P(\text{excedencia en } \Delta t \text{ años}) = 1 - [F(M)]^{\Delta t}$, donde $F(M)$ es la función de distribución acumulada evaluada en M . Los autores introducen además el coeficiente de variación (CV) del tiempo de recurrencia como métrica para cuantificar la regularidad temporal de los eventos, demostrando que $CV < 0.5$ indica procesos cuasi-periódicos (apropiados para modelos de renovación), mientras que $CV \geq 1.0$ sugiere procesos cercanos a Poisson (memoria temporal negligible).

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 significación estadística. El artículo analiza cómo estos elementos se relacionan entre sí y cómo influyen en la interpretación de los resultados, destacando el papel crítico del tamaño de la muestra en la validez de las conclusiones y proporcionando un marco metodológico útil para la toma de decisiones basadas en la evidencia dentro del ámbito clínico.

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. Se evalúan distribuciones como la Normal, Cauchy, Chi-Cuadrada, Normal estándar y t de Student, aplicando criterios de selección como el Criterio de Información de Akaike (AIC) para identificar la distribución que mejor se ajuste a los datos.

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 predicciones de modelos probabilísticos previos. Los autores, varios de ellos investigadores del Servicio Sismológico Nacional de la UNAM, discuten por qué el evento fue considerado “inesperado” según modelos tradicionales de Poisson pero resulta predecible cuando se emplean modelos de renovación que consideran el tiempo transcurrido desde el último evento significativo. El estudio enfatiza la importancia de incorporar memoria temporal en los modelos sísmicos y proporciona un caso de estudio valioso sobre validación retrospectiva de metodologías probabilísticas, demostrando que el análisis de brechas sísmicas y tiempos de recurrencia puede mejorar sustancialmente la identificación de regiones de alto riesgo.

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. El modelo se construye a partir de registros de sismos y réplicas en la región y permite relacionar la magnitud y localización de los eventos con la intensidad del movimiento esperado en distintos sitios, obteniendo una estimación fiable de los parámetros del movimiento del suelo.

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 una falla activa. El estudio usa diagramas de caja y parámetros meteorológicos para identificar patrones en las concentraciones de radón, los cuales se relacionan con procesos sísmicos subterráneos previos a la ocurrencia de terremotos.

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 comparan el ajuste de cuatro distribuciones de valores extremos (Gumbel, Weibull, GEV y Fréchet) para cada una de las regiones sismotectónicas del país, empleando el criterio BIC para la selección objetiva del modelo óptimo por región. Implementan bootstrap paramétrico con 10,000 iteraciones para construir intervalos de confianza robustos de los niveles de retorno, y calculan probabilidades de excedencia para magnitudes críticas ($M \geq 6.5$, $M \geq 7.0$) en horizontes temporales de 10, 20 y 50 años. Los resultados se integran en un índice de peligrosidad sísmica regional que combina ponderadamente la magnitud esperada, frecuencia histórica y proximidad temporal al último evento significativo.

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 que abarca desde 1900 hasta 2021. Los autores emplean teoría de valores extremos para caracterizar el comportamiento sísmico de cada región identificada, aplicando distribuciones Gumbel, Weibull y GEV según la naturaleza de los datos en cada zona. Desarrollan un índice compuesto de peligrosidad sísmica que integra cuatro componentes normalizados: nivel de retorno a 50 años, probabilidad de exceder $M \geq 7.0$ en 10 años, índice de proximidad temporal y magnitud promedio histórica. Los autores identifican y caracterizan las regiones de Oaxaca, Guerrero, Michoacán y Chiapas como zonas de particular interés debido a su elevada actividad sísmica y potencial de generación de eventos destructivos, proporcionando parámetros específicos de recurrencia y estimaciones de peligrosidad para cada segmento de la costa del Pacífico mexicano.

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, distribuciones de valores extremos como Gumbel, Weibull y GEV, así como bibliografía publicada que ahonda en la aplicación práctica para solución de problemas basados en fórmulas y cálculos derivados de la probabilidad y la estadística, y sus variables dependientes necesarias para realizar los cálculos.

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 riesgos para las personas que viven en áreas consideradas como altamente sísmicas (Abebe, Kebede, Kevin y Demissie 2023, Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022). Es por esto que constantemente se realizan análisis estadísticos y geográficos para este tipo de eventos naturales (Barrientos, Fernandez y Rivero 2007), así como evaluaciones probabilísticas de la peligrosidad sísmica (Sawires, Peláez y Santoyo 2023), con la finalidad de disminuir o prevenir los desastres que estos ocasionan. Dichos estudios permiten estimar magnitudes en la escala de Richter (Richter 1935) o calcular probabilidades de ocurrencia de futuros terremotos (Abebe, Kebede, Kevin y Demissie 2023, Al Banna, Taher, Kaiser, Mahmud, Rahman, Hosen y Cho 2020, Ferrás 2005, Jena, Pradhan, Naik y Alamri 2021, Matsumoto, Yaoyama, Lee, Hida y Itoi 2023, Singh, Arroyo, Srinivas y Suresh 2021, Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique 2019, Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022).

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, o también técnicas derivadas de la IA como los algoritmos transformer aplicados a series temporales (Abebe, Kebede, Kevin y Demissie 2023). De igual forma, los enfoques que se basan en deep learning o aprendizaje profundo (Abebe, Kebede, Kevin y Demissie 2023, Jena, Pradhan, Naik y Alamri 2021, Matsumoto, Yaoyama, Lee, Hida y Itoi 2023) se emplean para realizar mapeos de riesgo y la predicción de eventos sísmicos. También se aplican técnicas basadas en el uso de

redes neuronales convolucionales (Jena, Pradhan, Naik y Alamri 2021), las cuales evalúan la probabilidad de ocurrencia de un terremoto.

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 realizar estimaciones de futuros eventos sísmicos, tanto para profundidades relativamente bajas y magnitudes pequeñas (Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano 2021) como para desarrollar modelos de generación de movimiento del suelo mediante GANs (Matsumoto, Yaoyama, Lee, Hida y Itoi 2023) y ecuaciones de predicción del movimiento del suelo (GMPE) específicas para Taiwán y el Himalaya occidental (Phung, Loh, Chao, Chiou y Huang 2020, Singh, Arroyo, Srinivas y Suresh 2021). La teoría de valores extremos (extreme value theory, EVT) ha emergido como uno de los marcos teóricos más robustos para el análisis de eventos sísmicos raros de gran magnitud (Coles 2001, Coles, Bawa, Trenner y Dorazio 2021, Mignan, Danciu y Giardini 2021), proporcionando fundamentación matemática sólida para la predicción probabilística de terremotos.

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) demuestra que, para conjuntos de datos históricos con menos de 150 años de registros (como es común en sismología), el método de máximos anuales utilizando distribuciones GEV, Gumbel o Weibull resulta más robusto y confiable que el enfoque POT. Este hallazgo respalda la selección metodológica adoptada en numerosos estudios sísmicos que emplean magnitudes máximas anuales como variable de análisis (Mignan, Danciu y Giardini 2021, Yaghmaei-Sabegh y Neekmanesh 2022, Ramírez-Gaytán, Aguirre-González y Núñez-Cornú 2021).

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 modelos no Poisson como Brownian Passage Time (BPT), Weibull, distribución gamma y el modelo Epidemic Type Aftershock Sequence (ETAS) (Convertito, Ebrahimian, Amoroso, Jalayer, De Matteis y Capuano 2021, Ferrás 2005). La selección de la distribución probabilística más apropiada para cada región se realiza mediante criterios estadísticos rigurosos como el Criterio de Información Bayesiano (BIC) y el Criterio de Información de Akaike (AIC), complementados con pruebas de bondad de ajuste como Kolmogorov-Smirnov, Anderson-Darling y Lilliefors (Bommier 2023, Yaghmaei-Sabegh y Neekmanesh 2022). La evaluación de la adecuación de diferentes distribuciones a datos de valores extremos, incluso en áreas distintas a la sismología como los niveles máximos de inundación, se realiza mediante análisis comparativos (Shobanke, Olayemi y Olajide 2024). Igualmente, se han empleado la distribución de Venn para determinar el peligro de un sismo futuro (Jena, Pradhan, Naik y Alamri 2021), los estadísticos descriptivos de series temporales de radón en el suelo (Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique 2019), y el aprendizaje automático bayesiano (Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022). Un aspecto crucial de todos estos análisis es la aplicación de pruebas de hipótesis y la construcción de intervalos de confianza para validar las conclusiones y estimar la incertidumbre de los parámetros (Shreffler y Huecker 2025), empleando técnicas de bootstrap paramétrico que han demostrado ser particularmente efectivas para la cuantificación de incertidumbre en estimaciones de valores extremos (Papalexiou y Koutsoyiannis 2020, Rootzén y Zholud 2021).

La implementación de bootstrap paramétrico para la construcción de intervalos de confianza en análisis de valores extremos ha sido objeto de investigación metodológica específica. Papalexiou y Koutsoyiannis (2020) demuestran mediante experimentos de Monte Carlo que 10,000 iteraciones bootstrap proporcionan un equilibrio óptimo entre precisión estadística y eficiencia computacional para distribuciones de valores extremos, con mejoras marginales al incrementar a 20,000 iteraciones. Rootzén y Zholud (2021) complementan estos hallazgos proporcionando fundamentación matemática sobre la convergencia

de estimadores bootstrap y expresiones analíticas para el error de Monte Carlo, validando que 10,000 iteraciones son suficientes para alcanzar errores estándar menores al 0.5 % en los percentiles extremos que definen los intervalos de confianza al 95 %.

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 metodologías para el cálculo de niveles de retorno en horizontes temporales de 10, 50, 100 y 475 años (este último utilizado en códigos de construcción internacional), implementando intervalos de confianza mediante bootstrap paramétrico. Serinaldi y Kilsby (2021) formalizan la expresión para la probabilidad de que al menos un evento exceda una magnitud M en los próximos Δt años: $P(\text{excedencia en } \Delta t \text{ años}) = 1 - [F(M)]^{\Delta t}$, donde $F(M)$ es la función de distribución acumulada, y demuestran su aplicabilidad a procesos de renovación con memoria temporal. Estos enfoques permiten estimar no solo la magnitud esperada en diferentes horizontes temporales, sino también la probabilidad de que eventos específicos ocurran dentro de ventanas temporales de interés para la planificación y mitigación del riesgo.

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 que calculan un índice de proximidad temporal (IPT) definido como la razón entre el tiempo transcurrido desde el último evento significativo y el tiempo medio de recurrencia histórica, estableciendo umbrales para clasificación de riesgo. Ramírez-Gaytán, Aguirre-González y Núñez-Cornú (2021) aplican esta metodología específicamente a la brecha sísmica de Michoacán utilizando datos del SSN desde 1900, introduciendo además el coeficiente de variación ($CV = \sigma/\mu$) para cuantificar la regularidad temporal: $CV < 0.5$ indica recurrencia cuasi-periódica apropiada para modelos de renovación, mientras que $CV \geq 1.0$ sugiere procesos cercanos a Poisson. Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir (2020) validan retrospectivamente estos enfoques con el análisis del sismo de Puebla 2017, demostrando que eventos considerados “inesperados” bajo modelos de Poisson resultan predecibles cuando se incorpora memoria temporal mediante modelos de renovación.

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 índice que combina la magnitud esperada según modelos de valores extremos, probabilidades de excedencia en diferentes horizontes temporales, tiempo desde el último evento significativo y tasa histórica de actividad, validando retrospectivamente sus predicciones con eventos ocurridos entre 2016 y 2019 en Italia. Zúñiga, Suárez, Figueroa-Soto y Mendoza (2022) proponen una regionalización sismotectónica de México que emplea un índice similar, normalizando cuatro componentes (nivel de retorno a 50 años, probabilidad de exceder $M \geq 7.0$ en 10 años, índice de proximidad temporal y magnitud promedio histórica) para generar clasificaciones objetivas de peligrosidad por región. Estos índices compuestos permiten comparaciones cuantitativas entre regiones y facilitan la priorización de recursos para mitigación de riesgo.

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 identificar patrones en la concentración de gas radón en el subsuelo previos a la ocurrencia de sismos y plantea el que estos puedan actuar como posibles precursores sísmicos.

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 (comprendido por varios países, entre ellos Etiopía, Sudán y Uganda) (Abebe, Kebede, Kevin y Demissie 2023), Suiza en contexto de sismicidad inducida (Mignan, Danciu y Giardini 2021), Irán con análisis multi-regional (Yaghmaei-Sabegh y Neekmanesh 2022), Taiwán (Phung, Loh, Chao, Chiou y Huang 2020), Japón y otras regiones de Asia (Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022), el Arco del Himalaya occidental (Singh, Arroyo, Srinivas y Suresh 2021), Pakistán (Tareen, Nadeem, Kearfott, Abbas, Khawaja y Rafique 2019), Italia con validación

retrospectiva de modelos (Convertito, Catalli y Emolo 2020) y regiones de Norteamérica, Sudamérica y China (Velasco Herrera, Rossello, Orgeira, Arioni, Soon, Velasco y Vera 2022).

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). Particularmente relevante es la revisión de Coles, Bawa, Trenner y Dorazio (2021), que actualiza las recomendaciones metodológicas sobre teoría de valores extremos veinte años después del trabajo seminal de Coles (2001), incorporando avances computacionales y estadísticos recientes e integrando aplicaciones a cambio climático, desastres naturales y eventos sísmicos.

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) se enfocan en la brecha sísmica de Michoacán utilizando más de un siglo de datos del SSN, mientras que Zúñiga, Suárez, Figueroa-Soto y Mendoza (2022) proponen una regionalización sismotectónica completa del país que caracteriza específicamente las zonas de Oaxaca, Guerrero, Michoacán y Chiapas mediante parámetros de recurrencia y estimaciones de peligrosidad. Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir (2020), con autores pertenecientes al SSN-UNAM, proporcionan validación retrospectiva de modelos probabilísticos mediante el análisis del sismo de Puebla 2017, enfatizando la importancia de incorporar memoria temporal en modelos sísmicos mexicanos. Estos trabajos demuestran la madurez creciente de la aplicación de teoría de valores extremos al contexto sísmico mexicano y proporcionan parámetros regionales específicos que fundamentan estimaciones de peligrosidad actualizadas.

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 avanzadas de valores extremos (Gumbel, Weibull, GEV), enfoques de inteligencia artificial y aprendizaje profundo, así como indicadores geofísicos como el gas radón. Los trabajos revisados muestran una tendencia hacia la integración de métodos estadísticos rigurosos con técnicas modernas de machine learning, lo que permite mejorar la precisión de las estimaciones y adaptarlas a contextos regionales específicos. Un hallazgo relevante es la consolidación de la teoría de valores extremos (Coles 2001, Coles, Bawa, Trenner y Dorazio 2021) como marco teórico robusto para el análisis de eventos sísmicos raros de gran magnitud, particularmente mediante el método de máximos anuales que ha demostrado ser más confiable que enfoques alternativos para tamaños de muestra típicos en sismología (Bommier 2023). Este panorama refleja un campo en constante evolución, donde la combinación de enfoques tradicionales y emergentes constituye la base para avanzar hacia predicciones más confiables y útiles para la mitigación del riesgo sísmico.

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 como variable de análisis para conjuntos de datos con menos de 150 años de registros (Bommier 2023), (2) selección de distribución óptima por región mediante criterios BIC o AIC validados con pruebas de bondad de ajuste múltiples (Yaghmaei-Sabegh y Neekmanesh 2022), (3) estimación de parámetros mediante máxima verosimilitud (Mignan, Danciu y Giardini 2021), (4) cuantificación de incertidumbre mediante bootstrap paramétrico con 10,000 iteraciones (Papalexiou y Koutsoyiannis 2020), (5) cálculo de niveles de retorno para períodos estándar (10, 20, 50, 100 años) con intervalos de confianza al 95 %, (6) incorporación de análisis temporal mediante índices de proximidad y coeficientes de variación de recurrencia (Sánchez-Silva y Riaño 2020, Ramírez-Gaytán, Aguirre-González y Núñez-Cornú 2021), y (7) validación retrospectiva con eventos recientes (Singh, Reinoso, Arroyo, Ordaz, Cruz-Atienza, Pérez-Campos, Iglesias y Hjörleifsdóttir 2020). La adopción sistemática de estas prácticas en el contexto

mexicano, particularmente para las regiones de mayor actividad sísmica en la costa del Pacífico, representa una oportunidad para mejorar significativamente la robustez y confiabilidad de las estimaciones de peligrosidad a nivel nacional.

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.

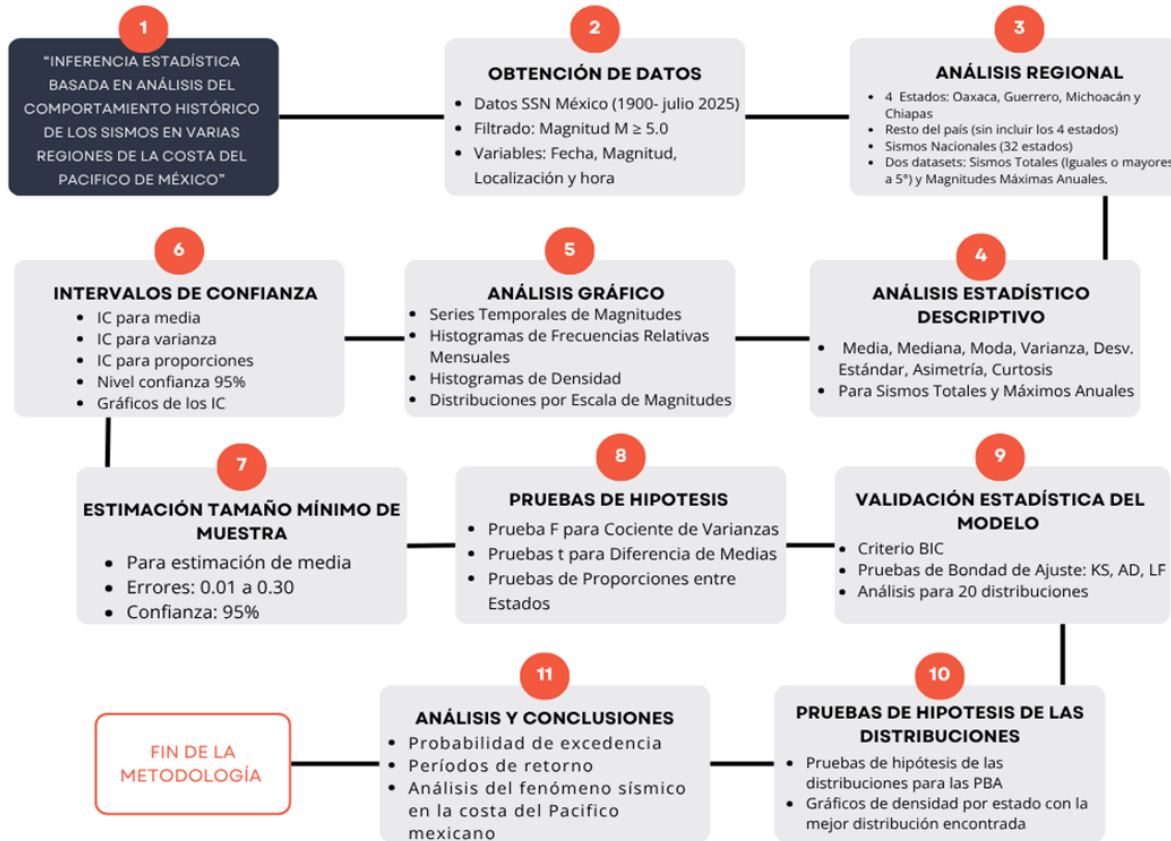


Figura 1: Diagrama de flujo de la aplicación de la metodología propuesta.

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 $A^2 > A^2_{\text{crítico}}(\alpha)$, 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 probabilista que rige a este proyecto, se hace el cálculo de los periodos de retorno, probabilidades de excedencia e intervalos de confianza.

4.2.20. Teoría de valores extremos y distribuciones empleadas

La teoría de valores extremos (Coles 2001) proporciona el marco teórico para modelar eventos raros de gran magnitud. Según el teorema de Fisher-Tippett, las distribuciones asintóticas de máximos (o mínimos) convergen a una de tres familias de distribuciones:

- **Gumbel (Tipo I):** Para distribuciones con cola exponencial
- **Fréchet (Tipo II):** Para distribuciones con cola pesada
- **Weibull (Tipo III):** Para distribuciones con cola limitada

Estas tres familias se unifican en la distribución generalizada de valores extremos (GEV).

Distribución de Gumbel

La distribución de Gumbel se utiliza cuando los datos extremos provienen de una distribución con decaimiento exponencial en la cola.

Función de densidad:

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \quad (4.23)$$

Función de distribución acumulada:

$$F(x; \mu, \sigma) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \quad (4.24)$$

donde μ es el parámetro de localización, $\sigma > 0$ es el parámetro de escala, $x \in (-\infty, +\infty)$. Esta distribución se utiliza para el estado de Oaxaca, donde los datos extremos muestran comportamiento de tipo Gumbel según los resultados obtenidos por las pruebas de bondad de ajuste.

Distribución de Weibull

La distribución de Weibull se emplea cuando existe un límite superior para los valores extremos:

Función de densidad:

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \quad (4.25)$$

Función de distribución acumulada:

$$F(x; k, \lambda) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \quad (4.26)$$

donde $k > 0$ es el parámetro de forma, $\lambda > 0$ es el parámetro de escala y $x \geq 0$. Los resultados indican que si $k < 1$ hay tasa de falla decreciente, si $k = 1$ la tasa de falla es constante (equivalente a distribución exponencial) y si $k > 1$ la tasa de falla es creciente. Esta distribución se utiliza para los estados de Guerrero y Chiapas, donde los datos muestran mejor ajuste a la distribución Weibull.

Distribución generalizada de valores extremos

La distribución GEV unifica las tres familias de valores extremos mediante un parámetro de forma:

Función de distribución acumulada:

$$F(x; \mu, \sigma, \xi) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\} & \text{si } \xi \neq 0 \\ \exp \left\{ - \exp \left[-\frac{x-\mu}{\sigma} \right] \right\} & \text{si } \xi = 0 \end{cases} \quad (4.27)$$

donde μ es el parámetro de localización, $\sigma > 0$ es el parámetro de escala y ξ es el parámetro de forma que determina el tipo de distribución, que puede ser si $\xi = 0$ Gumbel, $\xi > 0$ Fréchet (cola pesada) y si $\xi < 0$ Weibull (cola acortada). Esta distribución se utiliza para el estado de Michoacán donde la distribución GEV proporciona el mejor ajuste según las pruebas de bondad realizadas.

Estimación de parámetros

Los parámetros de las distribuciones se estiman mediante el método de máxima verosimilitud (Fisher 1922), que consiste en encontrar los valores de los parámetros que maximizan la función de verosimilitud:

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) \quad (4.28)$$

En la práctica, se maximiza la log-verosimilitud:

$$\ell(\theta) = \sum_{i=1}^n \ln f(x_i; \theta) \quad (4.29)$$

4.2.21. Periodos de retorno

El periodo de retorno T es el tiempo esperado entre eventos que exceden una magnitud específica x (Coles 2001):

$$T(x) = \frac{1}{1 - F(x)} \quad (4.30)$$

donde $F(x)$ es la función de distribución acumulada de la magnitud máxima anual. Aquí, por ejemplo, un periodo de retorno $T = 50$ años para una magnitud $M = 7.0$ significaría que en promedio se esperaría un sismo de magnitud ≥ 7.0 una vez cada 50 años.

Cabe aclarar para lo establecido en el párrafo anterior que el periodo de retorno es un promedio de largo plazo, no significa que después de un evento de periodo de 50 años el siguiente evento ocurriría después de otros 50 años. La ocurrencia sísmica tiene naturaleza estocástica.

Niveles de retorno

El nivel de retorno x_T es la magnitud asociada a un periodo de retorno T específico. Se obtiene invirtiendo la relación:

$$x_T = F^{-1} \left(1 - \frac{1}{T} \right) \quad (4.31)$$

Cálculo según cada distribución:

Gumbel:

$$x_T = \mu - \sigma \ln \left(-\ln \left(1 - \frac{1}{T} \right) \right) \quad (4.32)$$

Weibull:

$$x_T = \lambda \left(-\ln \left(1 - \frac{1}{T} \right) \right)^{1/k} \quad (4.33)$$

GEV:

$$x_T = \begin{cases} \mu + \frac{\sigma}{\xi} \left[\left(-\ln \left(1 - \frac{1}{T} \right) \right)^{-\xi} - 1 \right] & \text{si } \xi \neq 0 \\ \mu - \sigma \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] & \text{si } \xi = 0 \end{cases} \quad (4.34)$$

Obtener el cálculo de los niveles de retorno para $T \in \{10, 20, 50, 100\}$ años, proporcionando estimaciones de las magnitudes esperadas en diferentes horizontes temporales.

4.2.22. Probabilidades de excedencia

La probabilidad de que al menos un evento de magnitud $\geq M$ ocurra en los próximos Δt años se calcula como (Serinaldi y Kilsby 2021):

$$P(\text{excedencia en } \Delta t \text{ años}) = 1 - [F(M)]^{\Delta t} \quad (4.35)$$

donde $F(M)$ es la probabilidad de que la magnitud máxima anual no exceda M en un año, $[F(M)]^{\Delta t}$ es la probabilidad de no exceder M en Δt años consecutivos (suponiendo independencia entre años) y $1 - [F(M)]^{\Delta t}$ es la probabilidad de exceder M al menos una vez en Δt años.

Si el periodo de retorno de un evento es T , entonces la probabilidad de excedencia en n años es aproximadamente:

$$P \approx 1 - \exp \left(-\frac{n}{T} \right) \quad (4.36)$$

donde esta aproximación será válida cuando n/T es pequeño.

4.2.23. Intervalos de confianza mediante bootstrap paramétrico

Dado que las estimaciones puntuales de periodos de retorno y probabilidades tienen incertidumbre asociada, se utiliza bootstrap paramétrico para cuantificar esta incertidumbre mediante intervalos de confianza (Papalexiou y Koutsoyiannis 2020, Rootzén y Zholud 2021):

Algoritmo de bootstrap paramétrico:

1. **Ajustar el modelo:** se estiman los parámetros $\hat{\theta}$ de la distribución (Gumbel, Weibull o GEV) a partir de los datos originales de magnitudes máximas anuales.
2. **Generar muestras bootstrap (repetir $B = 10,000$ veces):**
 - Generar una muestra sintética de tamaño n de la distribución ajustada $F(x; \hat{\theta})$
 - Estimar los parámetros $\hat{\theta}_b^*$ de esta muestra bootstrap
 - Calcular los niveles de retorno x_T^{*b} para los periodos deseados ($T = 10, 20, 50, 100$ años)
3. **Construir intervalos de confianza,** para cada periodo de retorno T , calcular:
 - Límite inferior: percentil 2.5 % de $\{x_T^{*1}, x_T^{*2}, \dots, x_T^{*B}\}$

- Estimación central: mediana de $\{x_T^{*1}, x_T^{*2}, \dots, x_T^{*B}\}$
- Límite superior: percentil 97.5 % de $\{x_T^{*1}, x_T^{*2}, \dots, x_T^{*B}\}$

donde existe un 95 % de probabilidad de que el verdadero nivel de retorno esté contenido en el intervalo calculado.

4.2.24. Análisis temporal y tiempo desde el último evento

Para evaluar la proximidad a un posible evento futuro, se analiza el tiempo transcurrido desde el último sismo fuerte en cada región (Ramírez-Gaytán, Aguirre-González y Núñez-Cornú 2021, Sánchez-Silva y Riaño 2020):

Índice de proximidad temporal

Se define el índice de proximidad temporal como:

$$IPT = \frac{t_{\text{transcurrido}}}{t_{\text{medio}}} \quad (4.37)$$

donde $t_{\text{transcurrido}}$ es el tiempo (en años) desde el último sismo fuerte en la región y t_{medio} es el tiempo medio de recurrencia calculado como la media de los intervalos entre eventos históricos. El resultado indicaría que si $IPT < 0.5$ es una fase temprana del ciclo sísmico, si está en el intervalo $0.5 \leq IPT < 0.8$ indica que se está en una fase intermedia del ciclo sísmico, si $IPT \geq 0.8$ hay proximidad al tiempo medio y si $IPT > 1.0$ se excede el tiempo medio del ciclo sísmico lo cual indicaría que la zona se encuentra en riesgo sísmico constante.

Coefficiente de variación del tiempo de recurrencia

El coeficiente de variación (CV) cuantifica la regularidad de la recurrencia sísmica (Serinaldi y Kilsby 2021):

$$CV = \frac{\sigma_{\text{recurrencia}}}{\mu_{\text{recurrencia}}} \quad (4.38)$$

donde $\sigma_{\text{recurrencia}}$ es la desviación estándar de los intervalos entre eventos y $\mu_{\text{recurrencia}}$ es la media de los intervalos entre eventos. El resultado indicaría que si $CV < 0.5$ existe una recurrencia sísmica bastante regular, si $0.5 \leq CV < 1.0$ existe una recurrencia moderadamente variable y si $CV \geq 1.0$ la recurrencia es altamente variable (es cercana a un proceso de Poisson). Estos resultados indicarían que si CV es bajo un modelo de renovación (que considera el tiempo transcurrido) es más apropiado que un modelo de Poisson (que asume eventos independientes).

4.2.25. Índice de peligrosidad sísmica regional

Se construye un índice compuesto que integra múltiples factores para comparar el nivel de peligrosidad entre regiones (Convertito, Catalli y Emolo 2020, Zúñiga, Suárez, Figueroa-Soto y Mendoza 2022):

$$IPS = w_1 \cdot \frac{x_{50}}{x_{50, \text{máx}}} + w_2 \cdot \frac{P_{M \geq 7, 10 \text{ años}}}{P_{\text{máx}}} + w_3 \cdot IPT + w_4 \cdot \frac{\mu}{\mu_{\text{máx}}} \quad (4.39)$$

donde x_{50} es el nivel de retorno a 50 años, $P_{M \geq 7, 10 \text{ años}}$ es la probabilidad de exceder $M = 7.0$ en 10 años, IPT es el índice de proximidad temporal, μ es la magnitud promedio histórica, los términos se normalizan

4. Metodología

dividiendo por el máximo observado entre todas las regiones, w_1, w_2, w_3, w_4 son pesos que suman 1 (por defecto 0.3, 0.3, 0.2, 0.2). El resultado indicaría que si el índice se encuentra entre $0.0 \leq \text{IPS} < 0.3$ hay peligrosidad baja, si está entre $0.3 \leq \text{IPS} < 0.6$ la peligrosidad sísmica es media, si está entre $0.6 \leq \text{IPS} < 0.8$ hay peligrosidad alta y si está entre $0.8 \leq \text{IPS} \leq 1.0$ la peligrosidad es muy alta. Este índice permite realizar comparaciones entre regiones considerando múltiples dimensiones del riesgo sísmico.

Capítulo 5

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 pueden identificar patrones y análisis útiles.

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 reciente.

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 función de sus coordenadas.

El análisis de los archivos individuales de sismos y todos los cálculos relacionados al análisis estadístico se realizan en la herramienta R (versión 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 2 con los resultados del análisis para las 6 regiones de estudio.

5. Análisis de resultados

Tabla 2: Estadísticos descriptivos consolidados.

Región	Min	Max	Media	Mediana	Moda	Desv.	Var.	Asim.	Curt.	Totales
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
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	-0.050	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 han considerado sismos de magnitudes máximas. Esto permite identificar diferencias importantes en la intensidad de los sismos según las regiones analizadas.

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 ($p\text{-valor} > 0.05$): 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

6. Conclusiones

niveles entre 6.8° y 7.2° para el mismo horizonte temporal. La implementación de bootstrap paramétrico con 10,000 iteraciones (Papalexiou y Koutsoyiannis 2020) 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 ($\text{IPT} > 1.0$), 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.
- 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.
- 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/>
- Papalexiou, S. M. y Koutsoyiannis, D. (2020). Random number generators for extreme values: A comparison study, *Journal of Hydrology* **590**: 125445.
- 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.
- Rootzén, H. y Zholud, D. (2021). Efficient estimation of the number of false positives in high-throughput screening, *Biometrika* **108**(1): 1–17.
- 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.