



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

APLICACIÓN DE UN MODELO DE
APRENDIZAJE PROFUNDO PARA
SIMULTÁNEAMENTE DETECTAR SISMOS Y
SELECCIONAR FASES SÍSMICAS P Y S

TESIS

PARA OBTENER EL TÍTULO DE:

Ingeniera Geofísica

PRESENTA:

Ana Daniela Sánchez Gutiérrez

DIRECTOR DE TESIS:

Dr. Miguel Enríquez Vargas

Ciudad Universitaria, Cd. Mx., 2024



Gracias a mi mamá Rosa Luisa Gutiérrez Anaya.

Has marcado nuestra vida con tu fuerza y valentía.

Continuaré honrando tu memoria y escuchando tu corazón.

Agradecimientos

Gracias a mi papá Jesús Sánchez Toral por el amor y generosidad con que apoyó mi formación. Me propongo aprovechar de la mejor manera la oportunidad que me diste. A mi hermana Verónica Sánchez y a su familia por su importante y valioso apoyo. Mi regalo es una familia incondicional.

A mi asesor el Dr. Miguel Enríquez por su tiempo y enseñanzas como profesor, por su confianza y comprensión en la amistad. Al M.C. Jonatán Arreola y al Dr. Jorge López por su dedicación durante la realización y revisión de este trabajo. Esta tesis no sería lo mismo sin su buen juicio.

A mis profesores, y también sinodales, M.C. David Escobedo, Dr. Martín Cárdenas y Dr. Gerardo Cifuentes, por impartir conocimiento a través de su pasión por la geofísica.

A Ludwig Luis, Enrique Pereira, Gabriela Vidal, Raúl Jiménez y Alejandro Reyes, por su gran apoyo a lo largo de estos años. Soy muy afortunada en haber coincidido con ustedes en nuestra etapa universitaria. A Arturo Romero, por escucharme y confiar siempre en mí.

Por último, a mi alma mater, la Universidad Nacional Autónoma de México y a la Facultad de Ingeniería. Me siento agradecida por la valiosa oportunidad de haber realizado este trabajo, de dar un paso más aprendiendo sobre lo que me encanta.

Resumen

La sismología ha experimentado un crecimiento significativo con la incorporación de equipos asequibles que generan rápidamente grandes volúmenes de datos. Este crecimiento exponencial en los datos sismológicos requiere el desarrollo de nuevas herramientas de procesamiento. En el contexto de la detección automática de señales sísmicas, especialmente en el monitoreo sísmico, los enfoques más recientes se basan en aprendizaje automático y aprendizaje profundo.

El presente trabajo evalúa el desempeño de Earthquake Transformer (Mousavi, Ellsworth, Zhu, Chuang, y Beroza, 2020), un modelo de código abierto basado en aprendizaje profundo, diseñado para detectar sismos y seleccionar fases sísmicas P y S de forma simultánea. Earthquake Transformer (EQT) se aplicó en datos de tres secuencias sísmicas distintas: un enjambre sísmico en las cercanías de Coalcomán, Michoacán (2017); otro en localidades de Los Cabos, Baja California Sur (2023); y un sismo M_W 7.2 y parte de sus réplicas al sur de Petatlán, Guerrero (2014). En cada secuencia se revisaron las predicciones del modelo y se realizó una evaluación cuantitativa del desempeño utilizando métricas estadísticas. El objetivo fue analizar el rendimiento general de EQT en datos sísmicos no incluidos en su conjunto de entrenamiento, contribuyendo al procesamiento de datos sísmicos mediante enfoques de aprendizaje profundo.

Los resultados mostraron altos porcentajes de predicciones correctas con respecto a los catálogos del SSN y CENAPRED, lo que se reflejó en valores altos de sensibilidad, indicando su capacidad para minimizar los falsos negativos. Sin embargo, la alta tasa de falsos positivos representó un desafío significativo, disminuyendo la exactitud y precisión. Por ende, es sumamente recomendable completar el procesamiento mediante el uso de un mayor número de estaciones y utilizando algoritmos de asociación de fases y localización epicentral.

Abstract

Seismology has experienced an accelerating growth due to the increase in cost-effective equipment capable of generating large data volumes. This continuing expansion of seismological data volumes demands new, efficient processing tools. In automatic earthquake detection methods, the state-of-the-art methods are machine learning and deep learning-based.

This study evaluate the performance of Earthquake Transformer (Mousavi, Ellsworth, Zhu, Chuang, and Beroza, 2020), a deep learning, open source earthquake signal detector and phase (P and S) picker based on attention mechanism. Applying Earthquake Transformer to three different seismic sequences: a 2017 earthquake swarm near Coalcomán, Michoacán, another occurred in 2023 in the area of Los Cabos, Baja California Sur; and a subset of the 2014 M_W 7.2 Petatlán, Guerrero aftershock sequence. All of the predictions were manually reviewed and quantitative evaluation metrics were computed. The overall aim was to analyze the model's predictive performance on datasets not included in its training set, in order to provide insights on deep learning approaches for seismic data processing.

The model's recall is supported by the high number of correctly predicted events that actually belong to the SSN and CENAPRED manual catalogs. However, false positive rate posed a major challenge, reducing accuracy and precision. To address this challenge, it is highly recommend the use of more seismic stations and employing phase association and location algorithms.

Índice general

1. Introducción	14
1.1. Antecedentes	14
1.2. Justificación	15
1.3. Objetivo general	16
1.4. Objetivos específicos	16
1.5. Reconocimiento institucional	16
2. Fundamentos teóricos de sismología	17
2.1. Ondas sísmicas	17
2.2. Fases sísmicas	18
2.3. Localización de un sismo	21
2.3.1. Métodos determinísticos	23
2.3.1.1. Métodos linealizados iterativos	24
2.3.1.2. Métodos no lineales	26
2.4. Estimación de magnitud de un sismo	27
2.5. Estadística de sismos	30
2.5.1. Ley de Omori	31
2.5.2. Ley de Gutenberg-Richter	32
2.5.3. Enjambres sísmicos	34
2.6. Métodos para la detección de señales sísmicas	35
2.6.1. STA/LTA	36

2.6.2. Correlación cruzada de formas de onda	37
2.7. Tectónica en la Costa del Pacífico Mexicano	38
2.7.1. Sismicidad en la zona de subducción	39
2.7.1.1. La Brecha de Guerrero	39
2.8. Sismotectónica en Baja California Sur	41
3. Fundamentos teóricos de aprendizaje automático	43
3.1. Aprendizaje supervisado	44
3.2. Aprendizaje no supervisado	45
3.3. Aprendizaje por refuerzo	45
3.4. Aprendizaje profundo	46
3.4.1. Perceptrones	46
3.4.2. Algoritmo de propagación reversa	48
3.4.3. Redes Neuronales Profundas	53
3.4.4. Redes Neuronales Convolucionales	54
3.4.5. Redes Neuronales Recurrentes	55
3.4.6. Redes LSTM	56
3.5. Redes Neuronales Transformadoras	57
3.5.1. Atención	58
3.5.2. Transformación	58
3.5.3. Datos de entrada	59
3.5.4. Mecanismo de atención	59
3.5.5. Atención de producto punto escalado	60
3.5.6. Atención de múltiples cabezas	62
3.5.7. Arquitectura codificador-decodificador	64
3.6. Aprendizaje profundo y sismología	66
3.7. Earthquake Transformer	67
3.7.1. Arquitectura	68
3.7.2. Datos y entrenamiento	70

3.7.3. Mecanismo de atención	70
4. Metodología	72
4.1. Earthquake Transformer	72
4.2. Diseño experimental	74
4.3. Obtención y características de los datos	76
4.3.1. Secuencia sísmica del 18 al 22 de agosto de 2017, Michoacán	76
4.3.2. Secuencia sísmica del 8 al 9 de agosto de 2023, Baja California Sur	77
4.3.3. Secuencia de réplicas, sismo M_w 7.2 del 18 de abril de 2014, Guerrero	78
4.4. Métricas de evaluación	79
5. Resultados y discusión	80
5.1. Secuencia sísmica del 18 al 22 de agosto de 2017, Michoacán	80
5.2. Secuencia sísmica del 8 al 9 de agosto de 2023, Baja California Sur	83
5.3. Secuencia de réplicas, sismo M_w 7.2 del 18 de abril de 2014, Guerrero	85
5.4. Discusión	87
6. Conclusiones y recomendaciones	90
A. Ejemplos de detecciones	91
Referencias	101

Índice de figuras

2.1. (A) Propagación de onda P armónica y (B) onda S (Haldar, 2018).	18
2.2. Posibles geometrías de rayos en la corteza y su nomenclatura estandarizada (Shearer, 2019).	20
2.3. Sismograma de tres componentes de un sismo de profundidad intermedia registrado en Kazajistán, alrededor de 100° de distancia de la fuente (distancias angulares superiores a 30° se considera telesismo). En este registro se pueden observar múltiples fases sísmicas profundas que son resultado de la interacción de las ondas con las discontinuidades del núcleo, manto y corteza (Kennett, 2007).	21
2.4. Decaimiento del número de réplicas a través del tiempo (Guglielmi, 2017).	32
2.5. Dos representaciones del modelo Gutenberg-Richter (Žalohar, 2018).	33
2.6. Tectónica de la costa oeste del Pacífico en México. Las flechas indican la dirección de movimiento de las placas. Los puntos representan sismos ocurridos entre 1974 y 2014 registrados por el Servicio Sismológico Nacional. En color rojo se indican sismos con menos de 50 km de profundidad, con morado aquellos sismos profundos (Plata-Martínez, 2015).	40
2.7. Sistema de fallas a lo largo de la península de Baja California Sur (Plattner et al., 2007).	42
3.1. Modelo básico de un perceptrón simple k y sus elementos principales (Haykin, S., 2009).	48
3.2. Red neuronal de avance hacia adelante totalmente conectada con una capa oculta.	49

3.3. Representación visual con los conceptos básicos del principio de descenso por gradiente.	51
3.4. Esquema de un ejemplo muy simple de propagación reversa (<i>Interactive Chaos</i> , s.f). 53	
3.5. Red Neuronal Convolutiva para el procesamiento de imágenes tales como reconocimiento de objetos (<i>MathWorks</i> , s.f).	55
3.6. Secuencia temporal de una Red Neuronal Recurrente (LeCun, Y., Bengio, Y., y Hinton, G., 2015).	56
3.7. La unidad g funciona como una compuerta para capturar la información (Alpaydin, 2021).	57
3.8. Representación visual de la operación de matrices en una capa de atención de producto punto escalado para generar una matriz de salida (Bishop y Bishop, 2023). 61	
3.9. Flujo de información en una capa de atención. Las etapas de multiplicación de matrices se indican en las cajas nombradas como <i>matmul</i> , así como el reescalamiento y normalización en las cajas <i>scale</i> y <i>softmax</i> (Bishop y Bishop, 2023)	62
3.10. Representación de la operación de matrices de la ecuación 3.19 (Bishop y Bishop, 2023).	63
3.11. Representación visual de las etapas en una capa de atención con <i>h</i> número de cabezas concatenadas y llevadas a una transformación lineal (Vaswani et al. 2017). 64	
3.12. Arquitectura del modelo transformer original por Vaswani et al. (2017).	65
3.13. Representación simplificada de la arquitectura de la red neuronal profunda de EQT. Adaptada de <i>Automatic earthquake detection and phase picking in Muong Te, Lai Chau region</i> (Cong et al., 2022).	69
3.14. Visualización de los pesos de atención. El panel a representa la salida de las capas <i>transformer</i> y de atención global, b y c son resultado de las capas de atención local dedicadas a la identificación de cada fase (Mousavi et al., 2020).	71

4.1. Ejemplo de una forma de onda en las tres componentes con un evento detectado. Se utiliza la Transformada de Tiempo Corto de Fourier (STFT) para mostrar la señal en el dominio de la frecuencia. El ejemplo corresponde a un evento de magnitud 3.1 de acuerdo con el catálogo del SSN. EQT realizó la detección con un 0.98 de probabilidad, la fase P se seleccionó con 0.82 de probabilidad y una relación señal/ruido estimada en 24.9 dB, la fase S se seleccionó con 0.61 de probabilidad y una relación señal/ruido de 10.4 dB.	75
5.1. Histogramas de probabilidades de los 433 casos Verdaderos Positivos (2017). . . .	83
5.2. Histogramas de la relación señal/ruido de los 433 casos Verdaderos Positivos (2017). . . .	83
5.3. Histogramas de probabilidades de los 81 casos Verdaderos Positivos (2023).	84
5.4. Histogramas de la relación señal/ruido de los 81 casos Verdaderos Positivos (2023). . . .	85
5.5. Histograma de los tiempos de diferencia medidos para el arribo de la fase P.	86
5.6. Histogramas de probabilidades de los 503 casos Verdaderos Positivos (2014). . . .	87
5.7. Histogramas de la relación señal/ruido de los 503 casos Verdaderos Positivos (2014). . . .	88
A.1. Ejemplo representativo de un evento VP. De acuerdo con información del catálogo del SSN, es un sismo de magnitud 3.8. Se detectó con 0.93 de probabilidad. La fase P se seleccionó con 0.82 de probabilidad y se estimó una relación señal/ruido de 10.1 dB. La fase S se seleccionó con 0.81 de probabilidad y se estimó una relación señal/ruido de 15.8 dB.	92
A.2. Evento clasificado como FP. El modelo logró detectarlo con 0.79 de probabilidad. Se descartó la fase S, mientras que la fase P tiene probabilidad de 0.6 y relación señal/ruido de 6.4 dB.	93

A.3. Ejemplo de una ventana que contiene dos eventos clasificados como FN y VP, respectivamente. El primer evento fue clasificado como FN, ya que no hay datos sobre el arribo de la fase S. La probabilidad de detección es 0.91. La fase P tiene una probabilidad de 0.86 y se calculó una relación señal/ruido de 27.3 dB. El segundo evento fue confirmado por el catálogo con una magnitud de 3.5. Se detectó con 0.97 de probabilidad. La fase P se seleccionó con 0.83 de probabilidad y se estimó una relación señal/ruido de 27.3 dB. La fase S se seleccionó con 0.76 de probabilidad y no se estimó una relación señal/ruido.	94
A.4. Ejemplo de una ventana que contiene dos eventos clasificados como VP. En el primer caso, la probabilidad de detección es 0.94; con probabilidad de selección de fases (P): 0.92, (S): 0.88. El segundo evento se detectó con 0.92 de probabilidad; con probabilidad de selección de fases (P): 0.87, (S): 0.89., respectivamente. En ambos casos, relación señal/ruido (P): 20.9 dB, (S): 2.8 dB.	95
A.5. Ejemplo de evento no sísmico, se clasificó como FP y no se dispone de datos sobre el arribo de la fase P. El modelo lo detectó con una probabilidad de 0.16. La llegada de la fase S se determinó con probabilidad de 0.08, y la relación señal/ruido se estimó en 0.1 dB.	96
A.6. Ejemplo de un evento verdadero que fue clasificado como FN porque el modelo seleccionó únicamente la fase P. Tiene una probabilidad de detección de 0.4 y probabilidad de selección de fase (P): 0.45 con relación señal/ruido de 7.2 dB. . . .	97
A.7. Dos eventos que el modelo detectó correctamente (VP). El catálogo indica que sus magnitudes son 2.7 y 2.6, respectivamente. En el primer evento, la probabilidad de detección es 0.96; probabilidad de selección de fases (P): 0.74, (S): 0.77; relación señal/ruido (P): 9.5 dB, (S): 12.6 dB. El segundo evento se detectó con 0.98 de probabilidad; con probabilidad de selección de fases (P): 0.78, (S): 0.65; y relación señal/ruido (P): 10.1 dB, (S): 12.7 dB.	98

A.8. Ejemplo de dos detecciones incorrectas (FP). En el primer caso, la probabilidad de detección es de 0.65, sólo se seleccionó la fase P con 0.49 de probabilidad y un valor de 8.3 dB en la relación señal/ruido. El segundo evento tiene como probabilidad de detección: 0.7; probabilidad de selección de fases (P): 0.42, (S): 0.68; relación señal/ruido (P): 8.3 dB. 99

A.9. Ejemplo de un evento clasificado como FN. No cuenta con información sobre la llegada de la fase S. Probabilidad de detección: 0.9; probabilidad de selección de fase (P): 0.76, relación señal/ruido (P): 13.8 dB. 100

Índice de tablas

4.1. Parámetros de Earthquake Transformer necesarios para realizar la detección y selección.	73
4.2. Parámetros utilizados para la predicción con la estación MMIG (Michoacán, 2017).	77
4.3. Parámetros utilizados para la predicción con la estación MIRA (BCS, 2023).	78
4.4. Tabla 5.7 Parámetros utilizados para la predicción con la estación ZIIG (Guerrero, 2014).	78
5.1. Matriz de confusión obtenida para la secuencia de Michoacán en 2017.	81
5.2. Valores de métricas obtenidos para la secuencia de Michoacán en 2017.	82
5.3. Matriz de confusión obtenida para la secuencia de BCS en 2023.	84
5.4. Valores de métricas obtenidos para la secuencia de BCS en 2023.	84
5.5. Matriz de confusión obtenida para la secuencia de BCS en 2023.	87
5.6. Valores de métricas obtenidos para la secuencia de réplicas, Guerrero 2014.	87
5.7. Métricas obtenidas en las tres pruebas.	89
5.8. Duración de la traza procesada, número total de predicciones realizadas y tiempo de procesamiento que el modelo requirió para completar esas predicciones.	89

Capítulo 1

Introducción

1.1. Antecedentes

Un sismograma constituye el registro en tiempo del movimiento del terreno, medido por un instrumento sísmico en función del desplazamiento, velocidad o aceleración. La interpretación de un sismograma implica una secuencia de procesamiento y análisis, requiriendo conocimientos sólidos en fuente sísmica, estructura de la Tierra, propagación de ondas, patrones en la sismicidad, entre otros. En la década de los 80, la sismología se revolucionó con la incorporación de computadoras e instrumentos digitales. Surgieron sismómetros digitales de banda ancha con amplios rangos dinámicos capaces de generar grandes volúmenes de datos que podían procesarse en un tiempo relativamente corto ([Kulhanek y Persson, 2011](#)).

Recientemente, el desarrollo de sensores más accesibles y tecnologías emergentes en el registro del movimiento del terreno ha provocado un aumento exponencial en la cantidad de datos sísmicos. Este crecimiento demanda una nueva generación de herramientas de procesamiento. Identificar ondas sísmicas y determinar tiempos de arribo de las fases P y S en registros continuos con presencia de ruido, son tareas fundamentales en sismología que demandan tiempo y experiencia en el procesamiento manual por parte de los analistas.

Por ende, han surgido técnicas automáticas para la detección de señales sísmicas. En esta línea, los métodos más difundidos se basan en la amplitud y contenido de energía, como el

método STA/LTA (Allen, 1978), en la similitud de la forma de onda, como el método de correlación cruzada o incluso la técnica FAST (Yoon, O'Reilly, Bergen, y Beroza, 2015), un algoritmo para identificar sismos similares a partir de patrones. En la vanguardia de los métodos automáticos se encuentran los enfoques basados en aprendizaje automático (*Machine Learning*) (Jiang, Fang, Fan, y Li, 2021), que han evolucionado hacia el aprendizaje profundo (*Deep Learning*). La densidad de las bases de datos sismológicas y la necesidad de analizarlos e interpretarlos han impulsado la aplicación del aprendizaje profundo.

Estos modelos son particularmente útiles para procesar datos de alta dimensión, como los registros sísmicos, debido a su capacidad para extraer automáticamente características a partir de los datos observados, aprender relaciones no lineales entre las entradas y salidas, procesar datos secuenciales y reducir la dimensionalidad del problema. La implementación de métodos basados en aprendizaje profundo ha proliferado en el monitoreo sísmico, abordando las tareas de rutina que incluyen la detección y clasificación de eventos, selección de fases sísmicas, determinación de la polaridad del primer arribo, y asociación de fases. Además, pueden orientarse al procesamiento e interpretación de imágenes sísmicas, así como el modelado directo e inverso (Mousavi y Beroza, 2022).

1.2. Justificación

La disponibilidad de extensos catálogos de sismos ha proporcionado los conjuntos de datos etiquetados ideales para entrenar algoritmos de aprendizaje profundo diseñados para detectar ondas sísmicas y seleccionar tiempos de arribo de las fases. Esto ha impactado en el desarrollo de catálogos sísmicos de alta resolución. Estos nuevos catálogos ofrecen eficiencia y precisión, son capaces de contener un mayor número de eventos que los catálogos convencionales y además proporcionar picados de fases tan precisos como los de un analista experimentado. No obstante, la generación de modelos confiables y la interpretación de resultados obtenidos a través de aprendizaje profundo son algunos de los principales desafíos. En esta línea, es necesario explorar en estos métodos de vanguardia, medir la confiabilidad de los modelos, comprender el funcionamiento e interpretar los resultados, así como buscar áreas para mejoras futuras.

1.3. Objetivo general

Aplicar un modelo de código abierto, basado en aprendizaje profundo, denominado *Earthquake Transformer* (EQT), para evaluar su capacidad en la detección de sismos y la selección de fases sísmicas P y S de forma simultánea.

1.4. Objetivos específicos

1. Seleccionar secuencias sísmicas en México con señales continuas de tres componentes registradas en estaciones locales, respaldadas por un reporte o catálogo de sismos que valide el número de eventos.
2. Implementar EQT en los datos en una estación y analizar el impacto de la calibración de parámetros en los resultados de cada secuencia, buscando optimizar el rendimiento.
3. Realizar una revisión exhaustiva de las predicciones de EQT, clasificándolas en matrices de confusión y calculando métricas como exactitud, precisión, sensibilidad y puntuación F1 para una evaluación cuantitativa del desempeño del modelo.
4. Comparar y analizar el rendimiento de EQT en cada prueba, identificando posibles variaciones y evaluando la eficiencia del modelo en diferentes escenarios sísmicos.

1.5. Reconocimiento institucional

El catálogo de sismos del Servicio Sismológico Nacional es posible gracias a todo su personal y producto de los cálculos realizados por su área de Análisis e Interpretación de Datos Sísmicos.

Se agradece a la Dirección de Instrumentación y Cómputo del Centro Nacional de Prevención de Desastres por proporcionar los datos sísmicos solicitados.

Capítulo 2

Fundamentos teóricos de sismología

2.1. Ondas sísmicas

Un sismo es el resultado de una liberación súbita de energía acumulada en la litósfera de la Tierra. Esta energía es resultado de la acumulación de tensión entre las placas tectónicas de la litósfera, secciones rígidas de roca que se desplazan sobre las rocas deformables de la astenósfera. El calor de las rocas en la astenósfera crea corrientes de convección que hacen que las placas tectónicas interactúen entre sí a través de procesos que generan esfuerzos, como deformación, fricción, subducción, entre otros. Con el tiempo, estos esfuerzos superan la resistencia elástica de las rocas, causando una ruptura a lo largo de un plano de falla ([Kulhánek, 2012](#)).

A medida que la ruptura se propaga, varios tipos de ondas sísmicas se emiten desde el foco del sismo. Los dos tipos principales de ondas sísmicas son las ondas de cuerpo y las ondas de superficie. Las ondas de cuerpo son llamadas así porque se propagan a través del interior de la Tierra, son las primeras en liberarse después de un sismo. Por otro lado, las ondas de superficie se generan por la interacción de las ondas de cuerpo con discontinuidades del medio como la superficie terrestre y se propagan a lo largo de la superficie libre de la Tierra o a lo largo de otras discontinuidades en su interior.

Las ondas de cuerpo se dividen en dos tipos: ondas de compresión y ondas de corte. Las ondas de compresión, también llamadas ondas longitudinales u ondas primarias (P), se propagan

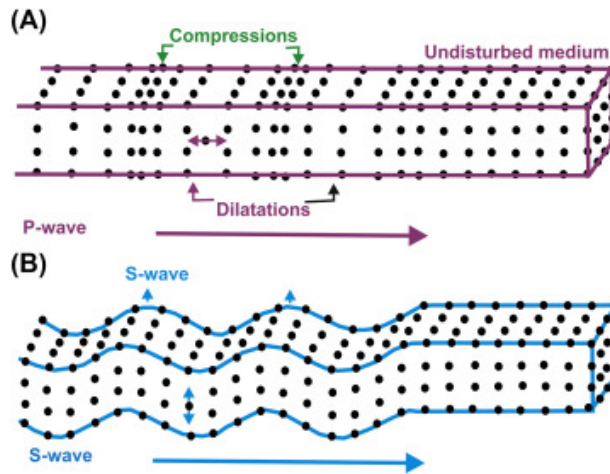


Figura 2.1: (A) Propagación de onda P armónica y (B) onda S (Haldar, 2018).

mediante compresión y dilatación en la dirección de su propagación a través del medio en el que se desplazan. La Figura 2.1 ilustra el movimiento de partículas en una onda P armónica. Estas ondas producen cambios en el volumen del material.

Las ondas de corte, también llamadas ondas de cizalla u ondas secundarias (S), llegan después de las ondas P pero antes que las ondas de superficie. Estas ondas tienen un movimiento de corte, sin cambios en el volumen del material. El movimiento de las partículas es perpendicular a la dirección de propagación de la onda. El desplazamiento de las ondas de corte se describe en dos polarizaciones ortogonales: con movimiento en el plano vertical al vector de propagación (ondas SV), como se ilustra en la Figura 2.1 para una onda S armónica, y con movimiento horizontal en la dirección perpendicular a este plano (ondas SH) (Shearer, 2019). El esfuerzo de corte no es posible en fluidos, las ondas S solo pueden propagarse a través de las capas sólidas.

Las diferencias de polarización producen que las ondas P se observen con mayor claridad en la componente vertical de los sismogramas, mientras que las ondas S tienden a observarse en las componentes horizontales de dichos sismogramas.

2.2. Fases sísmicas

En comparación con las ondas superficiales, las ondas de cuerpo resultan más afectadas por fenómenos como la refracción, la reflexión, la dispersión y la conversión de modos que ocurren

en las discontinuidades debido a los contrastes de impedancia entre las diferentes capas de la Tierra. Esto da lugar a una variedad de posibles geometrías de rayos, conocidas como fases sísmicas, que pueden ser identificadas en un sismograma.

A diferencia de las ondas superficiales, las ondas de cuerpo no presentan dispersión en el rango de frecuencias considerado, lo que facilita la observación de sus fases a diferentes distancias: local, regional y telesísmica. La visibilidad de las fases depende de su amplitud, polarización y contenido de frecuencias. Estas características que ayudan a identificar a las fases sísmicas derivan de las propiedades geométricas y físicas de las capas terrestres que atraviesan las ondas. Por lo tanto, es necesario identificarlas adecuadamente para el análisis de datos que se realiza en observatorios sismológicos. Un análisis correcto de las fases sísmicas aporta la precisión y exactitud indispensables en la localización de los sismos (Bormann P, 2012).

En el caso de los sismos registrados a distancias locales y regionales, las ondas sísmicas se propagan desde la fuente a través de la corteza y el manto superior que le subyace. Por lo tanto, un modelo aproximado que considera una capa horizontal sobre el semiespacio suele ser suficiente en la práctica. La Figura 2.2 representa un modelo global y simplificado con la discontinuidad de Mohorovičić, conocida como Moho, en la base de la corteza terrestre.

Las fases sísmicas esperadas a nivel local y regional reciben una nomenclatura especial: las ondas P y S que viajan a través de la estructura cortical superior se llaman P_g y S_g (o simplemente P y S), respectivamente. Por otro lado, las ondas P_b y S_b (o P^* y S^*) se desplazan por la capa inferior de la corteza. Las ondas P_mP y S_mS son ondas reflejadas en el Moho, mientras que P_n y S_n son ondas refractadas a lo largo del Moho.

La corteza oceánica tiene un espesor promedio de 6 km y un rango de 30 a 50 km de espesor para la corteza continental (Shearer, 2019). Esta diferencia en espesor conlleva a la dispersión de las ondas de cuerpo en el rango de frecuencias utilizado para detectar sismos cercanos (de 0.5 a 50 Hz). Como consecuencia, generalmente los sismogramas de estaciones en distancias locales y regionales presentan primeras fases seguidas de ruido generado por la señal o por ondas de coda, dificultando la observación de fases sísmicas posteriores que han sido reflejadas o refractadas en discontinuidades intra-corticales más débiles (Bormann P, 2012).

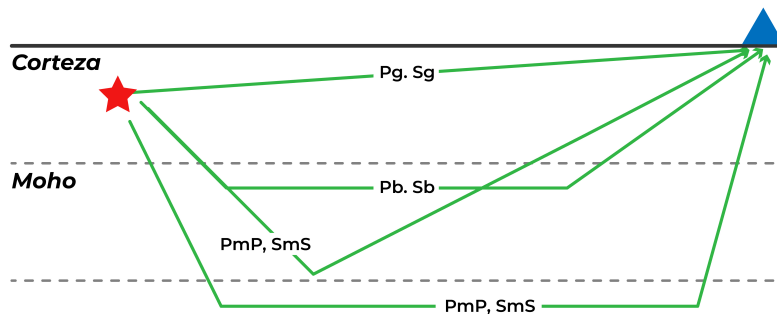


Figura 2.2: Posibles geometrías de rayos en la corteza y su nomenclatura estandarizada (Shearer, 2019).

Otro aspecto importante es que la velocidad de las ondas sísmicas aumenta significativamente en la discontinuidad de Moho, en un orden cercano al 20%. Esto facilita la identificación tanto de las primeras fases como de las fases posteriores, ya que logran ser detectadas a pesar del nivel de ruido ambiental o del ruido generado por la señal misma (Bormann P, 2012).

El suelo responde a las ondas de cuerpo y de superficie moviéndose en todas las direcciones. En la actualidad, los sismógrafos registran las tres componentes de movimiento del suelo en un amplio rango de frecuencias mediante dos sensores horizontales ortogonales y uno vertical. Esto produce tres señales, donde las ondas P suelen identificarse fácilmente en la componente vertical, mientras que las ondas S se registran con mayor amplitud en las componentes horizontales. La Figura 2.3 ilustra un sismograma que muestra diferentes tipos de fases y su nomenclatura, las cuales se reflejan con diferentes niveles de energía según la componente (Kennett, 2007).

En el análisis e interpretación de un sismograma es fundamental determinar el tiempo de llegada de las fases, es decir, medir el momento en que se produce el primer movimiento discernible asociado a una fase sísmica. Una de las operaciones rutinarias de los centros de información sismológica es registrar y reportar los tiempos de arribo de fases de manera eficiente, ya sea por analistas o mediante algún método automático, para lograr localizar sismos. Además, el análisis de la diferencia entre los tiempos de arribo observados y las predicciones basadas en

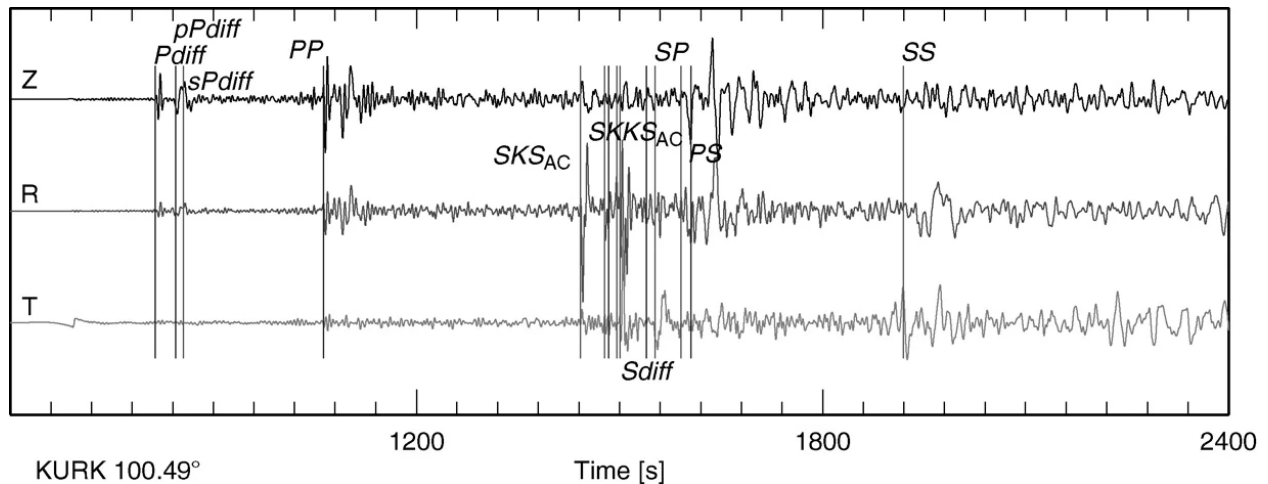


Figura 2.3: Sismograma de tres componentes de un sismo de profundidad intermedia registrado en Kazajistán, alrededor de 100° de distancia de la fuente (distancias angulares superiores a 30° se considera telesismo). En este registro se pueden observar múltiples fases sísmicas profundas que son resultado de la interacción de las ondas con las discontinuidades del núcleo, manto y corteza (Kennett, 2007).

modelos de velocidad, permitió a los especialistas elaborar curvas de tiempo de viaje para las fases principales, lo que a su vez permite inferir la estructura de velocidad radial promedio de la Tierra (Bormann P, 2012).

2.3. Localización de un sismo

La localización en espacio y tiempo, así como la determinación de magnitud son fundamentales en la caracterización inicial de un sismo. La localización de un sismo es útil en la evaluación de peligro sísmico y en la respuesta ante emergencias porque permite identificar áreas que podrían resultar afectadas y ayuda a determinar la región de generación de un posible tsunami. Además, es útil para calcular la magnitud o el momento sísmico. Conocer la localización de un evento sísmico es indispensable en estudios de sismicidad, en la caracterización de los cambios en esfuerzo y deformación producidos por el sismo, en la estimación del peligro, también en estudios de sismología básica y aplicada (Lomax, Michelini, Curtis, y Meyers, 2009).

Un sismo tiene origen en un instante determinado del tiempo, en un punto a profundidad llamado foco o hipocentro. El procedimiento para localizar un sismo es un problema inverso

no lineal. Estimar la posición del hipocentro es un problema inverso que implica determinar el tiempo origen del evento mediante el cálculo de los tiempos de viaje y los tiempos de arribo, los cuales son problemas directos. El tiempo de viaje se refiere al tiempo que le toma a una onda sísmica propagarse entre dos puntos, la fuente y el receptor, en un medio a lo largo de una trayectoria, mientras que los tiempos de arribo representan la primera medición de energía de una fase sísmica que aparece en un sismograma.

Se considera que la localización de un sismo es un problema inverso no lineal porque los tiempos de arribo de las ondas sísmicas tienen una relación no lineal con las variaciones en la posición del hipocentro y con el modelo de velocidades. A partir del modelo de velocidades se calculan predicciones de los tiempos de arribo con base en la velocidad con la que viajan las ondas entre la fuente y las estaciones sísmicas. Habitualmente, los algoritmos de localización emplean modelos de velocidades de una dimensión (1D), donde los cambios de velocidad se presentan únicamente en profundidad. Descartar variaciones laterales de la velocidad es útil principalmente en estudios locales o regionales. A escala global, se emplean modelos de velocidades con estructuras más complejas en 2D o 3D, con el objetivo de representar con mayor precisión los cambios de velocidad al interior de la Tierra, a menudo estos enfoques no son eficientes y por esa razón aún predominan los modelos de una dimensión.

En la revisión de ([Karasözen y Karasözen, 2020](#)) los métodos de localización de sismos se clasifican en función de (1) la precisión con la que se describe el espacio, (2) del número de eventos a localizar, y (3) del método para resolver el problema inverso de localización. (1) La localización absoluta indica la posición dentro de un sistema coordenado fijo y un estándar de tiempo (comúnmente UTC), mientras que la localización relativa se da en referencia a otro punto espacio-temporal. (2) Los métodos pueden ser aplicados para localizar un solo evento, o bien, para localizar simultáneamente múltiples eventos. (3) Los métodos principales para abordar el problema de localización son de tipo lineal, no lineal y probabilístico.

2.3.1. Métodos determinísticos

La localización de un sismo es un problema de cuatro incógnitas: las coordenadas (x_0, y_0, z_0) del hipocentro y el tiempo de origen t_0 . Para ello, se utiliza un modelo de velocidades de ondas sísmicas, que puede ser seleccionado entre los propuestos para la región de estudio o estimado específicamente. Este modelo permite calcular las trayectorias de propagación y, a partir de ellas, hacer predicciones de los tiempos de viaje. La localización del evento se determina minimizando la diferencia entre los tiempos de llegada observados y los estimados.

Si se asume que el medio es homogéneo y tiene una velocidad v conocida y constante para cada tipo de onda sísmica (P o S), y una lentitud u definida como el inverso $u = 1/v$, entonces cada fase sísmica tiene un tiempo de llegada t_{obs} , que se mide en estaciones localizadas en la posición $\mathbf{x} = (x_{obs}, y_{obs}, z_{obs})$, estas observaciones arriban al receptor después del instante en que se originó t_0 , y desde la localización de la fuente en el punto $\mathbf{x}_0 = (x_0, y_0, z_0)$,

$$t_{obs} = t_0 + u[(x_{obs} - x_0)^2 + (y_{obs} - y_0)^2 + (z_{obs} - z_0)^2]^{\frac{1}{2}}. \quad (2.1)$$

La expresión anterior cambia para un medio no homogéneo donde se presenta variación en el espacio de la velocidad $v(\mathbf{x})$, y por tanto de la lentitud $u(\mathbf{x}) = 1/v(\mathbf{x})$,

$$t_{obs} = t_0 + \int_{\mathbf{r}_0(s)} u(\mathbf{r}_0) d(s), \quad (2.2)$$

donde $\mathbf{r}_0(s)$ representa un punto posicionado a distancia s a lo largo del trayecto \mathbf{r}_0 entre la fuente y el receptor. La ecuación 2.2 es no lineal porque cualquier cambio en la localización de la fuente afecta la trayectoria \mathbf{r}_0 con la que se define la integral.

Las variables del problema se definen como los parámetros del modelo $\mathbf{m} = (x, y, z, t)^T$, y los datos como un vector de N tiempos de arribo registrados, $\mathbf{d}_{obs} = (t_{obs}^1, \dots, t_{obs}^N)^T$. Con base en las ecuaciones 2.1 y 2.2 se plantea el problema directo, que consiste en una función no lineal $g(\mathbf{m})$ para obtener los datos observados a partir de proponer un modelo conocido,

$$\mathbf{d}_{obs} = g(\mathbf{m}). \quad (2.3)$$

2.3.1.1. Métodos linealizados iterativos

Por otro lado, la solución de un problema inverso implica que dadas las observaciones, se deben encontrar los parámetros de mejor ajuste. Para lo cual se propone un modelo inicial, $\mathbf{m}^o = (x_0, y_0, z_0, t_0)^T$, que es una primera estimación de la solución y puede estar basado en información a priori del problema.

Mediante aproximaciones sucesivas de la forma

$$\mathbf{m}_1 = \mathbf{m}_0 + \delta\mathbf{m}_0, \quad (2.4)$$

los parámetros se actualizan en cada iteración hasta ser más cercanos a la solución.

Para obtener un valor aproximado de la función 2.3 se puede recurrir a su representación como expansión en series de Taylor manteniendo sólo el término lineal

$$g(\mathbf{m}_1 = g(\mathbf{m}_0 + \delta\mathbf{m}_0) \approx g(\mathbf{m}_0) + \left. \frac{\partial g}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} \delta\mathbf{m}_0, \quad (2.5)$$

de donde el primer término es la estimación para la primera iteración, la cual parte de una estimación inicial \mathbf{m}_0 y g es la función para obtener los datos observados de la ecuación 2.3. Así, se define un residual como la diferencia entre los valores observados y los estimados

$$\mathbf{d}_{obs} = \mathbf{d}_{calc} + \delta\mathbf{d}. \quad (2.6)$$

Combinando las ecuaciones 2.3, 2.5 y 2.6 se tiene que

$$\delta\mathbf{d} = \left. \frac{\partial g}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} \delta\mathbf{m}_0, \quad (2.7)$$

donde las derivadas parciales de la función g con respecto a los parámetros del modelo, \mathbf{m} , pueden escribirse en la forma matricial G , o kernel, de dimensiones i, j , donde i va de 1 al N

número de datos observados y j de 1 a 4, que es el número de parámetros establecidos,

$$G_{i,j} = \frac{\partial g_i}{\partial \mathbf{m}_j}.$$

Convirtiendo a la ecuación 2.7 en un producto matricial

$$\delta \mathbf{d} = G \delta \mathbf{m}. \quad (2.8)$$

La ecuación 2.8 representa un sistema de ecuaciones lineales. La solución es determinar el cambio en el modelo $\delta \mathbf{m}$, que al multiplicarse por el kernel G , produzca el cambio esperado $\delta \mathbf{d}$ en los datos. Esta combinación lineal corresponde a un problema inverso que, contrario al problema directo, busca determinar los parámetros a partir de los datos observados.

En términos de notación, la matriz de derivadas parciales G , tiene i filas, que corresponde al número de datos observados y j columnas que es igual al número de parámetros. En la práctica, se cuenta con un gran número de tiempos observados, conformando un vector que contiene datos de múltiples estaciones y que por lo tanto supera en tamaño al vector de parámetros,

$$\begin{pmatrix} \delta d_1 \\ \delta d_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta d_N \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ G_{N1} & G_{N2} & G_{N3} & G_{N4} \end{pmatrix} \begin{pmatrix} \delta m_1 \\ \delta m_2 \\ \delta m_3 \\ \delta m_4 \end{pmatrix}.$$

$G \in \mathbb{R}^{N \times 4}$ indica que se trata de un sistema sobredeterminado, lo cual dificulta hallar parámetros \mathbf{m} que cumplan con la ecuación 2.8. Sobre esta ecuación, es común que sea escrita como $\mathbf{d} = G\mathbf{m}$, para hacer la notación más simple. En el problema inverso, se requieren parámetros \mathbf{m} que minimicen el valor de $\|G\delta \mathbf{m} - \delta \mathbf{d}\|^2$, que es la solución general al problema por mínimos

cuadrados. Tal solución está dada por el inverso izquierdo de la matriz,

$$\delta \mathbf{m}_{est} = (G^T G)^{-1} G^T \delta \mathbf{d}. \quad (2.9)$$

2.3.1.2. Métodos no lineales

El método de mínimos cuadrados es útil para solucionar un problema inverso, el método consiste en estimar los parámetros del modelo que minimizan un tipo particular de error. Este error se mide mediante la norma euclidiana de los residuales entre los datos observados, t_{obs}^i , y los datos calculados, t_{calc}^i , para N mediciones $i = 1, \dots, N$,

$$e = (t_{obs}^i - t_{calc}^i).$$

El objetivo es encontrar los parámetros del modelo que generen los datos estimados más cercanos posibles a los datos observados, reduciendo al mínimo la discrepancia entre ellos.

En métodos no lineales, el error total se define como una diferencia cuadrática, o bien, medición del error a través de la norma L_2 ,

$$E = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (t_{obs}^i - t_{calc}^i)^2. \quad (2.10)$$

La ecuación 2.10 es la solución por mínimos cuadrados y corresponde al mínimo error, una mejor aproximación de la solución óptima del problema inverso. Es común asumir que los datos tienen una distribución gaussiana y por lo tanto, la longitud del error es medida con la norma L_2 . Sin embargo, cuando existen múltiples datos fuera de la tendencia general, (*outliers*), la norma L_1 es considerada más robusta.

$$E = \sum_{i=1}^N |e_i|. \quad (2.11)$$

En una distribución de números, la norma L_2 provee el valor promedio, mientras que la norma L_1 produce la mediana. Una desventaja de la norma L_1 son las complicaciones que implica el valor absoluto en las ecuaciones (Shearer, 2019), además de que minimizar el error utilizando la norma L_1 es un problema no suavizado, lo cual no es recomendable en algoritmos

de localización estándar por el alto costo computacional (Karasözen y Karasözen, 2020). Para controlar los efectos de los valores atípicos se recomienda ponderar los residuales del problema de mínimos cuadrados a través de un método iterativo (Shearer, 2019).

2.4. Estimación de magnitud de un sismo

La magnitud, junto con la información de su localización, son parámetros esenciales para caracterizar y comprender un sismo. Es una medida logarítmica que cuantifica el tamaño de un sismo. Es ampliamente utilizada para proveer información rápida al público. Con la introducción de la escala de magnitud local de Charles Richter en 1935, múltiples estudios han propuesto una variedad de escalas de medición. Estas escalas capturan diferentes aspectos de las ondas sísmicas, como la distribución de la energía entre las bajas y altas frecuencias, o la diferencia entre ondas de cuerpo y ondas de superficie. Los métodos de las diferentes escalas se enfocan en diferentes propiedades, sin embargo son apropiadas para el rango de magnitudes y distancias epicentrales que se pueden presentar (Mousavi y Beroza, 2020).

La escala de Richter cuantifica la magnitud local M_L , la cual está definida en función de la amplitud o máximo desplazamiento de la señal registrada en un instrumento estándar de la época, el sismógrafo de torsión Wood-Anderson,

$$M_L = \log_{10} A(X) - \log_{10} A_0(X), \quad (2.12)$$

Richter originalmente definió a la magnitud como la diferencia entre el logaritmo de la amplitud máxima A , registrada durante el sismo en cuestión, medida en micras (10^{-6} m), y el logaritmo de la amplitud, $A_0(X)$, de un sismo tomado como referencia y que se encuentra a 100 km del epicentro. Ambos en función de la distancia epicentral X , en kilómetros.

Existen nuevas aproximaciones empíricas, como la propuesta por Bullen y Bolt (1985), basadas en tablas de los valores de amplitud del sismo de referencia a diferentes distancias $A_0(X)$, que Richter elaboró,

$$M_L = \log_{10} A + 2,56 \log_{10}(X) - 1,67. \quad (2.13)$$

La magnitud local de Richter M_S , es válida para sismos locales a una distancia de hasta 600 km. Una característica común a esta y a todas las escalas posteriores es su comportamiento logarítmico debido al amplio rango en el que se puede presentar la energía sísmica. En esta escala, los sismos inferiores a 3.0 se clasifican como generalmente imperceptibles, mientras que los sismos de 5.5 en adelante pueden causar daños significativos en estructuras del sur de California. Sin embargo, M_L tiene limitaciones cuando se utiliza en otras regiones, ya que se requiere calibrar la escala a otros modelos de atenuación de ondas sísmicas distintos al del sitio donde originalmente se definió.

Tanto la magnitud de ondas de cuerpo m_b , como la magnitud de ondas de superficie M_S , son ampliamente utilizadas en estudios globales de eventos telesísmicos. La magnitud de ondas de cuerpo se define como

$$m_b = \log_{10}(A/T) + Q(\Delta, h), \quad (2.14)$$

donde A es el desplazamiento del suelo, medido en micras, durante la fase de onda P, T es el periodo fundamental de la señal, Δ es la distancia epicentral medida en grados y Q es una función empírica que depende de la distancia angular y la profundidad focal. Se calcula a partir de los primeros ciclos de la fase P, donde se presentan períodos de oscilación de un segundo, en instrumentos de componente vertical, lo que la hace útil para medir sismos con epicentros alejados del sitio de registro, ya que la atenuación de la onda P es menor que la de las ondas superficiales a medida que aumenta la distancia.

La magnitud de ondas superficiales está basada en la mayor amplitud observada en estas ondas, de forma general se define como

$$M_S = \log_{10}(A/T) + 1,66 \log_{10}(\Delta) + 3,3. \quad (2.15)$$

La idea de una magnitud basada en ondas superficiales fue inicialmente propuesta por Gutenberg (1945). A partir de esto se han derivado otras ecuaciones, como la de Vanek et al. (1962) que definió M_S usando ondas de Rayleigh con períodos de oscilación de 20 ± 2 s,

$$M_S = \log_{10} A_{20} + 1,66 \log_{10}(\Delta) + 2,0. \quad (2.16)$$

En ambas expresiones, A se refiere al desplazamiento del suelo medido en micras durante la fase de la escala, T es el período fundamental de la señal y Δ la distancia en grados. A diferencia de M_L y m_b , utilizar fases de ondas superficiales permite determinar magnitudes a distancias epicentrales mayores. Sin embargo, la amplitud de las ondas superficiales decrece al aumentar la profundidad, lo que limita su uso a únicamente a sismos someros.

Un obstáculo para utilizar las magnitudes basadas en amplitud radica en que son medidas empíricas que se fundamentan en lo observado directamente del sismograma, sin considerar características de la fuente sísmica. Otra dificultad es que las diferentes escalas generalmente no coinciden en sus valores. Este fenómeno se conoce como saturación de las escalas de magnitud y ocurre a partir de valores de $m_b > 6,2$ y $M_S > 8,3$. Esto significa que las magnitudes de amplitud no podrán medir con precisión el tamaño de sismos grandes.

La amplitud de las ondas sísmicas está relacionada con momento sísmico, con el ángulo azimutal del sismómetro en relación con la geometría de la falla, así como con la distancia epicentral y la profundidad focal. Además, debido a la duración finita de la función de la fuente en el dominio del tiempo, que depende de las dimensiones de la falla y el tiempo de inicio de la ruptura, las amplitudes también muestran variaciones en función de la frecuencia. Estas variaciones de frecuencia impiden que las escalas se ajusten para todas las magnitudes (Stein y Wysession, 2003).

El momento sísmico M_0 (Aki, 1966), es una alternativa para cuantificar el tamaño de un sismo y la energía liberada durante el proceso. Este concepto ha sido ampliamente aceptado por su estrecha relación con la física de una fuente sísmica general. La expresión escalar de M_0 es

$$M_0 = \mu \bar{D} S, \quad (2.17)$$

donde μ es el módulo de cizalla, o rigidez del medio, \bar{D} es el deslizamiento promedio en el área, S , de superficie de falla. Tiene unidades de energía dina-cm o N-m.

La escala de magnitud de momento, M_W , está basada en el momento sísmico. La magnitud de momento fue desarrollada por Kanamori (1977) y por Hanks y Kanamori (1979), se propuso

como una solución al problema de saturación de escalas y se define como

$$M_W = \frac{2}{3} [\log_{10} M_0 - 9,1]. \quad (2.18)$$

La magnitud de momento es aplicable tanto a sismos pequeños como a los más grandes, sin sufrir los efectos de saturación. No obstante, la determinación de la frecuencia de esquina ¹ en sismos de baja magnitud presenta dificultades debido al alto nivel de ruido, especialmente en redes regionales. Como resultado, el cálculo de M_W se limita a eventos con magnitudes superiores a 4 que se registran a nivel local. Aunque el momento no se estima para todos los sismos, es ampliamente utilizado en la investigación sismotectónica moderna y el análisis de riesgos sísmicos, donde las magnitudes son de gran importancia (Manzunzu y cols., 2021).

El Servicio Sismológico Nacional (SSN) reporta la magnitud de la longitud de coda M_c , para sismos de $M \leq 4,5$. Esta medición de magnitud es una relación promedio de los valores individuales de todas las estaciones que registraron el evento. Es una escala empírica que utiliza datos de estaciones mexicanas, fue propuesta por Havskov y Macías (1983) y está basada en el método de Lee et al. (1972) (Cárdenas, Chávez-García, y Gusev, 1997)

$$M_c = 0,09 + 1,85 \log_{10}(T) + 0,0004(D), \quad (2.19)$$

donde T es la longitud de coda en segundos y D la distancia epicentral en kilómetros.

2.5. Estadística de sismos

Aunque los sismos son fenómenos complejos, existen leyes empíricas que explican sus propiedades estadísticas. Dos de las aplicaciones de la física estadística en sismología son la ley de Gutenberg-Richter, que describe la distribución de la magnitud de sismos, y la ley de Omori, que describe el decaimiento en frecuencia de réplicas. Estas leyes son potenciales y se mantienen invariantes en escala, surgieron del análisis de un gran número de eventos.

¹Valor de frecuencia en la que se presenta un cambio abrupto de pendiente en el espectro sísmico teórico.

2.5.1. Ley de Omori

La ley de Omori, descubierta en 1894 por el sismólogo F. Omori, representa la primera ley empírica en sismología. En orden cronológico, precede a la ley de Gutenberg-Richter. Esta ley describe el comportamiento de la aparición de réplicas, que son eventos desencadenados por sismos anteriores y que se agrupan en el tiempo y en el espacio alrededor de los sismos causales, comúnmente conocidos como sismos principales (USGS, 2020).

La ley de Omori establece que la frecuencia de las réplicas disminuye a medida que transcurre el tiempo, siguiendo una tendencia hiperbólica representada por la ecuación

$$n(t) = \frac{k}{c + t} \quad (2.20)$$

Donde $n(t)$ es la frecuencia de réplicas en función del tiempo $t \geq 0$. La constante c define el decaimiento inicial en unidades de tiempo, mientras que k es una constante que establece la tasa global y es adimensional (Aaron, 2021).

En la práctica, la ley de Omori comúnmente se aproxima a una función de potencia

$$n(t) = \frac{k}{(c + t)^p} \quad (2.21)$$

Aquí, el exponente p puede variar según la región geográfica y el escenario tectónico. Por ejemplo, observaciones de sismos en California han mostrado un parámetro p que oscila en un rango que va desde 0.5 hasta 1.5, con un valor promedio cercano a $p = 1.08$. Esta versión modificada de la ley de Omori es llamada la ley de Omori-Utsu o simplemente ley de potencia, mientras que la ecuación 2.20 recibe el nombre de ley de Omori hiperbólica.

En la Figura 2.4 se muestra un ejemplo de cómo evolucionan las réplicas. Para ese caso, se analiza un sismo principal de magnitud 6.6 ocurrido en noviembre de 1987 al sur de California. El gráfico de la relación frecuencia-tiempo muestra réplicas que se encuentran dentro de un radio de 0.5 grados de la zona y que se registraron durante 90 días después del sismo principal. Es importante destacar que estudios recientes han identificado variaciones en el período de tiempo en que se cumple la ley de Omori para diferentes fuentes sísmicas. Sin embargo, se ha

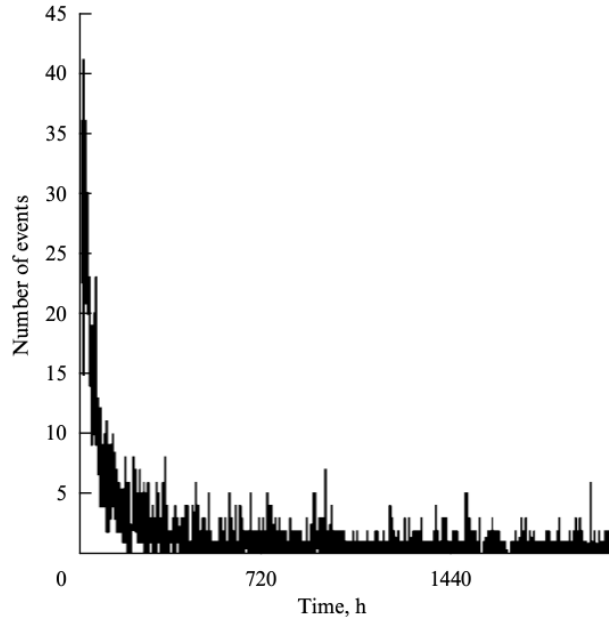


Figura 2.4: Decaimiento del número de réplicas a través del tiempo (Guglielmi, 2017).

establecido que el valor de 90 días es ampliamente aplicable a un grupo significativo de sismos principales (Guglielmi, 2017).

2.5.2. Ley de Gutenberg-Richter

Para la década de 1940, después de importantes contribuciones en el campo de la medición de la magnitud, Charles Richter en colaboración con Beno Gutenberg realizaron mediciones empíricas para cuantificar la relación entre la frecuencia de los sismos y su magnitud. Esta relación logarítmica, es conocida como la ley de Gutenberg-Richter o relación frecuencia-magnitud

$$\log N = a - bM \quad (2.22)$$

donde N es el número de sismos esperados con magnitud $\geq M$ ocurridos en un intervalo de tiempo. La relación entre variables es lineal, a y b son los parámetros que representan la intersección y la pendiente de la distribución. El valor de a es una constante que representa la actividad sísmica en una región y ventana de tiempo específicas. El parámetro b es resultado de la dependencia entre el número de sismos y la magnitud, es la pendiente de la recta y generalmente

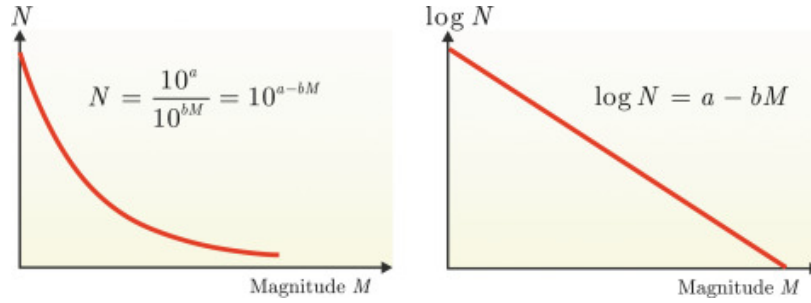


Figura 2.5: Dos representaciones del modelo Gutenberg-Richter (Žalohar, 2018).

tiene un valor de 1.

La expresión aplica para un número acumulativo de N sismos con magnitud $\geq M$, como se ilustra en la Figura 2.5. Sin embargo, la relación frecuencia-magnitud también puede escribirse para valores incrementales N_T en un rango de magnitud entre $M - \frac{1}{2}dM$ y $M + \frac{1}{2}dM$. Al despejar N de la ecuación 2.22

$$N = \frac{10^a}{10^{bM}}, \quad (2.23)$$

para luego derivar con respecto a M

$$\left| \frac{dN}{dM} \right| = N_T = \left| -b \ln 10 \left(\frac{10^a}{10^{bM}} \right) \right| \quad (2.24)$$

y aplicar el logaritmo

$$\log N_T = A - bM \quad (2.25)$$

con la nueva constante $A = b \ln 10 (10^a)$ se observa que a pesar del cambio en la intersección entre 2.22 y 2.25, el parámetro b no cambia.

La ley explica que la sucesiva disminución del tamaño de los sismos es inversamente proporcional a su ocurrencia. La frecuencia con la que se presentan sismos de menor magnitud es alrededor de 10 veces en comparación con los de mayor magnitud. Es decir, si anualmente se producen 10 sismos de $M > 7$, entonces se estima que ocurran 100 sismos de $M > 6$, 1000 sismos de $M > 5$, y así sucesivamente. El decremento logarítmico y el valor universal de $b \approx 1$ se ha observado en datos a nivel global y también en regiones sísmicas con características propias. Sin embargo, el valor del parámetro b puede fluctuar especialmente a menor escala, como es el

caso de los enjambres sísmicos, en estos fenómenos se puede presentar un valor de b de hasta 2.5, lo que significa una proporción notablemente mayor de sismos de pequeña magnitud.

2.5.3. Enjambres sísmicos

Un enjambre de sismos se refiere a un cúmulo de eventos moderados que se producen seguidamente en tiempo y espacio dentro de una región específica. No están antecidos por un sismo principal, sino son secuencias de actividad sísmica en periodos variables de tiempo. Las características generales que definen a un enjambre sísmico son:

- **Frecuencia:** La actividad sísmica de un enjambre es muy frecuente. Se trata de una cantidad anómala de sismos que ocurre en intervalos mucho más frecuentes de lo usual, pueden durar días, semanas, incluso meses.
- **Proximidad:** Los epicentros distribuidos en enjambre se encuentran geográficamente cercanos entre sí. El área epicentral está definida por una fuente en particular que origina la secuencia.
- **Magnitud:** El rango de magnitudes de los enjambres va de pequeñas a moderadas. Las magnitudes de cada evento del conjunto pueden cambiar pero son similares entre sí.
- **Ausencia de evento principal y secuencia de réplicas:** Los enjambres son distintos de secuencias que preceden a un sismo principal (*foreshocks*) o como consecuencia de este (réplicas), ya que son secuencias en las que no se identifica un sismo principal.
- **Duración:** Los enjambres pueden extenderse a lo largo de días, hasta varias semanas. Su duración es variable y depende del contexto geológico-tectónico de la región.
- **Relación volcano-tectónica:** Los enjambres sísmicos se observan principalmente en áreas volcánicas y otros sistemas con flujo de fluidos. En un ambiente volcánico, son un tipo de sismicidad que pueden indicar una inminente erupción porque se relacionan con la acumulación de magmas ascendentes en la corteza terrestre antes de su eventual erupción. Sin embargo, también pueden ocurrir enjambres puramente asociados a procesos

tectónicos, donde las fallas tectónicas son la fuente de los sismos sin la presencia de un volcán cercano a los epicentros ([Legrand y cols., 2023](#)).

Los estudios de enjambres sísmicos se centran en las particularidades de estas secuencias, en los procesos geológicos subyacentes y en las condiciones tectónicas específicas de la región. Debido a que los enjambres sísmicos pueden ser actividad precursora de erupciones u otros procesos, es fundamental investigarlos para comprender la actividad sísmica de una región en particular y, si es necesario, emitir alertas tempranas. En los sistemas de alarma en el monitoreo volcánico, la detección y el seguimiento de los enjambres sísmicos son un elemento crucial.

2.6. Métodos para la detección de señales sísmicas

Antes de 1960, en los inicios de la sismología observacional, los analistas interpretaban los primeros sismogramas analógicos impresos en papel. Hacia mediados de la década de 1980, el panorama se transformó con la llegada de sismómetros digitales de banda ancha, que aunado al acceso a computadoras rápidas y potentes, abrieron nuevas posibilidades en la interpretación de sismogramas ([Kulhanek y Persson, 2011](#)).

La adquisición continua de datos digitales está imponiendo retos para las herramientas, métodos y teorías que se usan para su almacenamiento y análisis. Así es como surgieron los algoritmos de automatización en el procesamiento de datos sísmicos, una invención y desarrollo de la comunidad científica para reemplazar el manejo manual de los datos por mediciones objetivas y más prácticas.

Los algoritmos de detección automática de eventos están basados en tres enfoques generales: (1) amplitud y contenido de energía, como el método de STA/LTA ([Allen, 1978](#)), (2) similitud de forma de onda, como el método de correlación cruzada o también el *software* de detección FAST que emplea la coincidencia de patrones ([Yoon y cols., 2015](#)) y (3) Inteligencia Artificial, más específicamente aprendizaje automático o *Machine Learning*. Con el paso del tiempo estos métodos empezaron a considerarse como los tradicionales en la detección de eventos, mostrando ventajas, desventajas y limitantes ([Jiang y cols., 2021](#)).

El ruido, como factor no deseado en la observación de señales sísmicas, es la razón por la que la detección y el picado de las fases son tareas demandantes en el monitoreo sísmico, sobre todo si la amplitud del ruido es lo suficientemente grande para ocultar señales sísmicas de menor amplitud. En el área de detección surgieron los algoritmos de activación para identificar señales sísmicas típicas, ya sea originadas por una fuente natural o por una fuente controlada, en presencia de ruido continuo. Se llaman de activación o disparo porque ponen en acción tareas de procesamiento cuando se asume la presencia de una señal sísmica.

Aunque hoy existen múltiples algoritmos de detección automática, sólo los relativamente simples son los que más se aplican. Los modelos complejos se caracterizan por un ajuste complicado de parámetros, evidentemente la estimación de tales parámetros debe ser la adecuada a pesar de las condiciones de ruido y señal en el sitio para evitar falsas detecciones que ocupan memoria y sobrecargan el análisis de datos, además de que generan el trabajo extra de eliminar tales detecciones falsas (Vaezi y Van der Baan, 2015).

2.6.1. STA/LTA

La relación entre amplitudes absolutas promedio a corto plazo y largo plazo (*short-term average/long-term average*: STA/LTA) (Allen, 1978) es un método de detección automática muy popular basado en el cálculo continuo de la energía promedio (envolvente o amplitud absoluta) de un registro sísmico a partir de dos ventanas de tiempo consecutivas. La razón STA/LTA se calcula para cada instante de tiempo de la señal de entrada, de donde el promedio de la amplitud absoluta para cada ventana están dadas por las ecuaciones 2.26 y 2.27

$$STA = \frac{1}{N_S} \sum_{n=1}^{N_S} y_{k,n}, \quad (2.26)$$

$$LTA = \frac{1}{N_L} \sum_{n=-N_L}^0 y_{k,n}, \quad (2.27)$$

N_S es el número de muestras del intervalo de corto plazo, mientras que N_L es el número de muestras correspondientes al intervalo de largo plazo. La variable $y_{k,n}$ representa la función

característica para cada instante de tiempo $y_t = g(x_t)$ y para cada canal k . Esta función característica está diseñada para atender cambios en la señal y comúnmente contempla la energía $y_t = x_t^2$, el valor absoluto $y_t = |x_t|$ y la función envolvente $y_t = \sqrt{x_t^2 + h(x_t)^2}$, donde h expresa una transformada de Hilbert, (Vaezi y Van der Baan, 2015).

La ventana corta (STA) representa el promedio de la amplitud absoluta de la señal en el intervalo a corto plazo, es decir, el nivel de amplitud de la señal sísmica en cada instante y su función es buscar los eventos. Por otra parte, la ventana larga (LTA) está al tanto de la media actual de la amplitud de ruido sísmico. Esta comparación constante de las dos ventanas en movimiento es la razón $R = \frac{STA}{LTA}$, una nueva serie de tiempo.

Si R excede el nivel de umbral de activación designado por el analista, se produce una detección. En el caso contrario, si R es inferior al nivel de umbral de desactivación, no se emite la detección, o bien, se considera que el evento ha concluido. Comúnmente los umbrales de activación y desactivación equivalen a un valor predeterminado llamado umbral de detección τ .

De manera que en el algoritmo STA/LTA los parámetros son las longitudes de las ventanas STA y LTA (N_S y N_L), así como el valor del umbral de detección (τ). La configuración de parámetros debe contemplar especialmente los casos en que la energía de la señal sísmica es comparable al nivel del ruido. Un bajo umbral de detección producirá falsas detecciones, mientras que un umbral de detección alto conducirá a la omisión de eventos de menor magnitud.

2.6.2. Correlación cruzada de formas de onda

La correlación cruzada de formas de onda es un método ampliamente utilizado para la caracterización de señales sísmicas y la detección de eventos sísmicos. El método consiste en evaluar la similitud de formas de onda con el propósito de determinar los tiempos relativos de arribo de las fases. Se aplica en el análisis de sismos correlacionados, aquellos que exhiben mayor similitud debido a que son eventos con una pequeña separación hipocentral y, en consecuencia, presentan mecanismos focales similares.

El análisis de correlación cruzada inicia con la selección una forma de onda representativa del grupo eventos esperados, también llamada plantilla, para luego compararla con las formas

de onda obtenidas de registros de interés. Posteriormente, se calcula la función de correlación cruzada entre la plantilla y cada forma de onda presente en un sismograma. Este procedimiento implica deslizar secuencialmente la plantilla sobre datos continuos para calcular el coeficiente de correlación en cada paso de tiempo. Cuando el máximo de la función de la correlación cruzada supera un umbral mínimo predefinido, se considera que la forma de onda coincide con la plantilla, y, por tanto, corresponde a un sismo. A partir de esta coincidencia, se procede a determinar los tiempos relativos de arribo de las fases entre los dos eventos en comparación.

El uso de la correlación cruzada para obtener diferencias de los tiempos de viaje entre eventos correlacionados resulta información útil para la localización relativa de hipocentros, donde uno de los métodos más utilizados es el algoritmo de relocalización de eventos por doble diferencia (HypoDD: Waldhauser and Ellsworth, 2000; Waldhauser, 2001). Una mayor precisión en la localización de eventos sísmicos cercanos es esencial para caracterizar la sismicidad de una región y sus implicaciones tectónicas.

2.7. Tectónica en la Costa del Pacífico Mexicano

México experimenta una intensa actividad sísmica debido a que se ubica sobre cinco placas tectónicas (Norteamérica, Pacífico, Rivera, Cocos y Caribe). En el país, la sismicidad se concentra a lo largo de la costa suroeste en el Océano Pacífico. La subducción de la placa oceánica de Cocos bajo la placa continental de Norteamérica, con velocidades de hasta 6 cm/año (DeMets et al. 1994), da lugar a la formación de la Fosa Mesoamericana, el rasgo geomorfológico que delimita el contacto entre ambas placas tectónicas y se extiende desde Colima hasta Chiapas. Los sismos más significativos registrados en el país en este siglo han ocurrido en la región entre la costa y la Fosa Mesoamericana ([Kostoglodov y Pacheco, 1999](#)).

La geometría compleja de la placa en subducción ha sido estudiada en términos de las variaciones en la pendiente con la que se sumerge hacia el manto. Los análisis basados en sismicidad y mecanismos focales sugieren que la placa no subduce a un ángulo constante, sino que se vuelve subhorizontal después de un ángulo inicial bajo (Suárez et al., 1990; Singh y Pardo, 1993; Pardo y Suárez, 1995). Los resultados de Pérez-Campos et al. (2008) detallan en

que la placa se adentra en el manto justo al sur del Cinturón Volcánico Trans-Mexicano con un ángulo de aproximadamente 75° , alcanzando una profundidad de 500 km a los 400 km de la costa, donde finaliza la placa. Bajo este contexto, los volcanes actualmente activos, como el Popocatepetl, se encuentran sobre la parte superior de la placa en subducción, entre los 80 y 210 km de profundidad (Gómez-Tuena et al., 2007). La historia del volcanismo guarda una estrecha relación con la evolución geométrica de la subducción (Pérez-Campos y cols., 2008).

2.7.1. Sismicidad en la zona de subducción

La zona de subducción mexicana se caracteriza por sismos de gran magnitud asociados a fallas inversas que ocurren a profundidades someras a menos de 25 km (Gardi et al., 2000). La Figura 2.6 muestra la distribución geográfica de la sismicidad desde 1974 hasta 2014 con datos del Servicio Sismológico Nacional. Dos eventos emblemáticos con estas características son el sismo del 3 de junio de 1932, que tuvo lugar en las costas de Jalisco y Colima, alcanzando una magnitud de 8.2 en la escala de Richter y una longitud de ruptura de 280 km. Asimismo, el sismo de Michoacán del 19 de septiembre de 1985 ($M_W 8$), con una longitud de ruptura estimada en 180 km que cubrió casi toda la costa del estado de Michoacán (Kostoglodov y Pacheco, 1999).

2.7.1.1. La Brecha de Guerrero

Es una brecha sísmica ubicada en la costa de Guerrero con aproximadamente 200 km de longitud. Es una región particular debido a que desde 1911 no ha experimentado sismos con $M > 7$ en su porción al noroeste. En áreas circundantes, se han registrado sismos de magnitudes relativamente menores. En la porción NW de la Brecha de Guerrero, los sismos más grandes desde entonces han sido dos eventos cercanos a la trinchera en el año 2002 ($M_W 6.7$, $M_W 5.9$) y dos réplicas del sismo $M_W 7.3$ de Papanoa en 2014 ($M_W 6.5$, $M_W 6.1$). En la zona de subducción de Guerrero, los sismos lentos son un fenómeno común, incluso en esta región se han registrado los sismos lentos más grandes en la historia, con magnitudes $M_W > 7.5$, que se presentan aproximadamente cada 4 años (Plata-Martínez y cols., 2021).

Pardo y Suárez (1995) infieren que, por las dimensiones de la interfaz interplaca acoplada, uno

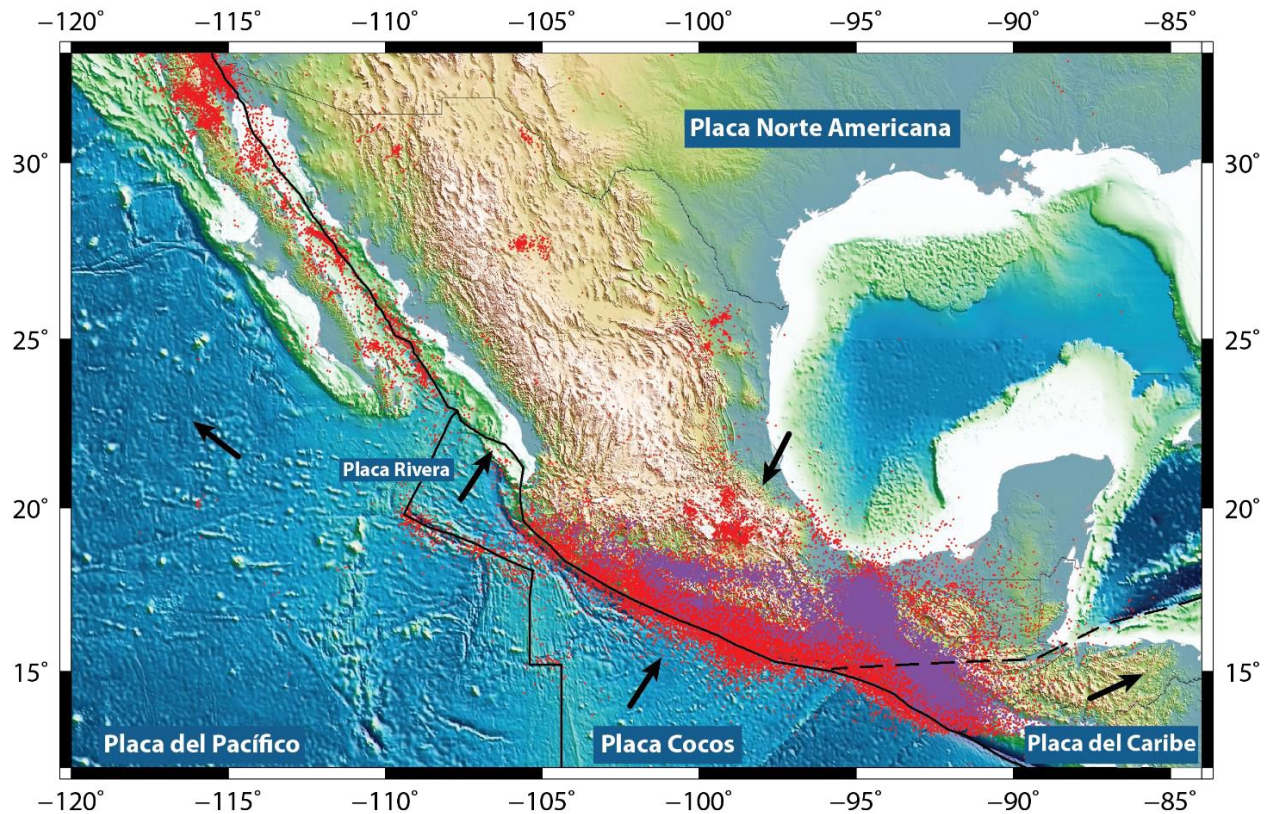


Figura 2.6: Tectónica de la costa oeste del Pacífico en México. Las flechas indican la dirección de movimiento de las placas. Los puntos representan sismos ocurridos entre 1974 y 2014 registrados por el Servicio Sismológico Nacional. En color rojo se indican sismos con menos de 50 km de profundidad, con morado aquellos sismos profundos (Plata-Martínez, 2015).

de los posibles escenarios en esta región podría ser la generación de un gran sismo con $M_W > 8$ (Kostoglodov y cols., 2003). Un sismo con esta magnitud tendría consecuencias devastadoras en ciudades cercanas, en la Ciudad de México podría generar aceleraciones del terreno hasta el doble que las experimentadas en el sismo del 19 de septiembre de 1985. Además, el potencial tsunami asociado a un evento de tal magnitud sería devastador para comunidades costeras como Acapulco y Zihuatanejo, entre muchas otras (Plata-Martínez y cols., 2021).

Un conocimiento preciso de la sismicidad y del estado de esfuerzo en las zonas de subducción es fundamental para desarrollar modelos de la evolución de la subducción y estimaciones realistas del peligro sísmico en esta región altamente vulnerable (Pacheco y Singh, 2010).

2.8. Sismotectónica en Baja California Sur

La provincia estructural del Golfo de California (GC) es un sistema de rift activo con deformación transtensional en el límite entre las placas tectónicas de Norteamérica y Pacífico, con una orientación NW-SE. En esta región se producen eventos con mecanismos de desplazamiento lateral derecho en fallas transformantes, que están interconectadas por una serie de cuencas de expansión oceánica. Estas zonas de formación de piso oceánico también presentan actividad sísmica pero de menor magnitud con mecanismos focales de tipo normal (Goff et al., 1987).

La mayoría de los sismos al sur del GC son generados por fallas transformantes. De hecho, esta región muestra mayor actividad sísmica en comparación con la porción central del GC. Por otro lado, la parte norte también es una zona activa, donde el origen principal de deformación proviene de fallas oblicuas (Castro, Stock, Hauksson, y Clayton, 2017).

En la Figura 2.7 se muestran las principales estructuras tectónicas de la península. Al oeste, se encuentra la zona de falla San Benito-Tosco-Abreojos, al este se extiende el sistema de fallas de tipo transforme y los centros de dispersión del golfo. El sur del estado se ve atravesado por un sistema de fallas con orientación NS. Destacan cinco estructuras principales que podrían estar activas: La Paz, El Carrizal, San Juan de los Planes, San Bartolo y San José del Cabo.

La actividad sísmica en la región de La Paz se caracteriza por la ocurrencia de enjambres sísmicos que son perceptibles en la zona urbana. Por ejemplo, entre febrero y abril de 2004, se registró un enjambre compuesto por 27 eventos, con magnitudes que variaron entre 1.2 y 3.3, y profundidades comprendidas entre 2 y 8 km (González, Mayer, y Aguirre, 2005).

Además, se han observado otros episodios de actividad relacionados con enjambres sísmicos en diferentes momentos y a lo largo de toda la península, (SSN, 2011), (SSN, 2016), (SSN, 2017), (SSN, 2018), (SSN, 2019). Al igual que en los últimos años, (Medrano, 2020), (Rondero, 2021), (Yáñez, 2022) y más recientemente en 2023, se registraron enjambres sísmicos en las cercanías del municipio de Loreto (Cervantes, 2023), así como otro en agosto de 2023 con epicentros en la zona norte del municipio de Los Cabos (BCS, 2023).

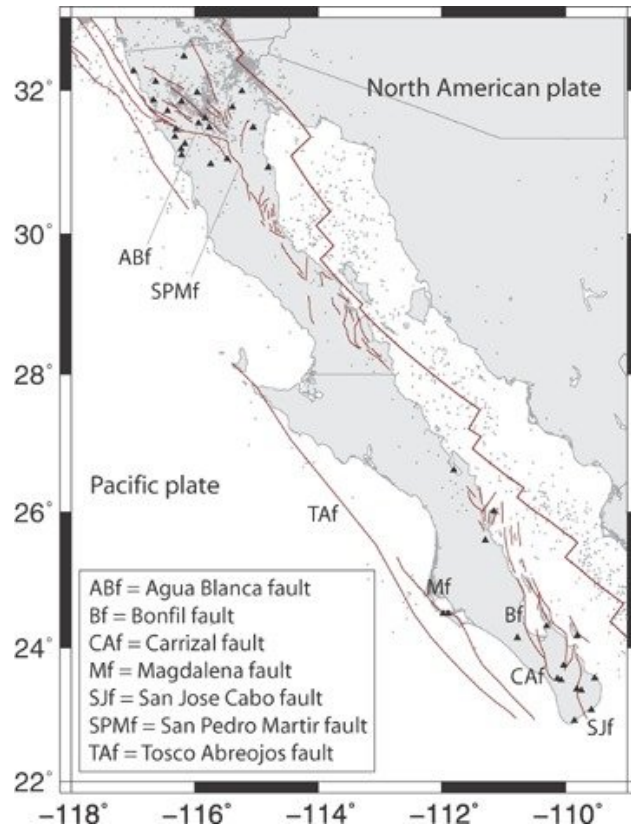


Figura 2.7: Sistema de fallas a lo largo de la península de Baja California Sur (Plattner et al., 2007).

Capítulo 3

Fundamentos teóricos de aprendizaje automático

La inteligencia artificial (IA), o en inglés *Artificial Intelligence* (AI), es un campo de la ciencia de la computación que busca crear modelos para realizar tareas avanzadas, imitando el razonamiento humano. Un desafío más para la IA ha surgido con la necesidad de manejar grandes volúmenes de datos que superan la capacidad humana. Con la creciente presencia de tecnologías digitales, la generación de datos es continua y se busca analizarlos para predecir las necesidades de los consumidores. Además del Internet, la ciencia es otra fuente importante de datos (Alpaydin, 2021). El aprendizaje automático, también conocido como *Machine Learning* (ML) en inglés, es una subcategoría popular de la Inteligencia Artificial (IA) que permite a los sistemas aprender de forma autónoma a partir de datos de entrenamiento.

El aprendizaje automático comienza con la recopilación de diversos tipos de datos, como cifras de mercados o industrias, imágenes para la clasificación de objetos, datos de series temporales, entre otros. A través de algoritmos de entrenamiento, el modelo de ML extrae patrones y relaciones relevantes del conjunto de datos de entrenamiento. Luego, el programador selecciona el modelo de ML apropiado para su objetivo, ingresa los datos de entrada y permite que el algoritmo genere respuestas basadas en su entrenamiento. Además, el programador puede ajustar el modelo utilizando algoritmos de optimización, los cuales modifican los parámetros del modelo

en cada iteración para mejorar su desempeño (Brown, 2021).

Es fundamental un conjunto de entrenamiento de calidad, ya que cuantos más datos reciba el algoritmo de ML, mayor será la precisión del modelo resultante. Además del conjunto de entrenamiento, se requieren conjuntos adicionales, llamados de validación y prueba, para la evaluación del modelo. El conjunto de validación se utiliza para ajustar y optimizar el algoritmo, comparando diferentes versiones y seleccionando la mejor según diversas métricas como la exactitud o la precisión. Por otro lado, el conjunto de pruebas se emplea para medir el rendimiento real del modelo en datos no vistos durante el entrenamiento o la validación, comprobando su capacidad de generalización y proporcionando resultados finales y errores del algoritmo.

Una característica distintiva del ML es su capacidad predictiva, mediante la cual el sistema realiza predicciones a partir de los datos observados en el entrenamiento. Por lo general, el ML se divide en tres categorías de aprendizaje: supervisado, no supervisado y por refuerzo.

3.1. Aprendizaje supervisado

El aprendizaje supervisado es una rama del ML enfocada en utilizar conjuntos de datos etiquetados para entrenar un algoritmo. El objetivo es desarrollar un modelo que pueda aplicarse a datos no etiquetados y hacer predicciones precisas. A través del aprendizaje supervisado, se realizan dos tareas principales de la minería de datos: regresión y clasificación. La regresión se utiliza para predecir valores numéricos, mientras que la clasificación asigna etiquetas a los datos.

En el caso de la regresión, se busca encontrar la relación entre un conjunto de datos de entrada con los de salida, es decir, la relación entre variables dependientes e independientes. El conjunto de datos de entrada contiene atributos, ya sean numéricos o categóricos, que el modelo utiliza para relacionar con las características de los datos de salida.

Por otro lado, la clasificación consiste en agrupar los datos en conjuntos que comparten al menos una característica común. Este proceso implica aplicar un discriminante a los datos. En algunos casos, el discriminante es desconocido y es necesario aprenderlo a partir de los datos. A diferencia de la regresión, donde el resultado es un valor numérico o continuo, en la clasificación la salida es una clase o valor discreto.

3.2. Aprendizaje no supervisado

El enfoque de aprendizaje no supervisado analiza conjuntos de datos sin etiquetar con la intención de encontrar, sin supervisión humana, patrones o relaciones en los datos. El aprendizaje no supervisado se divide en agrupación de clústeres, o *clustering*, y reducción de dimensionalidad. El *clustering* es útil para entender cómo se distribuyen los datos, en tanto que la reducción de dimensionalidad sirve para simplificar el manejo y la visualización de los datos.

El agrupamiento en clústeres consiste en segmentar los datos en grupos en función de similitudes que el modelo reconoció. Un algoritmo tradicional de *clustering* es el de *k-means* o agrupación por acercamiento de datos, el cual forma grupos con base en núcleos o centros de grupo. Comienza por aleatoriamente seleccionar dichos núcleos y medir la distancia a la que se encuentran sus puntos más cercanos, si estas distancias no son mínimas, se buscará un nuevo núcleo que ajuste cada vez más con el grupo que le rodea.

Cuando la dimensionalidad de los datos es inusualmente alta se procede a aplicar métodos para reducir el número de dimensiones sin comprometer la estructura y utilidad de los datos.

3.3. Aprendizaje por refuerzo

Esta área del ML se caracteriza por el uso de algoritmos que aprenden de manera autónoma a través de la experimentación y la retroalimentación. Un agente toma decisiones secuenciales dentro de un entorno dinámico, recibiendo retroalimentación con el objetivo de maximizar una métrica de recompensa a largo plazo por sus acciones. Una diferencia con aprendizaje supervisado, es que no requiere un conjunto de entrenamiento predefinido, sino que se basa en la experiencia adquirida a través de la interacción con el entorno. El objetivo es aprender un comportamiento óptimo para tomar acciones que conduzcan a obtener la máxima recompensa. El aprendizaje por refuerzo se aplica en la toma de decisiones secuenciales, es posible entrenar un agente para ganar en juegos de mesa o videojuegos, o bien en aplicaciones de robótica.

3.4. Aprendizaje profundo

Una de las aplicaciones más ampliamente difundidas del aprendizaje automático (ML) es el aprendizaje profundo, o en inglés *Deep Learning* (DL). A diferencia del enfoque convencional de ML, donde las características importantes para las predicciones deben ser extraídas o construidas previamente para producir la respuesta esperada, el DL puede automáticamente aprender estas características directamente de los datos de entrenamiento. Se denomina "profundo" debido a su estructura jerárquica compuesta de múltiples niveles de representación que se van aprendiendo de manera progresiva. Sin embargo, su aplicación se restringe a problemas que dispongan de grandes conjuntos de datos etiquetados para el entrenamiento.

La representación de las características consiste en convertir los atributos de los datos en un vector adecuado para que el algoritmo de aprendizaje detecte o clasifique patrones en los datos. El DL ha sido aplicado para obtener representaciones de alto nivel, esto quiere decir obtener una representación de las características de los datos más avanzada y/o abstracta que atravesó múltiples niveles de procesamiento. Los enfoques previos al DL adoptan un extractor de características diseñado por un experto y los datos se deben preparar antes de entrar al algoritmo, esto no ocurre en el DL porque este aprende por sí mismo las características de los datos.

Para comprender el aprendizaje profundo, es esencial tener un conocimiento sólido de las Redes Neuronales Artificiales (RNA), ya que el DL es un subcampo de la IA que utiliza estos modelos computacionales inspirados en la función cerebral de las neuronas biológicas. Las RNA han evolucionado desde modelos simples hasta redes que poseen numerosas capas de nodos, lo que ha dado lugar al surgimiento del aprendizaje profundo. Este aprendizaje a altos niveles de profundidad permite a los sistemas aprender características más abstractas y complejas de los datos. Los métodos de aprendizaje profundo se han desarrollado y refinado a través de múltiples arquitecturas de red para resolver una amplia gama de tareas por medio de IA.

3.4.1. Perceptrones

En 1943, el neurocientífico Warren McCulloch y el matemático Walter Pitts desarrollaron el primer modelo de red neuronal basado en la función de una neurona cerebral. En 1949, el

psicólogo Donald Hebb hizo una contribución clave con su libro "*The Organization of Behavior: A Neuropsychological Theory*". En él, propuso que las conexiones sinápticas entre neuronas se fortalecen cuando dos neuronas cercanas se activan aproximadamente al mismo tiempo. Esta teoría se convirtió fundamental en la neuropsicología y potenció los conceptos de aprendizaje y memoria en modelos de redes neuronales artificiales.

Posteriormente, en 1958 el neurocientífico Frank Rosenblatt introdujo una red neuronal de una sola capa a la que llamó perceptrón. Este modelo fue el primero con la capacidad de aprender a clasificar patrones simples en los datos a través de un entrenamiento supervisado. Básicamente, el perceptrón es un clasificador lineal gráficamente descrito en la Figura 3.1. En una clasificación binaria, el modelo clasifica a la salida como 0 o 1 comparando con un umbral de corte definido por una función de activación que se expresa como

$$y_k = \begin{cases} 1 & \text{si } v_k \geq 0 \\ 0 & \text{si } v_k < 0 \end{cases}, \quad (3.1)$$

donde v_k es la suma de los productos de las entradas por los pesos y añade un término de sesgo

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k. \quad (3.2)$$

Esto equivale al producto escalar entre la matriz de pesos \mathbf{W} de tamaño $n \times m$ y el vector de entrada $\mathbf{x} = [x_1, x_2, \dots, x_m]$, resultando en un vector de salida $\mathbf{v} = [v_1, v_2, \dots, v_n]$. Por lo tanto, la ecuación 3.2 se puede reescribir de la forma

$$\mathbf{v} = \mathbf{W}\mathbf{x} + \mathbf{b} \quad (3.3)$$

El entrenamiento supervisado en un perceptrón simple, se realiza con datos de entrada y sus correspondientes salidas deseadas, con el objetivo de ajustar los parámetros del modelo y minimizar la diferencia entre la salida real y la deseada. Sin embargo, está diseñado para problemas linealmente separables. Para superar esta limitación, se desarrolló el perceptrón multicapa (MLP, por sus siglas en inglés), también conocido como red neuronal de avance hacia

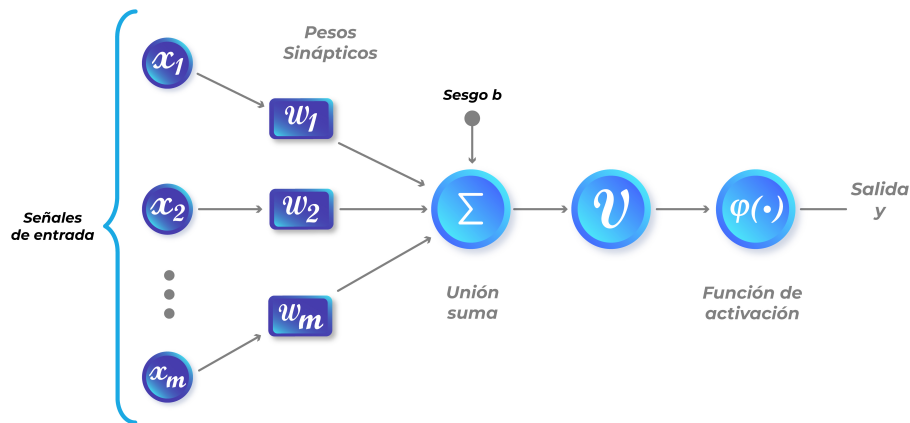


Figura 3.1: Modelo básico de un perceptrón simple k y sus elementos principales (Haykin, S., 2009).

adelante. Esta red consta de varios perceptrones simples conectados entre sí, y puede incluir una o varias capas ocultas que utilizan funciones de activación no lineales.

En un MLP totalmente conectado, como el de la Figura 3.2, las m neuronas de la capa de entrada proporcionan las entradas para las n neuronas de la capa oculta, mientras que las neuronas de la capa oculta son las entradas para las p neuronas de la capa de salida. Cada conexión entre dos neuronas tiene un peso sináptico. El nombre de “avance hacia adelante” se debe a que las unidades de la capa actual están conectadas a todas las anteriores.

Al inicio, se proporciona un vector de características en la capa de entrada, estableciendo los valores del vector como "salidas" para cada una de las unidades de entrada. Estos valores se propagan hacia las capas ocultas utilizando una función suma (producto matricial), estas salidas provenientes de las capas ocultas se transmiten a las capas de salida que proporcionan las salidas de toda la red aplicando una función de activación no lineal a cada elemento.

3.4.2. Algoritmo de propagación reversa

A partir de 1985 se popularizó un método para el entrenamiento de perceptrones multicapa llamado algoritmo de propagación reversa, o *backpropagation* en inglés. El método destacó por su eficiencia computacional y redireccionó el interés en las las redes neuronales (Haykin, 2009).

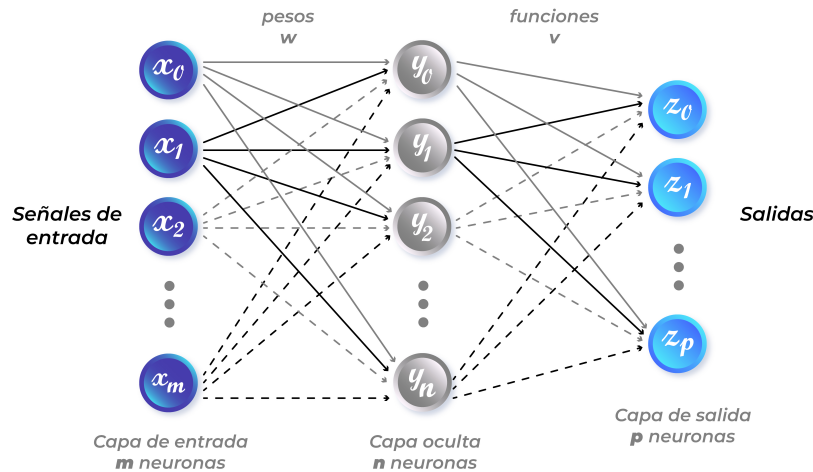


Figura 3.2: Red neuronal de avance hacia adelante totalmente conectada con una capa oculta.

Para minimizar el error de la red, el algoritmo de propagación reversa calcula el gradiente de la función de error, por ejemplo del Error Cuadrático Medio (ECM), recorriendo las capas desde atrás hacia adelante

$$E = \frac{1}{2} \sum_{k=1}^j (y_k - o_k)^2 \quad (3.4)$$

La ecuación 3.4 es una función de error que típicamente se utiliza como función de pérdida en el algoritmo de propagación reversa. En esta expresión, j es el número de neuronas de salida, y_k es el valor que debería devolver la red para la muestra k -ésima y o_k es la predicción que devuelve la red para la muestra actual.

El algoritmo de propagación reversa se relaciona con el algoritmo de descenso de gradiente, sin embargo, desempeñan roles diferentes. Mientras que *backpropagation* se refiere únicamente al método para calcular el gradiente de la función de error, o pérdida, el descenso de gradiente se utiliza para la optimización de parámetros que minimicen el error o pérdida. El algoritmo de descenso de gradiente consiste en la iteración de los siguientes pasos:

1. Se inicializan los parámetros del modelo con valores aleatorios.
2. Se ingresa a la red el conjunto de entrenamiento, es decir, pares de datos con una entrada y su correspondiente etiqueta de salida (x_k, y_k) .

3. Se calcula la función de pérdida con los valores de la salida obtenida o_k y la esperada y_k .
4. Se aplica el operador gradiente a la función de pérdida con respecto a cada uno de los parámetros del modelo. Es frecuente considerar al ECM 3.4 como función de pérdida.

$$J(W, b) = \frac{1}{2} \sum_{k=1}^j (y_k - o_k)^2, \quad (3.5)$$

que es una función multivariable porque asume que la salida del modelo, $o_k(W, b)$, depende de los parámetros de peso W y sesgo b .

Este gradiente de la función de pérdida $J(W, b)$ es la colección de todas las derivadas parciales posibles con respecto a cada uno de los parámetros. Para calcular dicho gradiente recurre al método de propagación reversa.

5. Se ajustan los parámetros del modelo en la dirección opuesta del gradiente. Esto a partir de que la derivada de la función de pérdida con respecto a un peso W , es igual a la pendiente de la recta tangente a la curva $J(W)$.

La Figura 3.3 es una representación del comportamiento de la función de pérdida al variar el parámetro W del modelo. Para optimizar el parámetro peso W el algoritmo cambia el valor de W con base en la pendiente: Si la pendiente es negativa, a la siguiente iteración se toma un valor hacia la derecha en sentido positivo. Si la pendiente es positiva, se avanza en sentido negativo hacia la izquierda. Esto se conoce como ajuste de los parámetros del modelo en la dirección opuesta al gradiente.

En cada iteración, los parámetros se actualizan según la expresión 3.6, que es la multiplicación del gradiente por una tasa de aprendizaje α y lo resta del valor actual W .

$$W = W - \alpha \frac{\partial J}{\partial W} \quad (3.6)$$

Y de la misma forma para otros parámetros del modelo, como son los sesgos b

$$b = b - \alpha \frac{\partial J}{\partial b} \quad (3.7)$$

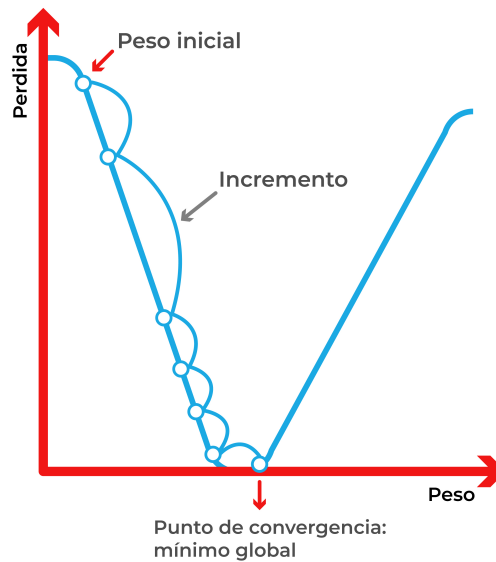


Figura 3.3: Representación visual con los conceptos básicos del principio de descenso por gradiente.

La tasa de aprendizaje α es un factor que indica la longitud del paso en que se decide cambiar el valor de los parámetros, asegurándose que en cada iteración se converge hacia un mínimo de la función de pérdida.

6. Se repite la secuencia anterior aplicándola a todas las muestras de entrenamiento hasta que la función de pérdida se minimice o se alcance un número determinado de iteraciones.

Para explicar cómo el método de propagación reversa se combina con el método de descenso de gradiente para utilizar la propagación de errores hacia atrás a través de la red, puede tomarse un ejemplo sencillo como el de la Figura 3.4 que representa una red con capas de entrada y salida, así como dos capas ocultas L y $L - 1$, con una neurona en cada una. En la primera capa oculta, la neurona N_{L-1} está determinada por el dato de entrada, x_i , también por el peso del enlace que la une con la capa de entrada, W_{L-1} , y por un sesgo b_{L-1} propio de la neurona. A la salida de la neurona N_{L-1} se espera un valor Z_{L-1} dado por la multiplicación del valor de entrada por el peso del enlace más el sesgo asociado

$$Z_{L-1} = W_{L-1}x_i + b_{L-1} \quad (3.8)$$

Este valor de salida es recibido como valor de entrada por la neurona N_L de la siguiente capa, la cual también está determinada por parámetros de peso sináptico, W_L , y de sesgo b_L . De igual manera, se espera un valor de salida Z_L

$$Z_L = W_L Z_{L-1} + b_L \quad (3.9)$$

De haber más capas ocultas, los valores continúan transmitiéndose por la red hasta analizar todo el conjunto de entrenamiento. De esta forma, se obtiene una predicción del modelo que puede compararse con la salida deseada mediante una función de pérdida J .

El proceso de propagación de errores hacia atrás comienza calculando el gradiente de la función de pérdida con respecto a la salida de la red

$$\frac{\partial J}{\partial Z_L} \quad (3.10)$$

Luego, se utiliza este gradiente para calcular el gradiente de la función de pérdida con respecto al peso y sesgo asociados a la neurona de la capa anterior, es decir, N_L ,

$$\frac{\partial J}{\partial W_L}, \frac{\partial J}{\partial b_L} \quad (3.11)$$

Este proceso se repite en la neurona de la capa previa, N_{L-1} ,

$$\frac{\partial J}{\partial W_{L-1}}, \frac{\partial J}{\partial b_{L-1}} \quad (3.12)$$

Estos gradientes se pueden resolver usando la regla de la cadena, considerando que los valores de salida son funciones derivables con respecto a los parámetros

$$\frac{\partial J}{\partial W_L} = \frac{\partial J}{\partial Z_L} \frac{\partial Z_L}{\partial W_L} \quad (3.13)$$

De forma análoga con el sesgo

$$\frac{\partial J}{\partial b_L} = \frac{\partial J}{\partial Z_L} \frac{\partial Z_L}{\partial b_L} \quad (3.14)$$

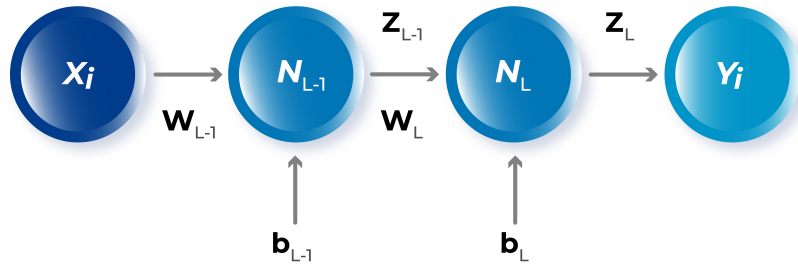


Figura 3.4: Esquema de un ejemplo muy simple de propagación reversa (*Interactive Chaos*, s.f).

El algoritmo propaga los errores desde atrás hacia adelante al calcular primero las derivadas parciales de la función de pérdida con respecto a los parámetros que unen la última capa oculta L con la de salida y después calcula las derivadas parciales de la función de pérdida con respecto a los parámetros que unen la capa de entrada con la oculta $L - 1$. Para reducir el error, se ajustan los pesos y sesgos de acuerdo con las expresiones 3.6 y 3.7 que hacen uso de las derivadas parciales 3.13 y 3.14, ya calculadas.

3.4.3. Redes Neuronales Profundas

Si se agregan múltiples capas a una red neuronal simple, se convierte en una red neuronal profunda. El aprendizaje profundo se basa en redes neuronales de numerosas capas que están conectadas a través de pesos sinápticos ajustables. Cada capa intermedia aplica transformaciones no lineales a las entradas y las transmite a capas subsecuentes hasta llegar a las salidas de la red. En el aprendizaje profundo, las redes pueden generar representaciones jerárquicas más complejas de las características presentes en los datos de entrada. Sin embargo, a medida que aumenta la profundidad de la red, su estructura se vuelve cada vez más abstracta, lo que puede disminuir la generalización del modelo. Por ello, es necesario implementar estrategias para abordar el sobreajuste de la red y maximizar la eficiencia en la etapa de entrenamiento. En este sentido, un aspecto fundamental ha sido el desarrollo de nuevas arquitecturas que utilizan diversos tipos de datos para realizar tareas específicas. Por ejemplo, las redes neuronales convolucionales (CNN) que se utilizan en tareas de visión artificial, las redes neuronales recurrentes (RNN) para tareas de procesamiento de datos secuenciales, y recientemente los

modelos de transformadores diseñados para procesamiento de lenguaje natural. El desarrollo de estas arquitecturas se ha llevado a cabo en relativamente poco tiempo y han sido diseñadas para mejorar la eficiencia de los modelos aprovechando los recursos computacionales del momento.

3.4.4. Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (CNN) son utilizadas principalmente en tareas de identificación de patrones. Están diseñadas para procesar datos que se presentan en forma de arreglos multidimensionales, como una imagen a color compuesta por tres arreglos 2D que representan los tres canales de color (RGB) y contiene los valores de intensidad de los píxeles (LeCun, Bengio, y Hinton, 2015). Una CNN se compone de capas convolucionales y de reducción, alternándose hasta llegar a una capa totalmente conectada, como se muestra en la Figura 3.5. Una descripción general de su funcionamiento consiste en los siguientes pasos:

1. Mapeo de características: Las capas convolucionales requieren datos de entrada y un filtro o kernel para realizar la operación de convolución. El kernel funciona como un detector de características que se aplica al arreglo de entrada mediante la convolución, generando valores de activación que forman un mapa de características o mapa de activación. A este mapa se le aplica una función de activación no lineal, como la función sigmoide.
2. Reducción de muestreo: Las capas convolucionales son seguidas por capas de reducción o submuestreo, las cuales disminuyen la resolución del mapa de características. Esta operación, también conocida como agrupación, se realiza con el objetivo de reducir el número de parámetros que la red debe aprender.
3. Capa totalmente conectada: Después de aprender características en las etapas anteriores, en la última capa se genera un vector que representa las clases a predecir y las probabilidades asociadas a cada clase en el arreglo que se está clasificando.

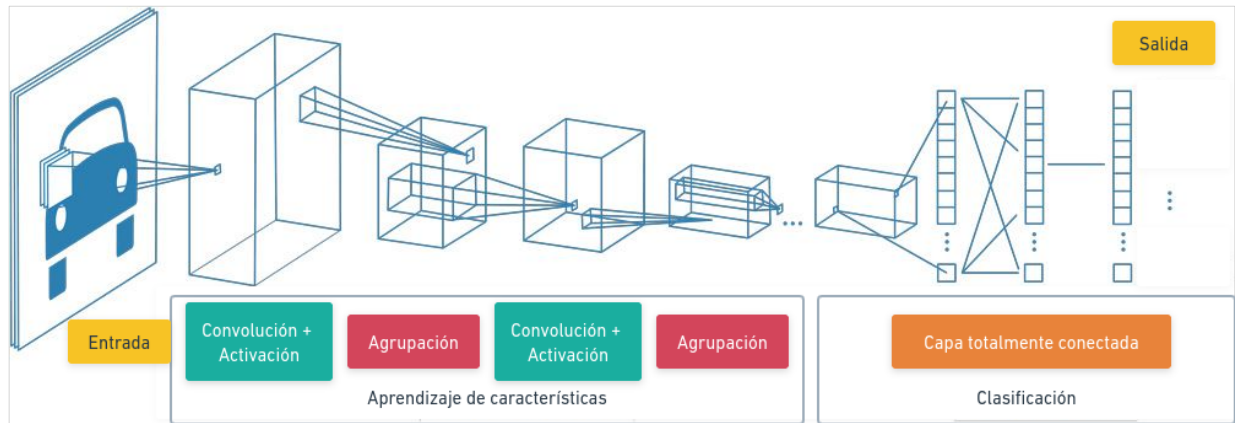


Figura 3.5: Red Neuronal Convolutional para el procesamiento de imágenes tales como reconocimiento de objetos (*MathWorks, s.f.*).

3.4.5. Redes Neuronales Recurrentes

Las Redes Neuronales Recurrentes (RNN) son especialmente útiles para el procesamiento de datos secuenciales, como entradas temporales, ya que tienen la capacidad de conectar la información previa con las acciones que se realizan en el presente. Se emplean en sistemas de traducción automática, predictores de palabras en teclados móviles, entre otros. No obstante, debido a su limitada capacidad se desarrollaron las Redes de Memoria Prolongada de Corto Plazo (LSTM), que incorporan una memoria explícita a la forma estándar de RNN.

En una RNN, los datos secuenciales se procesan elemento por elemento en la entrada, y luego a través de capas ocultas donde evoluciona el sistema. Estas capas ocultas contienen información tanto de las capas anteriores como de la capa actual. La Figura 3.6 muestra una representación de su arquitectura autorregresiva, donde las salidas de las capas ocultas en cada paso de tiempo se asemejan a las salidas de diferentes neuronas de un perceptrón multicapa. Los nodos o neuronas s con valores a determinado tiempo s_t se alimentan de salidas de neuronas anteriores, a medida que evoluciona la red ingresarán entradas x_t que se transformarán en una salida o_t que dependerá de todas las entradas previas $x_{t'}$ (para $t' \leq t$). Para cada paso de tiempo se mantienen los pesos U, V, W (LeCun y cols., 2015).

Un desafío para el entrenamiento de redes neuronales profundas es el desvanecimiento del gradiente. En la sección 3.4.2 se explicó detalladamente el proceso de entrenamiento mediante

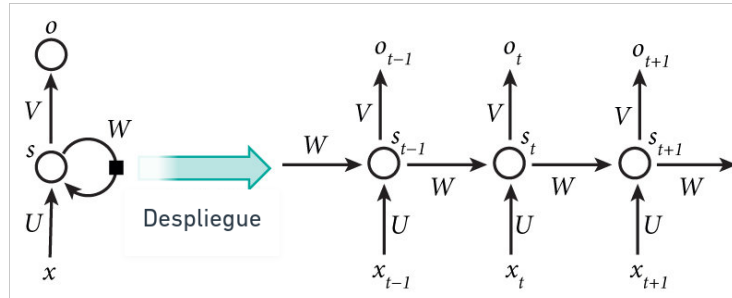


Figura 3.6: Secuencia temporal de una Red Neuronal Recurrente (LeCun, Y., Bengio, Y., y Hinton, G., 2015).

propagación reversa. La diferencia es que en las redes neuronales profundas, al constar de un mayor número de capas, existe mayor riesgo de que el gradiente se vuelva extremadamente pequeño. Esto implica que las actualizaciones de los parámetros serán mínimas y, como resultado, la precisión de las predicciones puede no mejorar significativamente.

3.4.6. Redes LSTM

Las Redes Neuronales Recurrentes en su forma estándar pueden presentar problemas debido a su limitada memoria a corto plazo al procesar secuencias largas de datos, lo que se conoce como desvanecimiento del gradiente. Para abordar esta limitación, así como problemas que requieran la captura de dependencias a largo plazo en secuencias de datos, surgieron versiones avanzadas de las RNN, como las redes de Memoria a Corto y Largo Plazo (LSTM) y las Unidades Recurrentes con Compuertas (GRU). En estas variantes se modifica la arquitectura para reducir el número de pasos de procesamiento en cada etapa de la red, lo que permite una mayor retención de información a lo largo de más tiempo, lo que ayuda a mitigar los problemas mencionados.

La red neuronal LSTM utiliza una celda de memoria y puertas de control adicionales, lo que le permite recordar o descartar información anterior y evitar el desvanecimiento de gradiente. Esto permite que la red aprenda a largo plazo, lo cual es especialmente útil en el procesamiento de secuencias extensas. En la Figura 3.7 la unidad de memoria g actúa como una neurona con compuertas que acumula entradas de A , B y C y su salida determina qué se tomará en cuenta en la conexión recurrente. Si la salida de g es 1, indica que la compuerta está cerrada y la predicción considera el pasado, mientras que si su valor es 0, la unidad decide borrar el contenido de la

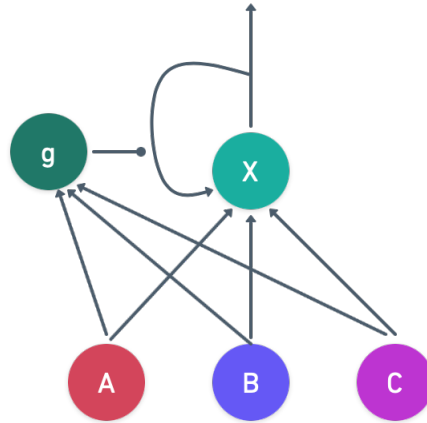


Figura 3.7: La unidad *g* funciona como una compuerta para capturar la información (Alpaydin, 2021).

memoria (Alpaydin, 2021).

3.5. Redes Neuronales Transformadoras

Previo al origen de las redes neuronales transformadoras, las redes neuronales recurrentes (RNN), incluyendo las redes LSTM y GRU, se habían convertido en los enfoques especializados en el procesamiento de datos secuenciales, como tareas de procesamiento de lenguaje natural (NLP). Sin embargo, a medida que aumenta la longitud de las secuencias, las estructuras recurrentes presentan restricciones de memoria y no es posible paralelizar operaciones. Estas limitaciones motivaron el desarrollo de una nueva arquitectura de red neuronal basada completamente en un mecanismo de atención (Vaswani y cols., 2017). En términos generales, la atención permite modelar dependencias mediante asignar diferentes pesos a las diferentes entradas, donde la ponderación depende específicamente de las entradas.

Los modelos de transformadores son sumamente versátiles, funcionan como modelos base que pueden adaptarse para resolver tareas en distintos dominios. Los transformadores han demostrado un desempeño superior a las RNN, e incluso combinados con capacidades de visión artificial, han logrado superar a las CNN en tareas de procesamiento de imágenes. Un avance significativo en esta arquitectura son los transformadores multimodales, que permiten el procesamiento e integración de múltiples tipos de datos, como texto, imágenes, audio y video.

Una de las mayores ventajas es que los transformadores aprovechan el cómputo en paralelo, lo que reduce significativamente los tiempos de entrenamiento y procesamiento, incluso en modelos con una cantidad masiva (del orden 10^{12}) de parámetros (Bishop y Bishop, 2023). Esta capacidad ha impulsado el desarrollo de grandes modelos de lenguaje como BERT (*Bidirectional Encoder Representations from Transformers*) y GPT (*Generative Pre-trained Transformer*).

3.5.1. Atención

Un ejemplo intuitivo del uso de la atención en el lenguaje natural es interpretar el significado de una palabra en diferentes contextos. Poniendo el caso de dos oraciones,

- La sierra se alzaba majestuosa sobre el valle, ofreciendo un espectáculo a los excursionistas.
- Yo utilicé la sierra para cortar la madera en la carpintería.

Cada oración ofrece contexto para comprender los dos significados de la palabra sierra. No obstante, en la primera oración las palabras “valle” y “excursionistas” son clave para entender que se habla de una montaña, mientras que en la segunda las palabras “carpintería” y “madera” indican que se trata de una herramienta. Esta variación en la atención ilustra la capacidad del mecanismo de atención para enfocarse en fragmentos más relevantes dentro de la secuencia. En otras palabras, el mecanismo le permite al modelo asignar pesos de atención totalmente con base en los datos de entrada.

3.5.2. Transformación

El nombre “transformador” proviene de que el modelo toma un conjunto de vectores representados en un determinado espacio para transformarlo en un conjunto correspondiente de vectores con las mismas dimensiones pero representados en un nuevo espacio. El objetivo de la transformación es crear una nueva representación más adecuada para resolver una tarea.

Un bloque de transformación se puede describir como una función que utiliza una matriz \mathbf{X} , conformada por el conjunto de vectores de entrada \mathbf{x}_n , o tokens, para crear una matriz

transformada $\tilde{\mathbf{X}}$ de la misma dimensión

$$\tilde{\mathbf{X}} = \text{Transformación}(\mathbf{X}) \quad (3.15)$$

A medida que aumenta el número de capas de transformación en una red neuronal, se generan representaciones internas de alta dimensión (Bishop y Bishop, 2023).

3.5.3. Datos de entrada

El preprocesamiento de los datos de entrada implica convertirlos en representaciones numéricas, específicamente en tokens, que son grupos de caracteres que representan segmentos individuales de la secuencia de entrada, como una palabra en un texto o un fragmento en una imagen. Luego, en una etapa conocida como incrustación, estos tokens se convierten en vectores de características. En el procesamiento de lenguaje natural, la incrustación de palabras captura la información semántica y sintáctica de una palabra en un vector. En un espacio de incrustación, cada palabra siempre se asigna al mismo vector y las palabras con significados similares están próximas entre sí. Aunque la creación de tokens y las incrustaciones no son exclusivas de la arquitectura *transformer*, suelen considerarse como las etapas iniciales.

3.5.4. Mecanismo de atención

Los transformadores introducen un mecanismo llamado auto-atención en para procesar datos secuenciales, permitiendo que el modelo examine y relacione diversas partes dentro de la misma secuencia de entrada para obtener una nueva representación. Como resultado, son altamente efectivos en una variedad de aplicaciones de procesamiento del lenguaje natural. Para implementar la operación de auto-atención se requieren tres vectores, denominados consulta \mathbf{q}_i (*query*), clave \mathbf{k}_i (*key*) y valor \mathbf{v}_i (*value*), para producir una salida. Esta salida se calcula como la suma ponderada de los valores, donde el peso asignado a cada valor se obtiene evaluando la compatibilidad entre la consulta y la clave, esto a través del producto punto de sus vectores $\mathbf{k}_i^T \mathbf{q}_i$ (Vaswani y cols., 2017).

Los vectores de consulta, clave y valor son conceptos asociados en el proceso de auto-atención, generados a partir de la misma secuencia de entrada.

- El vector consulta (\mathbf{q}_i) representa la información que se está buscando en la entrada, es decir, información relevante que el modelo necesita recuperar. Es utilizado para identificar las partes relevantes de la secuencia mediante compararlo con cada clave existente, evaluando a través del producto punto.
- El vector clave (\mathbf{k}_i) actúa como un identificador único de la información presente en la secuencia de entrada. Permite la interacción de información original con la consulta para determinar la importancia de cada parte.
- El vector valor (\mathbf{v}_i) contiene la información real de la entrada. Como se verá más adelante, cada uno de los vectores valor se ponderan según el grado de asociación entre consulta y clave para producir una salida.

3.5.5. Atención de producto punto escalado

Para simplificar el proceso de auto-atención, se abordará paso a paso cómo opera un solo módulo o cabeza de atención. En la arquitectura original (Vaswani, 2017), estos principios se generalizan a múltiples cabezas de atención. Este proceso, también llamado atención de producto punto escalado, implica los siguientes pasos generales:

1. Calcular los vectores de consulta, clave y valor para cada token en la secuencia de entrada. En la práctica se trabaja con las matrices \mathbf{Q} , \mathbf{K} , \mathbf{V} , conformadas por los vectores \mathbf{q}_i , \mathbf{k}_i , \mathbf{v}_i , respectivamente. Las cuales son obtenidas a partir de aplicar transformaciones lineales a los datos de entrada, representados por la matriz \mathbf{X} de dimensiones $N \times D$.

$$\begin{aligned}\mathbf{Q} &= \mathbf{XW}^{(q)} \\ \mathbf{K} &= \mathbf{XW}^{(k)} \\ \mathbf{V} &= \mathbf{XW}^{(v)}\end{aligned}\tag{3.16}$$

En estas transformaciones, la matriz \mathbf{X} se multiplica por las matrices de pesos $\mathbf{W}^{(q)}$, $\mathbf{W}^{(k)}$, $\mathbf{W}^{(v)}$. Estas matrices representan parámetros aprendidos durante el entrenamiento del modelo y contienen pesos que son ajustados por medio del algoritmo de propagación reversa. Las dimensiones de $\mathbf{W}^{(q)}$ y $\mathbf{W}^{(k)}$ son $D \times D_k$, donde D_k es la longitud de \mathbf{k}_i , mientras que la matriz $\mathbf{W}^{(v)}$ es de dimensiones $D \times D_v$, con D_v indicando la longitud de \mathbf{v}_i .

2. Calcular los puntajes de atención por medio del producto escalar de los vectores de consulta y clave, \mathbf{QK}^T . Este puntaje mide la relación entre los diferentes pares de tokens de entrada. Dado que el objetivo es utilizar tales puntajes de atención para ponderar los vectores de valor, es necesario normalizar dichos puntajes utilizando la función softmax. Esta función ajusta a 1 la suma de los productos escalares \mathbf{QK}^T , proporcionando una interpretación probabilística de los puntajes, donde los token de entrada con menor relevancia sobre la salida están asociados a puntajes de atención cercanos a cero, mientras que el valor aumenta para entradas mayormente relacionadas con la salida.

La Figura 3.8 ilustra la obtención de la matriz de valores de salida dadas \mathbf{Q} , \mathbf{K} , \mathbf{V} .

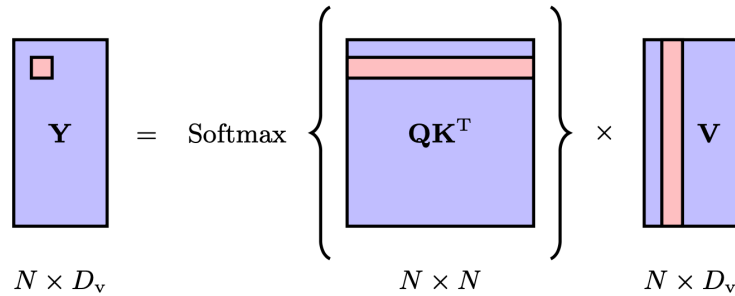


Figura 3.8: Representación visual de la operación de matrices en una capa de atención de producto punto escalado para generar una matriz de salida (Bishop y Bishop, 2023).

3. Reescalar los productos \mathbf{QK}^T dividiéndolos entre la raíz cuadrada de la longitud del vector clave, $\sqrt{D_k}$, antes de aplicar la función softmax. Con esto se estabiliza el gradiente de la función softmax para evitar el desvanecimiento del mismo.
4. Efectuar la multiplicación de los vectores de valor, \mathbf{V} , por los puntajes de atención normalizados y reescalados, es decir, los pesos de atención asociados, para sumar los productos y

obtener el vector de contexto que representa toda la secuencia de entrada.

$$Y = \text{Atención}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D_k}}\right)\mathbf{V} \quad (3.17)$$

Esta función determina la salida de una capa de auto-atención y por lo tanto representa el vector de contexto. La Figura 3.9 ilustra el flujo de los pasos descritos.

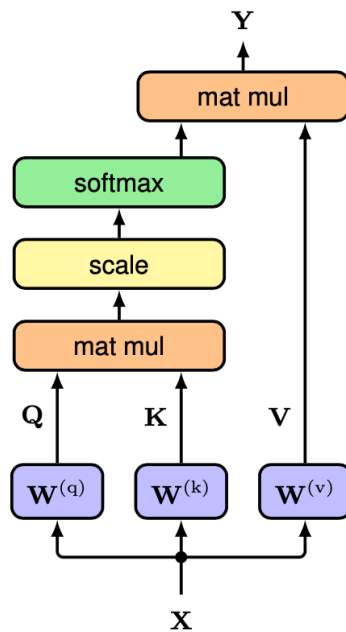


Figura 3.9: Flujo de información en una capa de atención. Las etapas de multiplicación de matrices se indican en las cajas nombradas como *matmul*, así como el reescalamiento y normalización en las cajas *scale* y *softmax* (Bishop y Bishop, 2023)

3.5.6. Atención de múltiples cabezas

La función de atención descrita permite al modelo relacionar cada token individual de la entrada con otros tokens en la misma secuencia. Por otro lado, en la práctica puede haber múltiples patrones en los datos que requieren atención simultáneamente. Por esta razón, la arquitectura transformer propone ejecutar la función de atención utilizando proyecciones lineales de los vectores consulta, clave y valor repetidamente a lo largo de H número de cabezas

indexadas como $h = 1, \dots, H$,

$$\begin{aligned}\mathbf{Q}_h &= \mathbf{X}\mathbf{W}_h^{(q)} \\ \mathbf{K}_h &= \mathbf{X}\mathbf{W}_h^{(k)} \\ \mathbf{V}_h &= \mathbf{X}\mathbf{W}_h^{(v)},\end{aligned}\tag{3.18}$$

a través de las matrices $\mathbf{W}_h^{(q)}, \mathbf{W}_h^{(k)}, \mathbf{W}_h^{(v)}$ con parámetros independientes entrenables.

Las cabezas se concatenan en una matriz y una vez más se aplica una transformación lineal al multiplicarla por la matriz \mathbf{W}^o , resultando de la forma

$$\mathbf{Y} = \text{Atención}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{Concatenar}(\mathbf{H}_1, \dots, \mathbf{H}_H)\mathbf{W}^o,\tag{3.19}$$

donde las cabezas \mathbf{H}_h son matrices de tamaño $N \times D_v$, las cuales se concatenan en una matriz de tamaño $N \times HD_v$. Nuevamente la matriz \mathbf{W}^o que se utiliza para la transformación lineal corresponde a parámetros aprendidos durante el entrenamiento del modelo y tiene dimensiones $HD_v \times D$, produciendo una matriz de salida de tamaño $N \times D$. Usualmente, la longitud del vector de valores D_v , es D/H para que la matriz de salida \mathbf{Y} tenga dimensiones $N \times D$ al igual que la matriz de datos originales \mathbf{X} (Bishop y Bishop, 2023). La Figura 3.10 muestra una representación del cálculo de la matriz de salida a lo largo de múltiples cabezas.

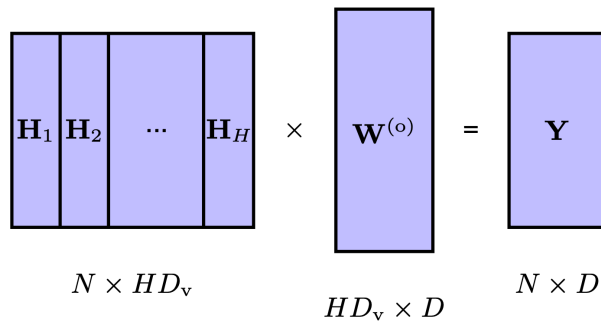


Figura 3.10: Representación de la operación de matrices de la ecuación 3.19 (Bishop y Bishop, 2023).

El uso de múltiples cabezas de atención permite al modelo analizar y establecer conexiones

entre la información representada en diferentes subespacios y en diferentes posiciones de manera simultánea, una capacidad que no se alcanza con una sola cabeza de atención. La Figura 3.11 muestra el flujo de información y la estructura en múltiples bloques de atención.

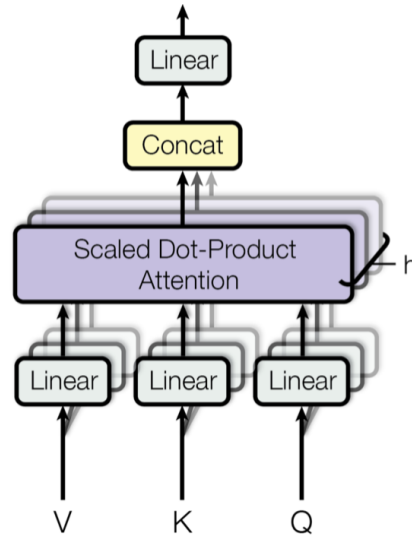


Figura 3.11: Representación visual de las etapas en una capa de atención con h número de cabezas concatenadas y llevadas a una transformación lineal (Vaswani et al. 2017).

3.5.7. Arquitectura codificador-decodificador

Aunque la piedra angular del transformer es el mecanismo de atención, su arquitectura se compone de dos partes principales: el codificador y el decodificador. A grandes rasgos, el codificador procesa la secuencia de entrada para producir una representación de la información, mientras que el decodificador utiliza esta representación para generar la salida final del modelo. Ambas partes consisten en una serie de capas idénticas con subcapas especializadas.

Las capas del codificador contienen dos subcapas principales: una de atención en múltiples cabezas y otra con una red neuronal totalmente conectada. El objetivo de la red neuronal es procesar en paralelo todos los vectores de la secuencia con base en los cálculos de atención producidos en la capa de múltiples cabezas para llegar a una única representación. Cada subcapa está seguida por una capa de conexión residual y normalización que evita el problema del desvanecimiento de gradiente y reduce el tiempo de entrenamiento del modelo.

Este proceso se repite en cada capa adicional del codificador. Por otro lado, las capas del decodificador son similares, pero incluyen una subcapa adicional de atención con enmascaramiento. Esta capa se diseñó específicamente para prevenir que los procesos del decodificador accedan a posiciones futuras que requieran de información con la que todavía no cuenta, afectando la generación de secuencias. Al igual que en el codificador, cada subcapa del decodificador se complementa con una capa de conexión residual y normalización.

La Figura 3.12 es una representación esquemática con los componentes de la primera arquitectura transformer. En el lado izquierdo se muestra el codificador y el decodificador en el lado derecho. La base de ambas partes son capas apiladas de auto-atención seguida de una red neuronal totalmente conectada para refinar la salida.

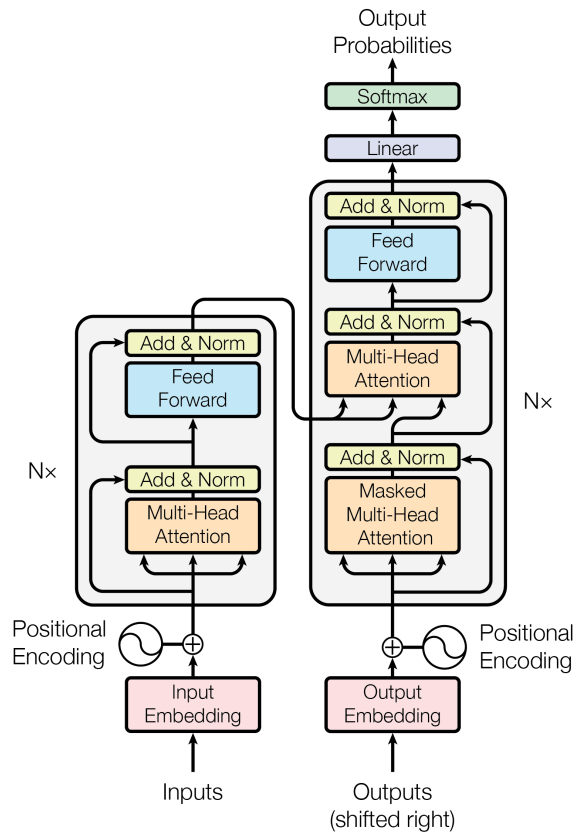


Figura 3.12: Arquitectura del modelo transformer original por Vaswani et al. (2017).

3.6. Aprendizaje profundo y sismología

En años recientes han surgido investigaciones enfocadas en el potencial de los modelos de Inteligencia Artificial (IA) para abordar datos geocientíficos. De hecho, ya se esboza el concepto de *Earth Artificial Intelligence*, como una combinación de sistemas para automáticamente monitorear y predecir eventos de la naturaleza en favor de la humanidad (Sun y cols., 2022). Las aplicaciones más recientes de aprendizaje automático (ML) en sismología están desarrollándose con métodos de aprendizaje profundo (DL) y están sobre todo orientados a remover el ruido de las señales, detectar señales sísmicas, picar el arribo de fases sísmicas, procesar e interpretar imágenes sísmicas, así como el modelado directo e inverso (Mousavi y Beroza, 2022).

En sismología, el aprendizaje automático halló la ventaja de contar con grandes cantidades de datos que a lo largo de la historia del monitoreo sísmico han sido analizados minuciosamente por analistas expertos, lo que proporciona bases de datos útiles para entrenar a sus algoritmos. Los enfoques basados en DL están superando a los métodos convencionales porque su alcance es mayor en términos de la capacidad para modelar y procesar conjuntos de datos en grandes volúmenes y de alta dimensión. No obstante, el DL es un área en desarrollo, sus teorías están en camino a consolidarse y para madurar requiere de estandarizar y validar sus métodos. La aplicación de Redes Neuronales Profundas en los métodos de monitoreo sísmico constituye una alternativa que puede reducir los efectos del riesgo sísmico a través de la estimación rápida y confiable de parámetros sísmicos que mejoren los sistemas de alerta temprana, así como la conformación de bases de datos más completas y basadas en métodos de DL que favorecen la evaluación de riesgos sísmicos a largo plazo (Mousavi y Beroza, 2018).

Los modelos con enfoque en DL están demostrando ser competentes herramientas para aplicarse en tareas de rutina en la sismología por su capacidad para aprender características de las formas de onda a partir de representaciones de alto nivel. Estos modelos pueden entrenarse con los diferentes tipos de aprendizaje: Supervisado, no supervisado, o semi-supervisado. Así como efectuarse en el dominio del tiempo, tiempo-frecuencia, o en combinación de tiempo y tiempo-frecuencia. Adicionalmente, en términos de rendimiento computacional, los modelos de aprendizaje profundo han proliferado en distintos sectores debido a que la capacidad

de procesamiento en paralelo de las unidades de procesamiento gráfico (GPU) ha reducido notablemente los tiempos de entrenamiento.

En el contexto del monitoreo sísmico, los modelos de aprendizaje profundo podrían aplicarse para elaborar catálogos más completos, con una descripción detallada de la sismicidad y mapas de mayor resolución, además de aportar eficiencia en las tareas de monitoreo en tiempo real. El potencial de las aplicaciones de aprendizaje profundo en la sismología no debe percibirse como un próximo reemplazo de analistas humanos, si no como una poderosa tecnología que puede llegar a potenciar la labor de los sismólogos.

3.7. Earthquake Transformer

Earthquake Transformer, referido como EQT a lo largo del texto, es un modelo de aprendizaje profundo utilizado para la detección de señales sísmicas y picado de fases. Su diseño está basado en una red neuronal profunda que fue entrenada con una base de datos global. En comparación con los algoritmos convencionales y modelos previos de aprendizaje profundo, EQT ha demostrado un mayor rendimiento gracias a su mecanismo de atención jerárquica que le permite realizar en conjunto las tareas de detección y selección de fases.

Este paquete de software de código abierto se publicó como *Earthquake Transformer: Un modelo de atención basado en aprendizaje profundo para detectar y seleccionar fases simultáneamente* (Mousavi, Ellsworth, Zhu, Chuang, y Beroza, 2020). Fue desarrollado en la Universidad de Stanford por Mousavi y colaboradores. Los resultados de su aplicación en el análisis de réplicas de sismos en Tottori, Japón, mostraron que el modelo fue capaz de detectar y localizar dos veces más eventos que el reporte de la Agencia Meteorológica de Japón (JMA) y para hacerlo requirió los datos de sólo una tercera parte de la red. Los autores de EQT argumentan que el modelo tiene resultados con una precisión cercana a las detecciones y picados de un analista humano. Debido a su eficiencia y sensibilidad, este modelo puede detectar y caracterizar eventos de menor magnitud, que a menudo son imperceptibles para el ojo humano.

3.7.1. Arquitectura

La arquitectura es un modelo de tipo transformador y se compone de un bloque codificador con 17 capas y un bloque decodificador dividido en tres ramas, cada una con entre 8 y 10 capas. Los bloques contienen procesos de convolución 1D, unidades LSTM unidireccionales y bidireccionales, bloques Network-in-Network (NiN), conexiones convolucionales-residuales (Res CNN), capas de retroalimentación y capas de auto-atención.

El diagrama de la Figura 3.13 describe en general la estructura de EQT. Los datos de entrada son sismogramas de tres componentes en el dominio del tiempo y los resultados de salida son las probabilidades de certeza de la detección de un evento y de sus respectivas fases P y S. De manera simplificada, los datos de entrada pasan por el codificador para generar representaciones de alto nivel y contexto de los datos, luego esta información se transfiere al bloque decodificador para convertirse en distribuciones de probabilidad de los tres escenarios posibles.

Los modelos de auto-atención requieren de un proceso de submuestreo para reducir el tamaño de los datos de entrada. En este caso, se aplicó una sección de submuestreo previa al bloque codificador, compuesta por capas convolucionales y de *pooling* máximo. Posteriormente, la información se procesa a través de Redes Neuronales Convolucionales Residuales (Res CNN) y de bloques LSTM del codificador para generar representaciones de alto nivel. A continuación, se utilizan capas de atención global para enfocar la atención de la red hacia las partes de la señal que son específicamente útiles para asociarlas con lo que caracteriza a la presencia de un evento sísmico en la señal continua, así como el arribo de las fases de onda P o de onda S.

Estas nuevas representaciones de alto nivel, obtenidas a partir de las capas de atención global, se separan con un vector de probabilidades para determinar si lo que se observa está asociado a una detección de señal sísmica o a las fases de ondas de cuerpo. En función de esta clasificación, se envía la información a una de las tres ramas del decodificador. Cabe mencionar que las ramificaciones dedicadas a la identificación de fase P y fase S presentan al inicio un bloque LSTM y uno de atención local, que dirigen la atención hacia características locales de la señal que son representativas de una u otra fase.

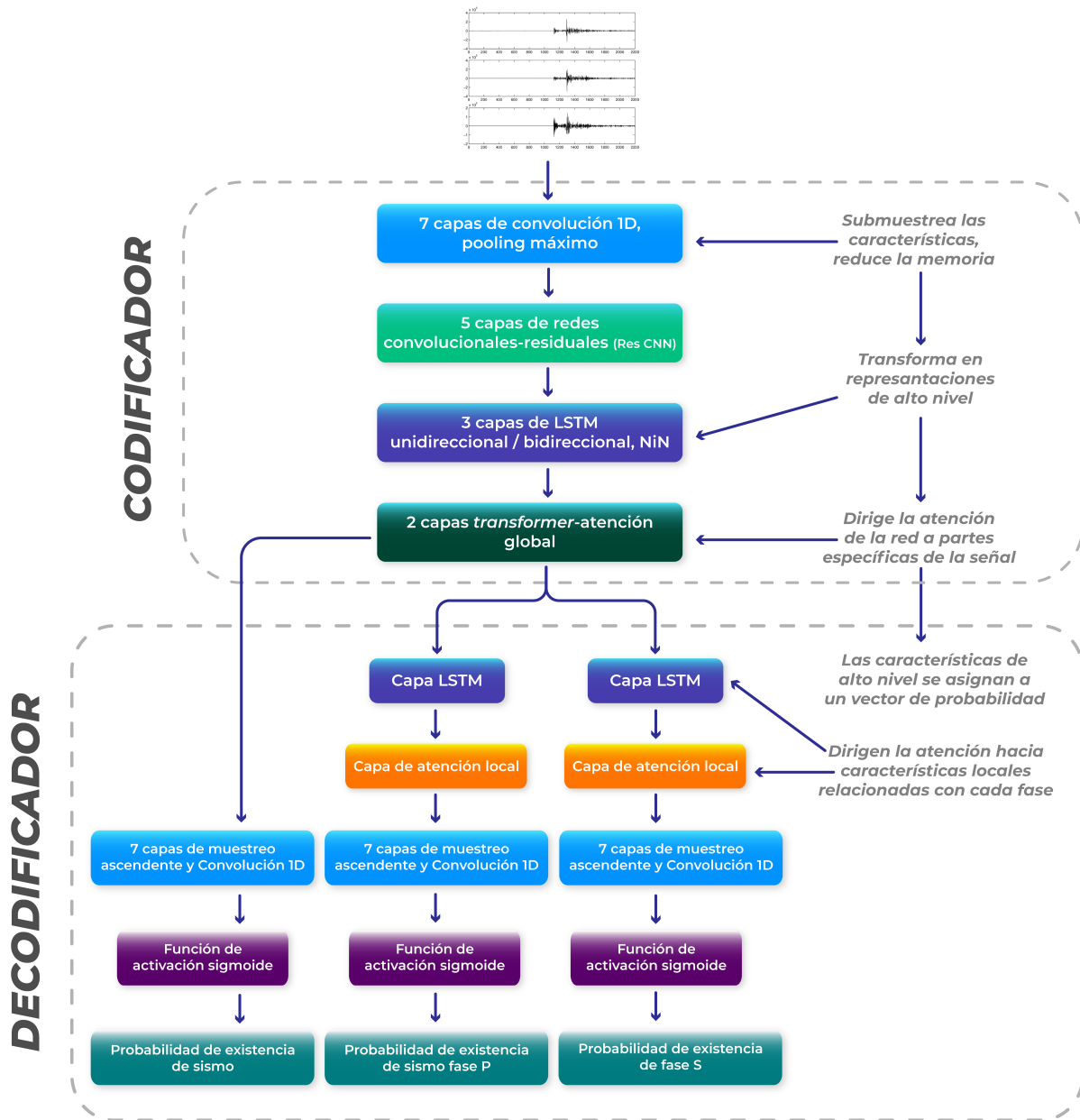


Figura 3.13: Representación simplificada de la arquitectura de la red neuronal profunda de EQT. Adaptada de *Automatic earthquake detection and phase picking in Muong Te, Lai Chau region* (Cong et al., 2022).

3.7.2. Datos y entrenamiento

El modelo fue entrenado con la base de datos global STEAD (*Stanford Earthquake Dataset*). Esta extensa base de datos contiene información de millones de señales, tanto sísmicas como no sísmicas. Incluye desde formas de onda generadas por sismos locales, registrados a distancias epicentrales locales de hasta 300 km, hasta formas de onda de ruido sísmico ambiental o cultural y que no están relacionadas con eventos sísmicos.

STEAD proporciona información detallada de cada señal, incluyendo si está asociada a un sismo, información de la estación, tiempo de origen, localización epicentral, profundidad, magnitud, tipo de magnitud, estimación del error, etc. En cuanto al registro de la señal, se incluye información sobre la relación señal-ruido para cada componente, la duración de coda y la distancia epicentral. En el caso de los registros de ruido la información se limita a información sobre la red, la estación y el instrumento utilizado (Mousavi, Sheng, Zhu, y Beroza, 2019).

Para desarrollar EQT se utilizó 1 M de formas de onda sísmicas y 300 K formas de onda de ruido. Estos datos están asociados a cerca de 450 K sismos distribuidos en diversas ubicaciones geográficas y bajo diversas condiciones tectónicas. El proceso de entrenamiento se completó en un total de 89 horas empleando 4 procesadores gráficos en paralelo.

3.7.3. Mecanismo de atención

La interpretación de un sismograma por parte de un analista se realiza por etapas. Lo primero es identificar la presencia de un sismo en la forma de onda. Una vez confirmada su existencia, se centra en detalles específicos dentro de la forma de onda para identificar las fases que aparecen en el sismograma, comúnmente las fases de las ondas P y S. Estos detalles a menudo involucran cambios en la frecuencia y/o en la amplitud de la señal.

Los autores de EQT explican que el mecanismo de atención jerárquica de su modelo se asemeja al enfoque de un analista experimentado. Esto se logra mediante un proceso de atención que otorga mayor peso a ciertas partes de la señal, permitiendo que el modelo se concentre en realizar una predicción precisa. En otras palabras, el modelo utiliza este mecanismo de atención jerárquica para aprender a reconocer qué partes son importantes para cada tarea, ya sea la señal

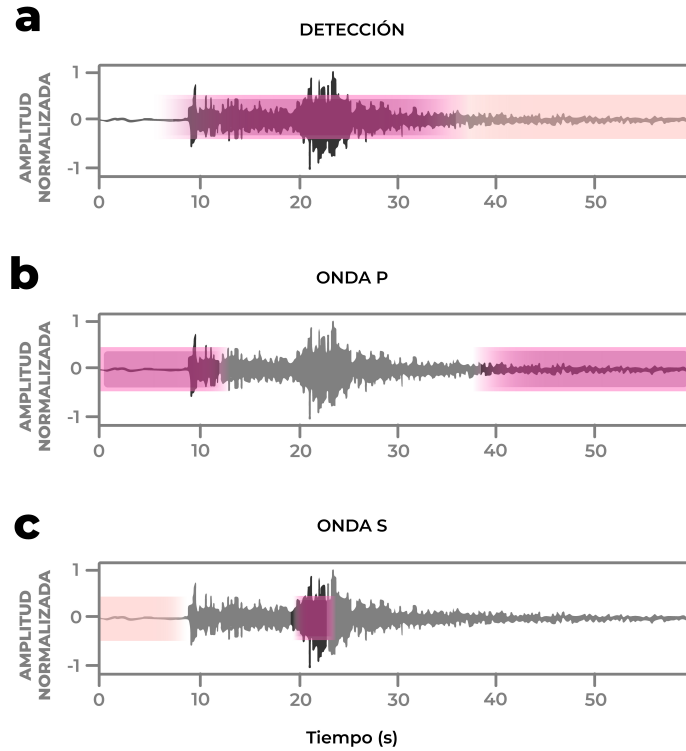


Figura 3.14: Visualización de los pesos de atención. El panel **a** representa la salida de las capas *transformer* y de atención global, **b** y **c** son resultado de las capas de atención local dedicadas a la identificación de cada fase (Mousavi et al., 2020).

en general para la detección del sismo, o en un fragmento específico de la señal para la selección de una fase en particular.

La Figura 3.14 es una representación visual de cómo se dirige la atención hacia partes específicas de la señal y que es considerada para la predicción. Esta visualización ofrece una comprensión útil de cómo funciona el mecanismo de atención en la arquitectura de la red.

Capítulo 4

Metodología

4.1. Earthquake Transformer

La secuencia de procesamiento con Earthquake Transformer (EQT), desde la instalación hasta la aplicación, involucra los siguientes pasos generales:

1. EQT es un paquete en Python versión 3.x que hace uso de bibliotecas como TensorFlow para implementar el aprendizaje automático, y ObsPy, diseñada para el procesamiento de datos sismológicos. EQT se puede instalar de varias maneras, ya sea a través del gestor de paquetes Conda, descargándolo desde el repositorio de software PyPI, o descargando el código fuente directamente desde la plataforma Github.
2. Como datos de entrada recibe registros sísmicos continuos en bruto en formato MiniSeed. Se pueden emplear archivos MiniSeed almacenados localmente en el equipo. Esta opción es recomendable para analizar periodos temporales más extensos, aunque demanda una mayor cantidad de memoria, así que requiere que los datos estén organizados en secuencias de alrededor de un mes ([Mousavi, 2020](#)).
3. Para efectuar la detección y selección automática, EQT requiere de un modelo pre-entrenado. En este sentido, existen dos opciones: el modelo original y el conservativo. Aunque comparten arquitecturas similares, presentan diferencias en el proceso de entrenamiento y en

los hiperparámetros utilizados en el aumento de datos.

Sin embargo, la elección del modelo depende del objetivo del análisis. El modelo original se utiliza comúnmente para maximizar el número de detecciones, aunque esto puede resultar en una mayor tasa de falsos positivos. Por otro lado, el modelo conservativo se introdujo posteriormente para reducir la tasa de falsos positivos, sin embargo esto puede llevar a la pérdida de algunas detecciones de sismos reales.

4. Lo siguiente es especificar una serie de parámetros para la predicción. Ver Tabla 4.1.

Tabla 4.1: Parámetros de Earthquake Transformer necesarios para realizar la detección y selección.

Parámetro	Descripción
Umbral de detección	Es un valor de umbral que determina qué detecciones tienen una probabilidad lo suficientemente alta como para ser considerados eventos sísmicos. Toma valores desde 0 a 1.
Umbral de fase P	Es un valor de umbral entre 0 y 1 para establecer a partir de qué valor de probabilidad de selección se considerará la llegada de la fase P.
Umbral de fase S	Es un valor de umbral entre 0 y 1 para establecer a partir de qué valor de probabilidad de selección se considerará la llegada de la fase S.
Traslape	Al configurar este parámetro, el modelo realizará la detección y selección utilizando ventanas que se superponen entre sí según un porcentaje.
Tamaño de lote	Es el número de muestras que se procesarán en cada iteración de la red. Se recomienda un valor entre 200 y 1000.

Para seleccionar los parámetros, se recomienda utilizar umbrales más altos para el modelo original y umbrales más bajos para el modelo conservativo. Es crucial destacar que el modelo original es ideal si se planea procesar las detecciones con un algoritmo de asociación de fases y luego con un software de localización de epicentros, lo que elimina las falsas detecciones y mejora los resultados al proporcionar una distribución geográfica.

5. Finalmente, los resultados de la predicción se generan en directorios de salida creados para cada estación. Estos directorios contienen un reporte con los parámetros utilizados, el tiempo de ejecución, el número total de detecciones, etc. También se genera un archivo

CSV con los resultados de la detección y selección de cada uno de los eventos. Adicionalmente, se crea una carpeta de figuras que presenta la señal en el dominio del tiempo o en tiempo y frecuencia como la Figura 4.1.

Las figuras incluyen ventanas de tiempo de un minuto a 100 muestras por segundo, con marcas verticales que indican el arribo de las fases. También un gráfico de curvas para representar la probabilidad con la que se predice la detección del evento, así como la selección de las fases P y S. Las figuras posibilitan una inspección práctica para descartar falsos positivos, así como analizar el impacto de los parámetros en el rendimiento.

Para el uso de EQT en este trabajo, se empleó Python 3.7, con las dependencias Tensorflow 2.5.0 y Obspy 1.3.1. Todas las pruebas se llevaron a cabo utilizando el entorno Jupyter Notebook. Para cada uno de los casos de estudio, se realizaron ajustes de los parámetros y se mantuvo el valor predeterminado de 500 para el tamaño del lote.

4.2. Diseño experimental

Comúnmente los modelos de detección automática son probados en estudios de microsismicidad, dado que la magnitud de los eventos y la calidad de la señal puede dificultar aún más el procesamiento manual de los sismogramas. Sin embargo, con el objetivo de obtener evidencia más completa del desempeño del modelo, se llevaron a cabo tres pruebas con diferentes características. La primera prueba consistió en analizar un enjambre sísmico ocurrido entre el 18 y 22 de agosto de 2017 en cercanías de Coalcomán, Michoacán. En segundo lugar se analizó una secuencia corta de microsismos en Los Cabos, Baja California Sur, ocurrida entre el 8 y 9 de agosto de 2023. Además, se analizó el sismo de magnitud 7.2 ocurrido el 18 de abril de 2014 en el sur de Petatlán, en el estado de Guerrero y sus réplicas en las primeras 36 horas posteriores.

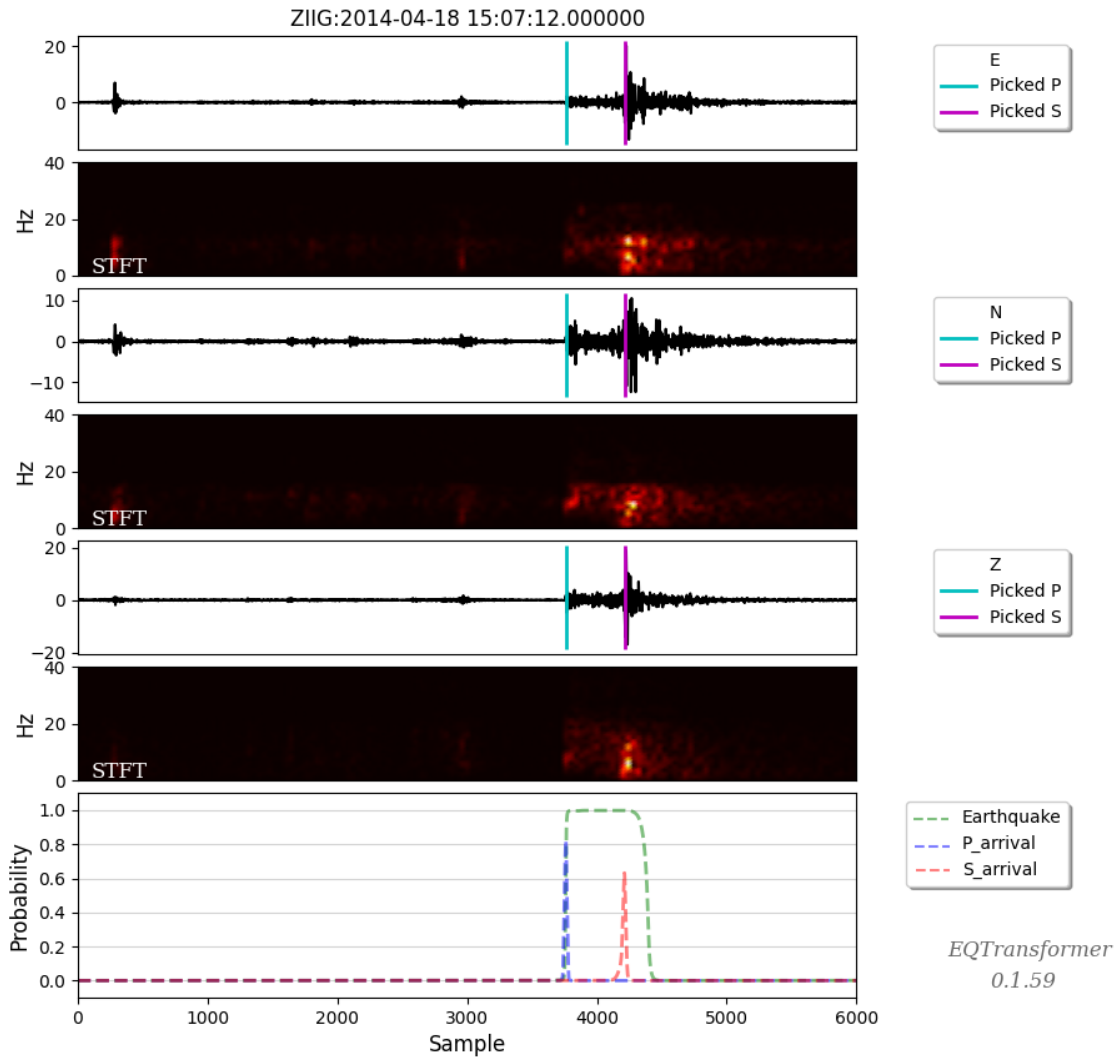


Figura 4.1: Ejemplo de una forma de onda en las tres componentes con un evento detectado. Se utiliza la Transformada de Tiempo Corto de Fourier (STFT) para mostrar la señal en el dominio de la frecuencia. El ejemplo corresponde a un evento de magnitud 3.1 de acuerdo con el catálogo del SSN. EQT realizó la detección con un 0.98 de probabilidad, la fase P se seleccionó con 0.82 de probabilidad y una relación señal/ruido estimada en 24.9 dB, la fase S se seleccionó con 0.61 de probabilidad y una relación señal/ruido de 10.4 dB.

4.3. Obtención y características de los datos

En el caso de la secuencia sísmica en Michoacán y el sismo en Guerrero, ambos analizados y reportados por el Servicio Sismológico Nacional (SSN), se seleccionaron datos de estaciones pertenecientes a la Red Sísmica de Banda Ancha del SSN. Las estaciones a cargo del SSN están equipadas con un sismómetro de tres componentes con respuesta de frecuencia plana desde 120 ó 240 s hasta 50 Hz, un acelerómetro de tres componentes y un digitalizador de 24 o 26 bits, también están provistos de un sistema global de navegación por satélite (GNSS) ([Servicio Sismológico Nacional, UNAM, 2023a](#)). Para preparar los datos se aplicó el siguiente proceso:

1. Se descargaron los catálogos de sismos del SSN de la ubicación geográfica y en el periodo de tiempo correspondientes a cada caso ([Servicio Sismológico Nacional, UNAM, 2023a](#)). Un catálogo del SSN proporciona para cada evento, su fecha y hora (en UTC y local), magnitud, latitud, longitud, profundidad, una referencia del epicentro y el estado de revisión.
2. Se descargaron los registros de las estaciones más cercanas a los epicentros utilizando el servidor SSNstp ([Servicio Sismológico Nacional, UNAM, 2023b](#)). SSNstp es un cliente en línea que permite acceder a los registros disponibles para la estación, canales y ventana de tiempo solicitados por el usuario. Estos datos se descargan en formato SAC en el equipo donde se ejecute el servidor.
3. Con el fin de ser compatibles con EQT, los datos fueron convertidos del formato SAC a MiniSeed mediante ObsPy.

En el caso de la pequeña secuencia de microsismos en Baja California Sur, los registros en tres componentes se solicitaron al Centro Nacional de Prevención de Desastres (CENAPRED). De igual forma, estos datos fueron convertidos a formato MiniSeed.

4.3.1. Secuencia sísmica del 18 al 22 de agosto de 2017, Michoacán

De acuerdo con el reporte especial publicado ([Servicio Sismológico Nacional, UNAM, 2017](#)) y el catálogo de sismos, esta secuencia de 86 eventos ocurrió entre el 18 y el 22 de agosto de 2017,

cerca del municipio de Coalcomán, Michoacán. El sismo más grande fue de magnitud 4.6 y el más pequeño de magnitud 2.8. El rango de profundidades va de 3.1 a 17.8 km.

Para evaluar el desempeño de EQT se utilizó el registro continuo de los cinco días de la secuencia obtenido en la estación Maruata (MMIG), de la Red de Banda Ancha, en Michoacán. La estación fue seleccionada debido a que es la más cercana a los epicentros de los eventos.

Para llevar a cabo la tarea de detección, se empleó el modelo conservativo, previamente descrito en la sección 4.1. Los parámetros seleccionados se presentan en la Tabla 4.2.

Tabla 4.2: Parámetros utilizados para la predicción con la estación MMIG (Michoacán, 2017).

Parámetro	Valor
Umbral de detección	0.45
Umbral de fase P	0.35
Umbral de fase S	0.35
Traslape	0.25

4.3.2. Secuencia sísmica del 8 al 9 de agosto de 2023, Baja California Sur

En el caso de esta secuencia sísmica más corta, los datos provienen de una estación temporal gestionada por el Centro Nacional de Prevención de Desastres (CENAPRED) como parte de una campaña de monitoreo dada la actividad sísmica reciente en zonas del municipio de Los Cabos, Baja California Sur. En esta adquisición de datos, se utilizó un sismómetro digital de tres componentes de banda ancha de la marca Güralp Systems, de la gama CMG y con modelo 6TD. Los datos sísmicos se muestrearon a 100 Hz y además se proporcionó una lista que incluye 91 sismos identificados junto con sus respectivos tiempos de arribo de la fase P, los cuales fueron determinados manualmente por el equipo encargado del monitoreo.

En el análisis de esta secuencia, nuevamente se empleó el modelo conservativo de EQT. Los parámetros seleccionados para esta prueba se pueden consultar en la Tabla 4.3. En comparación con la prueba anterior, los valores de los umbrales se redujeron para ampliar los rangos de confianza en la detección y selección. Este ajuste se realizó considerando que se trataba de sismos de menor magnitud con intervalos cortos entre la llegada de las fases P y S.

Tabla 4.3: Parámetros utilizados para la predicción con la estación MIRA (BCS, 2023).

Parámetro	Valor
Umbral de detección	0.1
Umbral de fase P	0.05
Umbral de fase S	0.05
Traslape	0.5

4.3.3. Secuencia de réplicas, sismo M_w 7.2 del 18 de abril de 2014, Guerrero

El SSN informó sobre un sismo ocurrido el 18 de abril de 2014, de magnitud 7.2 que tuvo lugar a 40 km al sur de Petatlán, Guerrero. El sismo ocurrió a las 09:27 hora local y se sintió con fuerza en varias regiones, incluyendo la Ciudad de México, Guerrero, Morelos, Hidalgo, Tlaxcala, Michoacán, Jalisco y Oaxaca, entre otras áreas ([Servicio Sismológico Nacional, UNAM, 2014](#)).

Después de las 14:27 del 18 de abril, hora UTC del sismo principal, y hasta el 20 de abril a las 00:00 horas, el catálogo de sismos del SSN reporta 109 eventos en el estado de Guerrero. La búsqueda se restringió al estado de Guerrero con el propósito de enfocarse en los eventos cercanos y posteriores al sismo principal. La réplica más grande tuvo una magnitud de 5.2.

La estación Zihuatanejo (ZIIG), perteneciente a la Red de Banda Ancha, en el estado de Guerrero, es el observatorio más cercano al epicentro del evento principal. Se trabajó con el registro continuo obtenido en ZIIG durante un periodo de 36 horas, que abarca desde el 18 de abril de 2014 a las 12:00 horas hasta el 20 de abril de 2014 a las 00:00 horas.

La detección de un sismo principal y sus réplicas se efectuó nuevamente con el modelo conservativo de EQT. Los parámetros seleccionados en esta prueba detallan en la [Tabla 4.4](#).

Tabla 4.4: Tabla 5.7 Parámetros utilizados para la predicción con la estación ZIIG (Guerrero, 2014).

Parámetro	Valor
Umbral de detección	0.6
Umbral de fase P	0.3
Umbral de fase S	0.3
Traslape	0.6

4.4. Métricas de evaluación

Una matriz de confusión es útil para visualizar el rendimiento del modelo, se puede identificar los tipos de errores y aciertos desglosando las predicciones en cuatro categorías posibles:

- **Verdaderos Positivos (VP):** Los casos en los que el modelo predijo correctamente una detección que sí es un sismo.
- **Falsos Positivos (FP):** Los casos en los que el modelo predijo incorrectamente una detección que no es un sismo. Se refiere a falsas alarmas.
- **Verdaderos Negativos (VN):** Los casos en los que el modelo predijo correctamente una detección que no es un sismo. Esta categoría no se considera porque el objetivo es analizar sismos verdaderos, no conjuntos negativos.
- **Falsos Negativos (FN):** Los casos en los que el modelo predijo incorrectamente una detección que sí es un sismo. Es decir, son sismos no detectados u omitidos. Esta categoría incluye aquellas detecciones en las que el modelo sólo determinó una de las dos fases del sismo, en otras palabras, una predicción incompleta.

Capítulo 5

Resultados y discusión

En este capítulo se muestran los resultados obtenidos en cada una de las pruebas diseñadas para evaluar el desempeño de EQT. Se presentan las características y necesidades más importantes, destacando los datos clave y los indicadores de rendimiento. Finalmente se discuten las observaciones generales sobre la eficiencia en términos de las métricas estadísticas.

5.1. Secuencia sísmica del 18 al 22 de agosto de 2017, Michoacán

La traza procesada abarcó cinco días completos, del 18 al 22 de agosto de 2017, considerados como la duración total del enjambre. En este período, el catálogo del Servicio Sismológico Nacional (SSN) reportó 86 sismos. En un tiempo de 1 minuto y 52.72 segundos, EQT generó un total de 568 detecciones, que incluyen el 100% de los sismos reportados por el catálogo.

El archivo de salida incluye el tiempo de llegada de ambas fases (o al menos una) con una probabilidad superior a los umbrales especificados para la detección y selección. Estas detecciones tienen una alta probabilidad de ser sismos y, por tanto, deben contrastarse con los eventos confirmados por el catálogo. Sin embargo, dado que las predicciones sobrepasan el número de sismos reportados, se optó por realizar una revisión manual de cada una de las detecciones generadas por el modelo.

La revisión manual de las 568 predicciones permitirá evaluar el rendimiento del modelo en

términos de detección de casos positivos (son sismos) y negativos (no son sismos o son ruido).

La matriz de confusión basada en esta revisión manual se presenta en la Tabla 5.1.

		Datos reales	
		Sismo	No sismo
Predicciones	Sismo	VP = 433	FP = 123
	No sismo	FN = 12	VN = 0

Tabla 5.1: Matriz de confusión obtenida para la secuencia de Michoacán en 2017.

Posteriormente, se pueden calcular métricas como exactitud, precisión, sensibilidad y puntuación F1, que son estadísticas útiles para interpretar los resultados del modelo.

- Con el cálculo de la exactitud manual, se determinó que EQT logró detectar correctamente el 76.23% de los casos:

$$(VP + VN)/(VP + FP + FN + VN) = 433/(433 + 123 + 12) = 433/568 = 0,7623$$

- La precisión es el porcentaje de sismos detectados correctamente entre todos los casos positivos:

$$VP/(VP + FP) = 433/(433 + 123) = 433/556 = 0,7788$$

- La sensibilidad mide la proporción de sismos detectados correctamente entre todos los casos reales en los datos. También se conoce como *recall*, en inglés, o tasa de verdaderos positivos:

$$VP/(VP + FN) = 433/(433 + 12) = 433/445 = 0,9730$$

- La puntuación F1, *F1 score* en inglés, es otra métrica útil como equilibrio por ser la media armónica de la precisión y sensibilidad:

$$F_1 = 2 \left[\frac{Precision * Sensibilidad}{Precision + Sensibilidad} \right] = 2 \left[\frac{(0,7788)(0,9730)}{0,7788 + 0,9730} \right] = 0,8651$$

En un problema de detección de sismos, la ocurrencia de eventos positivos depende de factores como la actividad sísmica en el sitio, la estación de monitoreo, entre otros. En regiones

con menor actividad sísmica, donde la mayor parte de los datos son eventos no sísmicos (ruido) y la proporción de casos positivos puede ser mucho más baja, se considera un problema desequilibrado. Por otra parte, en regiones de alta sismicidad, casos como enjambres sísmicos o secuencias de réplicas, la ocurrencia de eventos sísmicos puede ser mucho más frecuente y por lo tanto las clases están menos desequilibradas. Por esta razón es útil calcular varias métricas para proporcionar una imagen completa sobre el rendimiento del modelo. La Tabla 5.2 recoge los valores de las métricas calculadas.

Tabla 5.2: Valores de métricas obtenidos para la secuencia de Michoacán en 2017.

Exactitud	Precisión	Sensibilidad	Puntuación F1
0.7623	0.7788	0.9730	0.8651

Para ilustrar la calidad de las predicciones del modelo, las Figuras A.1, A.2, A.3 son ejemplos de formas de onda con eventos clasificados como VP, FP y FN, respectivamente. Cada ventana de tiempo tiene una longitud de 60 segundos a 100 muestras por segundo y un filtro pasa bandas de 1 a 45 Hz. Estos gráficos son útiles para visualizar el nivel de confianza de cada predicción a través de una distribución de probabilidad, lo que permite analizar el comportamiento del modelo en términos de la probabilidad de certeza de las predicciones.

En la Figura 5.1 se muestra la distribución de frecuencias de la probabilidad de detección y de selección de las fases P y S para las 433 predicciones correctas (VP). En el eje horizontal, los valores varían de 0 a 1, donde 0 indica una baja probabilidad de certeza y 1 indica una alta probabilidad de certeza. El eje vertical representa el número de eventos predichos en cada rango de probabilidad. En los tres histogramas se puede observar que los VP tienden a tener probabilidades altas dado que la mayoría se identificó con probabilidades que superan el 0.7.

Además, EQT calcula la relación señal-ruido para cada detección al seleccionar las fases. La Figura 5.2 muestra la distribución de la relación señal-ruido para ambas fases. Los datos de la relación señal-ruido en la fase P presentan una mayor dispersión que alcanza cerca de los 50 dB. En la fase S los valores son inferiores y se concentran entre los 6 y 14 dB.

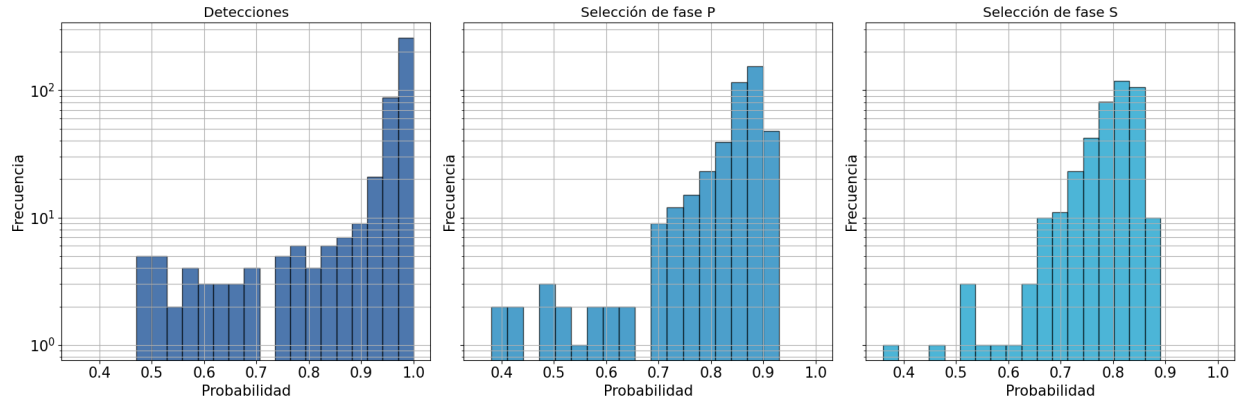


Figura 5.1: Histogramas de probabilidades de los 433 casos Verdaderos Positivos (2017).

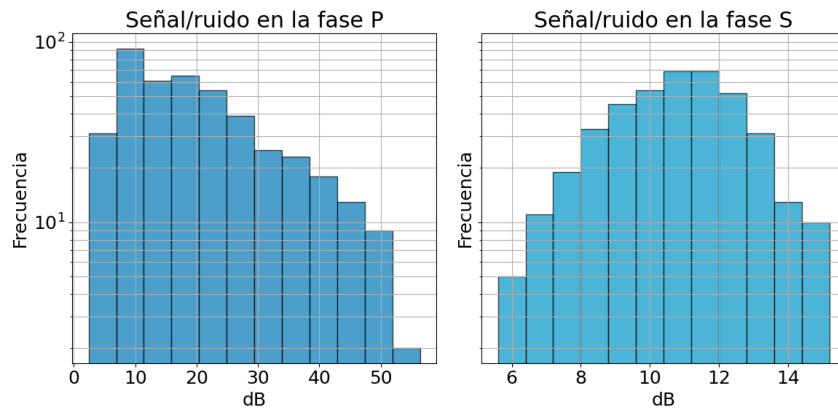


Figura 5.2: Histogramas de la relación señal/ruido de los 433 casos Verdaderos Positivos (2017).

5.2. Secuencia sísmica del 8 al 9 de agosto de 2023, Baja California Sur

Los datos analizados comprenden desde el 8 de agosto a las 19:30:54 hasta el 9 de agosto a las 22:36:18 de 2023, en horario UTC. En un tiempo de 40.91 segundos, los resultados de EQT incluyeron un total de 107 predicciones. El conjunto de datos reales consiste en un catálogo de 91 sismos proporcionado por el Centro Nacional de Prevención de Desastres (CENAPRED), de los cuales el modelo predijo correctamente el 81.32%. Como resultado, se identificaron 33 nuevos posibles sismos, de los cuales sólo siete fueron confirmados en la revisión manual.

Se conformó la matriz de confusión presentada en la Tabla 5.3. Los valores de sus métricas se

encuentran en la Tabla 5.4.

		Datos reales	
		Sismo	No sismo
Predicciones	Sismo	VP = 81	FP = 25
	No sismo	FN = 1	VN = 0

Tabla 5.3: Matriz de confusión obtenida para la secuencia de BCS en 2023.

Tabla 5.4: Valores de métricas obtenidos para la secuencia de BCS en 2023.

Exactitud	Precisión	Sensibilidad	Puntuación F1
0.7570	0.7642	0.9878	0.8617

Se presentan ejemplos de predicciones clasificadas como VP, FP y FN en las Figuras A.4, A.5 y A.6, respectivamente.

La Figura 5.3 muestra la distribución de valores de probabilidad para el conjunto de 81 eventos clasificados como VP, indicando que, aunque la muestra es pequeña, la selección de fases alcanzó predominantemente probabilidades superiores a 0.7. En el caso de la probabilidad de detección hay picos en todo el rango de datos, lo que indica que no hay una tendencia a las altas probabilidades a pesar de ser predicciones correctas. Por otro lado, la Figura 5.4 muestra la distribución de la relación señal-ruido, donde se observa una agrupación de valores para la fase P entre 0 y 20 dB, mientras que para la fase S, la mayoría se concentra entre 2.5 y 7.5 dB.

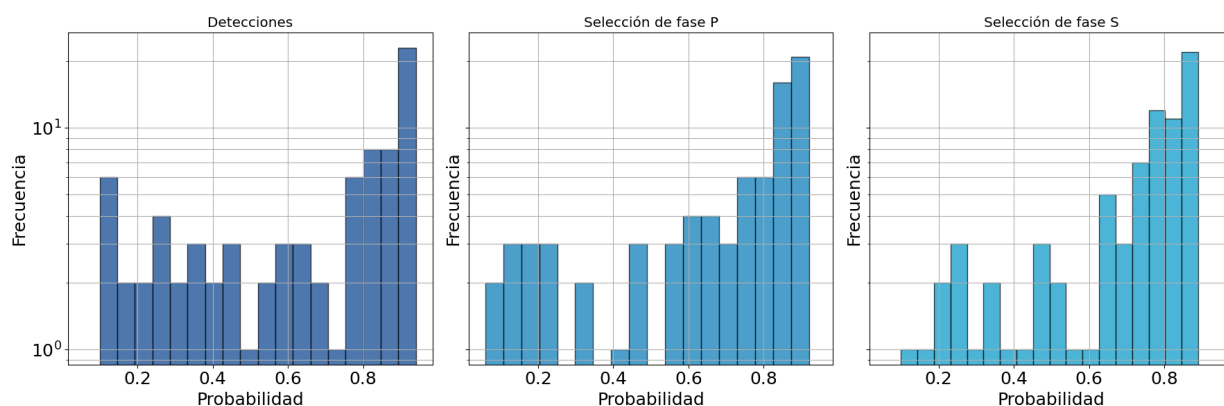


Figura 5.3: Histogramas de probabilidades de los 81 casos Verdaderos Positivos (2023).

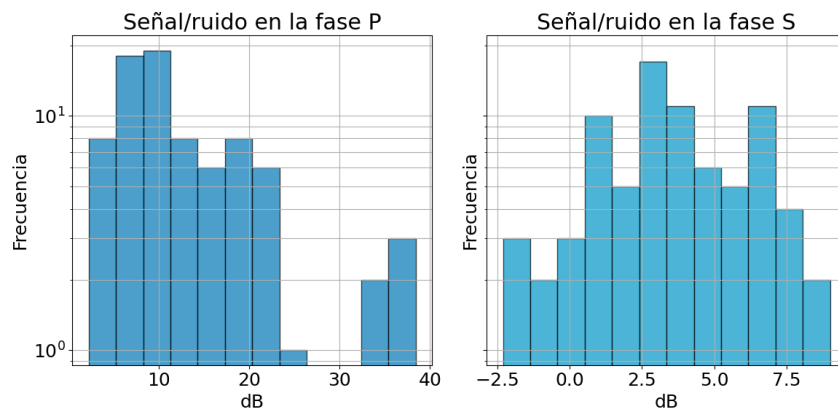


Figura 5.4: Histogramas de la relación señal/ruido de los 81 casos Verdaderos Positivos (2023).

Para comparar la selección manual realizada por los analistas de CENAPRED con la selección automática de EQT en los tiempos de arribo de la fase P, se calculó la diferencia entre las dos series de datos. La Figura 5.5 muestra la distribución de estas diferencias, cuya desviación estándar es de aproximadamente 0.06 segundos. El valor de la desviación estándar sugiere que la magnitud de las diferencias es similar, aunque hay cierta variabilidad, especialmente, y como se puede observar en la figura 5.5, en cuatro predicciones con un error significativo.

Otra medida que indica cuán cerca están las predicciones del valor real en promedio, es el error absoluto medio, una métrica que representa la magnitud promedio de las diferencias entre los valores predichos y los valores observados. Un error promedio bajo indica una alta precisión en las predicciones, mientras que un error promedio alto sugiere que las predicciones tienden a alejarse más de los valores reales. En este caso, un error promedio de 7×10^{-3} segundos indica que las predicciones automáticas de EQT difieren de las selecciones manuales por una cantidad extremadamente pequeña, lo que resalta la precisión de la predicción automática.

5.3. Secuencia de réplicas, sismo M_w 7.2 del 18 de abril de 2014, Guerrero

Se optó por aumentar el porcentaje de traslape debido a que son eventos de mayor duración, ver la Tabla 4.4. Con un menor traslape, cada ventana contiene información distinta, lo que

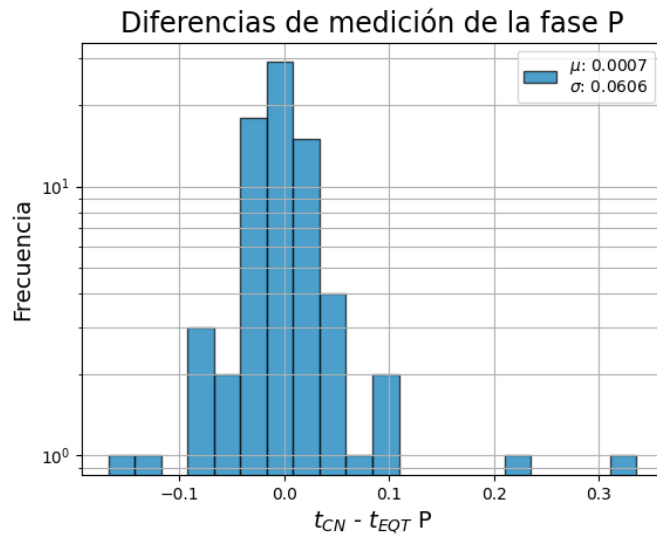


Figura 5.5: Histograma de los tiempos de diferencia medidos para el arribo de la fase P.

fragmenta la detección. Esto conduce a casos de eventos verdaderos donde se registra sólo una de las fases, o bien, son separadas en dos predicciones contiguas, una con el arribo de la onda P y la siguiente con el arribo de la onda S. Sin embargo, este aumento en el traslape implica procesar más datos y obtener más predicciones, por lo que también se incrementaron los umbrales para mitigar la tasa de falsos positivos. Esto no dificultó poder predecir a los eventos más grandes en magnitud, pero se omitieron sismos más pequeños, por ejemplo cuatro de ellos con magnitudes 2.6, 2.8, 2.9 y 3.

El periodo analizado abarcó desde el 18 de abril a las 12:00:00 hasta el 20 de abril a las 00:00:00 de 2014, en horario UTC. En un tiempo de 53.2 segundos, EQT arrojó un total de 1069 detecciones. En este lapso, el catálogo del Servicio Sismológico Nacional (SSN) registró 109 sismos, de los cuales EQT logró detectar el 96.33%. La revisión manual de las predicciones proporcionó los valores para una matriz de confusión que se presenta en la Tabla 5.5. Junto con las métricas asociadas que se presentan en la Tabla 5.6.

Las Figuras A.7, A.8 y A.9 presentan algunos ejemplos de predicciones clasificadas como Verdaderos Positivos (VP), Falsos Positivos (FP) y Falsos Negativos (FN).

En lo referente a la confianza de las predicciones, la Figura 5.6 muestra la distribución de frecuencias de las probabilidades en los 503 casos predichos correctamente. Se observa que

		Datos reales	
		Sismo	No sismo
Predicciones	Sismo	VP = 503	FP = 513
	No sismo	FN = 53	VN = 0

Tabla 5.5: Matriz de confusión obtenida para la secuencia de BCS en 2023.

Tabla 5.6: Valores de métricas obtenidos para la secuencia de réplicas, Guerrero 2014.

Exactitud	Precisión	Sensibilidad	Puntuación F1
0.4705	0.4951	0.9047	0.6399

los valores se concentran después del 0.6 de probabilidad. Asimismo, las tres distribuciones tienden a exhibir altas probabilidades. Otra característica es que en el caso de las fases, las predicciones son inferiores al 0.9 de probabilidad. La Figura 5.7 muestra la distribución de la relación señal/ruido en las fases, revelando que para la fase P los valores se encuentran entre 0 y 40 dB, mientras que para la fase S, los valores son inferiores y se agrupan en el rango de 5 a 15 dB.

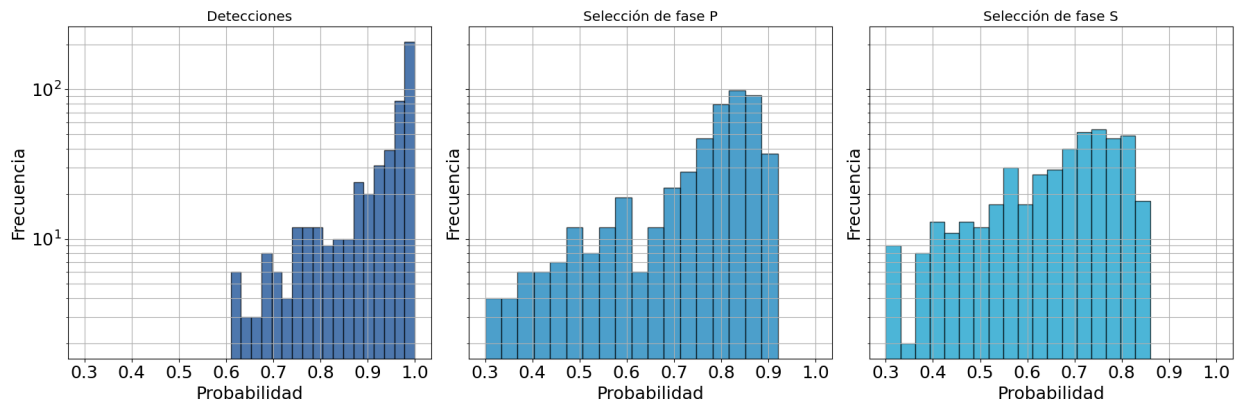


Figura 5.6: Histogramas de probabilidades de los 503 casos Verdaderos Positivos (2014).

5.4. Discusión

El objetivo de la revisión manual fue validar la detección de nuevos posibles sismos y comprender las limitaciones del modelo. Esta etapa proporcionó detalles sobre los aciertos y errores, así como las bases cuantitativas para calcular métricas de desempeño, resumidas en la Tabla 5.7. El criterio de selección de parámetros se enfocó en minimizar los falsos negativos, evitando

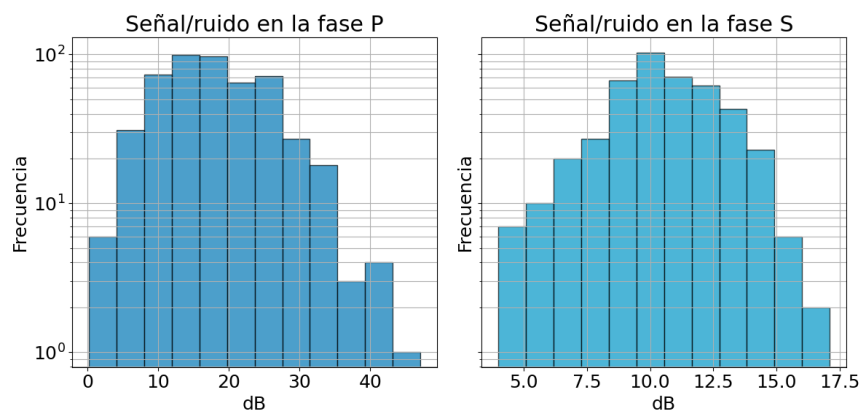


Figura 5.7: Histogramas de la relación señal/ruido de los 503 casos Verdaderos Positivos (2014).

omitir eventos del catálogo, lo que se reflejó en altos porcentajes de predicciones correctas y altos valores de sensibilidad. Esto revela que el modelo es eficiente para detectar un sismo verdadero. Sin embargo, la tasa de falsos positivos fue significativa, lo que redujo la exactitud y precisión. Esta situación indica que la calibración de parámetros es un área crítica para futuras mejoras. Además, la puntuación F1 se vio afectada por los valores bajos de precisión.

Los resultados de las tres pruebas reflejaron variaciones en el desempeño del modelo en diferentes contextos sísmicos, donde la prueba 3 presentó el menor desempeño, especialmente en exactitud, precisión y puntuación F1. En este caso, a pesar de procesar con umbrales más altos en comparación con las otras pruebas, la tasa de falsos positivos no disminuyó, lo que sugiere que el modelo encontró mayores desafíos para efectuar detecciones. Esto podría atribuirse a que en una secuencia de réplicas la magnitud y duración de los eventos se distribuyen en un rango más amplio en comparación con un enjambre sísmico.

En la prueba 2, el archivo de fases proporcionado fue clave para evaluar la precisión del tiempo de arribo de la fase P. La evaluación demostró consistencia en las diferencias entre ambos métodos. La baja magnitud del error promedio indica alta precisión en la selección de fase P.

La Tabla 5.8 muestra que el procesamiento se efectúa en tiempos cortos, lo que es una ventaja para realizar múltiples experimentos variando los parámetros y por ende abordar la calibración de los mismos para mejorar los resultados. Todas las predicciones fueron generadas utilizando un procesador Intel(R) Core(TM) i5-5350U a 1.80GHz de dos núcleos.

Tabla 5.7: Métricas obtenidas en las tres pruebas.

	Aciertos con respecto al catálogo (%)	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Puntuación F1 (%)
Prueba 1	100	76.23	77.88	97.30	86.51
Prueba 2	81.32	75.70	76.42	98.78	86.17
Prueba 3	96.33	47.05	49.51	90.47	63.99

Tabla 5.8: Duración de la traza procesada, número total de predicciones realizadas y tiempo de procesamiento que el modelo requirió para completar esas predicciones.

	Duración de la traza (h)	Número total de predicciones	Tiempo de procesamiento (hh:mm:ss)
Prueba 1	120	568	00:01:52.72
Prueba 2	27	107	00:00:40.91
Prueba 3	36	1069	00:00:53.2

EQT requiere de un preprocesamiento muy simple que solo implica convertir los datos de formato SAC a MSEED, sin necesidad de aplicar filtros adicionales. La descarga e instalación de EQT resulta relativamente sencilla siempre y cuando se gestionen correctamente las versiones de bibliotecas, como Keras o TensorFlow, que deben ser compatibles con la versión de Python recomendada para EQT (3.6 o 3.7).

Capítulo 6

Conclusiones y recomendaciones

Los aspectos más significativos sobre el desempeño del modelo en las tres pruebas incluyen que EQT se desempeñó con variaciones al procesar conjuntos de eventos con diferentes características. Aunque en general se observaron altos porcentajes de predicciones correctas, lo que benefició a la sensibilidad, la tasa de falsos positivos fue considerable, afectando la exactitud, la precisión y la puntuación F1. Por esta razón, se subraya la importancia del ajuste de parámetros. Además, en una de las pruebas fue posible demostrar la alta precisión de las predicciones en términos del tiempo de llegada de la fase P con respecto a un catálogo manual.

El número total de predicciones evidenció la capacidad de EQT para detectar un gran número de posibles eventos que no están presentes en el catálogo manual, lo que implicó una revisión exhaustiva de las predicciones. Para agilizar la eliminación de falsos positivos y evitar revisiones manuales extensas, lo más recomendable en problemas de detección automática es avanzar hacia estrategias que optimicen el manejo de las predicciones. Esto incluye procesos adicionales como el uso de algoritmos de asociación de fases y de localización de epicentros. También se sugiere ampliar el análisis temporal y aumentar el número de estaciones para una cobertura más completa, aprovechando la eficiencia en términos de rendimiento computacional y tiempo.

Finalmente se sugiere explorar la capacidad de EQT como *software* de código abierto para desarrollar un nuevo modelo que pueda ser entrenado con datos propios, aprovechando al máximo su poderosa arquitectura y adaptándolo a las características de los sismos en México.

Apéndice A

Ejemplos de detecciones

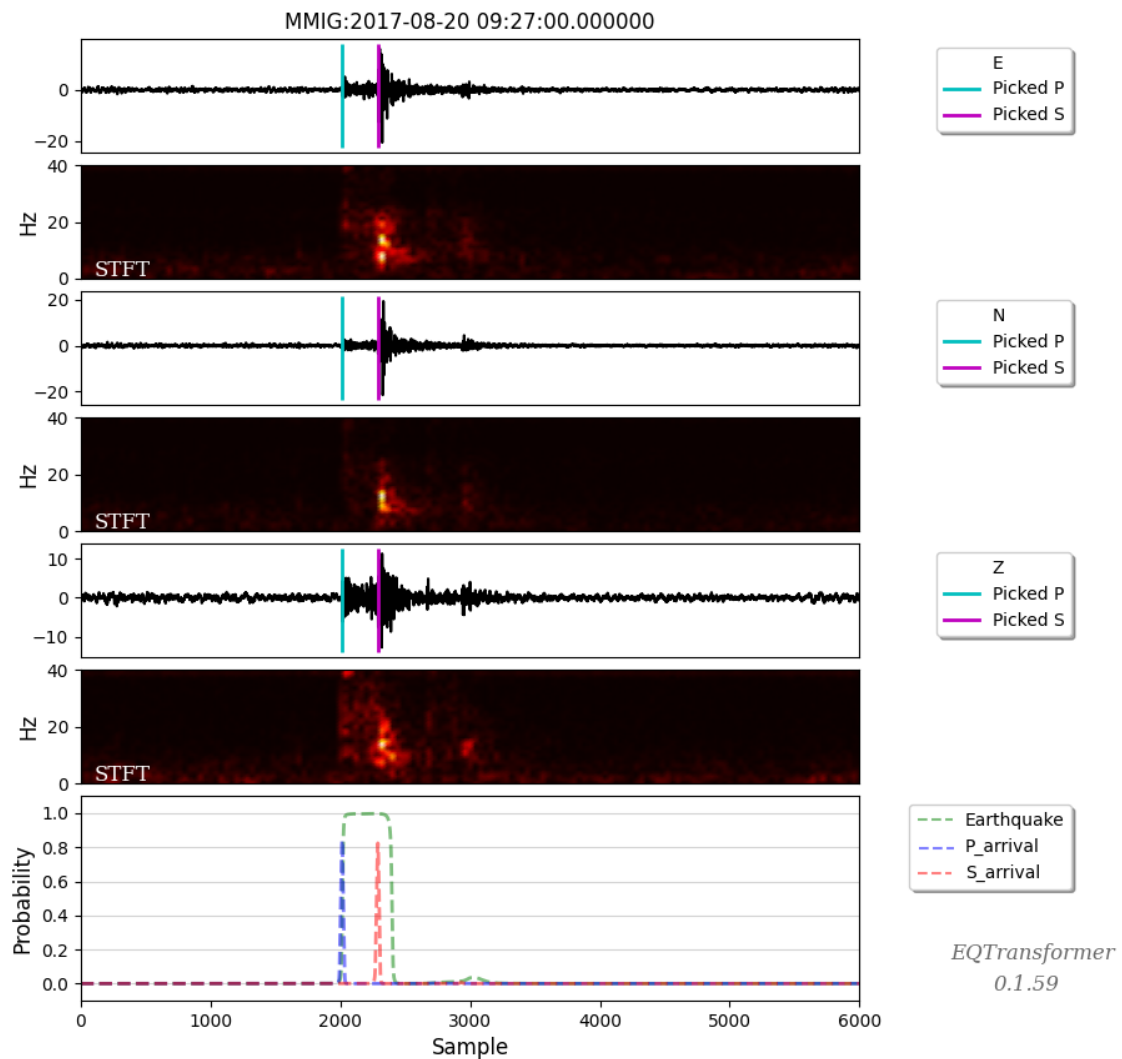


Figura A.1: Ejemplo representativo de un evento VP. De acuerdo con información del catálogo del SSN, es un sismo de magnitud 3.8. Se detectó con 0.93 de probabilidad. La fase P se seleccionó con 0.82 de probabilidad y se estimó una relación señal/ruido de 10.1 dB. La fase S se seleccionó con 0.81 de probabilidad y se estimó una relación señal/ruido de 15.8 dB.

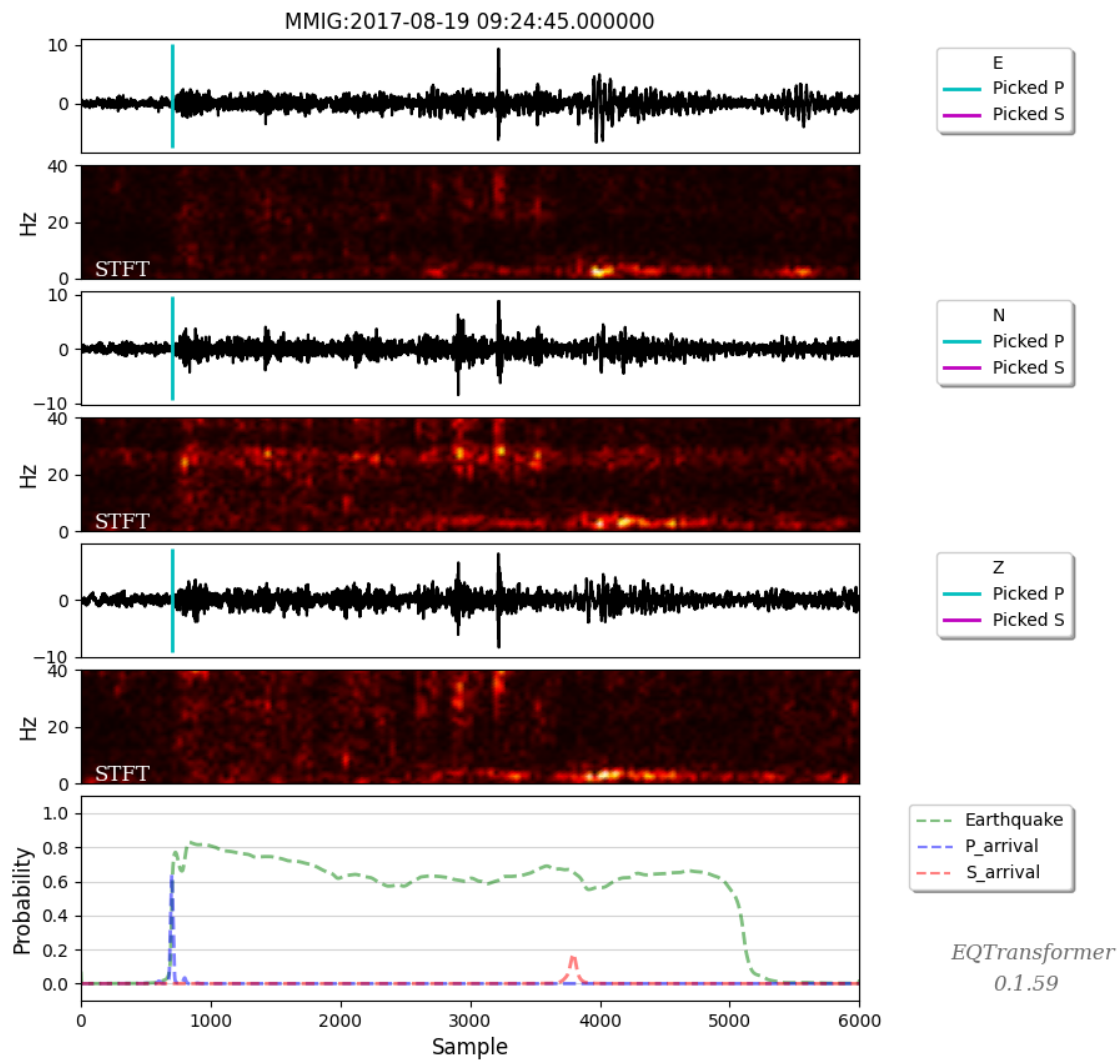


Figura A.2: Evento clasificado como FP. El modelo logró detectarlo con 0.79 de probabilidad. Se descartó la fase S, mientras que la fase P tiene probabilidad de 0.6 y relación señal/ruido de 6.4 dB.

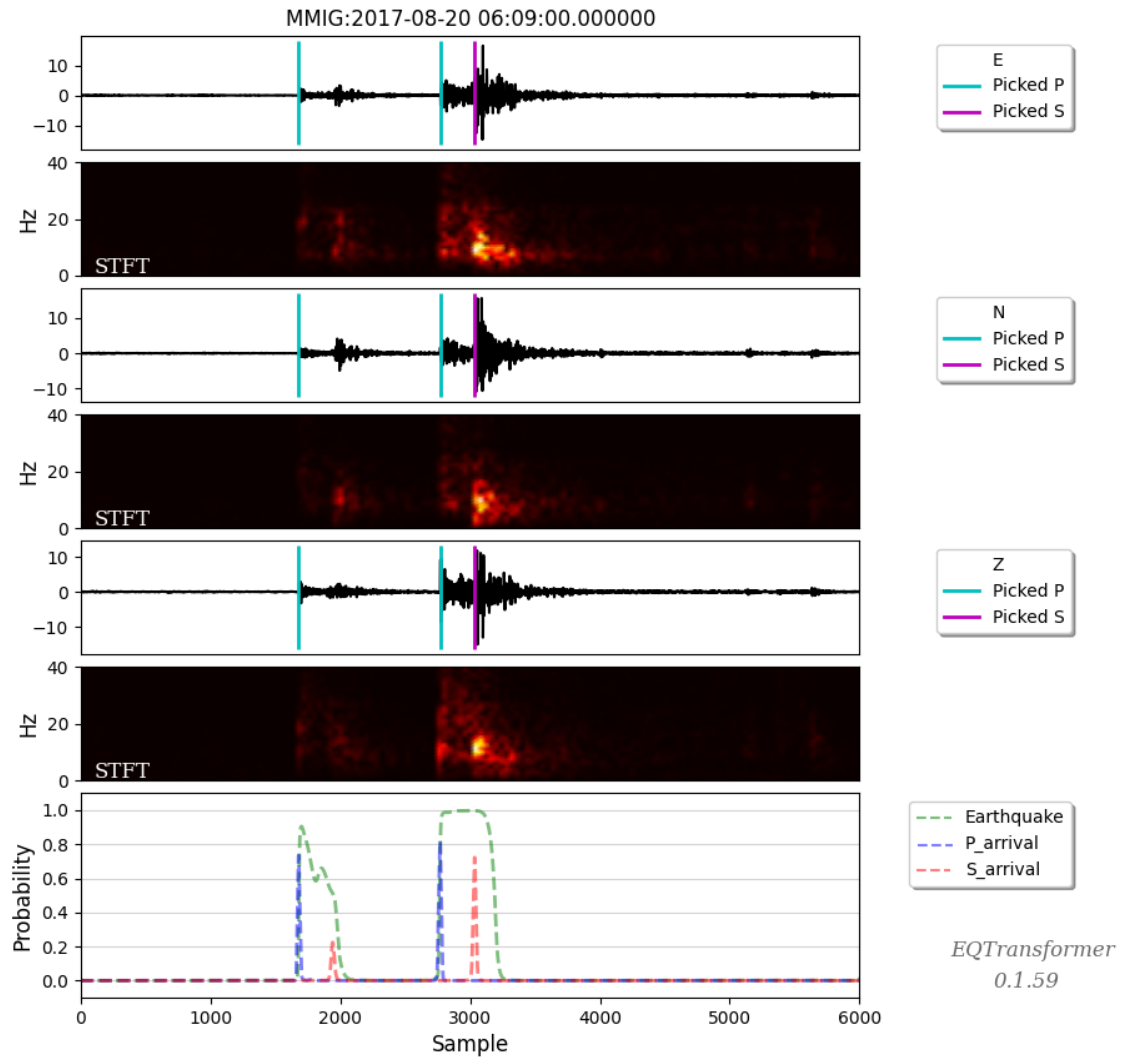


Figura A.3: Ejemplo de una ventana que contiene dos eventos clasificados como FN y VP, respectivamente. El primer evento fue clasificado como FN, ya que no hay datos sobre el arribo de la fase S. La probabilidad de detección es 0.91. La fase P tiene una probabilidad de 0.86 y se calculó una relación señal/ruido de 27.3 dB. El segundo evento fue confirmado por el catálogo con una magnitud de 3.5. Se detectó con 0.97 de probabilidad. La fase P se seleccionó con 0.83 de probabilidad y se estimó una relación señal/ruido de 27.3 dB. La fase S se seleccionó con 0.76 de probabilidad y no se estimó una relación señal/ruido.

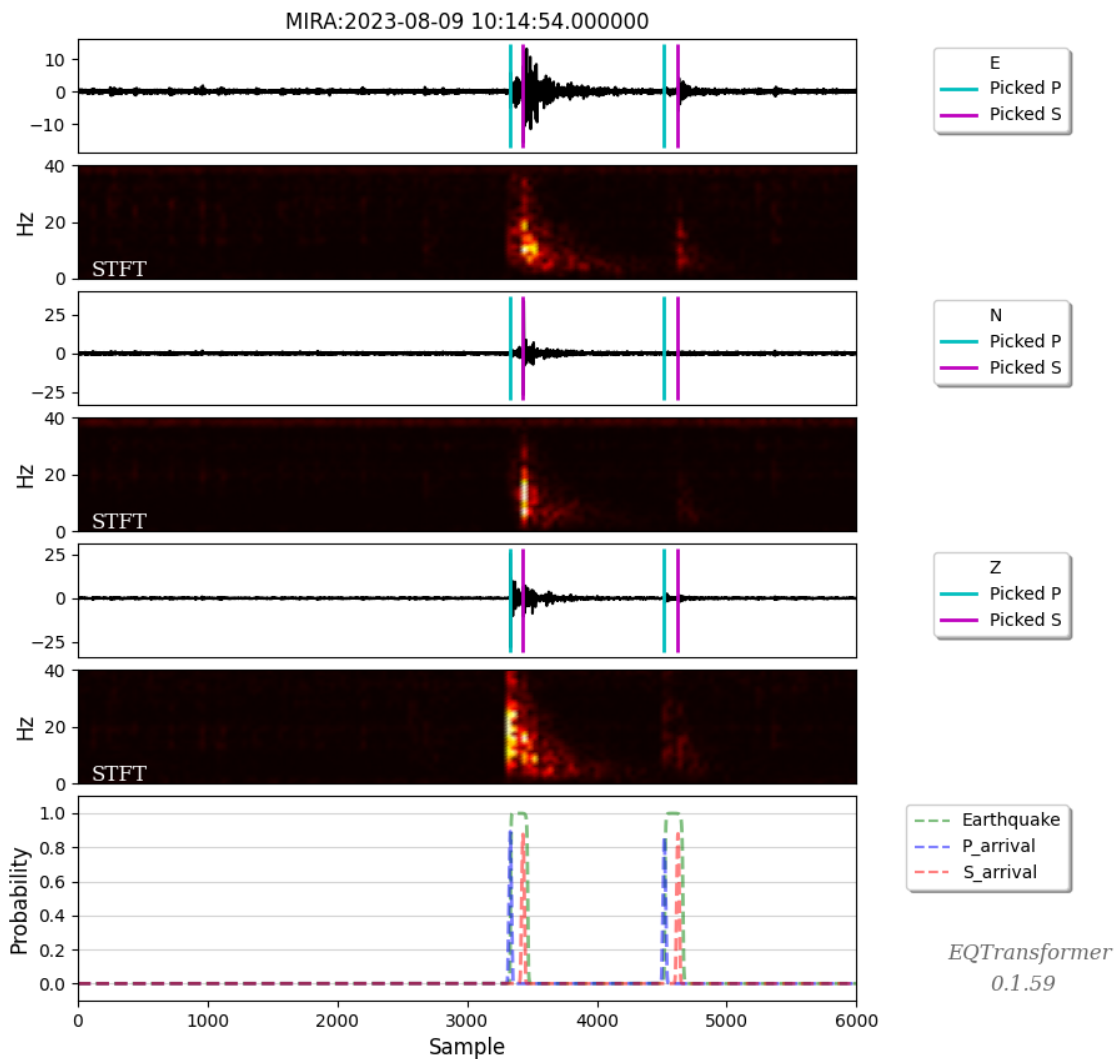


Figura A.4: Ejemplo de una ventana que contiene dos eventos clasificados como VP. En el primer caso, la probabilidad de detección es 0.94; con probabilidad de selección de fases (P): 0.92, (S): 0.88. El segundo evento se detectó con 0.92 de probabilidad; con probabilidad de selección de fases (P): 0.87, (S): 0.89., respectivamente. En ambos casos, relación señal/ruido (P): 20.9 dB, (S): 2.8 dB.

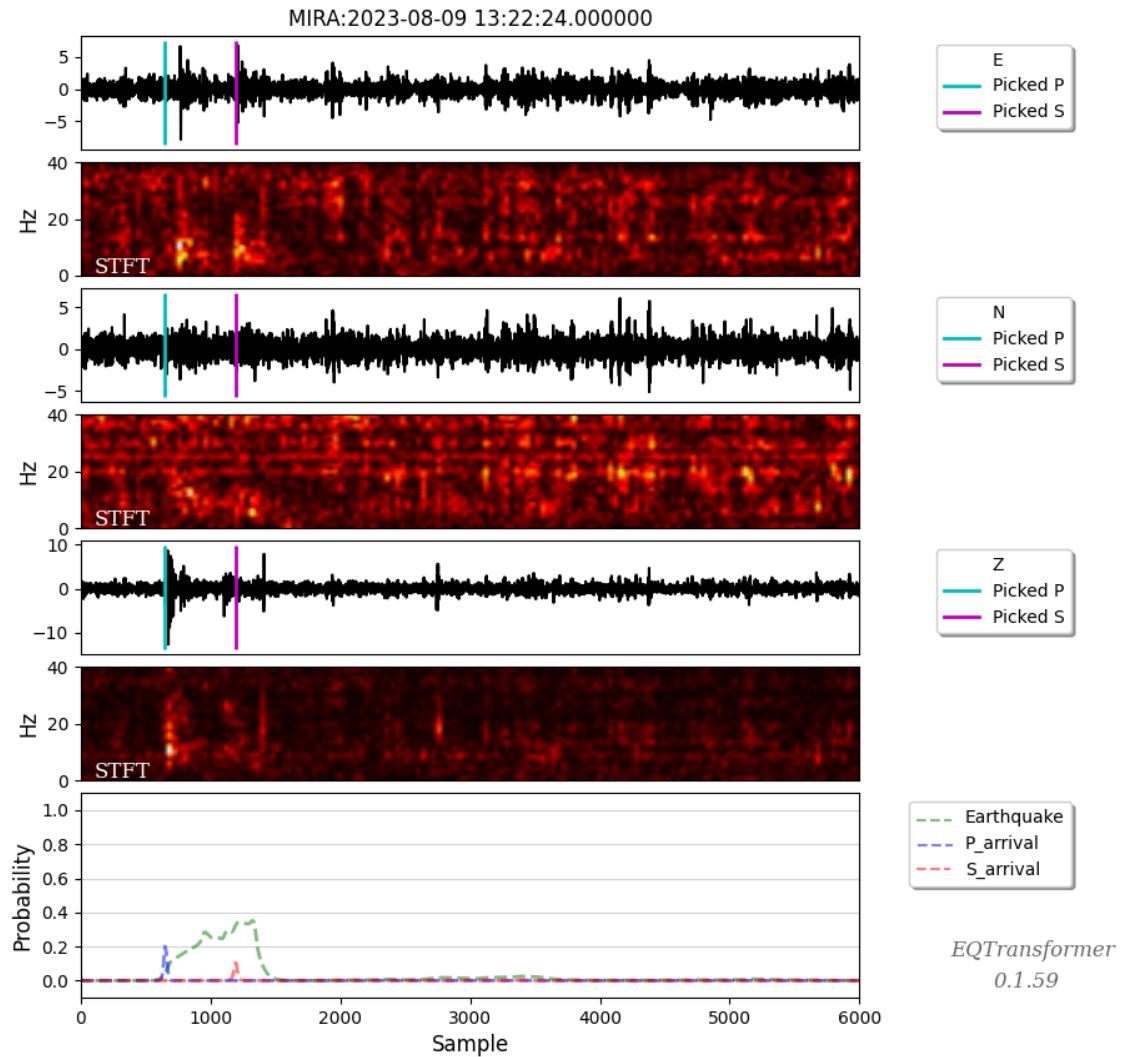


Figura A.5: Ejemplo de evento no sísmico, se clasificó como FP y no se dispone de datos sobre el arribo de la fase P. El modelo lo detectó con una probabilidad de 0.16. La llegada de la fase S se determinó con probabilidad de 0.08, y la relación señal/ruido se estimó en 0.1 dB.

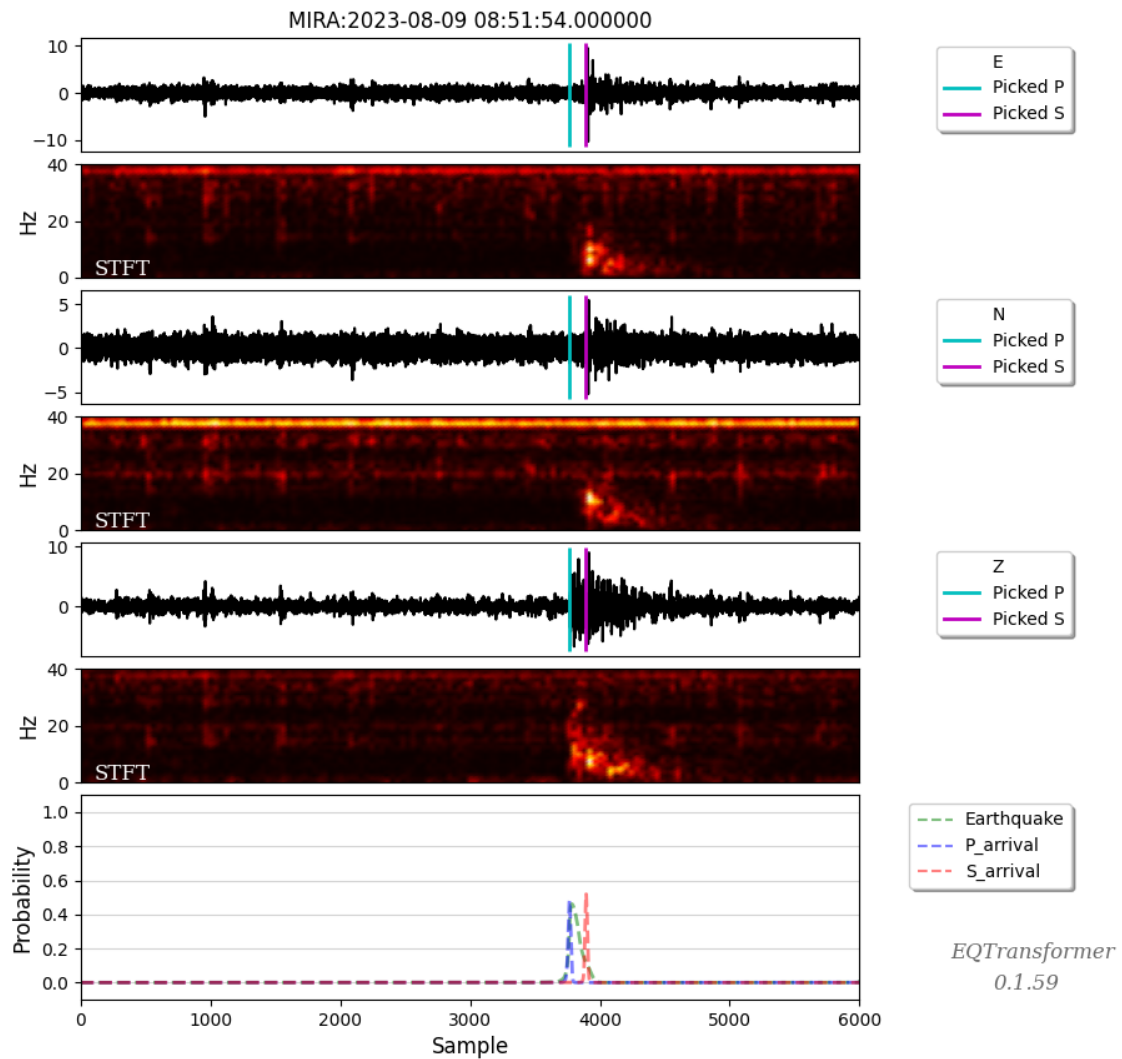


Figura A.6: Ejemplo de un evento verdadero que fue clasificado como FN porque el modelo seleccionó únicamente la fase P. Tiene una probabilidad de detección de 0.4 y probabilidad de selección de fase (P): 0.45 con relación señal/ruido de 7.2 dB.

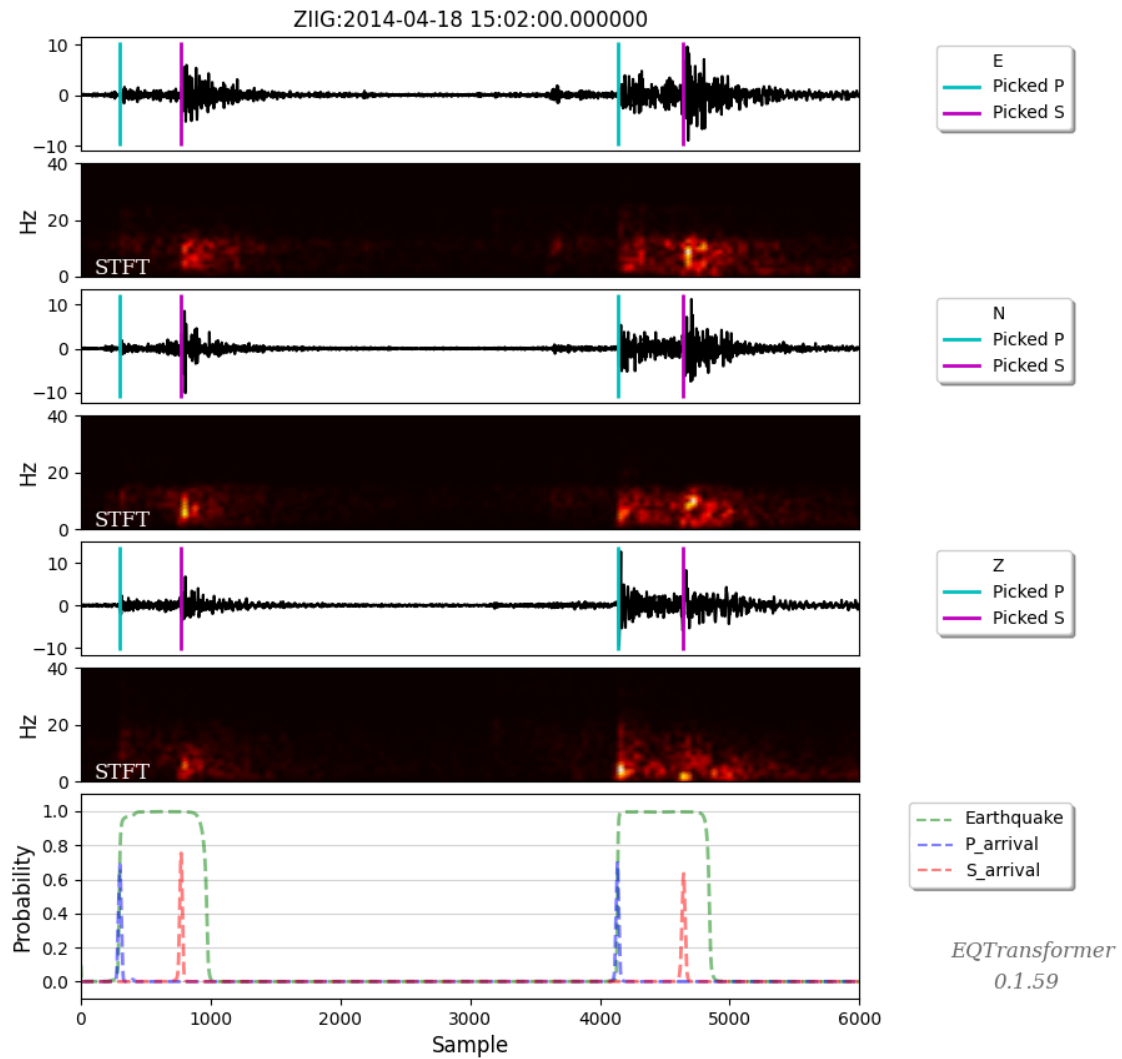


Figura A.7: Dos eventos que el modelo detectó correctamente (VP). El catálogo indica que sus magnitudes son 2.7 y 2.6, respectivamente. En el primer evento, la probabilidad de detección es 0.96; probabilidad de selección de fases (P): 0.74, (S): 0.77; relación señal/ruido (P): 9.5 dB, (S): 12.6 dB. El segundo evento se detectó con 0.98 de probabilidad; con probabilidad de selección de fases (P): 0.78, (S): 0.65; y relación señal/ruido (P): 10.1 dB, (S): 12.7 dB.

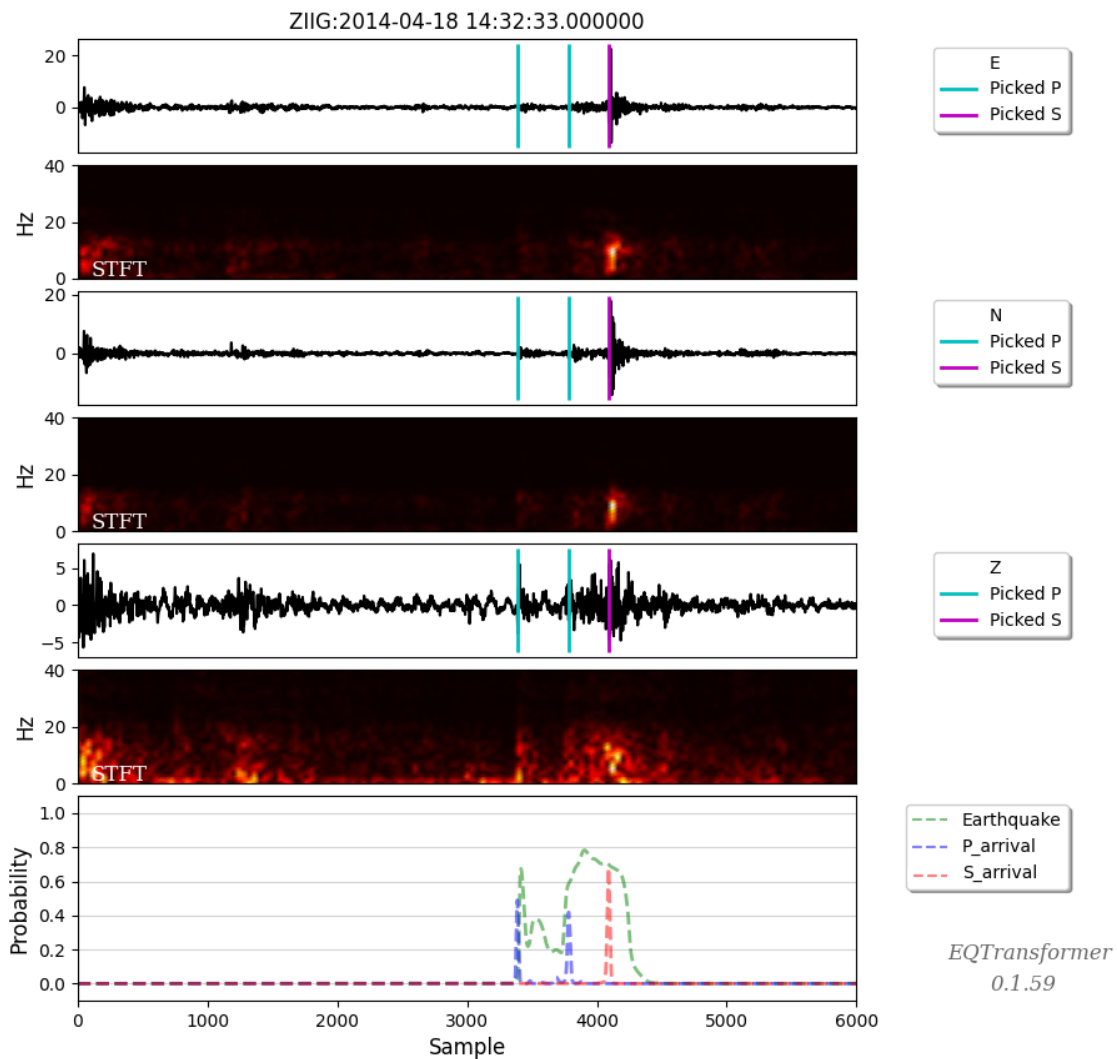


Figura A.8: Ejemplo de dos detecciones incorrectas (FP). En el primer caso, la probabilidad de detección es de 0.65, sólo se seleccionó la fase P con 0.49 de probabilidad y un valor de 8.3 dB en la relación señal/ruido. El segundo evento tiene como probabilidad de detección: 0.7; probabilidad de selección de fases (P): 0.42, (S): 0.68; relación señal/ruido (P): 8.3 dB.

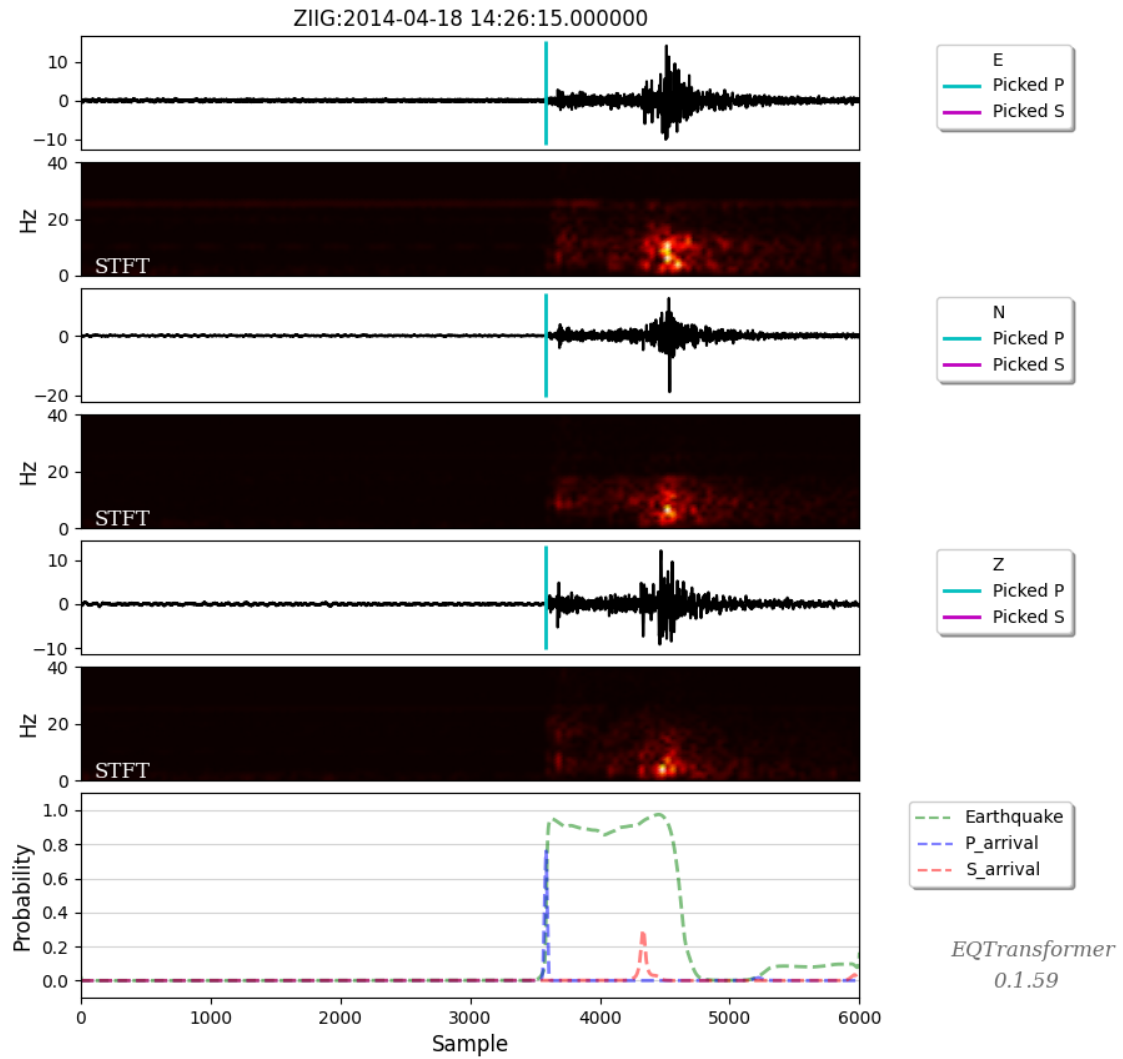


Figura A.9: Ejemplo de un evento clasificado como FN. No cuenta con información sobre la llegada de la fase S. Probabilidad de detección: 0.9; probabilidad de selección de fase (P): 0.76, relación señal/ruido (P): 13.8 dB.

Referencias

- Aaron, M. M. (2021). *Ley de Omori para sismos premonitorios y réplicas ocurridos en el valle de Atrato: Sismo de Murindó-Antioquia 1992*. <https://manglar.uninorte.edu.co/bitstream/handle/10584/11512/1001913482.pdf?sequence=1&isAllowed=y>. (Recuperado el 15 de septiembre de 2023)
- Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, 68(5), 1521–1532.
- Alpaydin, E. (2021). *Machine Learning*. Mit Press.
- BCS, N. (2023, Sep). Reportan enjambre sísmico en zonas de San José del Cabo; registran hasta 15 temblores por día. *BCSNoticias*. Descargado de <https://www.bcsnoticias.mx/reportan-enjambre-sismico-en-zonas-de-san-jose-del-cabo-registran-hasta-15-temblores-por-dia/>
- Bishop, C. M., y Bishop, H. (2023). Transformers. En *Deep learning*. Springer Cham.
- Bormann P, E. B. y. K. R. (2012). New manual of seismological observatory practice 2 (nmsop2). En (cap. Seismic Wave Propagation and Earth models). Potsdam : Deutsches GeoForschungsZentrum GFZ.
- Brown, S. (2021). *Machine learning, explained*.
- Cárdenas, M., Chávez-García, F. J., y Gusev, A. (1997). Regional amplification of ground motion in central Mexico. Results from coda-length magnitude data and preliminary modeling. *Journal of Seismology*, 1, 341–355.
- Castro, R., Stock, J., Hauksson, E., y Clayton, R. (2017). Active tectonics in the Gulf of California and seismicity ($M > 3.0$) for the period 2002–2014. *Tectonophysics*, 719-720, 4-16. Descarga-

do de <https://www.sciencedirect.com/science/article/pii/S0040195117300707>
(From continental to oceanic rifting in the Gulf of California) doi: <https://doi.org/10.1016/j.tecto.2017.02.015>

- Cervantes, A. (2023, May). Enjambre sísmico sacude Loreto, van 17 réplicas. *Tribuna de México*. Descargado de <https://tribunademexico.com/enjambre-sismico-sacude-loreto-van-17-replicas/>
- González, E., Mayer, S., y Aguirre, A. (2005). Reporte de un enjambre sísmico de febrero-abril de 2004, en la Bahía de La Paz, Baja California Sur, México. *Geos*, 24, 451–456.
- Guglielmi, A. V. (2017). Omori's law: A note on the history of geophysics. *Physics-Uspexhi*, 60(3), 319.
- Haykin, S. (2009). *Neural networks and learning machines*, 3/e. Pearson Education India.
- Jiang, C., Fang, L., Fan, L., y Li, B. (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science*, 34(5), 425–435.
- Karasözen, E., y Karasözen, B. (2020). Earthquake location methods. *GEM-International Journal on Geomathematics*, 11, 1–28.
- Kennett, B. (2007). Seismic phases. En D. Gubbins y E. Herrero-Bervera (Eds.), *Encyclopedia of geomagnetism and paleomagnetism* (pp. 903–908). Dordrecht: Springer Netherlands. Descargado de https://doi.org/10.1007/978-1-4020-4423-6_290 doi: 10.1007/978-1-4020-4423-6_290
- Kostoglodov, V., y Pacheco, J. F. (1999). *Cien años de sismicidad en México*. <https://usuarios.geofisica.unam.mx/vladimir/sismos/100a%F1os.html>.
- Kostoglodov, V., Singh, S. K., Santiago, J. A., Franco, S. I., Larson, K. M., Lowry, A. R., y Bilham, R. (2003). A large silent earthquake in the Guerrero seismic gap, Mexico. *Geophysical Research Letters*, 30(15).
- Kulhánek, O. (2012). *Anatomy of Seismograms: For the IASPEI/Unesco Working Group on Manual of Seismogram Interpretation*. Elsevier.
- Kulhanek, O., y Persson, L. (2011). Seismogram interpretation. En H. K. Gupta (Ed.), *Encyclopedia*

- of solid earth geophysics* (pp. 1315–1324). Dordrecht: Springer Netherlands. Descargado de https://doi.org/10.1007/978-90-481-8702-7_219 doi: 10.1007/978-90-481-8702-7_219
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436–444.
- Legrand, D., Pertou, M., Macías, J., Siebe, C., Pacheco, J., Chacón, F., ... Cisneros, G. (2023). Repeated seismic swarms near Parícutin volcano: precursors to the birth of a new monogenetic volcano in the Michoacán-Guanajuato volcanic field, México? *Bulletin of Volcanology*, 85(5), 30.
- Lomax, A., Michelini, A., Curtis, A., y Meyers, R. (2009). Earthquake location, direct, global-search methods. *Encyclopedia of complexity and systems science*, 5, 2449–2473.
- Manzunzu, B., Brandt, M., Midzi, V., Durrheim, R., Saunders, I., y Mulabisana, T. (2021). Towards a homogeneous moment magnitude determination for earthquakes in South Africa: Reduction of associated uncertainties. *Journal of African Earth Sciences*, 173, 104051. Descargado de <https://www.sciencedirect.com/science/article/pii/S1464343X20303022> doi: <https://doi.org/10.1016/j.jafrearsci.2020.104051>
- Medrano, G. (2020, Oct). La Paz, BCS, bajo los efectos de un enjambre sísmico. *Excelsior*. Descargado de <https://www.excelsior.com.mx/nacional/la-paz-bcs-bajo-los-efectos-de-un-enjambre-sismico/1403655>
- Mousavi, S. M. (2020). *Welcome to EQTransformer's documentation!* <https://eqtransformer.readthedocs.io/en/latest/index.html>. (Recuperado el 12 de abril de 2023)
- Mousavi, S. M., y Beroza, G. C. (2018). Evaluating the 2016 one-year seismic hazard model for the central and eastern United States using instrumental ground-motion data. *Seismological Research Letters*, 89(3), 1185–1196.
- Mousavi, S. M., y Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47(1), e2019GL085976.
- Mousavi, S. M., y Beroza, G. C. (2022). Deep-learning seismology. *Science*, 377(6607), eabm4470.
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., y Beroza, G. C. (2020). Earthquake Transformer—an attentive deep-learning model for simultaneous earthquake detection

- and phase picking. *Nature communications*, 11(1), 3952.
- Mousavi, S. M., Sheng, Y., Zhu, W., y Beroza, G. C. (2019). Stanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, 7, 179464–179476.
- Pacheco, J. F., y Singh, S. K. (2010). Seismicity and state of stress in Guerrero segment of the Mexican subduction zone. *Journal of Geophysical Research: Solid Earth*, 115(B1).
- Pérez-Campos, X., Kim, Y., Husker, A., Davis, P. M., Clayton, R. W., Iglesias, A., ... Gurnis, M. (2008). Horizontal subduction and truncation of the Cocos Plate beneath central Mexico. *Geophysical research letters*, 35(18).
- Plata-Martínez, R., Ide, S., Shinohara, M., Garcia, E., Mizuno, N., Dominguez, L., ... others (2021). Shallow slow earthquakes to decipher future catastrophic earthquakes in the Guerrero seismic gap. *Nature communications*, 12(1), 3976.
- Rondero, L. (2021, Jul). Aumentan réplicas del enjambre sísmico en BCS. *Metrópoli*. Descargado de <https://metropolimx.com/aumentan-replicas-del-enjambre-sismico-en-bcs/>
- Servicio Sismológico Nacional, UNAM. (2014). *Reporte de Sismo: Sismo del día 18 de abril de 2014, Guerrero (M 7.2)*. <http://www.ssn.unam.mx/>.
- Servicio Sismológico Nacional, UNAM. (2017). *Secuencia sísmica del 18 al 22 de agosto de 2017, Michoacán*. <http://www.ssn.unam.mx/>.
- Servicio Sismológico Nacional, UNAM. (2023a). *Catálogo de Sismos*. <http://www.ssn.unam.mx>. (Extraído de <http://www2.ssn.unam.mx:8080/catalogo/>) doi: <https://doi.org/10.21766/SSNMX/EC/MX>
- Servicio Sismológico Nacional, UNAM. (2023b). *MX Seismic Network*. <http://www.ssn.unam.mx>. doi: <https://doi.org/10.21766/SSNMX/SN/MX>
- Shearer, P. M. (2019). *Introduction to Seismology*. Cambridge University Press.
- SSN. (2011). *Nota informativa del enjambre sísmico del 11 de enero de 2011, Golfo de California*. <http://www.ssn.unam.mx/>.
- SSN. (2016). *Secuencia sísmica del 25 al 28 de marzo de 2016, Golfo de California*. <http://www.ssn.unam.mx/>.

- SSN. (2017). *Reporte de enjambre sísmico del día 23 y 24 de julio de 2017 en el Golfo de California*. <http://www.ssn.unam.mx/>.
- SSN. (2018). *Reporte de enjambre sísmico del día 27 al 28 de julio de 2018 en el Golfo de California*. <http://www.ssn.unam.mx/>.
- SSN. (2019). *Reporte de enjambre sísmico del día 10 al 12 de diciembre de 2019 en el Golfo de California*. <http://www.ssn.unam.mx/>.
- Stein, S., y Wysession, M. (2003). *An Introduction to Seismology, Earthquakes*. London: Bantam.
- Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., . . . others (2022). A review of earth artificial intelligence. *Computers & Geosciences*, 105034.
- Vaezi, Y., y Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 203(3), 1896–1908.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Yoon, C. E., O'Reilly, O., Bergen, K. J., y Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science advances*, 1(11), e1501057.
- Yáñez, B. I. (2022). *Continúan Los Enjambres Sísmicos en BCS*. <https://tribunademexico.com/enjambres-sismicos-bcs/>.