



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO



INSTITUTO TECNOLÓGICO DE LA PAZ
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
MAESTRÍA EN SISTEMAS COMPUTACIONALES

**CARACTERIZACIÓN DE PARÁMETROS
DE SEÑALES SÍSMICAS MEDIANTE
INTELIGENCIA ARTIFICIAL**

TESIS

QUE PARA OBTENER EL GRADO DE
MAESTRO EN SISTEMAS COMPUTACIONALES

PRESENTA:

HÉCTOR MICHELLE CARBAJAL CORTÉS

DIRECTOR DE TESIS:

DR. ISRAEL MARCOS SANTILLÁN MÉNDEZ

DRA. DANA DIERNA CARCIMARU

DR. ROBERTO ORTEGA RUIZ

Agradecimientos

Esta tesis se realizó dentro del marco de Ciencia de Frontera. Modalidad: Paradigmas y Controversias de la Ciencia 2022”, en el proyecto 319664, titulado "Fluctuaciones estacionales y cíclicas de los periodos dominantes del suelo del valle de México: un postulado de cambio". Agradecemos al CONHACYT por su apoyo.

Los datos utilizados en este trabajo fueron proporcionados amablemente por la "Red Sísmica del Valle de México (RSVM; Quintanar et al., 2018)", perteneciente al Servicio Sismológico Nacional. El mantenimiento de las estaciones, adquisición de datos y distribución de los mismos es gracias a su personal. El apoyo financiero es proporcionado por el Consejo Nacional de Ciencia y Tecnología, la Universidad Nacional Autónoma de México y el Gobierno de la Ciudad de México.

Índice general

INTRODUCCIÓN	5
1.1 Antecedentes	6
1.2 Descripción del problema	7
1.3 Justificación	7
1.4 Hipótesis	8
1.5 Objetivos	8
1.5.1 Objetivo general	8
1.5.2 Objetivos específicos	8
1.6 Alcances y limitaciones	8
MARCO TEÓRICO	9
2.1. Señales Sísmicas	9
2.1.1 Fases Sísmicas	9
2.2. Ruido Sísmico	10
2.3. Registro Sísmico	10
2.4. Tectónica de Placas	11
2.4.1. Tectónica Global	11
2.4.2. Tectónica en México	13
2.5. Sismos en México	15
2.6. Sismometría	18
2.6.1 Historia de la Sismometría	19
2.7. Servicio Sismológico Nacional	21
2.8.1. Técnica STA/LTA (Short-Term Average to Long-Term Average)	22
2.8.2. Técnicas Espectrales	24
2.8.3. Intensidad de Arias	25
2.8.4. Moving-Window Cross-Spectral Analysis	26
2.8.5. Visualización de los sismos mediante analistas expertos	27
2.9. Aprendizaje Automático	28
2.9.1. Tipos de Aprendizaje Automático	28

2.9.2. Perceptrón.....	30
2.9.3. Redes Neuronales	31
2.9.4. Redes Neuronales Transformadoras	31
2.9.5 Reducción de dimensionalidad.....	32
2.9.6 Métricas de clasificación	36
RECURSOS DE SOFTWARE.....	39
3.1. Herramientas Utilizadas	39
3.1.1. Python.....	39
3.1.2. Jupyter Notebook	39
3.1.3. Keras.....	40
3.1.4. Pandas	40
3.1.5. HDF5	40
3.1.6. NumPy	41
3.1.7. Matplotlib.....	41
3.2. Bibliotecas y Herramientas de Python para el procesamiento de registros sísmicos	41
3.2.1. EQTransformer	41
3.2.2. MSNoise	45
3.2.3. ObsPy	46
METODOLOGÍA	47
4.1. Cambios de velocidad de la propagación de las ondas sísmicas.....	47
4.2. Datos	50
PROCESAMIENTO DE LOS DATOS.....	55
5.1. Gráficos MSNoise.....	57
CONCLUSIONES.....	62
BIBLIOGRAFÍA.....	66
APÉNDICE 1	68
APÉNDICE 2	72
APÉNDICE 3	99

Capítulo 1

INTRODUCCIÓN

Un sismo, también conocido como terremoto o temblor, es un fenómeno natural que se produce debido a la liberación repentina de energía acumulada en el interior de la Tierra. Esta liberación de energía genera ondas sísmicas que se propagan a través de la corteza terrestre, causando vibraciones y movimientos en la superficie. Los sismos pueden tener diversas magnitudes y pueden ser clasificados en diferentes categorías, dependiendo de su intensidad y efectos.

La sismología es la rama de la geología y la geofísica que se dedica al estudio de los sismos, terremotos y las ondas sísmicas que se generan a raíz de ellos. Los sismólogos investigan una amplia gama de aspectos relacionados con los movimientos y las vibraciones de la Tierra. Los objetivos específicos de la sismología incluyen:

- Comprender la generación de sismos
- Caracterizar la estructura interna de la Tierra
- Determinar la magnitud y la intensidad de los sismos
- Predecir y mitigar riesgos sísmicos

El profundo impacto de la sismología en la comprensión de nuestro planeta se debe a que cada uno de sus hallazgos está estrechamente vinculado a los avances tecnológicos que han ocurrido a lo largo del tiempo.

Gran parte de los progresos alcanzados en las últimas dos décadas se fundamentan en gran medida en la implementación de redes de instrumentación y comunicación cada vez más extensas, tanto a nivel regional como global. Diversos aspectos clave en el desarrollo de la sismología están estrechamente relacionados con los avances tecnológicos.

A medida que la tecnología ha avanzado, se han desarrollado métodos más sofisticados y precisos para medir los terremotos, como los sismómetros modernos y las redes de monitoreo sísmico que utilizan estaciones distribuidas globalmente para registrar y analizar los eventos sísmicos en tiempo real.

Es interesante ver cómo la medición de los sismos ha evolucionado a lo largo de la historia, pasando de observaciones visuales y escalas cualitativas a instrumentos más precisos y sistemas de monitoreo avanzados.

1.1 Antecedentes

En 2019 el CICESE recibió un proyecto de CONACYT para estudiar los efectos temporales de sitio en el Valle de México. La idea principal era tratar de conocer si los periodos dominantes varían a lo largo del año. Un periodo dominante es una característica geotécnica muy importante para conocer la dinámica del sitio útiles para elaborar códigos de construcción sismorresistentes. El estudio consistía en utilizar la red sísmica del Valle de México para poder hacer análisis continuo durante varios años y conocer si las características dinámicas del sitio no cambian debido a que estas mediciones normalmente se hacen en cualquier momento del año y se desconoce completamente si algunas de estas mediciones pueden ser diferentes en época de verano o invierno especialmente en el Valle de México donde se encuentran depósitos arcillosos altamente saturados que en ocasiones aumenta su nivel de agua.

En un principio los resultados preliminares encontramos que los periodos sobre todo las velocidades tenían ciertas variaciones a lo largo del año. Al realizar los estudios con un mayor detalle, queríamos saber si había algún cambio en la profundidad o bien en los periodos dominantes que rigen el comportamiento del diseño sísmico eran cambios significativos. Afortunadamente los resultados arrojaron que los periodos dominantes no cambian, esto significa que los códigos de construcción se pueden utilizar en cualquier parte del año, sin embargo, las variaciones de velocidad y los espesores de los sedimentos están cambiando constantemente. Esta observación nos llevo a tratar de buscar cual era la razón por la cual estos valores están cambiando a lo largo del año. Sin embargo, aunque los resultados no eran completamente satisfactorios observábamos que a pesar de haber una tendencia de muy bajo orden en algunos momentos los cambios de velocidad dv/v no correspondían completamente a las correlaciones que esperábamos del aporte de lluvia o bien de recarga geohidrológica hacia la cuenca del Valle de México. Debido a esta dificultad observamos que los datos del Valle de México son datos que a pesar de tener una gran cantidad de nivel de ruido mucha de esta información corresponde a eventos transientes que representan exclusivamente eventos sísmicos de magnitud menor que en su mayoría son eventos de magnitud negativa que alteran fuertemente los resultados de la técnica que estábamos utilizando. Nuestro interés era precisamente tener datos de muy alta calidad, pero de ruido sísmico. En esta ocasión nos encontrábamos con un problema muy diferente al que se había presentado en la literatura de aprendizaje automático en la sismología: la segmentación.

En un estudio más detallado, observamos que la verdadera relación que hace que cambien estos parámetros está fuertemente ligado a los cambios hidrometeorológicos. Al mismo tiempo que realizábamos este estudio nos pusimos en contacto con grupos de la Universidad de Harvard y la Universidad de Washington en Estados Unidos de Norte América quienes estaban interesados en realizar un estudio muy similar. Ellos tenían la experiencia de haber utilizado este tipo de técnicas en el Sur de California y habían descubierto que tenían la capacidad de ver cambios geohidrológicos en zonas áridas de los Valles de California.

En general la segmentación se trata de poder identificar los límites en los cuales nuestro espacio de solución se va a poder identificar. En computación la segmentación para el caso de imágenes, se trata de encontrar los límites espaciales donde se puede hacer algún tipo de identificación basado en su forma, por ejemplo, poder identificar automóviles, personas, edificios, etc. En el caso de registros sísmicos la segmentación se refiere a poder identificar el tiempo en el cual ocurre un registro sísmico, es decir la ventana temporal donde inicia y termina el registro sísmico y por consiguiente donde continúa la parte de ruido ambiental. En nuestro caso la segmentación está enfocada en poder distinguir cuál es el registro de vibración ambiental que vamos utilizar para poder separar el ruido sísmico. No conocemos en la literatura ningún caso de segmentación en sismología para utilizar el ruido sísmico, mucho menos utilizando inteligencia artificial directamente mediante técnicas de aprendizaje automático. Después de varios intentos para poder construir una red neuronal que tenga la mejor capacidad de reconocimiento, descubrimos que las redes neuronales transformantes que además tienen partes convolucionales y partes de atención específica son las más exitosas. Por esa razón descartamos otro tipo de redes neuronales que no tienen el mismo desempeño que las redes transformantes. En la literatura existen diversos ejemplos de redes para esta detección. La red más exitosa es EQTransformer, pero tiene la desventaja que solo ha sido probada para localización de sismos por esta razón utilizamos algún tipo de análisis preliminar que pudiera identificar si nuestra detección es adecuada para los fines de reconocimiento de vibración ambiental. Los resultados que obtuvimos utilizando EQTransformer indicaban que tenía la capacidad de encontrar decenas de miles de pequeños eventos transientes lo cual garantizaba que estaba identificando las ondas de cuerpo. Precisamente las ondas de cuerpo son aquellas que se utilizan para reconocer cualquier tipo de sismo, ya sean microsismos o bien terremotos de características devastadoras. En esta tesis presentamos los resultados de esta investigación, y en las siguientes secciones describiremos cada una de las etapas en las que fuimos integrando los diferentes procedimientos para poder llegar a diferenciar y obtener los valores de los cambios de velocidad a través del tiempo.

1.2 Descripción del problema

Existe en la actualidad un problema ambiental alrededor del mundo debido a la escasez de agua y es importante utilizar todo tipo de recursos para poder analizar el volumen total que hay dentro de un acuífero. Mediante aprendizaje automático es posible aportar información del cambio de volumen de agua que ocurre en el subsuelo midiendo vibración ambiental, sin embargo, abordar este desafío utilizando algoritmos de aprendizaje automático se ve obstaculizado por la gran cantidad de datos que necesitan ser procesados, presentando así un componente crítico en la resolución de esta problemática.

1.3 Justificación

Los cambios de velocidad de las ondas de corte pueden servir para identificar características asociadas a fenómenos hidrológicos. Sin embargo, el manejo eficiente de grandes cantidades de datos sísmicos se presenta como un desafío desde el punto de vista computacional. La gestión de la enorme cantidad de datos generados requiere el uso de algoritmos avanzados para detectar eventos de manera precisa y procesar la información de manera efectiva.

1.4 Hipótesis

Usando algoritmos de aprendizaje automático es posible determinar cambios en las propiedades del subsuelo, especialmente aquellos relacionados con la presencia agua.

1.5 Objetivos

El propósito central de la investigación consiste en implementar un algoritmo de aprendizaje automático supervisado con el fin de clasificar las formas de onda sísmicas presentes en sismogramas. Este objetivo se centra específicamente en la distinción entre eventos sísmicos y ruido. El desarrollo de este objetivo se estructura en los siguientes pasos:

Obtención de los Datos y Preprocesamiento: En esta fase, se recopilan los datos sísmicos necesarios para el proyecto. Luego, se realiza un preprocesamiento de los datos para asegurar su calidad y relevancia.

Elección de la Estructura del Modelo: En esta etapa, se analiza y determina la estructura óptima del modelo de aprendizaje automático. Se selecciona el tipo de algoritmo a utilizar (por ejemplo, máquinas de soporte vectorial, redes neuronales, etc.) y se define la estructura del modelo, incluyendo el número de capas y nodos en una red neuronal o la configuración de parámetros en otros algoritmos.

Entrenamiento: Los datos previamente recopilados y preprocesados se utilizan para entrenar al algoritmo para distinguir entre eventos sísmicos y ruido ambiental. Durante el entrenamiento, el modelo busca patrones y relaciones en los datos para hacer predicciones precisas

Validación de los Resultados: Se centra en evaluar la calidad de los datos clasificados y la utilidad de la información proporcionada por el modelo, asegurándose de que las clasificaciones sean información de calidad y útil a la problemática que se aborda.

1.5.1 Objetivo general

Desarrollar un sistema de clasificación automático de formas de onda sísmica en sismogramas utilizando algoritmos de aprendizaje automático supervisado.

1.5.2 Objetivos específicos

Implementar un modelo de aprendizaje automático capaz de distinguir con precisión entre eventos sísmicos y ruido en las señales registradas en sismogramas.

1.6 Alcances y limitaciones

El estudio se realizará para las condiciones de la región la cuenca de México, con las estaciones de la Red Sísmica del Valle de México que cuenta con un total de 30 estaciones de monitoreo sísmico

Capítulo 2

MARCO TEÓRICO

2.1. Señales Sísmicas

Una señal sísmica es un registro gráfico o digital de las vibraciones y movimientos de la tierra causados por un sismo o terremoto. Cuando ocurre un terremoto, se generan ondas sísmicas que se propagan a través de la Tierra en forma de vibraciones. Estas vibraciones pueden ser detectadas y registradas por instrumentos llamados sismógrafos o acelerómetros. (figura 2.1.1)

La señal sísmica registrada por un sismógrafo muestra cómo cambia con el tiempo la intensidad de las vibraciones en diferentes frecuencias. Dependiendo de la magnitud y la distancia del terremoto, las señales sísmicas pueden variar en amplitud (la altura de las oscilaciones), en duración y en la frecuencia de las ondas.

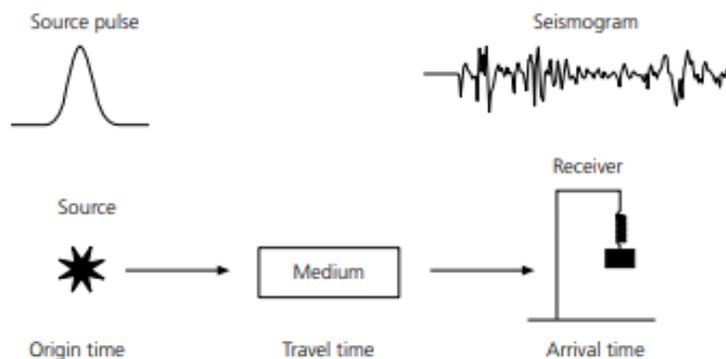


Figura 2.1.1 Diagrama esquemático de un sismo

2.1.1 Fases Sísmicas

Las fases sísmicas son las diferentes características distintivas de las ondas sísmicas que se propagan desde el epicentro de un terremoto. Estas fases son utilizadas en la sismología para describir y analizar los registros sísmicos y comprender mejor la naturaleza y el comportamiento de los terremotos. Las principales fases sísmicas son las siguientes:

- Onda P (Primaria): También conocida como onda compresional o longitudinal, es la primera en llegar a una estación sísmica después de un terremoto. Las ondas P se propagan más rápido que las demás y son capaces de viajar tanto a través de sólidos como de líquidos. Estas ondas se caracterizan por una oscilación en la dirección de propagación de la onda.
- Onda S (Secundaria): También conocida como onda de corte o transversal, sigue a las ondas P en su llegada a una estación sísmica. Las ondas S se propagan más lentamente que las ondas P y solo pueden viajar a través de sólidos, ya que requieren la rigidez del medio. Estas ondas se caracterizan por una oscilación perpendicular a la dirección de propagación de la onda.
- Onda superficial: Estas ondas son generadas cuando las ondas P y S alcanzan la superficie de la Tierra. Se propagan a lo largo de la superficie terrestre y son responsables de la mayoría de los daños en los terremotos. Las ondas superficiales tienen amplitudes más grandes y velocidades de propagación más lentas que las ondas P y S.

2.2. Ruido Sísmico

Llamamos indistintamente ruido sísmico o ruido ambiente al registro temporal continuo de un sismómetro que no contiene eventos sísmicos tales como explosiones o terremotos.

Previamente al año 2005, el registro sísmico sin eventos se consideraba inservible y se desechaba. Solamente se guardaba y analizaba la parte del registro con señales de eventos sísmicos. En ese año se demuestra experimentalmente que con el ruido sísmico se pueden obtener propiedades de la propagación de las ondas superficiales entre dos sensores (e.g., Sabra et al., 2005a; Shapiro et al., 2005). A pesar de la nueva aplicación y de que en la actualidad se utiliza como una señal sísmica, al registro sísmico continuo sin eventos, se ha continuado denominándola ruido. Así, en esta memoria el ruido sísmico se utiliza como señal a partir de la que se mide la velocidad de las ondas superficiales.

La utilización del registro sísmico completo ha supuesto un cambio en la grabación y almacenamiento de señales sísmicas. En la actualidad tanto el registro como el almacenamiento son continuos y se guarda el registro completo para el análisis, no sólo los eventos (e.g., explosiones o terremotos). Este cambio implica la necesidad de una mayor capacidad de almacenamiento en la memoria del sensor y en los centros de recepción de datos.

2.3. Registro Sísmico

Un sismograma, el registro del movimiento del suelo en un receptor llamado sismómetro, contiene información tanto sobre la fuente como sobre el medio. Esta información puede adoptar varias formas. Las ondas proporcionan información sobre la localización y la naturaleza de la fuente que las generó. [1]

Un sismógrafo es un instrumento utilizado para medir y registrar las ondas sísmicas generadas por los movimientos de la Tierra. Consiste en un sensor de movimiento llamado

sismómetro, que puede ser sensible a diferentes componentes de las ondas sísmicas, como las ondas P (primarias), las ondas S (secundarias) y las ondas superficiales.

Cuando ocurre un terremoto, el sismógrafo registra las vibraciones del suelo y las convierte en señales eléctricas. Estas señales se registran en un gráfico o se almacenan digitalmente como datos. El registro sísmico muestra cómo la amplitud de las vibraciones varía con el tiempo.

El eje horizontal del registro sísmico representa el tiempo, generalmente en segundos o minutos, y el eje vertical representa la amplitud de las vibraciones registradas. Dependiendo del sismógrafo y del evento sísmico, el registro sísmico puede mostrar una variedad de patrones y formas de onda, que corresponden a las características de las ondas sísmicas y a la magnitud y distancia del terremoto.

El análisis de los registros sísmicos es fundamental para el estudio de los terremotos y para comprender mejor los procesos que ocurren en el interior de la Tierra. Los sismólogos utilizan los registros sísmicos para determinar la ubicación y la magnitud de los terremotos, estudiar la propagación de las ondas sísmicas, caracterizar la estructura interna de la Tierra y evaluar la amenaza sísmica en diferentes regiones.

2.4. Tectónica de Placas

La Tectónica de Placas es una teoría unificadora que explica una variedad de características y acontecimientos geológicos. Se basa en un sencillo modelo de la Tierra que expone que la rígida litosfera se encuentra fragmentada, formando un mosaico de numerosas piezas de diversos tamaños en movimiento llamadas placas, que encajan entre sí y varían en grosor según su composición ya sea corteza oceánica, continental o mixta. [2]

El movimiento de las placas no se da en forma uniforme, se tienen zonas donde el movimiento es muy lento, del orden de una centésima de milímetro al año y otras en las cuales el movimiento es muy rápido, de más de 10 cm al año. De igual forma existen segmentos de la corteza que chocan entre sí y otros en que no existe este choque. Estos movimientos son llamados tectónicos y son los responsables de la aparición de montañas, volcanes, sismos, formación de plegamientos y fallas geológicas, expansión de océanos, desplazamiento de continentes y también está asociado a yacimientos minerales y petrolíferos. [2]

2.4.1. Tectónica Global

La formación teórica completa y el desarrollo de la Tectónica de placas fue desarrollada entre 1960 y 1965 por los geofísicos Tuzo Wilson (Teoría del Ciclo), Jason Morgan y Harry H. Hess. Según esta teoría, la corteza terrestre está formada por una serie de bloques rígidos denominados placas tectónicas, las cuales comprenden la corteza y la parte superior del manto que se mueven sobre la astenosfera.

La clave para entender la tectónica de placas es una comprensión de la estructura interna de la Tierra, la cual se ilustra en la figura 2.4.1 El núcleo de la Tierra consiste principalmente en hierro. El núcleo exterior está lo suficientemente caliente para que el hierro sea líquido.

El núcleo interno, aunque incluso más caliente, está bajo tanta presión que es sólido. El manto está compuesto por minerales de hierro y silicato de magnesio. La mayor parte del manto que rodea el núcleo exterior es roca sólida, pero es lo suficientemente plástica como para poder fluir lentamente. La parte más externa del manto es rígida. La corteza, compuesta principalmente de granito en los continentes y principalmente de basalto debajo de los océanos, también es rígida. La corteza y el manto rígido más externo forman juntos la litosfera.

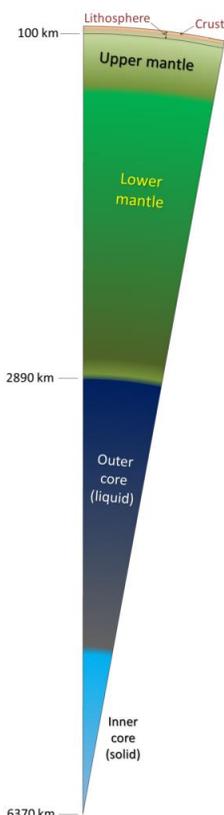


Figura 2.4.1: Los componentes interiores de la Tierra

Además, la litósfera está compuesta por 14 placas tectónicas principales y 46 placas tectónicas secundarias (figura 2.4.2). La diferencia entre las placas principales y las secundarias radica en el tamaño de las mismas, siendo las principales las de mayor tamaño. Las principales Placas Tectónicas son: Africana, Antártica, Arábiga, Caribe, Cocos, Euroasiática, Filipina, Indo australiana, Norteamericana, Sudamericana y del Pacífico.

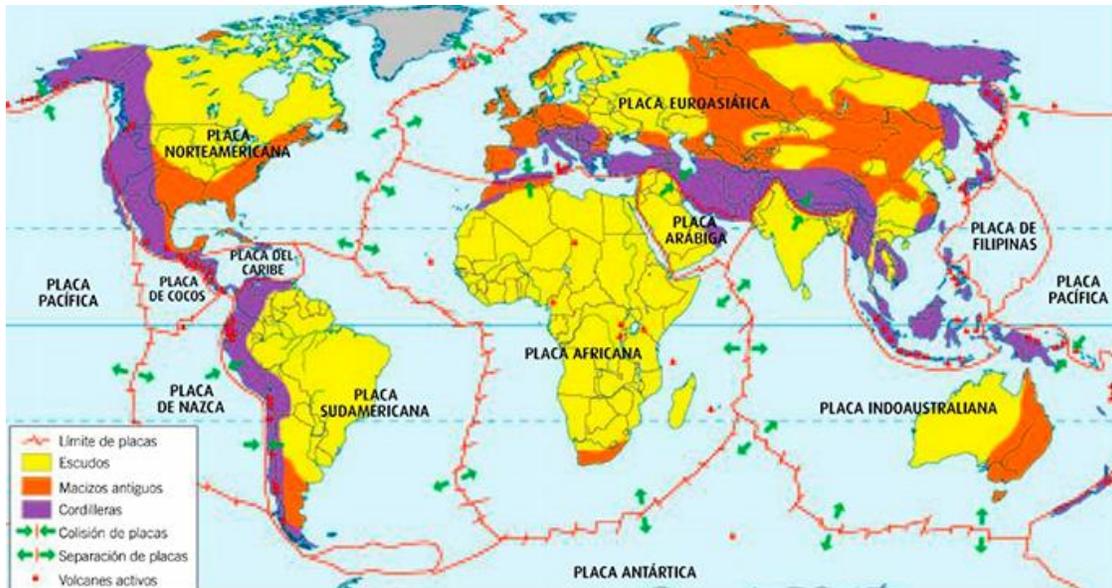


Figura 2.4.2 Placas Tectónicas en el mundo [2]

2.4.2. Tectónica en México

La estructura de la corteza continental de México, así como de gran parte de Centro-américa y el Caribe, ofrece un panorama geológico complejo que hace difícil la interpretación de su evolución tectónica y paleogeográfica.

Los primeros modelos globales propuestos en el pasado para interpretar la evolución tectónica de México tuvieron como marco de referencia dos conceptos fundamentales. El primero es el concepto del ciclo geotectónico inherente al desarrollo de las fajas orogénicas y el segundo es la consideración de que las masas de corteza continental han mantenido en general la misma configuración (Burkhardt, 1930; Álvarez, 1949; Flores, 1951; de Cserna, 1960). En los trabajos de interpretación posteriores se observa un alejamiento paulatino del concepto de ciclo geotectónico y un acercamiento hacia la causalidad de los fenómenos relacionados con la tectónica de placas, sin embargo, prevalece la idea de que la corteza continental correspondiente a la porción de México había presentado, en general, la misma distribución.

Más recientemente han cobrado importancia los modelos que interpretan la integración de la mayor parte de la corteza continental de México y parte de Centroamérica y el Caribe a partir del efecto combinado de la separación de Norteamérica y Sudamérica y el traslado de masas continentales del noroeste por medio de una o varias fallas laterales izquierdas (Pilger, 1978; Walper, 1980; Anderson y Schmidt, 1983). Las evidencias que se manejan en estos modelos están sustentadas en las propuestas de grandes corrimientos y en el conocimiento reciente de la estructura del Golfo de México. Por otro lado, los datos paleomagnéticos son generalmente congruentes con la idea de fallas laterales izquierdas (Urrutia-Fucugauchi, 1984). [3]

Los conceptos actuales sobre la movilidad de la litosfera en el espacio de México y el Caribe han obligado a la reinterpretación de la evolución paleogeográfica y tectónica de algunas

regiones a través de la formulación de nuevos modelos para problemas hasta ahora no resueltos.

La estructura geológica de México es notablemente diversa y compleja debido a su ubicación en una región donde convergen varias placas tectónicas (figura 2.4.3). Esto ha dado lugar a una serie de características geológicas sobresalientes en el país. En la costa oeste, la interacción entre la Placa de Cocos y la Placa Norteamericana ha originado imponentes cadenas montañosas como la Sierra Madre Occidental y la Sierra Madre Oriental. Estas áreas son propensas a terremotos y actividad volcánica. Además, México alberga una rica geología costera y marina incluyendo sistemas de cenotes y arrecifes de coral.

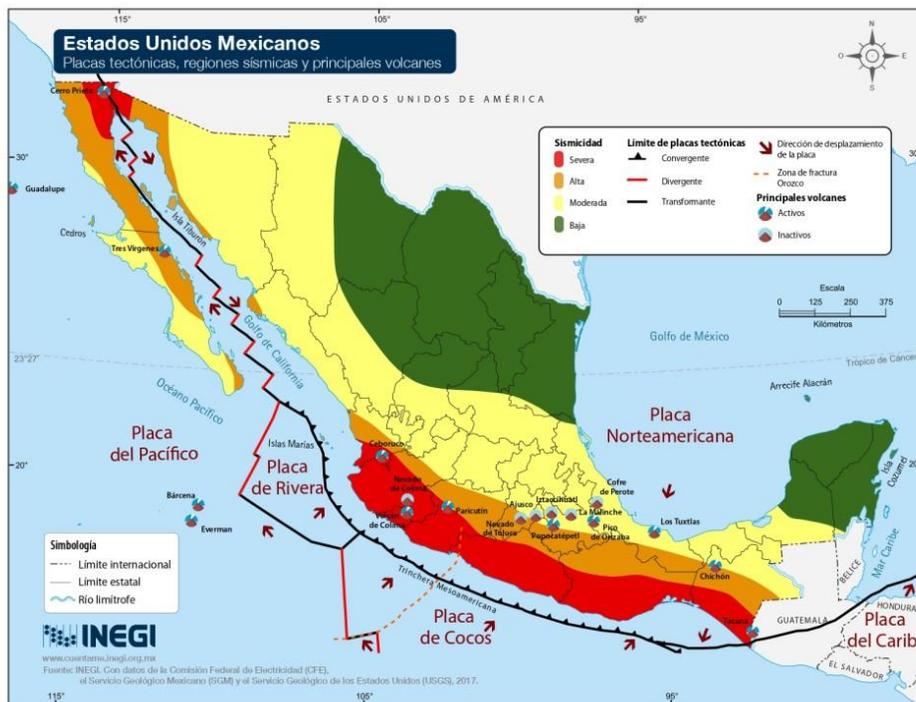


Figura 2.4.3 Placas tectónicas en México

2.5. Sismos en México

La investigación de los fenómenos sísmicos en México requiere un análisis detenido de su evolución histórica y las diversas perspectivas culturales y científicas que han moldeado la comprensión de los sismos a lo largo del tiempo. En este contexto, se explorará la representación y clasificación de los terremotos desde la época prehispánica hasta eventos sísmicos clave en la historia, abordando las transformaciones en los métodos de registro, interpretación y las consecuencias sociales de dichos eventos.

Época Prehispánica

Durante el México prehispánico, los antiguos habitantes de la cuenca de México registraron estos fenómenos en códices que pasaron a la posteridad para ayudar a entender cómo se vivieron estos fenómenos en esa época.

Tlalollín es la palabra en náhuatl que se utilizaba para describir los sismos, Tlalli significa tierra y Ollín movimiento, es así como nace el término movimiento de la tierra, ocupado en los códices que registraron esos movimientos. Al observar estos registros, resulta sumamente interesante observar cómo se clasificaban los sismos; dos glifos, dependiendo de cuando ocurrió el sismo, es decir, se ocupaba un glifo para los sismos del día y otro para los de noche.

Dependiendo de la posición del símbolo Ollín en relación con el Tlalli se ha calculado la intensidad y el tipo de sismo; en ocasiones el primero se encuentra arriba, otras en medio y algunas veces hay dos o más Tlallis (Figura 2.5.1). Gracias a estas representaciones pictográficas se sabe que el pueblo mexica sufrió al menos 8 terremotos de magnitud 8.1, la misma magnitud que el temblor de 1985.[4]

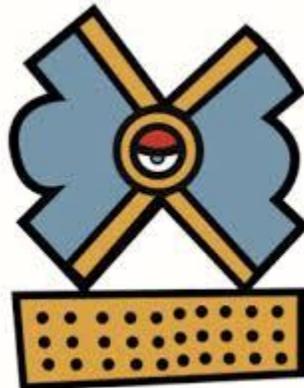


Figura 2.5.1: Símbolo de Tlaollin

Periodo Colonial

Tras el arribo de los españoles a Mesoamérica, la manera de registrar los sismos cambió diametralmente. En este caso el cristianismo jugó un papel importante, no sólo en la manera de registrar sino también de explicar los movimientos telúricos. Es así como de acuerdo con el santoral católico se hacía referencia a los años de los sismos, Ahora bien, si de registros se habla, una vez asentada la conquista espiritual, los sacerdotes cambiaron la manera de registrar los temblores; la duración se cuantificó en rezos y no en pictogramas.

En la noche del 24 de diciembre de 1545, a la llegada de fray Bartolomé de las Casas, el nuevo obispo de Chiapas a Ciudad Real (hoy San Cristóbal de las Casas); El fraile dominico Francisco Ximénez, nos relató lo siguiente: “Hizo un tan grande temblor en la tierra, que pensamos que se hundía el mundo y duró espacio de tres salmos de miserere que a todos nos puso en admiración”. Ejemplos como el anterior se pueden encontrar en diversas notas y apuntes de frailes evangelizadores que se dedicaron a realizar un seguimiento arduo de esta clase de fenómenos naturales [4]

Los Sismos de 1787

En 1787 ocurrió una secuencia de cuatro sismos enormes al sur de México, entre los cuales quizás se encuentre el más grande ocurrido en México del que se tenga conocimiento. Estos ocurrieron los días 28, 29 y 30 de mayo, y 3 de junio, y fueron seguidos por un gran número de réplicas en los días posteriores. Los primeros sismógrafos se empezaron a desarrollar poco antes de 1900, por lo tanto, en 1787 no existía la instrumentación necesaria para registrar las ondas sísmicas y así poder definir los epicentros y la magnitud de los sismos. Sin embargo, existen archivos históricos con descripciones y relatos que permiten conocer en qué lugares, qué tan fuerte, y por cuánto tiempo las personas sintieron que el suelo se sacudía. Además, existen archivos con descripciones de los daños ocurridos, por ejemplo, en iglesias, haciendas y palacios de gobierno. Con esta información se puede definir una intensidad para cada lugar de donde hay reportes.[5]

El Sismo de Madero 1911

El día 7 de junio de 1911, aconteció un terremoto asociado a la historia de México, cuando recién había dejado la silla presidencial Porfirio Díaz, y precisamente ocurriría a escasas 7 horas antes de la espectacular entrada de Francisco I. Madero a la Ciudad de México.

Fue mejor conocido como el “temblor maderista”. Tuvo su epicentro en Ciudad Lázaro Cárdenas Michoacán y la magnitud fue de 7.6.

Este terremoto generó daños en la Ciudad de México, fallecieron 30 personas del Regimiento de Artillería en Rivera de San Cosme. Provocó daños importantes en el altar de la iglesia de San Pablo.

Sismo de Acambay 1912

En el contexto de la Revolución Mexicana, exactamente un martes 19 de noviembre de 1912 ocurrió el terremoto de Acambay, considerado uno de los eventos sismológicos más importantes del país, debido a las afectaciones que provocó en el epicentro y zonas cercanas. Se trató de un sismo con magnitud aproximada de 6.9 en el municipio de Acambay, Estado de México, registrado a las 07:55 horas, de acuerdo con el Servicio Sismológico Nacional (SSN).

Este temblor sucedió dentro de la Placa Norteamericana, por lo que se le considera uno de tipo intraplaca cortical. Debido al tamaño del sismo, tanto por su magnitud y sus características tectónicas, se reportó un número significativo de réplicas, aproximadamente 60 entre el 19 de noviembre de 1912 al 15 de abril de 1913. El sismo provocó el derrumbe de varias estructuras, lo que a su vez dejó un saldo de 140 decesos. Las construcciones eran coloniales en su mayoría, hechas de ladrillo y adobe, un material muy frágil, por lo que la mayor parte de Acambay quedó en ruinas.

El Sismo de Xalapa de 1920

El sismo ocurrido cerca de la ciudad de Xalapa, Veracruz ocurrió el 3 de enero de 1920. Este sismo tuvo magnitud 6.4 y se originó dentro de la placa Norteamérica, aparentemente, en una falla que podría tener relación con la Faja Volcánica Transmexicana, donde existen varias fallas activas que han provocado sismos importantes en el pasado. En Xalapa se registraron daños en construcciones por el movimiento propio del sismo; sin embargo, la verdadera catástrofe ocurrió debido a otros eventos inducidos por el mismo. El movimiento brusco provocó una avalancha de lodo sepultando varias poblaciones. Se calcula que en algunas partes el lodo alcanzó una profundidad de más de 60 metros, matando a más de 400 personas. Además, el sismo también generó el nacimiento de manantiales, derrumbes en montañas, inundaciones de lodo, y hundimientos de suelo. Se estima que todo esto causó la muerte de alrededor de 650 personas. Esto lo convierte en el segundo evento sísmico que ha causado más muertes en México de que se tiene registro.[5]

El Sismo de 1985

Definitivamente, el sismo que ha causado más muertes en México fue el ocurrido el 19 de septiembre de 1985 en las costas de Michoacán, de magnitud de momento (Mw) 8.1. Existe mucha controversia acerca de la cifra real de muertos, pero el número de víctimas pudo haber alcanzado 40,000, principalmente en la Ciudad de México. A pesar de que el epicentro del sismo ocurrió a una distancia de más de 350 km de la Ciudad de México, ahí el suelo se sacudió fuertemente, alcanzando una intensidad de Mercalli Modificada (MMI) de 9. Estudios posteriores han mostrado que esto se debió en gran parte a cómo responde el suelo de la Ciudad de México al paso de las ondas sísmicas. Esta ciudad se construyó sobre una isla en la mitad del antiguo Lago de Texcoco. Este tipo de suelo, junto con la propia geometría del antiguo lago, causan que las ondas sísmicas se amplifiquen, además de que se genere un efecto de resonancia, lo que causa que el movimiento del suelo sea más brusco y dure por más tiempo [5]

2.6. Sismometría

La Sismometría es el campo de estudio y la práctica relacionada con la medición y el monitoreo de los terremotos y las vibraciones sísmicas. Esta disciplina es fundamental para comprender la actividad sísmica de la Tierra, prever terremotos, evaluar la amenaza sísmica y diseñar estructuras resistentes a sismos.[6]

En la actualidad, el término sismógrafo está reservado a los instrumentos que registran el movimiento del terreno en función del tiempo y trabajan en forma continua durante las 24 horas del día.

Un sismógrafo consta básicamente de cuatro partes (figura 2.6.1):

- El sismómetro que responde al movimiento del suelo, en la dirección vertical u horizontal, y lo convierte en una señal eléctrica.
- Un sistema de amplificación que permite aumentar la precisión del registro del movimiento.
- Un sistema de registro de la señal amplificada para graficar la variación del movimiento, a lo largo del tiempo.
- Un sistema de tiempo preciso que se incorpora al registro de la señal sísmica.

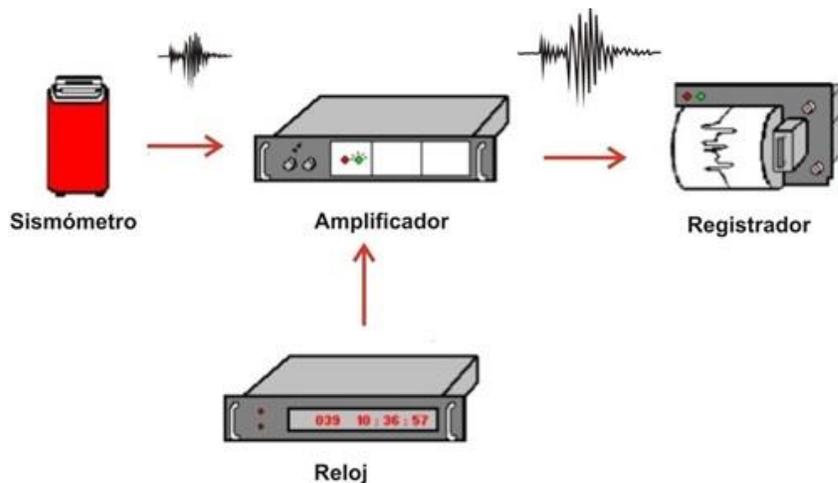


Figura 2.6.1: Esquema indicativo de los componentes de un sismógrafo [6]

2.6.1 Historia de la Sismometría

La historia de la Sismometría, o el estudio y la medición de los terremotos y las vibraciones sísmicas, se remonta a la antigüedad y ha evolucionado a lo largo de los siglos.

El primer instrumento específicamente construido para medir los sismos, que se tiene conocimiento, es el sismoscopio inventado, en el año 132 d.C., por el filósofo chino Chang Heng (figura 2.6.2). El mismo consistía en una jarra con ocho cabezas de dragón, cada una de las cuales tenían una bola metálica, que, ante la ocurrencia de un terremoto, un mecanismo interno liberaba la bola y caía sobre un sapo ubicado debajo de la cabeza, indicando de manera aproximada la dirección del sismo.



Figura 2.6.2 Primer sismoscopio inventado por el científico chino Heng Chen (Año 132 d.C.)

En 1703 el abate Jean de Hautefeuille (1647-1724) ideó un sismoscopio construido en madera con una base circular en el centro, de un pedestal, con un diámetro interno de aproximadamente 12 centímetros. En su parte superior contenía mercurio, con ocho pequeñas canaletas en su flanco, cuatro en correspondencia con los cuatro puntos cardinales y cuatro en puntos intermedios. Debajo de cada una de estas canaletas hay ocho envases pequeños, que fijados a la base son utilizados para recoger el mercurio que se vierte ante un movimiento sísmico (figura 2.6.3). Cada uno de estos ocho recipientes, se encuentran identificados con las letras: N, S, E, O, NE, NO, SE y SO. El instrumento, se debe orientar adecuadamente, en un plano perfectamente horizontal: analizando la taza en la que se derramó el mercurio indicará la dirección proveniente del sismo.



Figura 2.6.3 Sismoscopio construido por el abate Jean de Hautefeuille

Científicos como James D. Forbes (1841), Luigi Palmieri (1856), Verbeck (1873), el italiano Filippo Cecchi (1875) y Alfred Wegner (1880), hicieron su contribución con instrumentos para la medición de los sismos, a pesar de que ninguno de ellos funcionó adecuadamente.

En 1898, el científico alemán Emil Wiechert (1861-1928) desarrolla el primer sismógrafo con amortiguamiento viscoso, logrando con gran eficiencia el registro de los sismos en toda su duración (figura 2.6.4).

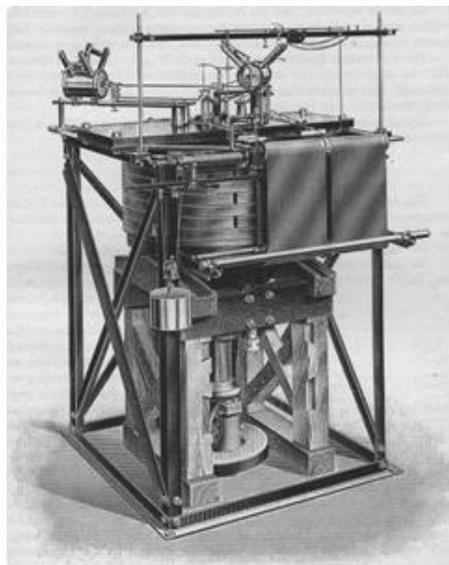


Figura 2.6.4 Sismógrafo Wiechert horizontal (Alemania 1904) Masa: 1.000 kg. Período: 8 s

La incorporación de la computadora en el campo de la sismología, durante la década del 1950-60, permitió ampliar notablemente el campo de la investigación sísmica.

El primer sismógrafo con grabación digital funcionó en el California Institute of Technology (CALTECH), alrededor de 1961.

A partir de 1970 se generaliza la instalación de los sismógrafos digitales a nivel mundial, comienzan a utilizarse los primeros sismógrafos digitales portátiles para estudios específicos (características del sitio, ruido sísmico, etc.), y se establecen los primeros archivos de datos sísmicos digitales.

Desde 1990 hasta la actualidad la tecnología electrónica ha podido desarrollar instrumentos cada vez más reducidos, compactos y livianos, con mayor sensibilidad y precisión, llegando a contener en un solo sismómetro tres sensores para la medición de las tres componentes: N-S, E-O y vertical (Z).

En general, las estaciones sismológicas actualmente incluyen sensores que registran las tres componentes del movimiento del suelo: dos direcciones horizontales perpendiculares, y la dirección vertical. Estas señales ingresan a un sistema digital de adquisición de datos (DAS, de sus siglas en inglés Digital Acquisition System) de 16 o 24 bits, que incluye un GPS para incorporar la señal horaria y las coordenadas del lugar. La información almacenada es transmitida en forma continua y en tiempo real a los centros de investigación, vía satelital o Internet, para su procesamiento, estudio y almacenamiento. [6]

2.7. Servicio Sismológico Nacional

Remonta al primero de abril de 1904, cuando se reunieron en Francia, dieciocho países, entre ellos México, con el fin de crear la Asociación Sismológica Internacional y mejorar la instrumentación sísmica a nivel mundial. Para cumplir con los compromisos adquiridos en esa reunión, el gobierno mexicano inició la instalación de una red sismológica en el territorio nacional. El 5 de septiembre de 1910 se inauguró la primera estación sismológica de la red, en Tacubaya. La red sismológica quedó a cargo del Instituto Geológico Nacional dependiente de la Secretaría de Minería y Fomento.

El SSN pasó a ser parte de la Universidad Nacional Autónoma de México (UNAM) en 1929 y desde 1948 quedó adscrito al Instituto de Geofísica de la UNAM.

Desde 1910 a 1923, se realizó la implementación de un total de nueve estaciones sismológicas a nivel nacional. La central de Tacubaya fue complementada con estaciones en diferentes puntos: Mazatlán, Oaxaca, Mérida, Chihuahua, Veracruz, Guadalajara, Monterrey y Zacatecas. Todas estas fueron equipadas con sismógrafos “Wiechert” de origen alemán, equipos de amortiguamiento viscoso.

La instalación de la Red Sísmica de Apertura Continental (RESMARC) se inició en la UNAM a mediados de los años setenta, con el fin de contar con estaciones telemétricas digitales en todo el territorio nacional. En esta red, la transmisión de las señales se realizaba

por medio de enlaces de microondas proporcionados por la Secretaría de Comunicaciones y Transportes. Las operaciones de detección y captura de eventos sísmicos se utilizaba una computadora PDP11-40.

En agosto de 1986, RESMARC pasó a formar parte del SSN. Así, se integraron las dos redes de cobertura nacional. Esto mejoró el monitoreo sísmico en el país y, por primera vez, permitió al SSN contar con registros en tiempo real desde diferentes puntos de la República Mexicana. Sin embargo, la cobertura nacional aún era insuficiente.

Hasta 1992 se inició la innovación de la Red Sismológica Nacional mediante instalación de equipos de vanguardia, siendo el origen de la red de Observatorios Sismológicos de Banda Ancha.

2.8. Métodos Convencionales de detección sísmica

2.8.1. Técnica STA/LTA (Short-Term Average to Long-Term Average)

Una de las técnicas utilizadas en sismología y monitoreo sísmico para detectar y caracterizar terremotos es la técnica STA/LTA, la cual basa en el análisis de las señales sísmicas registradas por un sismógrafo y se utiliza para identificar eventos sísmicos significativos en función de ciertos criterios de relación entre el promedio a corto plazo (STA) y el promedio a largo plazo (LTA) de la amplitud de la señal sísmica. STA es sensible a las fluctuaciones rápidas en la amplitud de la serie temporal y el LTA proporciona información sobre el ruido de fondo; por lo tanto, la relación STA/LTA puede utilizarse para la selección de llegadas. La metodología STA/LTA y la relación entre los distintos parámetros se describen a grandes rasgos (Figura 2.8.1.2) ilustra el método de promedio a corto plazo/promedio a largo plazo (STA/LTA). En este enfoque, la ventana de tiempo a corto plazo (longitud SS, el recuadro punteado en rojo) precede a la ventana de tiempo a largo plazo (longitud LL, el recuadro punteado en negro). La ventana a corto plazo (SS) tiene una duración de 1–2 veces el período dominante de los eventos micro sísmicos, mientras que la ventana a largo plazo (LL) abarca 5–10 veces la duración de la ventana a corto plazo (SS) (Han 2010). De esta manera, el período dominante se puede estimar a partir de la representación espectral (b), que es el espectrograma del evento mostrado en (a). En (c), se presentan la función de correlación cruzada R (curva azul) y su derivada R' (curva naranja). Utilizamos el valor máximo de R' como el punto de llegada de los eventos micro sísmicos.

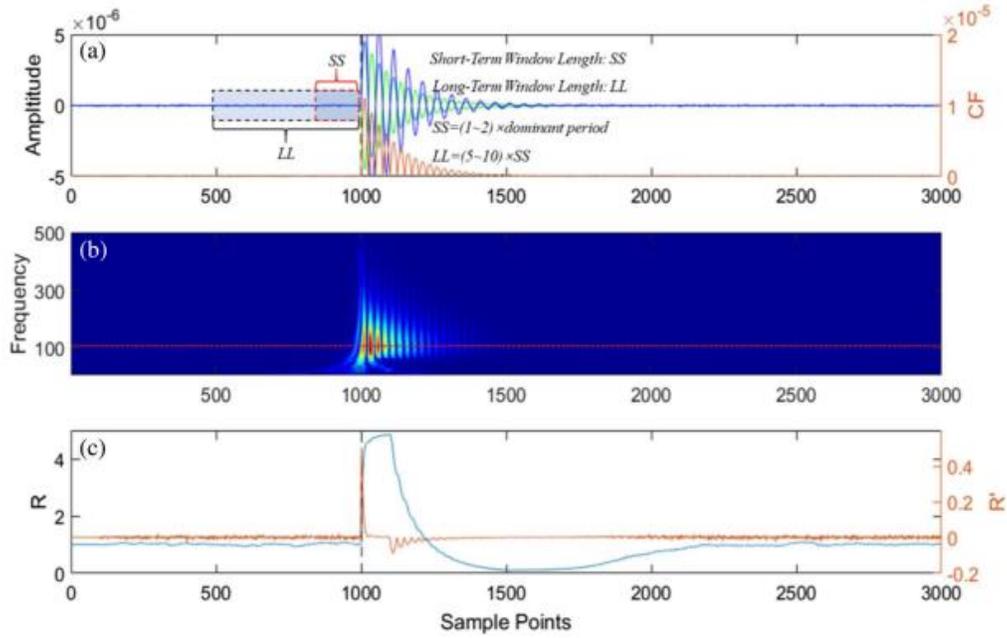


Figura 2.8.1 [7]

Se utilizan ecuaciones 1-4 como función característica de los datos sísmicos (ecuación a); para calcular STA y LTA (ecuación b); para asegurar que R -la relación entre STA y LTA- es una secuencia continua (ecuación c); y para derivar R', (cuyo valor máximo consideramos el punto de llegada de los eventos micro sísmicos), a partir de R (como se muestra en la Ecuación d).[7]

$$CF(i) = \sqrt{x(i)^2 + y(i)^2 + z(i)^2} \quad (a)$$

$$STA(i) = \frac{1}{SS} \sum_{j=i-SS+1}^i CF(j) \quad (b)$$

$$LTA(i) = \frac{1}{LL} \sum_{j=i-LL+1}^i CF(j)$$

$$R(i) = \begin{cases} \frac{STA(i)^{1, i \in [1, SS]}}{\frac{1}{i} \sum_1^i CF(i)}, & i \in [SS + 1, LL - 1] \\ \frac{STA(i)}{LTA(i)}, & i \in [LL, L] \end{cases} \quad (c)$$

$$R'(i) = diff(R(i)) \quad (d)$$

2.8.2. Técnicas Espectrales

Técnicas H/V

La idea de obtener la razón espectral entre las componentes horizontal y vertical (H/V) registradas en un solo instrumento (un único sitio de medición) fue primeramente introducida por Nogoshi & Igarashi (1971) quienes mostraron la relación entre esa razón H/V con la curva de elipticidad de las ondas Rayleigh e identificaron la coincidencia entre el máximo de la más baja frecuencia observada en H/V con la frecuencia fundamental de resonancia, para usarla como indicador de la estructura del suelo debajo del sitio de registro (Bard, 1999).

La razón H/V está básicamente relacionada con la elipticidad de las ondas Rayleigh, debido a la predominancia de ondas Rayleigh en la componente vertical.

Esta elipticidad es dependiente de la frecuencia y muestra un pico definido cerca de la frecuencia fundamental en sitios donde existe un contraste entre impedancias significativo (superior a la unidad), dado por la siguiente relación (Kramer, 1996):

$$\alpha = (\rho_2 v_2) / (\rho_1 v_1) \quad (1)$$

Donde: α representa la razón de impedancias, v_1 , v_2 las velocidades de onda cortante de las capas 1 (en superficie) y 2 (en profundidad) y ρ_1 , ρ_2 las densidades para los mismos materiales 1 y 2.[8]

El método sísmico HVSR (Horizontal to vertical Spectral Ratio) o método Nakamura es una técnica sísmica pasiva que utiliza las vibraciones naturales presentes en la naturaleza (figura 2.8.2). El registro de estas vibraciones permite obtener la frecuencia de resonancia fundamental del suelo (o periodo fundamental). Este parámetro es muy importante en ingeniería sísmica para comprender y predecir de manera correcta el efecto de sitio (amplificación) en una determinada región.

La técnica HVSR tiene un carácter no invasivo y se puede aplicar sin necesidad de realizar ningún tipo de perforación ni emplear fuentes de energía externa. Por lo tanto, es un método muy útil en ambientes urbanos altamente antropizados. Los registros sísmicos ambientales se caracterizan por leves vibraciones presentes en todas partes y que consiste principalmente en trenes de ondas superficiales (Rayleigh y Love), producidas por la interferencia de ondas P y S en las capas superficiales.

- Los resultados que se pueden obtener de los estudios HSRV son:
- Las características de frecuencia de la resonancia.
- La frecuencia de la resonancia de un edificio (siempre y cuando la medición se realice en su interior).
- Estimación de la velocidad de onda de corte V_s (cuando haya información adicional sobre las condiciones geológicas del subsuelo).

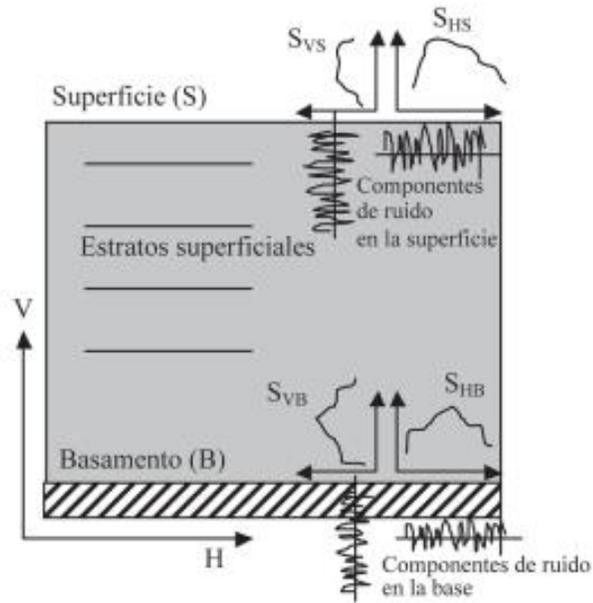


Figura 2.8.2: Esquema a partir del cual se deduce la técnica de Nakamura. [8]

2.8.3. Intensidad de Arias

Es una medida que se utiliza para cuantificar la energía liberada por un terremoto en función de la aceleración del suelo durante un evento sísmico. Esta medida proporciona una estimación de la agitación del suelo y puede ser útil para evaluar el impacto sísmico en estructuras y para comparar la intensidad de diferentes terremotos.

La Intensidad de Arias se utiliza a menudo en estudios de ingeniería sísmica y geotecnia para evaluar el riesgo sísmico y diseñar estructuras resistentes a terremotos. También puede ser útil para comprender la potencia relativa de diferentes terremotos y su impacto en una región determinada.

La Intensidad de Arias se calcula mediante la siguiente fórmula donde $(t)^2$ es la aceleración del movimiento del terreno elevada al cuadrado, g es la aceleración de la gravedad y t es la duración total del registro. Estos parámetros están relacionados con la energía del movimiento del terreno registrado que, a su vez, está relacionado con la magnitud y la distancia a la fuente, y las condiciones del sitio. [9]

$$I_A = \frac{\pi}{2g} \int_0^t a(t)^2 dt$$

Ecuación 2.8.3.1: Formula de Intensidad de Arias

2.8.4. Moving-Window Cross-Spectral Analysis

La técnica MWCS (Multichannel Wiener *deconvolution of Cross-correlation Spectra*) fue introducida por primera vez por Poupinet *et al.* (1984) para la recuperación de variaciones relativas de velocidad entre pares de terremotos. Más recientemente, Brenguier *et al.* (2008 a,b) aprovecharon esta técnica al aplicarla a registros de ruido sísmico, aprovechando la posibilidad de tratar las autocorrelaciones de ruido de manera análoga a los pares de terremotos.

Este análisis se aplica a series temporales que se calculan mediante la autocorrelación de las secuencias de ruido registradas en dos estaciones sísmicas diferentes, para todas las posibles combinaciones de estaciones. El paso preliminar para el análisis es construir al menos una autocorrelación de referencia y varias autocorrelaciones actuales. Dado que, por razones computacionales, los registros continuos de ruido se dividen en secuencias cortas (por ejemplo, una para cada día u hora), es necesario apilar un cierto número de autocorrelaciones individuales.

El primer paso consiste en el cálculo del retardo temporal entre las dos señales dentro de una serie de ventanas superpuestas. El segundo paso es la evaluación de la variación relativa de velocidad asociada a la función actual con respecto a la referencia. Es importante señalar que la primera operación se ejecuta en el dominio espectral, a través del estudio de la fase del espectro cruzado, lo que permite una selección precisa del ancho de banda de frecuencia en función de la coherencia entre las dos autocorrelaciones con ventanas. Por lo tanto, el segundo paso implica la evaluación de la tendencia, $\delta t/t$, de las estimaciones de tiempo de retardo a lo largo de toda la longitud de las señales (figura 2.8.4).

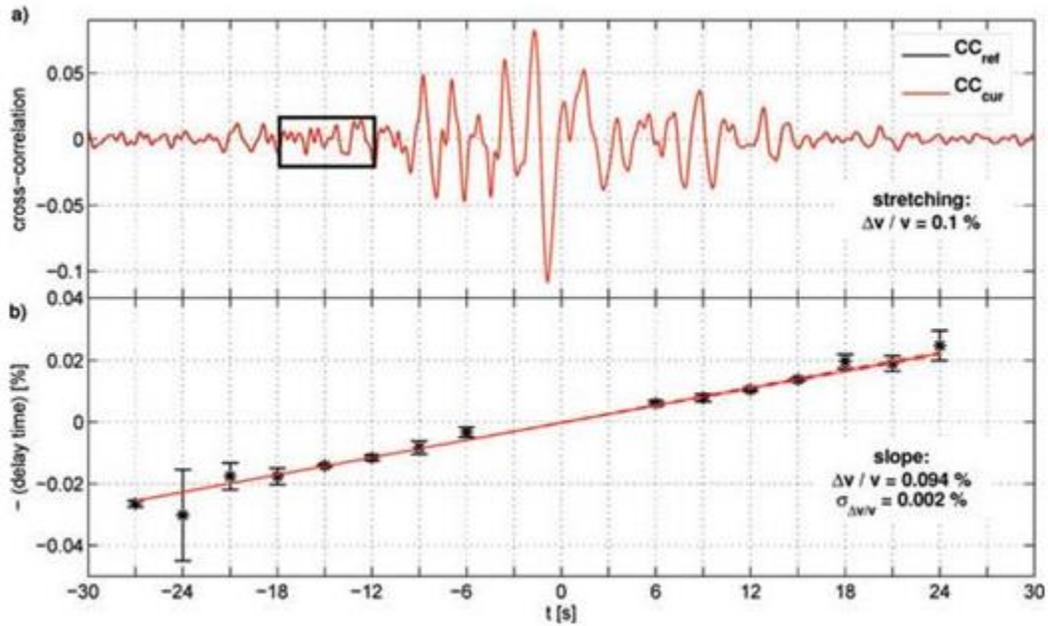


Figura 2.8.4: Par de una autocorrelación de referencia real y una autocorrelación actual sintética

2.8.5. Visualización de los sismos mediante analistas expertos

Aunque es imposible predecir en qué momento ocurrirá el próximo sismo de gran magnitud, si un país cuenta con infraestructura para conocer el comportamiento de este fenómeno natural, tendrá mayor posibilidad de contar con mejores medidas de prevención y protección civil. En México, contamos con un centro de monitoreo permanente que nos informa la localización y la magnitud de los sismos El Servicio Sismológico Nacional, a cargo del Instituto de Geofísica de la UNAM, recoge de manera permanente datos sobre la sismicidad en todo el territorio mexicano para informar a la población y promover la investigación.

Para cumplir con esta misión, opera una red de 61 observatorios sismológicos. Cada uno posee un sismómetro, un acelerómetro y un GPS, aparatos que registran la velocidad, la aceleración y el desplazamiento del suelo durante la propagación de las ondas sísmicas.

Luego del análisis de esos datos, los sismólogos establecen la localización y la magnitud del fenómeno.

Ejemplo Procesamiento de los Datos:

Turno Sismológico

En el Departamento de Sismología se ejecuta un Turno Sismológico de 24 horas al día, en el cual se integra el personal del departamento de Sismología y el grupo de Electrónica. La persona que realiza dicho turno tiene la capacidad de poder localizar cualquier sismo en menos de 20 minutos, para esto se cuenta con tres estaciones de trabajo (SUN) en red con

varias computadoras (PC IBM compatibles), que también sirven para el procesamiento de los datos.

El sismólogo de turno procesa todos los sismos poco después de haber ocurrido el evento. En caso de un sismo fuerte, la computadora principal del sistema, emite una alarma acústica para iniciar el procesamiento inmediato del evento. También, cuando ocurre alguna anomalía en cualquier volcán, ya sea un enjambre sísmico o erupción, el sismólogo de turno reacciona adecuadamente.

Procesamiento Diario

Los datos preliminares obtenidos por los sismólogos de turno se revisan diario por un técnico experimentado. Se eliminan errores y se hace una relocalización de los eventos.

Procesamiento Mensual

Finalmente, para producir el boletín sismológico mensual, se reprocesan detalladamente todos los datos obtenidos, se integra información adicional, como: descripción y encuesta macro sísmica de sismos sentidos, resultados de análisis de enjambres en fallas locales o en volcanes, informaciones sobre actividades volcánicas y sismos ocurridos dentro del territorio nacional. El boletín mensual se entrega normalmente 5 días después de haber finalizado el mes.

2.9. Aprendizaje Automático

El aprendizaje automático consiste en extraer conocimientos de los datos. Es un campo de investigación intersección de la estadística, la inteligencia artificial y la informática. también se conoce como análisis predictivo. La idea clave del uso de técnicas de aprendizaje automático es automatizar procesos o tareas aprendiendo patrones específicos de los datos. [10]

2.9.1. Tipos de Aprendizaje Automático

El aprendizaje automático se puede clasificar en varios tipos principales según la forma en que se realiza el entrenamiento y cómo se aplican los modelos resultantes. Los tipos de aprendizaje automático más comunes son (figura 2.9.1):

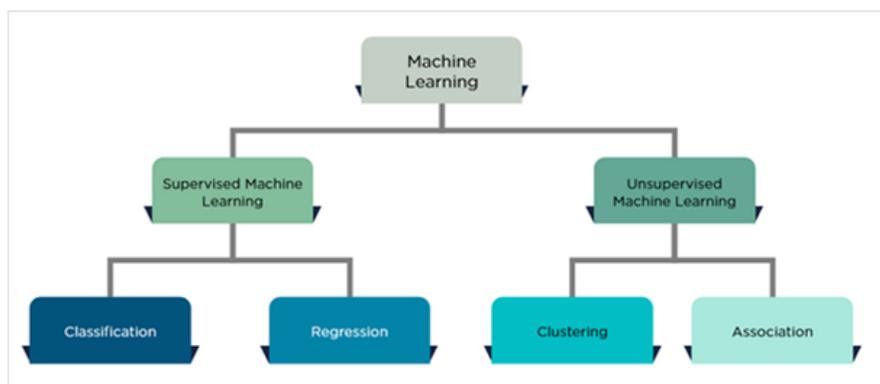


Figura 2.9.1 Diagrama de tipos de aprendizaje automático

Aprendizaje Supervisado (*Supervised machine learning*).

En este tipo de aprendizaje, se entrena un modelo utilizando un conjunto de datos etiquetado, que contiene ejemplos de entrada y la salida deseada. El modelo aprende a mapear las entradas a las salidas correctas y luego puede hacer predicciones sobre nuevas entradas. Existen dos tipos principales de problemas de aprendizaje automático supervisado:

Algoritmos de regresión (*regression*)

En las tareas de regresión, el programa de aprendizaje automático debe estimar y comprender las relaciones entre las variables. El análisis de regresión se enfoca en una variable dependiente y una serie de otras variables cambiantes, lo que lo hace particularmente útil para la predicción y el pronóstico. Existen varios algoritmos de aprendizaje automático en problemas de regresión dentro de las que se pueden destacar:

- Regresión lineal y Regresión no lineal
- Máquinas de vectores de soporte
- Árboles de decisión
- Bosques aleatorios
- Redes neuronales y Aprendizaje profundo

Algoritmos de Clasificación (*Classification*)

Esto típicamente abarca la lista de problemas o tareas en las que la máquina tiene que tomar puntos de datos o muestras y asignar una clase o categoría a cada muestra. Existen varios algoritmos de aprendizaje automático en problemas de clasificación dentro de las que se pueden destacar:

- Regresión logística
- Árboles de decisión
- Bosques aleatorios
- Redes neuronales y Aprendizaje profundo

Aprendizaje No Supervisado (*Unsupervised machine learning*)

En este caso, el modelo se entrena en un conjunto de datos sin etiquetas. El objetivo principal es encontrar patrones, estructuras ocultas o grupos dentro de los datos. A su vez, el aprendizaje no supervisado se puede dividir en dos categorías: en algoritmos de agrupamiento (*Clustering*) y de asociación.

Algoritmos de *Clustering*:

Los conglomerados o grupos suelen formarse a partir de muestras de datos de entrada de entrada haciendo que la máquina aprenda u observe patrones latentes inherentes, relaciones y similitudes entre los propios puntos de datos de entrada. Se utilizan para agrupar datos existentes de los que desconocemos sus características en común o queremos descubrirlas.

Estos métodos intentan crear “puntos centrales” y jerarquías para diferenciar grupos y descubrir características comunes por cercanía. Los más utilizados son:

- *K-Means*
- *K-Medians*
- *Hierarchical Clustering*

Algoritmos de Asociación (*Association*)

Son algoritmos utilizados para descubrir patrones y relaciones interesantes entre variables en grandes conjuntos de datos. Estos buscan encontrar asociaciones frecuentes, es decir, conjuntos de elementos que tienden a aparecer juntos en muchas instancias del conjunto de datos. Los algoritmos más comunes son:

- Apriori
- FP-Growth (Frequent Pattern Growth):
- Eclat:
- Maximal Patterns.
- Closed Patterns
- Algoritmos basados en restricciones
- Algoritmos basados en árboles de decisión

2.9.2. Perceptrón

La arquitectura del Perceptrón, conocida como mapeo de patrones (*pattern-mapping*), aprende a clasificar modelos mediante un aprendizaje supervisado. Los modelos que clasifica suelen ser generalmente vectores con valores binarios (0,1) y las categorías de la clasificación se expresan mediante vectores binarios. El Perceptrón presenta dos capas de unidades procesadoras (PE) y sólo una de ellas presenta la capacidad de adaptar o modificar los pesos de las conexiones. La arquitectura del Perceptrón admite capas adicionales, pero éstas no disponen la capacidad de modificar sus propias conexiones.

La figura 2.9.2 muestra la unidad procesadora básica del Perceptrón. Las entradas x_i llegan por la parte izquierda y cada conexión con la neurona tiene asignada un peso de valor w_{ji} .

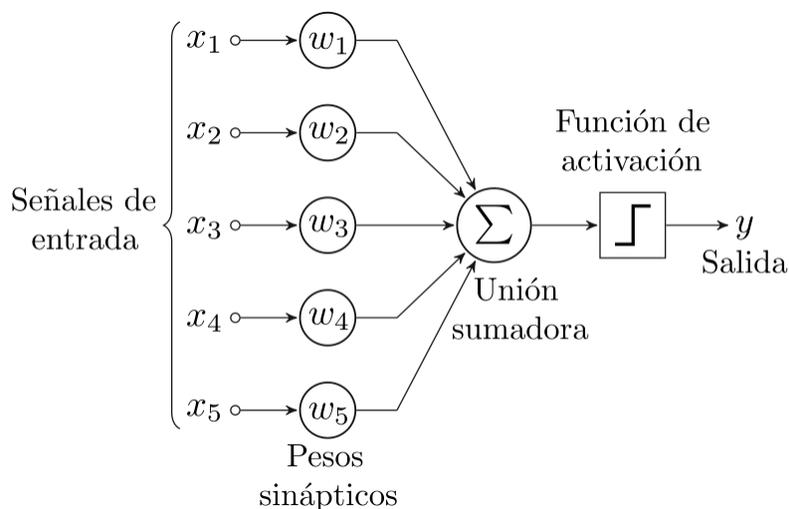


Figura 2.9.2 Unidad Procesadora Básica del Perceptrón

2.9.3. Redes Neuronales

Las redes neuronales son un tipo de modelo computacional inspirado en la estructura y el funcionamiento del sistema nervioso de los seres vivos, especialmente en el cerebro humano. Estas redes artificiales se utilizan para el procesamiento de información y el aprendizaje automático en campos como la inteligencia artificial.

2.9.4. Redes Neuronales Transformadoras

Las Redes Neuronales Transformadoras (TNR, por sus siglas en inglés) son un tipo de arquitectura de redes neuronales que han ganado popularidad en el campo del procesamiento del lenguaje natural. La TNR es una variante de las redes neuronales recurrentes (RNN) y las redes neuronales convolucionales (CNN) que se basa en la idea de la "atención" para procesar secuencias de datos de manera eficiente.

La atención es un mecanismo que permite a la red enfocarse en partes específicas de una secuencia durante el procesamiento. En una TNR, la atención se aplica de manera recursiva en múltiples capas, lo que permite que la red capture relaciones a largo plazo en la secuencia de entrada. Esta capacidad de atención en múltiples niveles es lo que diferencia a las TNR de otras arquitecturas de redes neuronales.

Las TNR han demostrado un rendimiento sobresaliente en una amplia gama de tareas de procesamiento del lenguaje natural, como la traducción automática, el reconocimiento de voz, la generación de texto y la respuesta a preguntas.

La arquitectura de las Redes Neuronales Transformadoras (TNR) se basa en el concepto de atención y consta de varios componentes clave. La arquitectura básica de una TNR se puede describir de la siguiente manera:

- Codificador (*Encoder*): El codificador es la parte de la red que procesa la secuencia de entrada y genera representaciones de alta dimensionalidad que capturan el contexto y las relaciones entre las palabras. El codificador consta de múltiples capas, cada una de las cuales contiene subcapas de atención y capas de alimentación hacia adelante (*feed-forward*).
- Autoatención (*Self-attention*): La autoatención es un mecanismo esencial en las TNR. Permite a la red asignar ponderaciones a diferentes palabras en la secuencia de entrada, enfocándose en las relaciones contextuales entre ellas. La autoatención se aplica tanto a nivel de palabras individuales como a nivel de secuencia completa, lo que permite capturar tanto la información local como la global.
- Capas de alimentación hacia adelante (*Feed-forward layers*): Después de la autoatención, se aplican capas de alimentación hacia adelante para procesar las representaciones generadas en cada capa. Estas capas consisten en capas completamente conectadas y funciones de activación no lineales para transformar y mezclar las características de las palabras.
- Decodificador (*Decoder*): En muchas aplicaciones de procesamiento del lenguaje natural, como la traducción automática, se utiliza un decodificador para generar la secuencia de salida. El decodificador también consta de múltiples capas que incluyen autoatención y capas de alimentación hacia adelante. El decodificador utiliza las representaciones generadas por el codificador y aprende a generar una secuencia de salida coherente y relevante.

2.9.5 Reducción de dimensionalidad

La reducción de dimensionalidad es una técnica utilizada para disminuir el número de características en un conjunto de datos, al mismo tiempo que se conserva la mayor cantidad de información importante posible. En aprendizaje automático, los datos de alta dimensionalidad se refieren a datos con un gran número de características o variables. Esta técnica ayuda a reducir la complejidad del modelo y mejorar su rendimiento.

La maldición de la dimensionalidad se refiere al hecho de que, en ausencia de suposiciones simplificadoras, el tamaño de la muestra necesario para estimar una función de varias variables hasta cierto grado de precisión (es decir, obtener una estimación con una varianza razonablemente baja) crece exponencialmente con el número de variables. Hay dos enfoques principales para la reducción de dimensionalidad (figura 2.9.3) [16]:

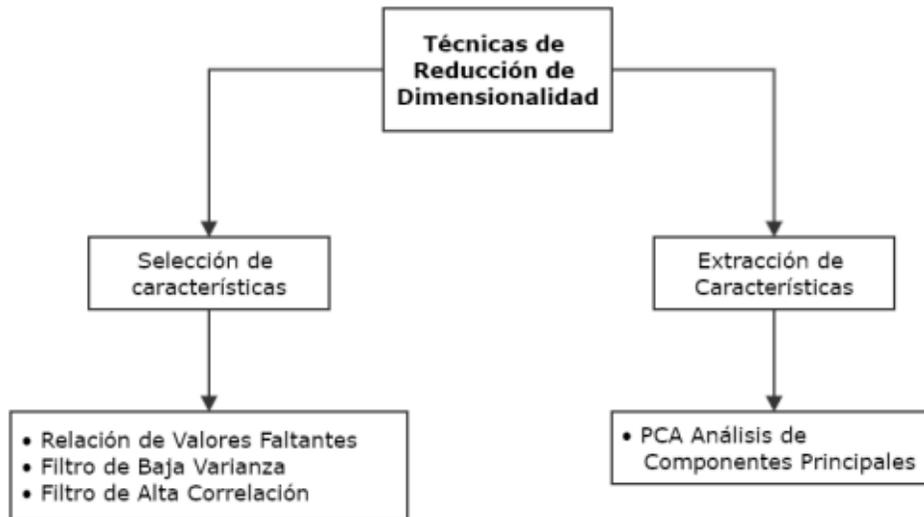


Figura 2.9.3 Técnicas para reducción de la dimensionalidad

2.9.5.1. Selección de características

La selección de características implica elegir un subconjunto de las características originales que son más relevantes para el problema en cuestión. El objetivo es reducir la dimensionalidad del conjunto de datos mientras se retienen las características más importantes.

La Proporción de Valores Ausentes es un método sencillo de selección de características que toma decisiones basadas en la cantidad de valores faltantes en una característica.

2.9.5.1.1. Relación de valores faltantes

Implica descartar aquellas características que tienen un porcentaje significativo de datos ausentes, superando un umbral predefinido. Este enfoque busca simplificar el conjunto de datos y evitar complicaciones relacionadas con la imputación de valores faltantes en esas variables.

2.9.5.1.2. Filtro de baja varianza

Es una técnica utilizada comúnmente para identificar aquellas variables que tienen una varianza muy baja en comparación con el resto del conjunto de datos. Esta ayuda a identificar características que no varían mucho a lo largo de las instancias del conjunto de datos, lo que sugiere que podrían no ser informativas o contribuir significativamente a la variabilidad del conjunto de datos.

2.9.5.1.3. Filtro de alta correlación

Sirve para identificar y eliminar características que exhiben una correlación significativa entre sí. En este método, se establece un umbral de correlación y aquellas características cuya correlación supera este umbral son consideradas altamente correlacionadas. La eliminación de características con alta correlación se realiza debido a que sus valores cambian de manera muy similar, lo que indica que proporcionan información redundante.

2.9.5.2. Extracción de características

La extracción de características implica crear nuevas características mediante la combinación o transformación de las características originales. El objetivo es crear un conjunto de características que capture la esencia de los datos originales en un espacio de menor dimensionalidad.

2.9.5.2.1. Análisis de componentes principales PCA

El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es un método estadístico versátil para reducir una tabla de datos de casos por variables a sus características esenciales, llamadas componentes principales. Estos componentes principales son unas pocas combinaciones lineales de las variables originales que explican de manera óptima la variabilidad de todas las variables. En el proceso, el método proporciona una aproximación de la tabla de datos original utilizando solo estos pocos componentes principales. [17]

Es útil como herramienta para la visualización, filtrado de ruido y extracción de características, ya que extrae un nuevo conjunto de variables de un gran conjunto de variables, estas variables extraídas se denominan componentes principales. Un componente principal es una combinación lineal normalizada. Estas nuevas características son ortogonales, lo que significa que no están correlacionadas, además se clasifican según su varianza explicada.

Una forma intuitiva de entender el proceso de PCA consiste en interpretar las componentes principales desde un punto de vista geométrico (figura 2.9.4). Supóngase un conjunto de observaciones para las que se dispone de dos variables (X_1 , X_2). El vector que define la primera componente principal (Z_1) sigue la dirección en la que las observaciones varían más (línea roja). La proyección de cada observación sobre esa dirección equivale al valor de la primera componente para dicha observación (*principal component scores*, z_{i1}).

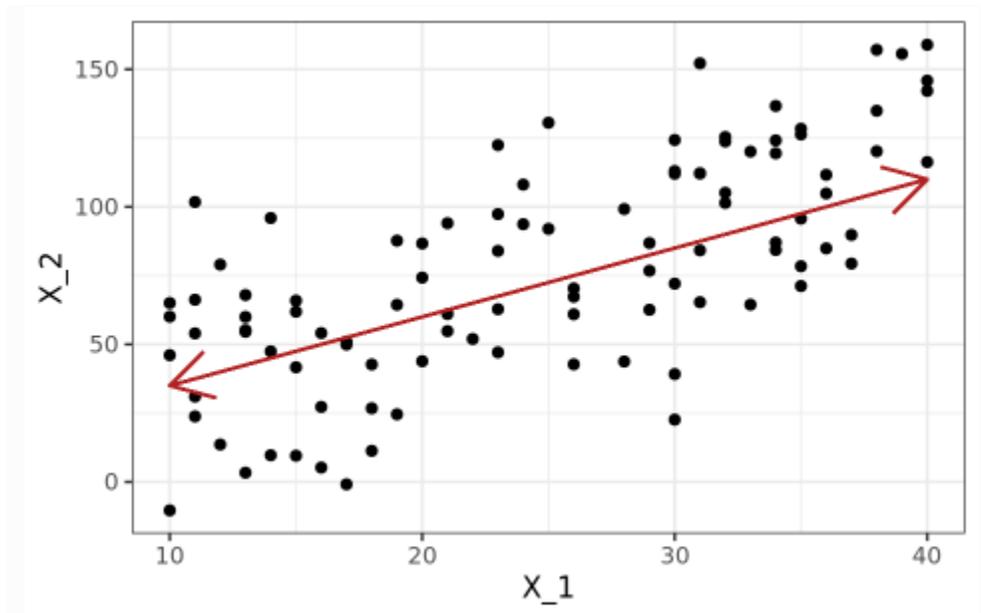


Figura 2.9.4 Interpretación geométrica de PCA con vector Z_1

La segunda componente (Z_2) sigue la segunda dirección en la que los datos muestran mayor varianza y que no está correlacionada con la primera componente (figura 2.9.5). La condición de no correlación entre componentes principales equivale a decir que sus direcciones son perpendiculares/ortogonales.

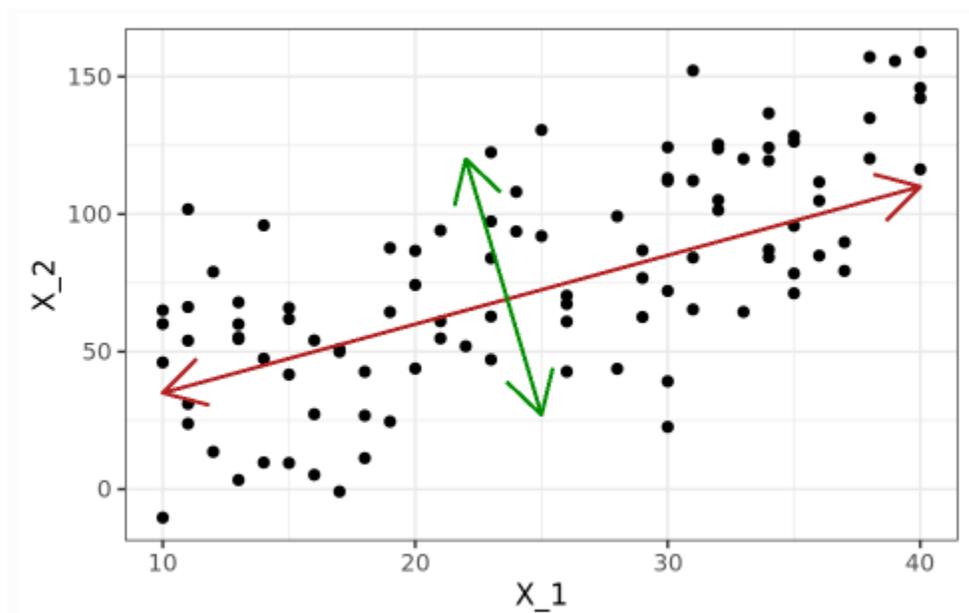


Figura 2.9.5 Interpretación geométrica de PCA con vector Z_1 y Z_2

2.9.6 Métricas de clasificación

Las métricas de evaluación son medidas utilizadas para evaluar el rendimiento de los modelos de aprendizaje automático. Estas métricas se utilizan para comparar diferentes modelos y seleccionar el que tenga un mejor rendimiento en base a las necesidades y objetivos del problema a resolver.

2.9.6.1. Matriz de confusión

Una matriz de confusión es una herramienta utilizada en el campo de la clasificación en el aprendizaje automático para evaluar el rendimiento de un modelo predictivo. La matriz muestra la cantidad de predicciones correctas e incorrectas hechas por el modelo en comparación con los valores reales en un conjunto de datos de prueba (figura 2.9.6)

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 2.9.6 Matriz de confusión

Dado un modelo de aprendizaje automático y una instancia, existen 4 resultados posibles:

- Verdaderos Positivos (TP): es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (TN): es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FN): es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FP): es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Métricas obtenidas a partir de la matriz de confusión

Exactitud (*Accuracy*): La precisión simplemente mide con qué frecuencia el clasificador predice correctamente. Podemos definir la precisión como la proporción entre el número de predicciones correctas y el número total de predicciones.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precisión (*Precision*): La precisión explica cuántos de los casos correctamente predichos resultaron ser positivos. La precisión es útil en situaciones donde la preocupación por los Falsos Positivos es mayor que la de los Falsos Negativos.

$$Precision = \frac{TP}{TP + FP}$$

Recuperación (*Recall*)- Sensibilidad (*Sensitivity*): Explica cuántos de los casos positivos reales pudimos predecir correctamente con nuestro modelo. El recuerdo es una métrica útil en casos donde la preocupación por los Falsos Negativos es mayor que la de los Falsos Positivos.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: Ofrece una idea combinada de las métricas de *Precision* y *Recall*. Alcanza su valor máximo cuando la *Precision* es igual al *Recall*.

$$F1\ Score = \frac{(Precision)(Recall)}{Precision + Recall}$$

2.9.6.2. Curva AUC-ROC

La Curva Característica del Operador Receptor (ROC) es una curva de probabilidad que representa la Tasa de Verdaderos Positivos (TPR) frente a la Tasa de Falsos Positivos (FPR) en varios valores de umbral y separa la 'señal' del 'ruido'. El Área Bajo la Curva (AUC) es la medida de la capacidad de un clasificador para distinguir entre clases. A partir del gráfico, simplemente decimos que el área bajo la curva ABDE y los ejes X e Y.

A partir de la figura 9.2.7, cuanto mayor sea el AUC, mejor será el rendimiento del modelo en diferentes puntos de umbral entre las clases positivas y negativas. Esto simplemente significa que cuando el AUC es igual a 1, el clasificador puede distinguir perfectamente entre todos los puntos de clase Positiva y Negativa. Cuando el AUC es igual a 0, el clasificador estaría prediciendo todos los Negativos como Positivos y viceversa. Cuando el AUC es 0.5, el clasificador no puede distinguir entre las clases Positiva y Negativa.

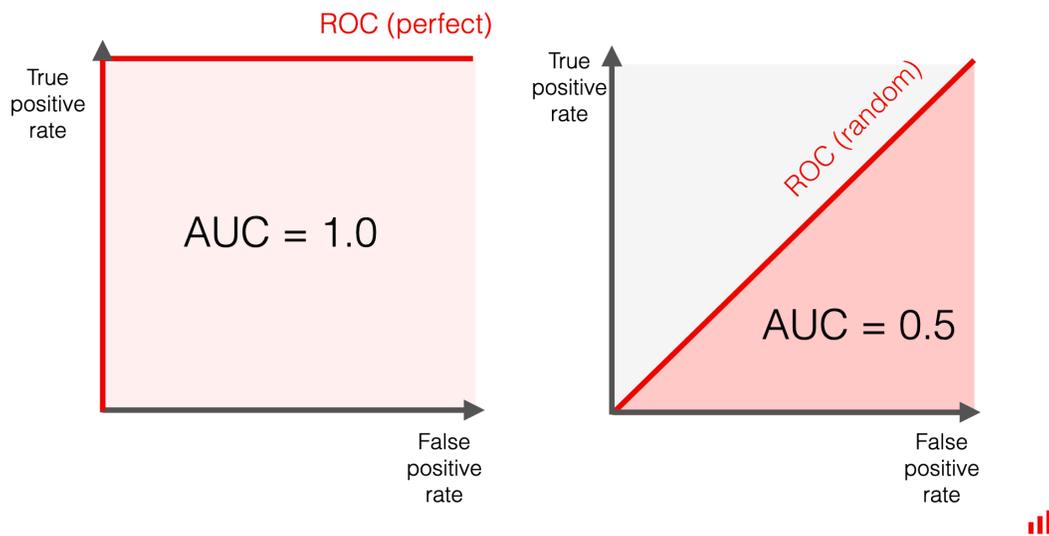


Figura 9.2.7 Curva AUC/ROC

2.9.6.3. Pérdida Logarítmica

La pérdida logarítmica también llamada pérdida de regresión logística o pérdida de entropía cruzada, mide el rendimiento de un modelo de clasificación donde la entrada de predicción es un valor de probabilidad entre 0 y 1. El objetivo del modelo de aprendizaje automático es minimizar este valor. Un modelo perfecto tendría una pérdida logarítmica de 0 y aumenta a medida que la probabilidad prevista difiere de la etiqueta real. Para un buen algoritmo predictivo la pérdida debe estar cerca de cero. El objetivo de un modelo de aprendizaje automático es minimizar este valor. [15].

Capítulo 3

RECURSOS DE SOFTWARE

3.1.Herramientas Utilizadas

3.1.1. Python

Python es un lenguaje de programación multiparadigma. Esto significa que más que forzar a los programadores a adoptar un estilo particular de programación, permite varios estilos: programación orientada a objetos, programación imperativa y programación funcional. Otros paradigmas están soportados mediante el uso de extensiones.

Una característica importante de Python es la resolución dinámica de nombres; es decir, lo que enlaza un método y un nombre de variable durante la ejecución del programa (también llamado enlace dinámico de métodos).

Otro objetivo del diseño del lenguaje es la facilidad de extensión. Se pueden escribir nuevos módulos fácilmente en C o C++. Python puede incluirse en aplicaciones que necesitan una interfaz programable. [11]

3.1.2. Jupyter Notebook

Jupyter Notebook es un entorno computacional interactivo basado en web para crear documentos de cuaderno Jupyter. Es compatible con varios lenguajes como Python (IPython), Julia, R, etc. y se utiliza principalmente para análisis de datos, visualización de datos y más computación exploratoria interactiva.

Proporciona un entorno donde puede registrar código, ejecutar código, ver resultados, visualizar datos y ver resultados de salida. Estas características lo convierten en una herramienta conveniente para realizar flujos de trabajo de ciencia de datos de extremo a extremo, que se pueden utilizar para la limpieza de datos, el modelado estadístico, la creación y capacitación de modelos de aprendizaje automático, la visualización de datos y muchos otros propósitos.

3.1.3. Keras

Keras es una biblioteca de código abierto (con licencia MIT) escrita en Python, que se basa principalmente en el trabajo de François Chollet, un desarrollador de Google, en el marco del proyecto ONEIROS (*Open-ended Neuro-Electronic Intelligent Robot Operating System*).

El objetivo de la biblioteca es acelerar la creación de redes neuronales: para ello, Keras no funciona como un *framework* independiente, sino como una interfaz de uso intuitivo (API) que permite acceder a varios *frameworks* de aprendizaje automático y desarrollarlos. Entre los *frameworks* compatibles con Keras, se incluyen Theano, Microsoft Cognitive Toolkit (anteriormente CNTK) y TensorFlow.

3.1.4. Pandas

La biblioteca de software de código abierto Pandas está diseñada específicamente para la manipulación y el análisis de datos en el lenguaje Python. Es potente, flexible y fácil de usar.

Con esta biblioteca se puede utilizar el lenguaje Python para cargar, alinear, manipular o incluso fusionar datos. El rendimiento es realmente impresionante cuando el código fuente del *back-end* está escrito en C o Python.

Esta herramienta permite importar y exportar datos en distintos formatos, como CSV o JSON. Además, Pandas también ofrece la funcionalidad de *Data Cleaning*. Esta biblioteca es muy útil para trabajar con datos estadísticos, datos tabulares como tablas SQL o Excel, con datos de series temporales y con datos matriciales arbitrarios con etiquetas de filas y columnas.

3.1.5. HDF5

El paquete h5py es una interfaz de Python para el formato binario de datos HDF5, este permite almacenar grandes cantidades de datos numéricos y manipular fácilmente esos datos desde NumPy. Se pueden almacenar miles de conjuntos de datos en un solo archivo, categorizados y etiquetados como *desees*. [15]

Un archivo HDF5 es un contenedor para dos tipos de objetos: conjuntos de datos (*datasets*) y grupos (*groups*).

- Conjuntos de datos (*datasets*): Estos son objetos similares a matrices que contienen datos. Los conjuntos de datos pueden almacenar datos numéricos, cadenas y otros tipos de datos. Son los objetos principales para almacenar y manipular datos reales dentro de un archivo HDF5.
- Grupos (*groups*): Estos son contenedores similares a carpetas que pueden contener conjuntos de datos y otros grupos. Los grupos te permiten organizar y estructurar tus datos dentro del archivo HDF5. Proporcionan una forma de categorizar y agrupar conjuntos de datos relacionados.

3.1.6. NumPy

NumPy es el paquete fundamental para la computación científica en Python. Es una biblioteca de Python que proporciona un objeto de matriz multidimensional, varios objetos derivados (como matrices enmascaradas y matrices) y una variedad de rutinas para realizar operaciones rápidas en matrices. Esto incluye operaciones matemáticas, lógicas, manipulación de formas, ordenamiento, selección, entrada/salida, transformadas discretas de Fourier, álgebra lineal básica, operaciones estadísticas básicas, simulación aleatoria y mucho más. [12]

3.1.7. Matplotlib

Matplotlib es una librería Python *open source*, es posible crear trazados, histogramas, diagramas de barra y cualquier tipo de gráfica con ayuda de algunas líneas de código. Se trata de una herramienta muy completa, que permite generar visualizaciones de datos muy detalladas.

Esta librería es particularmente útil para las personas que trabajan con Python o NumPy. Se emplea particularmente en servidores de aplicación *web*, *shells* y *scripts* de Python. Con las API de matplotlib, también es posible que los desarrolladores integren gráficas con aplicaciones de interfaz gráfica.

3.2. Bibliotecas y Herramientas de Python para el procesamiento de registros sísmicos

3.2.1. EQTransformer

EQTransformer es una herramienta utilizada en sismología y análisis de señales sísmicas. Específicamente, EQTransformer es un método y software de procesamiento de datos sísmicos que utiliza redes neuronales convolucionales para detectar y localizar eventos sísmicos, así como para realizar la clasificación de señales sísmicas.

EQTransformer está basada en una arquitectura de redes neuronales convolucionales bidimensionales (2D-CNN) y redes neuronales recurrentes (RNN). La arquitectura general de se compone de los siguientes componentes principales (figura 3.2.1):

- Bloque de extracción de características: Este bloque está formado por capas de redes neuronales convolucionales bidimensionales (2D-CNN). Estas capas convolucionales se utilizan para extraer características importantes de las formas de onda sísmicas, capturando patrones y estructuras relevantes en los datos.
- Bloque de atención temporal: Después del bloque de extracción de características, se aplica una capa de atención temporal basada en redes neuronales recurrentes (RNN). La atención temporal se utiliza para resaltar partes importantes de la secuencia de datos sísmicos y ayudar en la detección y localización precisa de los eventos sísmicos.

- Bloque de detección y clasificación: Este bloque utiliza capas de redes neuronales totalmente conectadas para realizar la detección y clasificación de los eventos sísmicos. Utiliza información extraída de las etapas anteriores para identificar las llegadas de los eventos sísmicos y clasificar las señales sísmicas en diferentes categorías.

La combinación de las capas de redes neuronales convolucionales bidimensionales (2D-CNN) y redes neuronales recurrentes (RNN) en EQTransformer permite un procesamiento eficiente y una extracción precisa de características en las señales sísmicas. Esto facilita la detección, localización y clasificación automática de eventos sísmicos en grandes volúmenes de datos sísmicos. [13]

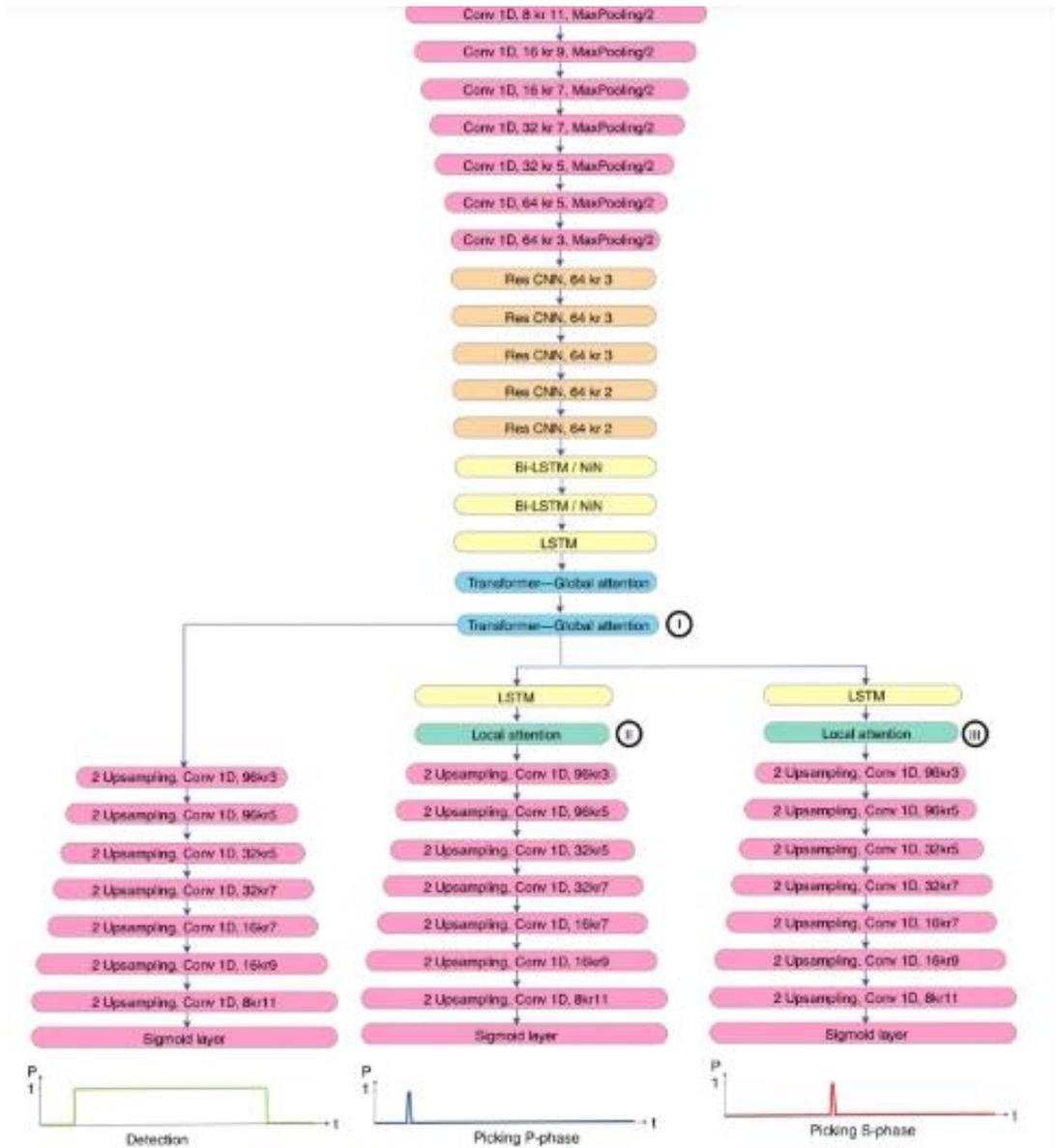


Figura 3.2.1 Arquitectura de EQTransformer

Para entrenar el EQTransformer se utiliza el conjunto de datos STEAD (*Stanford Earthquake Dataset*). STEAD es una base de datos mundial a gran escala de señales sísmicas y no sísmicas etiquetadas. Aquí utilizamos formas de onda de terremotos de 1 M y de ruido de 300 K (incluyendo tanto ruido ambiental como cultural) registradas por ~ 2600 estaciones sísmicas a distancias epicentrales de hasta 300 km. Las formas de onda sísmicas están asociadas a unos 450 K terremotos con una distribución geográfica diversa en todo el mundo. La mayoría de estos terremotos son menores de M 2,5 y se han registrado a menos de 100 km del epicentro. Las formas de onda tienen una duración de 1 minuto, una frecuencia de muestreo de 100 Hz y un filtro de paso de banda causal de 1,0-45,0 Hz.

Se utilizaron más de 113,000 formas de onda de prueba (ejemplos de terremotos y ruido) para evaluar y comparar el rendimiento de detección y selección de EQTransformer con otros métodos de aprendizaje profundo y métodos tradicionales. Los modelos de aprendizaje profundo utilizados para las comparaciones son modelos preentrenados basados en diferentes conjuntos de entrenamiento, y todos se aplican a un conjunto de pruebas común

La comparación de rendimientos, aplicaron tres detectores de aprendizaje profundo (DetNet5, Yews4 y CRED7) y un detector tradicional (STA/LTA11) al mismo conjunto de pruebas (Tabla 3.2.1) [13]

EQTransformer y CRED se entrenaron con el conjunto de datos STEAD, mientras que los resultados de DetNet y Yews se basan en modelos preentrenados en diferentes conjuntos de datos. Aquí se utiliza el algoritmo STA/LTA recursivo. Los valores en negrita representan el mejor rendimiento. (tabla 3.2.1) [13]

Tabla 3.2.1 Pr, Re y F1 representan precisión, recuperación y puntuación. [13]

<i>Model</i>	<i>Pr</i>	<i>Re</i>	<i>F1</i>	<i>Training data</i>	<i>Training size</i>
EQTransformer	1.0	1.0	1.0	Global	1.2M
CRED	1.0	0.96	0.98	Global	1.2M
DetNet	1.0	0.89	0.94	China	30K
Yews	0.84	0.85	0.85	Taiwan	1.4M
STA/LTA	0.91	1.0	0.95	—	—

μ y σ son la media y la desviación estándar de los errores (verdad conocida - predicción) en segundos, respectivamente. Pr, Re y F1 representan la precisión, la recuperación y la puntuación F1, respectivamente. MAE y MAPE son el error absoluto medio y el error porcentual absoluto medio, respectivamente (tabla 3.2.2 y tabla 3.2.3). [13]

Tabla 3.2.2 Identificación de Fase P [13]

<i>Model</i>	μ	σ	<i>Pr</i>	<i>Re</i>	<i>F1</i>	<i>MAE</i>	<i>MAPE</i>	<i>Training data</i>	<i>Training size</i>
EQTransformer	0.00	0.03	0.99	0.99	0.99	0.01	0.00	Global	1.2M
PhaseNet	-0.02	0.08	0.96	0.96	0.96	0.07	0.01	North California	780K
GPD	0.03	0.10	0.81	0.80	0.81	0.08	0.01	South California	4.5M
PickNet	0.00	0.09	0.81	0.49	0.61	0.07	0.02	Japan	740K
PpkNet	-0.01	0.15	0.90	0.90	0.90	0.10	1.90	Japan	30K
Yews	0.07	0.13	0.54	0.72	0.61	0.09	0.02	Taiwan	1.4M
Kurtosis	-0.03	0.09	0.94	0.79	0.86	0.08	0.01	—	—
FilterPicker	-0.01	0.08	0.95	0.82	0.88	0.14	0.02	—	—
AIC	-0.04	0.09	0.92	0.83	0.87	0.09	0.01	—	—

Tabla 3.2.3 Identificación de Fase S [13]

<i>Model</i>	μ	σ	<i>Pr</i>	<i>Re</i>	<i>F1</i>	<i>MAE</i>	<i>MAPE</i>	<i>Training data</i>	<i>Training size</i>
EQTransformer	0.00	0.11	0.99	0.96	0.98	0.01	0.00	Global	1.2M
PhaseNet	-0.02	0.11	0.96	0.93	0.94	0.09	0.01	North California	780K
GPD	0.03	0.14	0.81	0.83	0.82	0.10	0.01	South California	4.5M
PickNet	0.08	0.17	0.75	0.75	0.75	0.10	0.03	Japan	740K
PpkNet	0.02	0.15	1.00	0.91	0.95	0.10	1.85	Japan	30K
Yews	-0.02	0.13	0.83	0.55	0.66	0.11	0.01	Taiwan	1.4M
Kurtosis	-0.10	0.13	0.89	0.39	0.55	0.11	0.01	—	—
FilterPicker	-0.05	0.13	0.61	0.41	0.49	0.10	0.01	—	—
AIC	-0.07	0.15	0.87	0.51	0.64	0.12	0.02	—	—

3.2.2. MSNoise

El uso del ruido sísmico ambiental ha demostrado recientemente su eficacia en diversos contextos, desde la obtención de imágenes hasta la monitorización. La respuesta al impulso (o función de Green [GF]) entre dos sensores puede reconstruirse a partir de la correlación del ruido sísmico registrado (Campillo y Paul, 2003). Este método ha proporcionado resultados excelentes en la obtención de imágenes del interior de la Tierra, desde escalas globales hasta regionales o locales.

La automatización de la detección es importante, ya que, por ejemplo, las corrientes de datos provenientes de estaciones telemétricas en tiempo real contienen lagunas que podrían completarse más adelante y proporcionar datos importantes para su análisis. Tal herramienta debe ser capaz de trabajar con cualquier formato sísmico común, ser razonablemente rápida y estar optimizada cuando sea posible. Debería interactuar con un archivo de datos y una base de datos con funciones auxiliares de alto nivel para poder conectarse y ampliarse. Finalmente, debe producir datos exportables, ya sea en formato de forma de onda, archivos de texto tabulares o figuras de alta calidad. Este es el propósito de *Monitoring using Seismic Noise* (es decir, MSNoise).

MSNoise es el primer paquete de software completo para calcular y monitorear variaciones relativas de velocidad utilizando ruido sísmico ambiental este nos proporciona gráficos de $\delta v/v$ a lo largo del tiempo. Estos datos se pueden representar en varios gráficos, uno por par de estaciones, uno promedio para todos los pares de estaciones, etc. MSNoise incluye varios ejemplos de gráficos que se pueden duplicar o modificar según las necesidades del usuario. MSNoise utiliza pandas para el análisis de datos (McKinney, 2012), ya que proporciona funciones auxiliares excelentes para analizar datos de series temporales (promedio móvil, remuestreo, eliminación de tendencias en series temporales que contienen NaN, etc.) y matplotlib (Hunter, 2007) para generar salidas gráficas estáticas (figura 3.2.2).

Para trabajar requiere Python y solo algunos paquetes adicionales para ejecutarse en cualquier sistema operativo (SO). El archivo de formas de onda debe estar en un formato conocido, como la Estructura de Datos SeisComP (SDS) o la Estructura de Datos Uniforme de Buffer (BUD), o uno puede definir el suyo propio. MSNoise utiliza una base de datos para almacenar bits de configuración, metadatos de formas de onda y trabajos. Esta base de datos puede ser tanto MySQL como sqlite, aunque recomendamos utilizar MySQL por razones de rendimiento. El paquete incluye un script de instalación que inicializará la conexión a la base de datos y almacenará el nombre de usuario y la contraseña localmente para permitir que el Configurador y todos los procesos interactúen con la base de datos.

El Configurador GUI (construido utilizando *Enthought Tools Suite* [Enthought, 2008]) de MSNoise muestra tres paneles diferentes: la configuración general de MSNoise, la configuración de estaciones y la configuración de filtros.

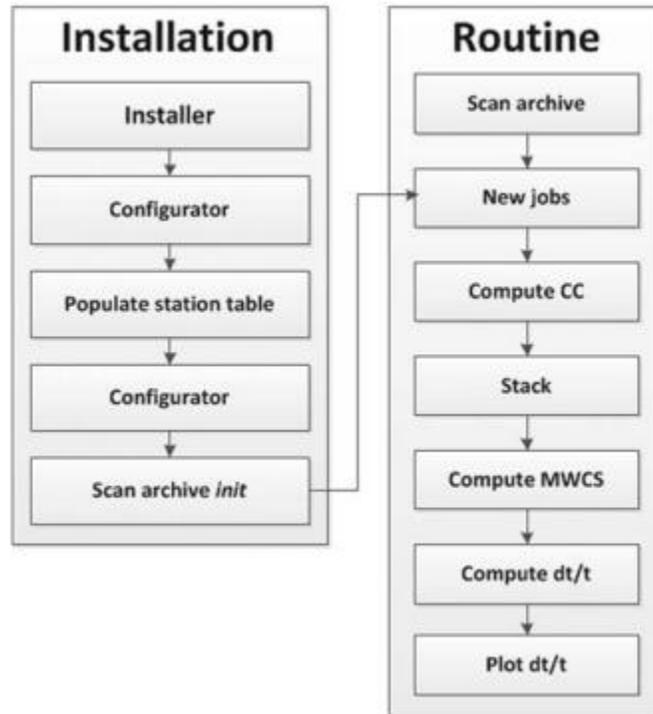


Figura 3.2.2 Vista esquemática del flujo de trabajo de MSNoise, la parte de instalación única entra en la rutina del flujo de trabajo antes del paso de definición de trabajos

3.2.3. ObsPy

El objetivo de ObsPy es proporcionar una caja de herramientas de Python que simplifique el uso de la programación en Python para sismólogos. En ObsPy se implementan y están listas para usar las siguientes rutinas esenciales de procesamiento sismológico: lectura y escritura de datos solo en SEED/MiniSEED y Dataless SEED y SAC, así como filtrado, simulación de instrumentos, activación y trazado.

Python le proporciona al usuario todas las características de un lenguaje de programación completo, incluida una gran colección de módulos científicos de código abierto. ObsPy extiende Python al proporcionar acceso directo a las series temporales reales, lo que permite el uso de potentes módulos de programación de arrays numéricos como NumPy o SciPy. Los resultados se pueden visualizar utilizando módulos como matplotlib (2D). [14]

Capítulo 4

METODOLOGÍA

Existen varios programas de cómputo para analizar el ruido sísmico, entre ellos está la suite de códigos en Python denominada MSNoise. Esta herramienta está basada en una secuencia de procesados que son: apilado, crosscorrelacion, promedios móviles espectrales y graficación de resultados. El programa tiene un flujo de trabajo establecido de tal forma que el propósito final es tener una curva de variación de velocidad como función del tiempo. Esta grafica es la que sirve para interpretar cualquier cambio o fluctuación estacional.

El programa realiza la correlación por año, Los diagramas de dv/v se usaron para analizar el porcentaje de variación anual de la velocidad relativa, así como para asociar dicha variación con fenómenos hidrológicos (Clements y Denolle, 2018).

Debido a que el objetivo general del presente trabajo es estudiar los cambios temporales, presentamos a continuación un análisis que conecta los cambios de velocidad y de amplitud de H/V con fenómenos hidrológicos. Varios autores (Clements y Denolle, 2018; Tsai, 2011; etc.) describen detalladamente diferentes procesos que involucran los cambios de dv/v con respecto a termoelasticidad, poroelasticidad y cambios hidrológicos, en nuestro caso los cambios poro elásticos e hidrológicos los consideramos en el mismo fenómeno y la parte termo elástica no será incluida debido a que sólo afecta regiones muy áridas donde el coeficiente de conductividad térmica controla las características elásticas de las rocas (Tsai, 2011).

4.1. Cambios de velocidad de la propagación de las ondas sísmicas

Existen modelos teóricos que pueden relacionar los cambios de velocidad con características físicas del medio poroso utilizando las ecuaciones de Berger (1975). Esta relación nos permite modelar analíticamente la correlación de la cantidad de agua que hay en un acuífero con la curva final que presentamos en esta tesis. Es decir, como es bien sabido una simple correlación de datos entre cantidad de agua o de lluvia y los cambios de velocidad en el tiempo no son suficientes para demostrar causalmente la relación de agua con el suelo. Se debe de tener bases físicas para que la correlación que encontremos entre dv/v y cantidad de lluvia tengan sentido.

Aunque no se observa una clara correlación entre precipitación y los cambios mínimos estimados además de no contar con información de pozos, la tendencia de las figuras muestra una posible relación entre la precipitación y los cambios de velocidad. La variación entre

pico y pico es similar a la reportada por otros autores (Hillers et al., 2015), los valores se encuentran en un rango de $\pm 0.2\%$.

Un análisis hidrológico completo de la cuenca del Valle de México está fuera del alcance de la presente tesis, existe trabajo en proceso para integrar los resultados que hemos obtenido con el Departamento de Recursos Naturales del Instituto de Geofísica de la UNAM, quienes se han acercado para realizar estudios interdisciplinarios que nos permitan modelar los flujos subterráneos de la cuenca del Valle de México a través de la información hidrológica que se puede inferir mediante la metodología que hemos presentado en este trabajo (Carciumaru y Ortega, 2022).

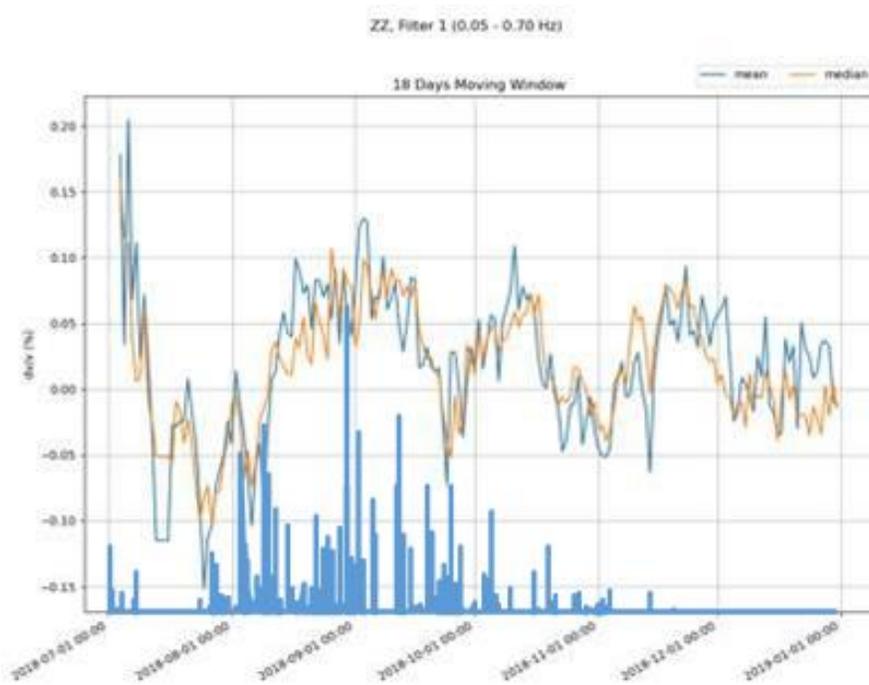


Figura 4.1.1 Series de tiempo de dv/v para la cuenca del Valle de México y valores de precipitación de la estación CU (2018) [19].

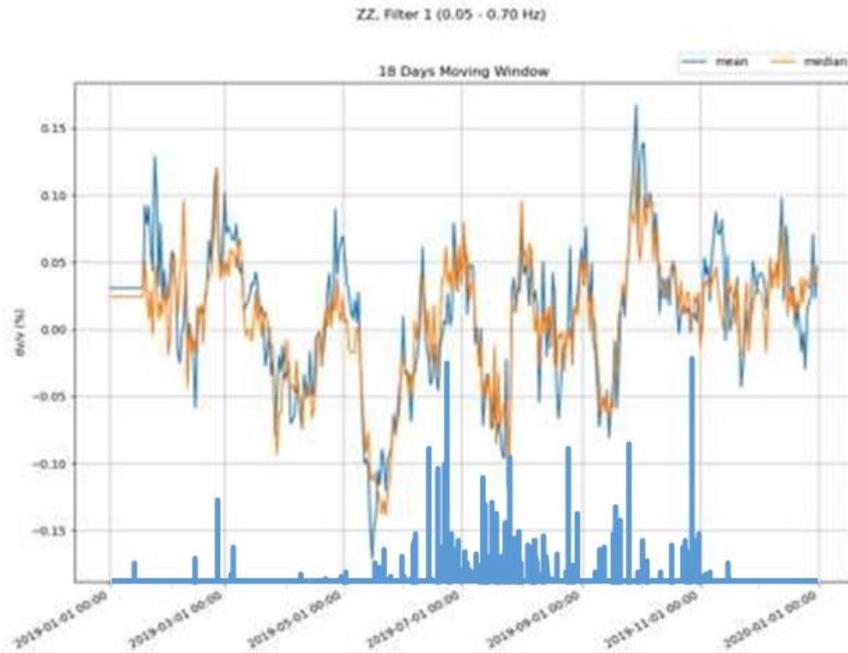


Figura 4.1.2 Series de tiempo de dv/v para la cuenca del Valle de México y valores de precipitación de la estación CU (2019) [19].

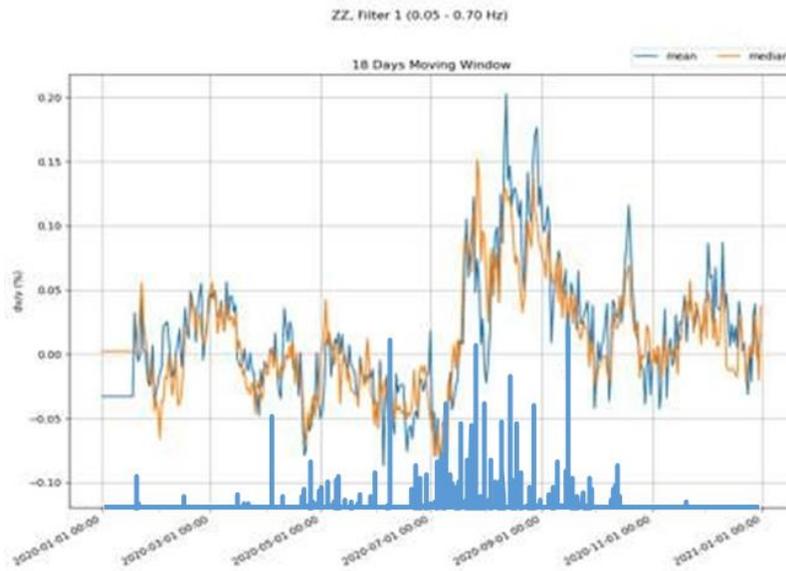


Figura 4.1.3 Series de tiempo de dv/v para la cuenca del Valle de México y valores de precipitación de la estación CU (2020) [19].

4.2. Datos

Los datos que se utilizaron son registros sísmicos de la red sísmica del Valle de México de la Universidad Autónoma de México (Figura 4.2.1). Esta red sísmica fue instalada para registrar los sismos que ocurren dentro de la ciudad de México, sin embargo, es posible utilizarlo para diversos fines incluyendo el ruido sísmico que sea utilizado para diversos estudios de otra índole. Es muy importante destacar que todo tipo de estaciones sísmicas requieren de un estudio minucioso de tipo geotécnico. En el Apéndice 2 se hace una descripción completa de las características geotécnicas de la red sísmica del Valle de México. Las características geotécnicas son una descripción del sitio en las que se encuentran ubicadas las estaciones sísmicas, en general cualquier tipo de registro sísmico tiene una influencia directa del lugar donde sea instalado el sismógrafo.

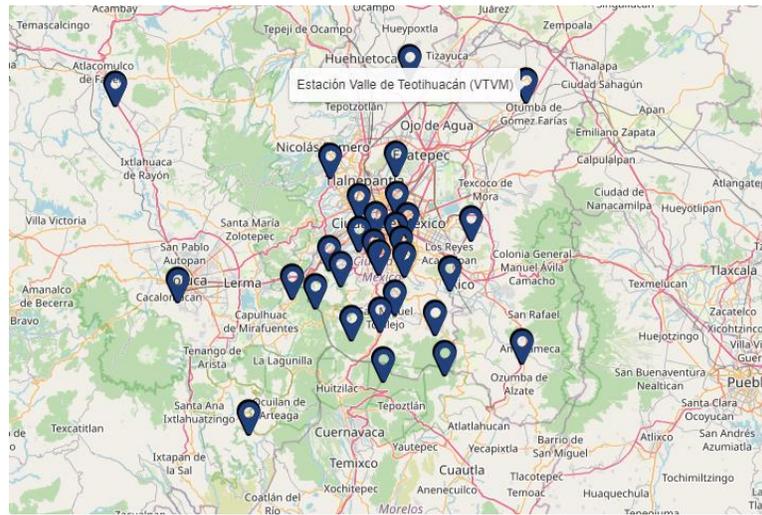


Figura 4.2.1: Distribución de estaciones de La Red Sísmica del Valle de México [20]

La red está conformada por varias estaciones sísmicas estratégicamente ubicadas en diferentes partes del Valle de México:

Tabla 4.2.1 Estaciones de la Red sísmica del Valle de México [20]

Código	Nombre	Estado	Latitud	Longitud	
AMVM	Amecameca	Estado de México	19.131894	-98.785873	VM
AOVM	Álvaro Obregón	Ciudad de México	19.267567	-99.3219	VM
APVM	Azcapotzalco	Ciudad de México	19.488017	-99.20865	VM

Código	Nombre	Estado	Latitud	Longitud	
ATVM	Atlacomulco	Estado de México	19.761847	- 99.841048	VM
AZVM	Atizapán	Estado de México	19.584833	- 99.282932	VM
BJVM	Benito Juárez	Ciudad de México	19.374986	-99.17069	VM
CHVM	Chichinautzin	Morelos	19.088367	- 99.147483	VM
CIVM	Cerro de Cilcuayo	Ciudad de México	19.10355	- 98.988117	VM
CJVM	Cuajimalpa	Ciudad de México	19.361647	- 99.285089	VM
COVM	Coyoacán	Ciudad de México	19.3511	- 99.156167	VM
CTVM	Cuauhtémoc	Ciudad de México	19.443032	- 99.165502	VM
GMVM	Gustavo A. Madero	Ciudad de México	19.493042	- 99.111033	VM
ICVM	Iztacalco	Ciudad de México	19.384483	-99.09895	VM
INVM	ININ	Estado de México	19.291225	- 99.382735	VM
IPVM	Iztapalapa	Ciudad de México	19.3461	- 99.091467	VM
MAVM	Malinalco	Estado de México	18.957208	-99.49409	VM
MCVM	Magdalena Contreras	Ciudad de México	19.320633	- 99.255317	VM
MHVM	Miguel Hidalgo	Ciudad de México	19.407983	- 99.209083	VM

Código	Nombre	Estado	Latitud	Longitud	
MPVM	Milpa Alta	Ciudad de México	19.201013	- 99.011447	VM
MZVM	Mezontepec	Ciudad de México	19.189182	- 99.229317	VM
PBVM	Peñón de los Baños	Ciudad de México	19.440729	- 99.083231	VM
PTVM	Pico Tres Padres	Estado de México	19.591275	- 99.112503	VM
THVM	Tláhuac	Ciudad de México	19.311011	- 98.973198	VM
TLVM	Tlalpan	Ciudad de México	19.209389	- 99.153727	VM
TOVM	Toluca	Estado de México	19.282842	- 99.677688	VM
TXVM	Texcoco	Estado de México	19.43369	- 98.917862	VM
VRVM	Venustiano Carranza	Ciudad de México	19.41785	-99.1144	VM
VTVM	Valle de Teotihuacán	Estado de México	19.770101	- 98.774508	VM
XCVM	Xochimilco	Ciudad de México	19.252731	-99.11685	VM
ZUVM	Zumpango	Estado de México	19.826459	- 99.077543	VM

En las figuras 4.2.2 y la 4.2.3 se representan ejemplos de registros de un día completo, este tipo de registros tienen el nombre de helicorders, debido a que en un principio los registros provenían de unos tambores en forma helicoidal donde se hacían originalmente los registros analógicos. Ahora esta representación es muy útil para poder ver la secuencia completa de un día, y de esta forma visualizar el contenido completo de lo que tiene la información sísmica. En algunos casos los helicorders nos ayudan a distinguir rápidamente sismos de gran importancia.

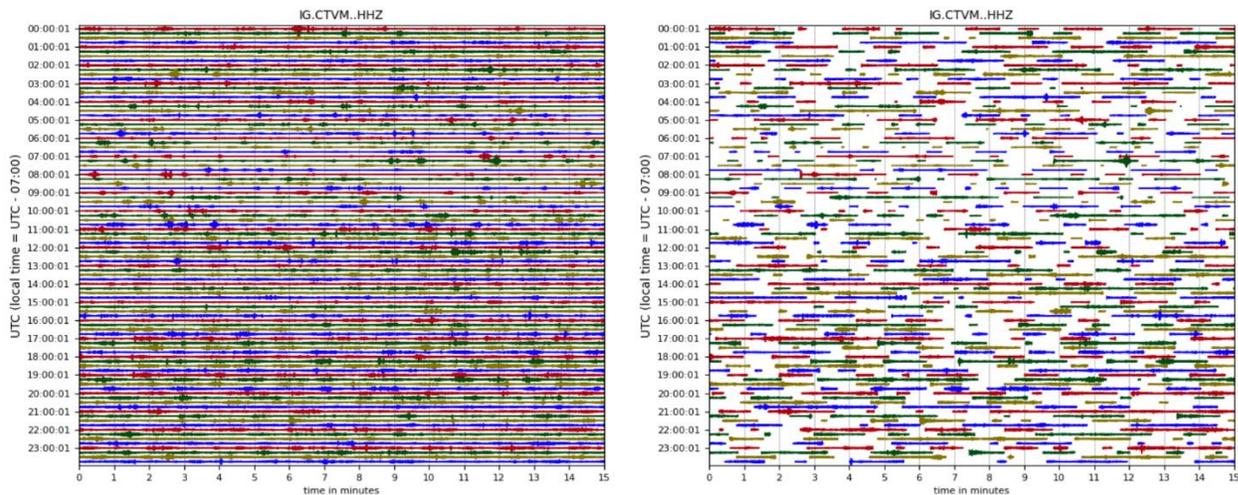


Figura 4.2.2 Helicorder de estación CTVM

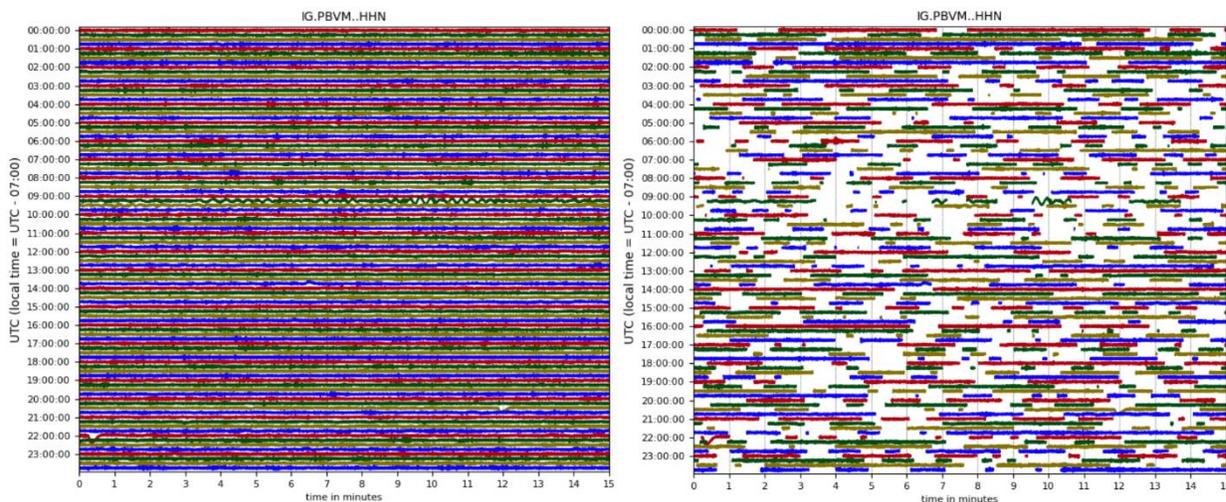


Figura 4.2.3 Helicorder de estación PBVM

Cada estación sísmica tiene características muy especiales debido a su cercanía con diferentes estructuras civiles, ya sean edificios, puentes, pozos de agua o cualquier otro tipo de edificaciones que puedan alterar, incluso en el Valle de México hay casos en los cuales se encuentran las estaciones exactamente arriba de túneles de la Ciudad de México. Por esta razón es importante hacer un análisis detallado de los datos que al mismo tiempo que comparar su calidad dependiendo la aplicación que se le vaya a dar. En nuestro caso tratamos de utilizar la mayor cantidad de información que se encuentra disponible durante el año de 2020. Es importante destacar que es posible seleccionar las estaciones que se encuentren en ciertas regiones específicas, por ejemplo, se pueden solamente seleccionar las estaciones que se encuentran en lo que antes era el lecho del Lago de Texcoco, como también se podrían escoger exclusivamente las estaciones que se encuentran en las zonas de Lomas y Lomeríos que pertenecen a antiguos depósitos volcánicos. En nuestro caso no hicimos distinción de ningún tipo, pero en el Apéndice 2 tenemos una descripción más detallada de donde se encuentran estas estaciones y sus características en términos de comportamiento dinámico para construcción y para sus características en frecuencia.

El objetivo de este trabajo es encontrar las características de variaciones de velocidad a lo largo de un año sin embargo en un futuro existe la posibilidad de refinar estos estudios y tratar de entender cuáles son las características en cada una de estas zonas específicas que se encuentran descritas en el Apéndice 2.

Capítulo 5

PROCESAMIENTO DE LOS DATOS

El propósito de llevar a cabo la segmentación de los registros sísmicos radica en la eliminación selectiva de todos los eventos sísmicos contenidos en el registro, permitiendo así un enfoque exclusivo en el análisis del ruido ambiental. Los resultados obtenidos a través de la aplicación del algoritmo EQTransformer revelaron la identificación precisa de los puntos de inicio y finalización de las fases P y S en un conjunto de 21,941,220 registros. Nuestro algoritmo, diseñado con este fin, utiliza estos marcadores temporales para segmentar el registro original, conservando únicamente los intervalos de tiempo correspondientes al ruido ambiental.

El proceso de manipulación de datos en MSNOISE se desglosó en dos fases distintas. Inicialmente, los datos fueron sometidos a un procesamiento integral sin previa segmentación, a partir de esto se obtuvieron graficas de relación cruzada (Figuras 5.1.2 a 5.1.5), estas brindaron información sobre los posibles filtros a aplicar para la graficar los cambios de velocidad (Figuras 5.1.6 a 5.1.9). Posteriormente, se replicó el mismo procedimiento de procesamiento, pero utilizando los datos que previamente habían sido segmentados por nuestro algoritmo y el filtro más adecuado previamente seleccionado, dando como resultado el gráfico de cambio de velocidad (Figura 5.1.10). Este enfoque estratégico permitió una comparación detallada de los resultados entre ambas instancias, contribuyendo a una evaluación más completa de la eficacia de la segmentación en el contexto del análisis de ruido ambiental mediante MSNOISE.

En el transcurso del procesamiento de datos desde EQTransformer hasta MSNOISE, se enfrentaron desafíos computacionales que abarcaban tanto aspectos de hardware como de software. Para hacer eficiente la ejecución de este trabajo, fue necesario realizar mejoras sustanciales en el hardware, culminando en la adquisición de un equipo con un procesador M1 de 10 núcleos, 256 GB de memoria RAM y una GPU con 16 núcleos. La necesidad de este robusto hardware se originó en la abrumadora cantidad de datos que requerían procesamiento, una carga que resultaba inviable para los equipos previamente utilizados.

Los inconvenientes de software se manifestaron principalmente en los algoritmos empleados para el procesamiento. En el caso de EQTransformer, la entrada de registros correspondientes a un año y a 22 estaciones, con tres componentes cada una, resultó en un conjunto aproximado de 21,941,220 registros (Figura 5.1.1). El tiempo de procesamiento asociado con esta tarea específica se realizó en una semana completa.

Por otro lado, en la fase inicial de MSNoise, al trabajar con registros completos, el procesamiento se completaba en cuatro días. Sin embargo, surgió una problemática significativa en la segunda fase, especialmente al introducir el algoritmo diseñado para la segmentación de datos. Inicialmente, el enfoque se centraba en el uso de dataframes, pero la complejidad de esta tarea se reflejó en un tiempo de procesamiento de dos horas por cada registro de una estación. La totalidad de los datos, se proyectaba un periodo de aproximadamente cuatro meses de trabajo continuo para llevar a cabo la segmentación completa. Para abordar este desafío, se implementó una base de datos, logrando reducir drásticamente los tiempos de segmentación y culminando en la finalización de la tarea en tan solo siete días.

En el caso de MSNoise, utilizando los datos no segmentados, el tiempo de trabajo se estimó en aproximadamente tres días, mientras que, con los datos segmentados, se experimentó un ligero incremento en el tiempo de trabajo, alcanzando los cinco días. Estas consideraciones proporcionan una visión integral de los desafíos enfrentados y las estrategias implementadas para superarlos en el proceso de procesamiento de datos sísmicos.

5.1. Gráficos MSNoise



Figura 5.1.1 Data Availability MSNOISE

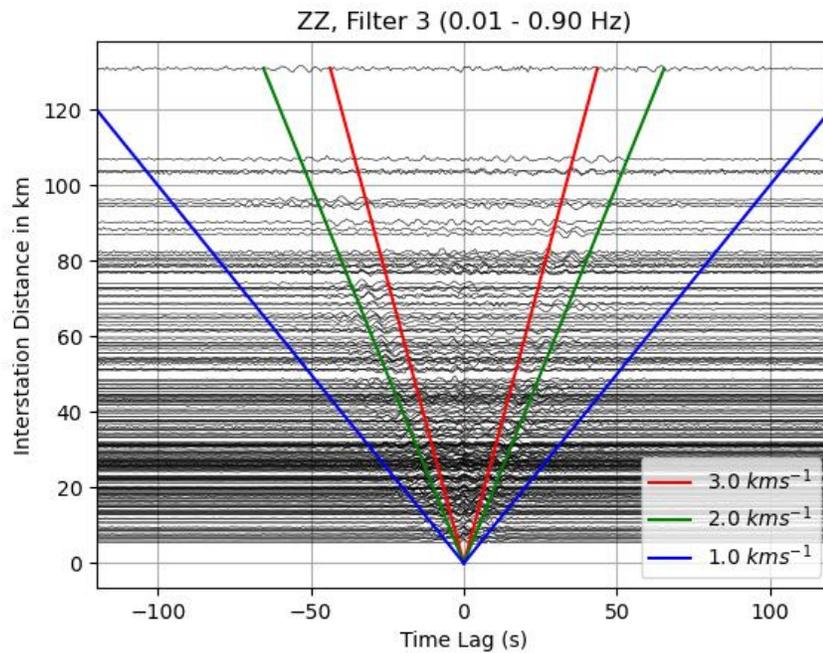


Figura 5.1.2 Grafica de correlación cruzada utilizando datos no segmentados para filtro de (0.01-0.90Hz)

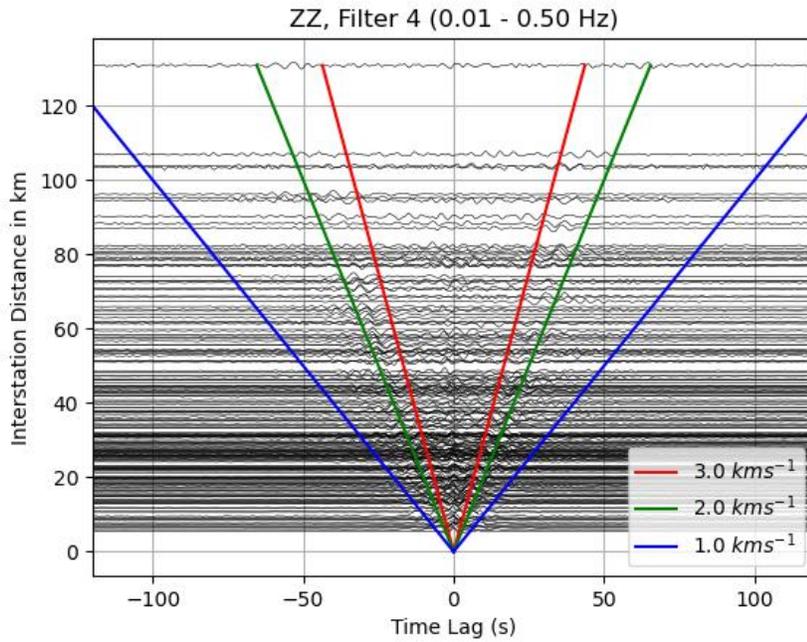


Figura 5.1.3 Grafica de correlación cruzada utilizando datos no segmentados para filtro de (0.01-0.50Hz)

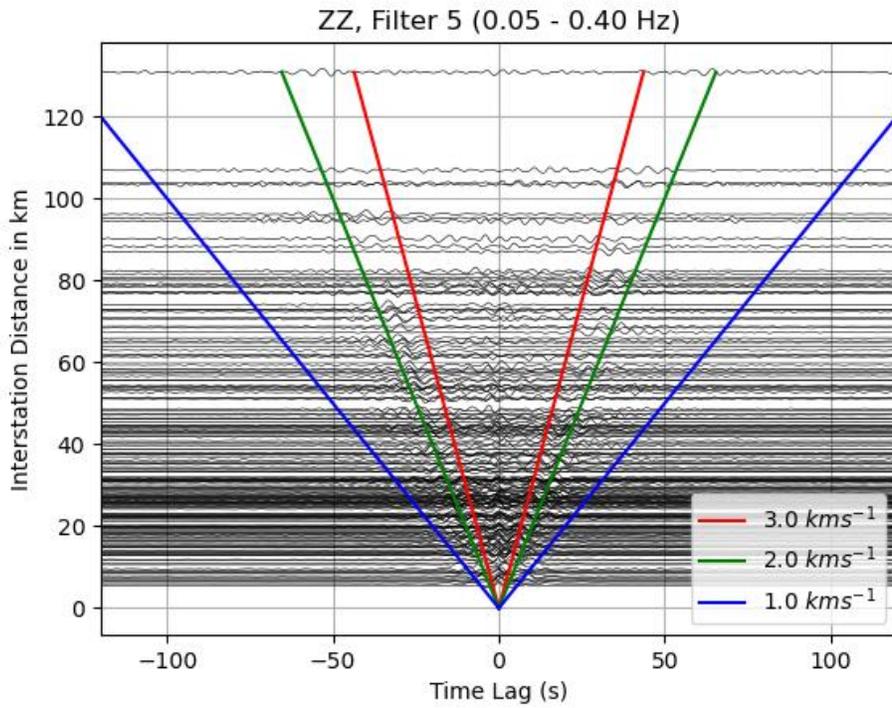


Figura 5.1.4 Grafica de correlación cruzada utilizando datos no segmentados para filtro de (0.05-0.40Hz)

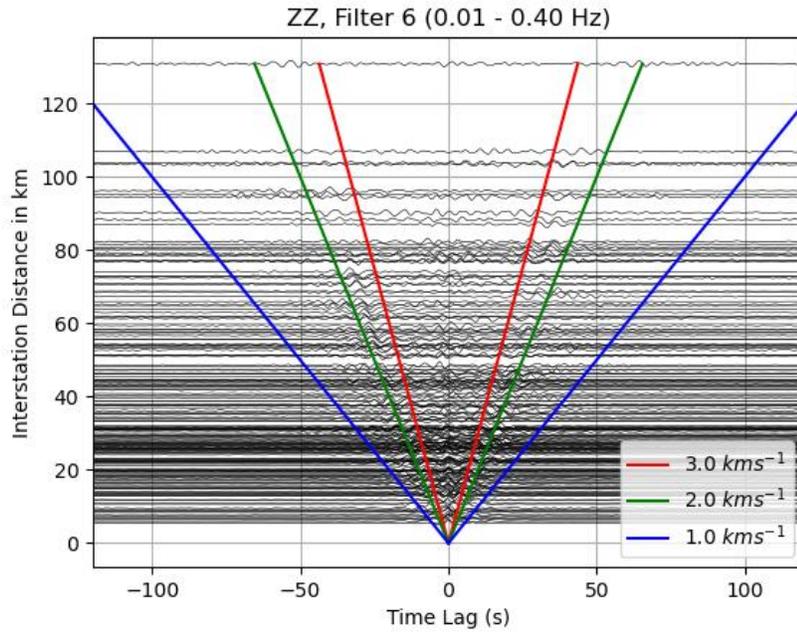


Figura 5.1.5 Grafica de correlación cruzada utilizando datos no segmentados para filtro de (0.01-0.40Hz)

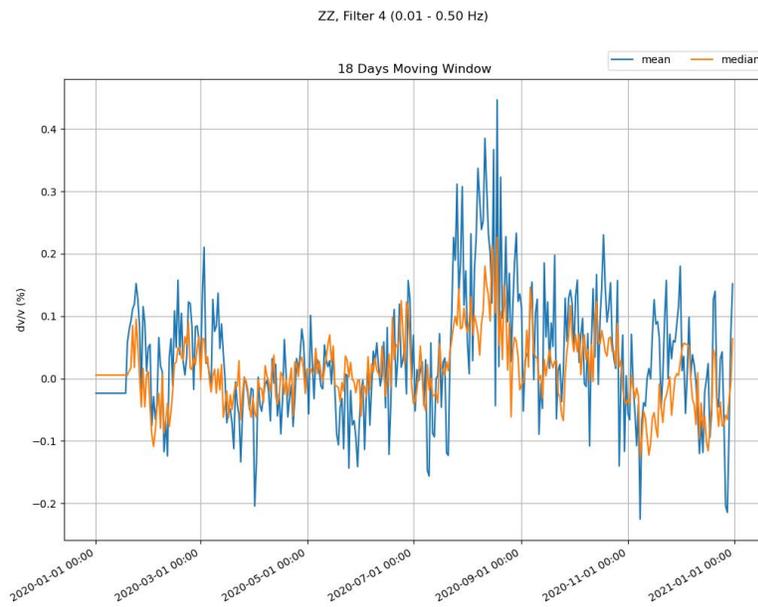


Figura 5.1.6 Grafica de Cambios de Velocidad dv/v utilizando datos no segmentados para filtro (0.01-0.50Hz)

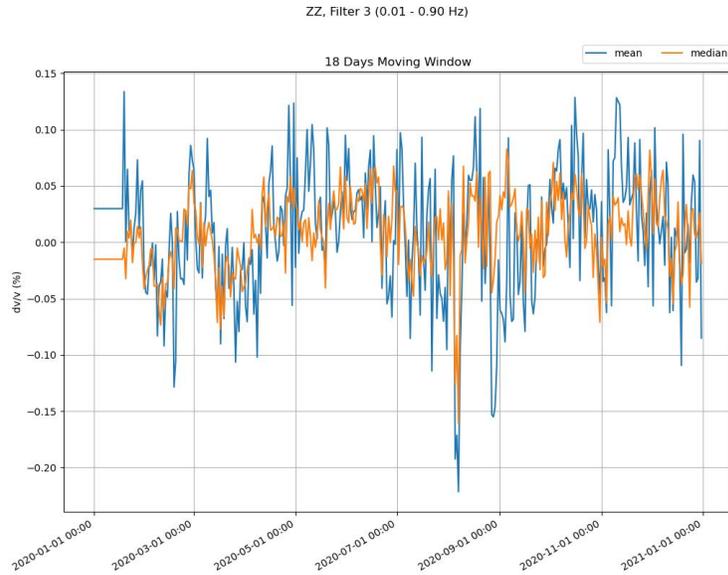


Figura 5.1.7 Grafica de Cambios de Velocidad dv/v utilizando datos no segmentados para filtro (0.01-0.90Hz)

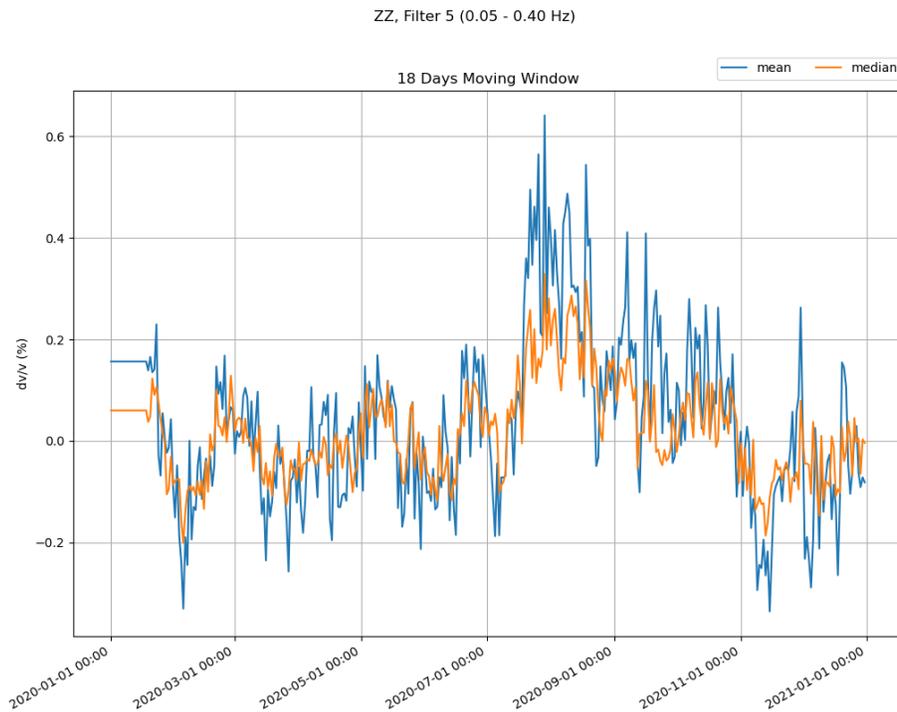


Figura 5.1.8 Grafica de Cambios de Velocidad dv/v utilizando datos no segmentados para filtro (0.05-0.40Hz)



Figura 5.1.9 Grafica de Cambios de Velocidad dv/v utilizando datos no segmentados para filtro (0.01-0.40Hz)

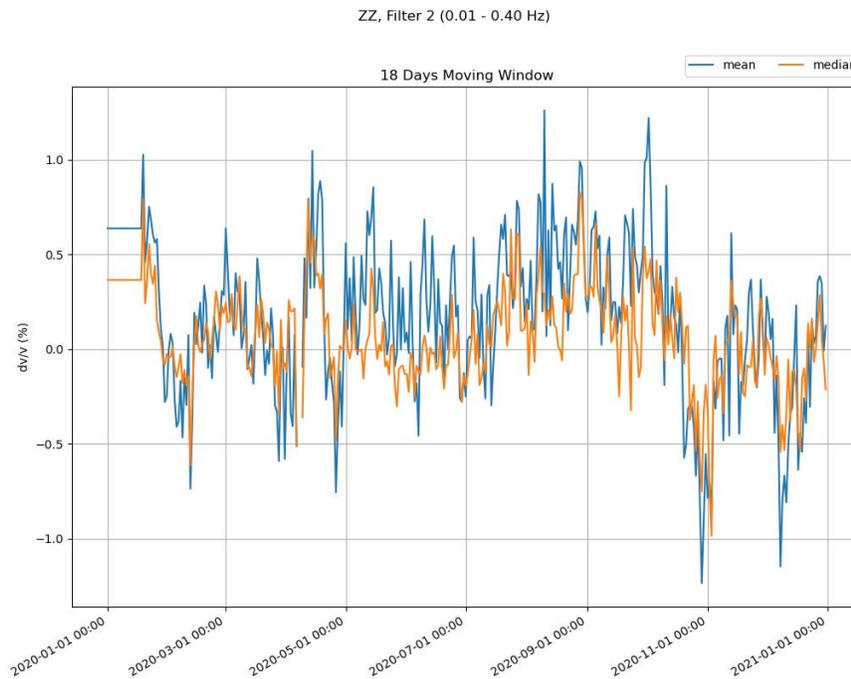


Figura 5.1.10 Grafica de Cambios de Velocidad dv/v utilizando datos segmentados para filtro (0.01-0.40Hz)

Capítulo 6

CONCLUSIONES

En general demostramos que las variaciones de velocidad correlacionan con la cantidad de lluvia en el Valle de México, este resultado es esperado debido a que otros autores han encontrado resultados similares. En la figura 6.1.1 se presenta los cambios de velocidad a lo largo de 18 días y además la cantidad de precipitación en el mismo intervalo utilizando un modelo numérico para el valle de México. Los valores de precipitación fueron procesados mediante un filtro de promedios móviles por 18 días a fin de representar una tendencia más suave y tratar de coincidir con los 18 días en los cuales fueron apiladas las trazas de ruido sísmico. Aunque los resultados obtenidos son muy favorables, es común observar un retraso entre la curva que refleja los cambios de velocidad y la curva de precipitación. Este fenómeno se debe a que, en un sistema hidrológico, la lluvia constituye la principal contribución a los mantos acuíferos, pero suele experimentar demoras, que varían entre 15 días y 2 meses según la transmisividad hidráulica. Este retraso debería reflejarse en desfases entre ambas curvas, sin embargo, en el valle de México, esta sincronización aparente sugiere la existencia de mecanismos singulares. Por tanto, se plantea la necesidad de futuros estudios que aborden este fenómeno. Estas investigaciones, además, enfrentarán retos computacionales similares a los presentados en este trabajo.

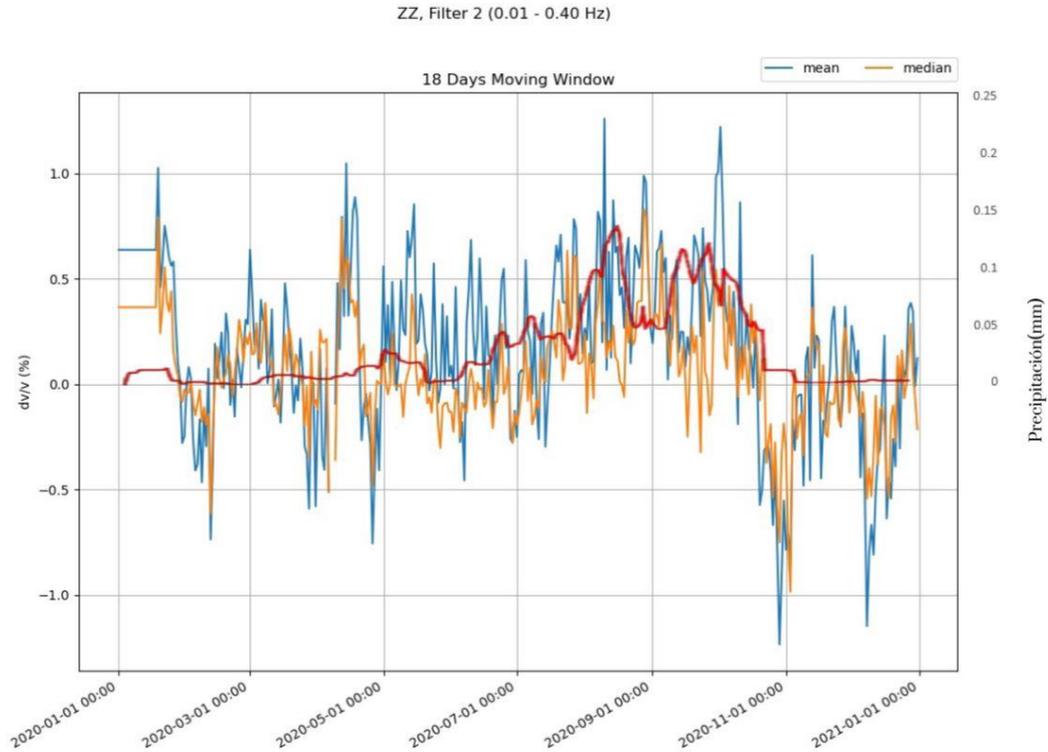


Figura 6.1.1: Comparación de los datos de velocidad obtenido en el procesamiento de ruido sísmico representados en variación porcentual sobre una media de referencia a lo largo de todo el año. La línea azul representa la media y la naranja representa la mediana de las variaciones de velocidad mientras que la línea roja representa los valores de precipitación medidos en milímetros después de aplicar un filtro de promedios móviles de 18 días. Se observa una gran similitud en la tendencia de todas estas graficas.

Una de las posibles teorías que hacen que el Valle de México tenga estos cambios al mismo tiempo son:

1. Cambios poro elásticos, los cuales representan cambios en la velocidad debido a las ecuaciones elásticas cuando se representa un medio continuo con ecuaciones hidrológicas. En ese caso los parámetros elásticos se reflejan inmediatamente simplemente por la saturación de agua, pero no incluye ningún fenómeno de recarga en el acuífero es decir simplemente están relacionados con la presión de poro que se afecta directamente al aumentar el contenido de agua en un medio.
2. Cambios termo elásticos, estos cambios son comunes simplemente porque la temperatura a lo largo del año cambia en la atmosfera y en el suelo. En este caso la variación de velocidad no se afecta con el contenido de agua si no simplemente con los cambios de temperatura. En nuestro caso es posible que en el momento que ocurren las lluvias

intensas la temperatura del suelo cambie inmediatamente, sobre todo en casos extremos como caída de granizo o cuando el suelo se ha calentado de forma dramática y al simple contacto de la lluvia cambia fuertemente su temperatura.

3. Variaciones de velocidad por sismos lejanos. Esta posibilidad es la menos aceptada debido a que hemos hecho un trabajo intenso en limpiar los registros de ruido y además ocurriría exclusivamente cuando existan sismos importantes en el mundo, pero no estarían ligados a cambios estacionales.

El presente trabajo tiene una importancia fundamental en el procesado de los datos de ruido sísmico debido a que principalmente se acostumbra utilizar todo el registro de un solo día y en algunos casos es común que utilicen algoritmos muy sencillos para solamente limpiar parcialmente los datos cuando se enfrenten a pequeños eventos que pueden alterar los resultados. No conocemos ningún artículo en la literatura que utilice algoritmos de aprendizaje automático para procesar el ruido sísmico y obtener así de esta manera datos de primera calidad donde estamos convencidos que se ha eliminado de una forma significativa cualquier pequeño sismo que interfiera en el análisis del ruido sísmico.

Se ha encontrado que a pesar de que los algoritmos de aprendizaje automático son muy eficientes el procesado del ruido tiene otro tipo de dificultades computacionales, por ejemplo, la segmentación. Durante el pre procesado de los datos es importante segmentar los registros sísmicos para eliminar cualquier indicio de pequeños sismos que afecten el análisis, sin embargo, el tiempo de cómputo es un parámetro muy importante en poder realizar este tipo de trabajos. La combinación de una gran cantidad de datos que supera varias decenas de millones junto con las necesidades de escritura en los discos duros genero un problema significativo. Nos encontramos que un pre proceso de ruido sísmico para un año podría tardar más de seis meses. Al investigar donde se encuentra el problema nos dimos cuenta que es importante utilizar algoritmos muy eficientes en todas las etapas incluyendo las bases de datos, en general los algoritmos de aprendizaje automático utilizan sistemas sencillos de administración de datos por ejemplo en el lenguaje Python las bibliotecas de manejo de datos más comunes son los diccionarios internos y la librería de pandas. En nuestro caso estas herramientas no se pudieron utilizar debido a que tienen una limitación en la memoria y cuando exceden varios millones de datos el proceso es sumamente lento. La solución es utilizar bases de datos relacionales que tienen un motor de acceso muy eficiente, además son herramientas que han sido probadas durante varias décadas y funcionan de manera confiable en casos de tener millones de datos, como por ejemplo en consultas bancarias.

Otra herramienta importante es poder paralelizar las partes del código que no necesitan estar serializadas por ejemplo en la parte de segmentación no se requiere haber segmentado un día en especial para poder segmentar otro día diferente de esta manera pudimos paralelizar con el número total de procesadores todas las estaciones de forma independiente.

El algoritmo de identificación de sismos ha sido plenamente probado durante tres años, pero no conocemos ninguna aplicación de identificación mediante aprendizaje automático donde el fin último sea un análisis hidrológico. Esta es una nueva línea de la sismología que apenas

se encuentra en sus etapas iniciales y debe ser explotada mediante nuevos algoritmos de inteligencia artificial por ejemplo combinar resultados de lectura de pozo y los cambios que se observan de dv/v . Es importante destacar que todavía existen problemas técnicos y científicos diversos por ejemplo la instrumentación de sensores que durante todas estas décadas han sido exclusivamente diseñados para detectar eventos sísmicos. En nuestro caso es importante distribuir las redes sísmicas de forma tal que optimicen las lecturas de los cambios hidrológicos y se ubiquen las estaciones en puntos estratégicos alrededor de una cuenca hidrológica a diferencia de las redes sísmicas que tratan de ubicar las estaciones alrededor de fallas activas. Por esta razón los datos aun no tienen la calidad suficiente para poder estudiar con mayor detalle los cambios hidrológicos.

Debemos de abrir la oportunidad para nuevas posibilidades científicas en estas líneas, por ejemplo, el estudio de posibles desastres naturales que puedan ocurrir después de la incidencia de un huracán está fuertemente ligados a los cambios de velocidad sísmica. Existen sistemas de alerta temprana que miden en tiempo real los cambios de velocidad durante fuertes tormentas y estiman la posibilidad de deslaves cuando el coeficiente de fricción no puede sustentar los esfuerzos cortantes, los cuales se miden a través de los cambios de velocidad.

La conclusión general de este trabajo se resume en lo siguiente:

1. Los algoritmos de aprendizaje automático son muy eficientes para el pre procesamiento de datos sísmicos a fin de tener señales de muy alta calidad.
2. El análisis muestra una fuerte correlación entre la cantidad de lluvia y las variaciones de velocidad sísmica, lo cual abre una línea de investigación novedosa y con alto potencial.
3. A pesar de que el procesamiento de ruido sísmico tiene diferentes retos computacionales es posible realizar este tipo de estudios en pocos días, a diferencia de algoritmos tradicionales que pueden causar serios retardos en el procesamiento de los datos.
4. Nuestro trabajo abre la oportunidad de una serie de investigación más profunda en diferentes elementos que todavía quedan sin resolver, por ejemplo, la interacción con huracanes, o sequías importantes.
5. Es importante colaborar con grupos multidisciplinarios como geo hidrólogos, geofísicos, científicos de la computación e incluso dentro de las ciencias sociales para tener un panorama más completo de toda la gama de información que podemos aprovechar mediante este tipo de procesamiento.

BIBLIOGRAFÍA

- [1]. SETH STEIN, MICHAEL E. WYSESSION. AN INTRODUCTION TO SEISMOLOGY, EARTHQUAKES, AND EARTH STRUCTURE. BLACKWELL PUBLISHING, 2003
- [2]. SERVICIO GEOLÓGICO MEXICANO, TECTÓNICA DE PLACAS, 2017, RECUPERADO DE: [HTTPS://WWW.SGM.GOB.MX/WEB/MUSEOVIRTUAL/RIESGOS-GEOLOGICOS/TECTONICA-DE-PLACAS.HTML](https://www.sgm.gob.mx/web/museoVirtual/riesgos-geologicos/TECTONICA-DE-PLACAS.HTML)
- [3]. D. J. MORAN Z. GEOFÍSICA INTERNACIONAL, BREVE REVISIÓN SOBRE LA EVOLUCIÓN TECTÓNICA EN MÉXICO. VOL. 25, 1986.
- [4]. RAÚL GONZÁLEZ. LOS SISMOS, DESDE TENOCHTITLAN HASTA NUESTROS DÍAS. REVISTA CONSIDERACIONES, 2017
- [5]. HECTOR GONZÁLEZ HUIZAR. LA OLIMPIADA XXIV DE CIENCIAS DE LA TIERRA: LOS GRANDES TERREMOTOS EN MÉXICO. GEOS, VOL. 39, NO. 1 (2019)
- [6]. INSTITUTO NACIONAL DE PREVENCIÓN SÍSMICA –INPRES. INTRODUCCIÓN A LA SISMOMETRIA
- [7]. A FIRST ARRIVAL PICKING METHOD OF MICROSEISMIC DATA BASED ON SINGLE TIME WINDOW WITH WINDOW LENGTH INDEPENDENT. JOURNAL OF SEISMOLOGY. 2018
- [8]. VÍCTOR SCHMIDT-DÍAZ. APLICACIÓN DE LA TÉCNICA H/V EN DISTINTOS EMPLAZAMIENTOS A PARTIR DE LA MEDICIÓN DE VIBRACIONES AMBIENTALES Y DE SISMOS. UNIVERSIDAD DE COSTA RICA. 2015
- [9]. LUIS G. PUJADES, LUIS A. PINZÓN, RODRIGO E. ALVA INTENSIDAD DE ARIAS Y DURACIÓN SIGNIFICATIVA EN ANÁLISIS DINÁMICO DE ESTRUCTURAS. UNIVERSIDAD DE COSTA RICA. 2021
- [10]. ANDREAS C. MÜLLER, SARAH GUIDO. INTRODUCTION TO MACHINE LEARNING WITH PYTHON. FIRST EDITION, 2017.
- [11]. PYTHON SOFTWARE FOUNDATION. PYTHON 3.7 DOCUMENTATION
- [12]. COPYRIGHT 2008-2022, NUMPY DEVELOPERS
- [13]. MOUSAVI, S.M., ELLSWORTH, W.L., ZHU, W., CHUANG, L.Y., BEROZA, G.C., “EARTHQUAKE TRANSFORMER: AN ATTENTIVE DEEP-LEARNING MODEL FOR SIMULTANEOUS EARTHQUAKE DETECTION AND PHASE PICKING “. NATURE COMMUNICATIONS, (2020)
- [14]. MORITZ BEYREUTHER, ROBERT BARSCH, LION KRISCHER, TOBIAS MEGIES, YANNIK BEHR, AND JOACHIM WASSERMANN, OBSPY: A PYTHON TOOLBOX FOR SEISMOLOGY, JUNE 2010
- [15]. ANALYTICS VIDHYA. BINARY CROSS ENTROPY AKA LOG LOSS-THE COST FUNCTION USED IN LOGISTIC REGRESSION, 2020
- [16]. ANALYTICS VIDHYA. THE ULTIMATE GUIDE TO 12 DIMENSIONALITY REDUCTION TECHNIQUES (WITH PYTHON CODES)

- [17]. GREENACRE, M., GROENEN, P.J.F., HASTIE, T. ET AL. PRINCIPAL COMPONENT ANALYSIS. NAT REV METHODS PRIMERS 2, 100 (2022).
- [18]. QUINTANAR L., A. CÁRDENAS-RAMÍREZ, D. I. BELLO-SEGURA, V. H. ESPÍNDOLA, J. A. PÉREZ-SANTANA, C. CÁRDENAS-MONROY, A. L. CARMONA-GALLEGOS, AND I. RODRÍGUEZ-RASILLA (2018); A SEISMIC NETWORK FOR THE VALLEY OF MEXICO: PRESENT STATUS AND PERSPECTIVES; SEISMOLOGICAL RESEARCH LETTERS, VOLUME 89, NUMBER 2A, PP. 356-362, MARCH/APRIL 2018; DOI: 10.1785/0220170198.
- [19]. HAIDEE MARIA ASSI HAGMAIER (2021). "ESTUDIO DE LAS VARIACIONES DE LA VELOCIDAD Y LOS PERIODOS DE VIBRACIÓN EN EL VALLE DE MÉXICO, MEDIANTE EL PROCESADO DE RUIDO SÍSMICO". TESIS DE MAESTRÍA, CICESE.
- [20]. SERVICIO SISMOLÓGICO NACIONAL (S.F.), 2023. RECUPERADO DE [HTTP://WWW.SSN.UNAM.MX](http://www.ssn.unam.mx)

APÉNDICE 1

Scripts de EQTransformer

Conversión de MiniSeed a HDF5:

Utilizando la siguiente secuencia de comandos, puedes convertir archivos MiniSeed para cada estación en un solo archivo HDF5 y un archivo CSV que contiene la lista de trazas en el archivo HDF5. Esto generará archivos "station_name.hdf5" y "station_name.csv" para cada estación en el directorio especificado.

```
import glob
from EQTransformer.core.mseed_predictor import mseed_predictor

mseed_dir = "directorio_de_datos_mseed"
mseed_hdfs_dir = "directorio_de_salida_hdfs"

mseed_predictor(mseed_dir=mseed_dir,
                 mseed_hdfs_dir=mseed_hdfs_dir)
```

Detección y Picking:

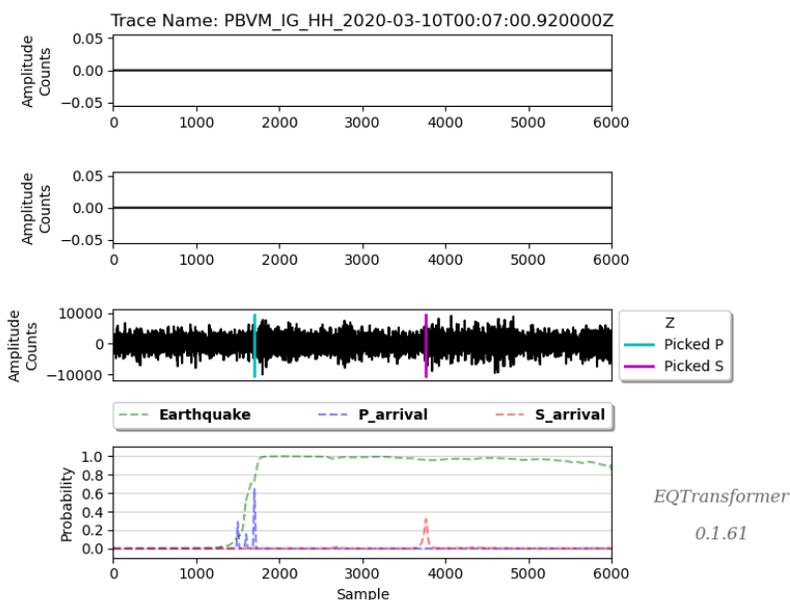
Utilizando EQTransformer, realiza la detección y el picking con los parámetros mostrados en la siguiente figura:

```
from EQTransformer.core.predictor import predictor
meses=["01","02","03","04","05","06","07",
       "09","10","11","12","13","14","15","16","17","18","19",
       "20","21","22","23","24","25","26","27","28","29",
       "30","31","32","33","34","35","36","37"]
for me in meses[27:28]:
    input_dir="/Users/roberto/EQTransformer/examples/downloads_vm"+me+"_processed_hdfs"
    output_dir="detections"+me
    predictor(input_dir=input_dir,
              input_model='/Users/roberto/EQTransformer/ModelsAndSampleData/EqT_original_model.h5',
              output_dir=output_dir,
              estimate_uncertainty=False,
              output_probabilities=False,
              number_of_sampling=5,
              loss_weights=[0.02, 0.40, 0.58],
              detection_threshold=0.3,
              P_threshold=0.3,
              S_threshold=0.3,
              number_of_plots=10,
              plot_mode='time',
              batch_size=500,
              number_of_cpus=4,
              keepPS=False,
              splimit=60)
```

Interpretación de Resultados

Los resultados de la detección de eventos sísmicos y la selección de fases P y S revelan una comprensión profunda de la actividad sísmica en el conjunto de datos analizado. La configuración cuidadosa de los parámetros de detección, como el umbral de detección, ha minimizado la probabilidad de falsos positivos, garantizando así la robustez del análisis. La identificación precisa de las fases P y S en los eventos sísmicos detectados proporciona información valiosa sobre la naturaleza y la localización de dichos eventos.

Los gráficos temporales generados ofrecen una visualización clara de la distribución temporal de los eventos sísmicos, destacando patrones temporales que podrían ser indicativos de variaciones en la actividad sísmica. (Figura n). La evaluación de estadísticas de rendimiento, como la sensibilidad y la especificidad, proporciona una comprensión cuantitativa de la eficacia del modelo.



Código de segmentación

Introducción al Script

La función del script consiste en segmentar los registros sísmicos eliminando las ventanas de tiempo correspondiente a las detecciones de las fases P y S. Esto asegura que en el registro de tiempo final solo permanezcan los segmentos de ruido ambiental. La función `identificar_ventana_tiempo` opera sobre un DataFrame de Pandas que contiene información sobre eventos sísmicos, filtrando aquellos con una probabilidad de detección superior a un umbral establecido. La información resultante, que incluye detalles sobre el evento, las ventanas de inicio y fin, así como los nombres de los archivos asociados, se almacena en un nuevo DataFrame. La función `restar_ventana` elimina las ventanas de tiempo

correspondientes a las detecciones de fases P y S asegurando que solo permanezcan los segmentos de ruido ambiental en el registro.

```
# Funciones para restar ventana de tiempo y identificar ventanas de tiempo
from obspy.core import read
import os
import pandas as pd
from datetime import datetime, timedelta

def restar_ventana(registro, ventana=20):
    # Implementación de La función restar_ventana

def identificar_ventana_tiempo(dataframe, probabilidad_umbral=0.8):
    # Implementación de La función identificar_ventana_tiempo
```

Scripts de MSNOISE construye dv/v

MSNoise realiza un estudio de ruido sísmico ambiental en dos etapas fundamentales. En primer lugar, identifica y registra redes y estaciones presentes en los archivos de datos, gestionando la inserción o actualización de información en la base de datos. Posteriormente, realiza cálculos esenciales como funciones de correlación cruzada, funciones de transferencia y análisis de variación temporal, generando resultados visuales que permiten comprender las propiedades del subsuelo. Este enfoque metodológico integral proporciona una base sólida para interpretar y caracterizar el entorno sísmico.

```
#Paso 1: Escanea Los archivos de datos, identifica Las Redes/Estaciones e
#inserta en La tabla de estaciones en La base de datos.
$ msnoise populate

# Paso 2: examina La marca de tiempo de modificación de Los archivos en Los archivos
#comprimidos para identificar y gestionar La inserción o actualización
#de archivos nuevos o modificados en una tabla de disponibilidad de datos
$ msnoise -t 2 scan_archive

#Paso 3: busca en La base de datos archivos marcados como "N"uevos o "M"odificados, y
#para cada fecha dentro del rango configurado, verifica La disponibilidad en otras estaciones y
#define Los nuevos trabajos a procesar, Los cuales son Luego insertados en La tabla de trabajos de La base de datos.
$ msnoise new_jobs

# Paso 4: Calcula Las funciones de correlación cruzada.
$ msnoise compute_cc

# Paso 5: Cálculo de Las funciones de transferencia
msnoise.ftan()

# Paso 6: calcular Las funciones de correlación cruzada utilizando una función
#de referencia definida por un intervalo de fechas
$ msnoise stack -r
$ msnoise reset STACK
$ msnoise stack -m

# Paso 7: Analisis de La ventana móvil (MWCS) para estudiar el desfase relativo entre
#conjuntos de datos en ventana móvil y una referencia mediante el análisis cruzado espectral en ventana móvil.
$ msnoise compute_mwcs

#Paso 8:Cálculo de dt/t utilizando Los resultados de Las operaciones de análisis cruzado espectral
#en ventana móvil (MWCS).
msnoise compute_dtt

# Paso 9: Generar gráficos dv/v y resultados
msnoise plot dvv
```

APÉNDICE 2

El preprocesamiento de las formas de onda sísmicas comienza con la selección de todos los registros potenciales de un día en cada estación. Las formas de onda se fusionan y dividen para obtener secciones y promedios ininterrumpidos, se eliminan las tendencias y se realiza un filtrado cónico coseno. Las secciones cortas se completan con ceros y las largas se truncan. A continuación, se realiza la iteración para pares de estaciones y componentes diferentes. Las componentes radial y transversal se calculan a partir de la rotación de las componentes este y norte mediante una fórmula especial. La correlación cruzada se realiza en el dominio de la frecuencia.

Durante el procesamiento se aplican diversos filtros y configuraciones, que se definen en tablas especiales. El tiempo de análisis es de 86,400 s y el solapamiento es del 50 %. Los resultados, como las funciones de correlación cruzada (CCF), se generan y guardan a intervalos de 30 minutos. Los ajustes generales incluyen parámetros como el tiempo de análisis, el formato de archivo, el tipo de correlación cruzada y la frecuencia de muestreo.

Los resultados de la zona lacustre muestran un espectro claro, pero hay desviaciones en los mapas de periodos dominantes: la estación CTVM en Cuauhtémoc está afectada por factores antropogénicos, como túneles subterráneos y edificios cercanos. En la zona de transición, la estación APVM en Azcapotzalco refleja una litología compleja con amplitudes bajas; en la zona de Lomas, se esperan espectros planos, pero se observan variaciones, como en la estación AMVM en Amecameca; MHVM en Miguel Hidalgo. Algunas estaciones, como la estación, muestran efectos estacionales, donde las variaciones están relacionadas con el flujo subsuperficial.

El análisis revela formas de onda sísmicas complejas influidas por factores geológicos, antropogénicos y estacionales en diferentes regiones, lo que destaca la importancia de un análisis detallado en la evaluación del riesgo sísmico.

El siguiente extracto corresponde a la tesis titulada “Estudio de las variaciones de la velocidad y los periodos de vibración en el Valle de México, mediante el procesado de ruido sísmico” presentada por Haidee Maria Assi Hagmaier para obtener el grado de Maestro en Ciencias en Ensenada, Baja California, México, en 2022.

Se incorpora al apéndice con el objetivo de respaldar y contextualizar la investigación, utilizando los hallazgos y conclusiones de las tesis de Haidee Maria Assi Hagmaier como referencia. La mencionada tesis describe detalladamente las características geotécnicas de cada estación en el Valle de México, **ofreciendo una valiosa contribución al marco conceptual de este trabajo.**

Preprocesamiento de forma de onda

Para cada estación, se utilizan todos los registros que potencialmente contienen todos los datos del día. A continuación, las trazas se fusionan y se dividen para obtener las secciones sin interrupciones. Luego se les eliminará la media, se elimina la tendencia, se le aplica un filtrado taper cosenoidal y se fusionan nuevamente en una ventana de un día de duración. Si es más corto que un día, la traza se complementa con ceros. Si es más largo, se corta para que coincida con el inicio/final del día. Posteriormente, cada traza de un día de duración se filtrará dependiendo cada caso y luego se diezmará o se reducirá.

Procesado

Una vez que todas las formas de onda están cargadas en la memoria, el cálculo se realiza por iteración en los pares de estaciones, en los diferentes componentes a procesar y luego en los diferentes filtros para cada ventana definida con cortes de 30 minutos por defecto, configurables en la Tabla 2 que aparece más adelante.

Los componentes radiales (R) y transversal (T) se calculan a partir de la rotación de los componentes este (E) y norte (N). El ángulo de rotación se define como el azimut (Az) entre los dos sensores.

$$R = N \cos(Az) + E \sin(Az) \quad (13)$$

$$T = N \cos(Az) - E \sin(Az) \quad (14)$$

La correlación cruzada se realiza en el dominio de frecuencia. Sean $x(t)$ y $y(t)$ dos series de tiempo y $X(f)$

y $Y(f)$ sus transformadas de Fourier. La operación de correlación se define entonces como:

$$C(f) = X^*(f) \times Y(f) \quad (15)$$

siendo $X^*(f)$ el complejo conjugado de $X(f)$. Si $x(t) = y(t)$, entonces la operación se llama autocorrelación. La FCC $c(t)$ es la transformada de Fourier inversa de $C(f)$. Si se desea guardar el $c(t)$, entonces para cada segmento de 30 minutos se puede almacenar en un directorio de salida en la computadora. Nuevamente, si se desea, se almacena la pila de distintos días que no contengan ceros, infinitos o NaNs. Esta pila es de 48 segmentos cada una de 30 minutos y finalmente se almacenan los valores de $c(t)$.

Tabla 2. Configuración general de MSNoise.

Nombre	Valor	Nombre	Valor
analysis_duration	86400	maxlag	120
archive_format		mov_stack	18
autocorr	N	network	*
cc_sampling_rate	100	output_folder	CROSS_CORRELATIONS
cc_type	CC	overlap	0.5
cc_type_single_station_AC	CC	plugins	
cc_type_single_station_SC	CC	preprocess_highpass	0.01
channels	*	preprocess_lowpass	8
components_to_compute	ZZ	preprocess_max_gap	10
components_to_compute_single_station		preprocess_taper_length	20
corr_duration	1800	pws_power	2
crondays	600	pws_timegate	10
data_folder	/INGVDATA/RSVMS DS/	ref_begin	01/01/20
data_structure	SDS	ref_end	01/01/21
dtc_lag	static	remove_response	N
dtc_maxdt	0.1	resampling_method	Lanczos
dtc_maxerr	0.1	response_format	dataless
dtc_mincoh	0.65	response_path	inventory
dtc_minlag	5	response_prefilt	(0.005, 0.006, 30.0, 35.0)
dtc_sides	right	sac_format	doublets
dtc_v	1	stack_method	linear
dtc_width	40	startdate	01/01/20
enddate	31/12/20	stretching_max	0.01
export_format	MSEED	stretching_nsteps	1000
hpc	N	whitening	A
keep_all	N	whitening_type	B

Resultados de H/V en la Zona del Lago

En general los espectros de la Zona III denominada Zona de Lago, tienen un pico muy claro, pero no ocurre todo el tiempo. En los mapas de periodos dominantes se observan duraciones mayores a un segundo. En la mayoría de los casos observados el periodo fundamental coincide con el periodo dominante, pero en algunos sitios los periodos que hemos calculado están muy lejos de los reportados en el mapa de periodos dominantes del Valle de México.

Estación CTVM (Cuauhtémoc)

Esta estación se encuentra ubicada dentro de las instalaciones de un centro de cultura, la edificación es una capilla colonial protegida denominada “Capilla Británica”. Se observaron letreros de alta tensión que al parecer se encuentran enterrados a muy pocos metros del sensor. Es importante destacar que el túnel del metro de la línea 2 está a unos 20 metros del sitio, por lo tanto, los resultados espectrales están afectados por todos estos elementos.

Por lo antes mencionado, interpretamos este comportamiento como una estación que está controlada exclusivamente por cambios antropogénicos, es decir, únicamente lo afectan actividades humanas. En primera, esta estación se encuentra exactamente arriba de un túnel del metro de la línea 2, la cual se encuentra en un ambiente artificial controlado por bombas que continuamente extraen el exceso de agua, además es importante destacar la presencia de un puente de grandes dimensiones que inevitablemente resuena a lo largo del año en un periodo fijo. Por otro lado, los cambios a frecuencias altas reflejan las diferentes actividades humanas que ocurren en su entorno, por ejemplo, se observó un letrero de alta tensión que pertenece a la casa de cultura y posiblemente se energiza en ciertas épocas del año. Además, es un sitio de tráfico pesado y continuo no es raro que ocurran actividades de mantenimiento, construcciones asociadas y reparaciones. Observamos que este tipo de estaciones sísmicas son comunes cuando se hacen trabajos de esta naturaleza en zonas urbanas como la Ciudad de México. Es importante destacar que en los mapas de periodos dominantes este sitio debe tener su periodo máximo a los 1.5 seg. (Figura 3), sin embargo, el periodo fundamental parece estar a los 1.17 seg. Esto se debe probablemente a la influencia de las estructuras grandes como el puente de concreto que se encuentra a muy pocos metros del sitio. En cambio, existe un pequeño pico en 1.5 seg. el cual interpretamos precisamente como el periodo dominante del sitio y claramente no es tan evidente como el periodo fundamental de 1.17 seg., esto demuestra la dificultad que existe en elegir los periodos dominantes de sitio.

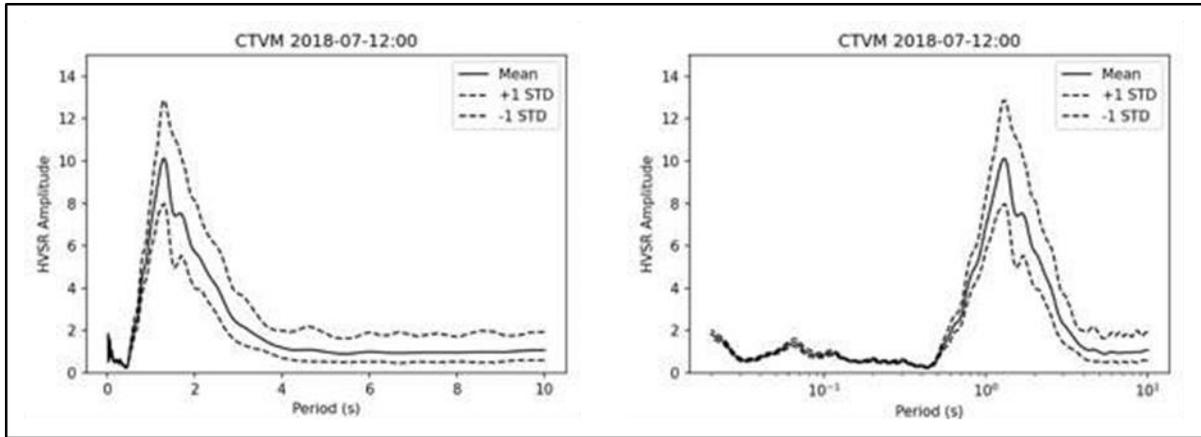


Figura 16 CTVM_2018-07-12/00.



Figura 17 Estación CTVM. La estación se encuentra cerca de una capilla británica donde se aprecian letreros de alta tensión, en la calzada podemos observar un puente y a pocos metros la estación del metrobús Revolución.

Estación ICVM (Iztacalco)

Se encuentra localizada dentro de las instalaciones de la preparatoria número 3 de la Universidad Nacional Autónoma de México. La estación está muy cerca de una estructura de un piso. Los espectros de HVSR son de alta calidad y muestran un periodo dominante muy bien definido.

Aparentemente esta es una estación que se puede interpretar con un modelo sencillo, sin embargo, al hacer un análisis más detallado existen algunas incongruencias que debemos puntualizar, por ejemplo, el periodo dominante debería estar en 2.5 segundos, sin embargo,

el periodo fundamental de este sitio está a los 3 segundos. Aunque este error podría ser mínimo, debido a la resolución de nuestro instrumento, creemos que es poco probable que sea un error de medición. En cambio, esta diferencia parece deberse a cambios locales, por ejemplo, es probable que haya zonas de canalización en Iztacalco, sobre todo en la parte central donde hace poco continuaban activos los canales de desagüe.

Podemos observar que hay un ligero pico a los 2 segundos el cual es muy cercano a los 2.5 segundos que esperábamos, por lo tanto, es posible que existan dos periodos característicos. Uno con una profundidad de 43 metros de un sedimento blando de muy baja velocidad de aproximadamente 70 m/s, el cual probablemente esté limitado por alguna estructura geológica discordante (sean cenizas volcánicas o piroclásticos) seguidos por otro estrato de baja velocidad con un espesor de 50 metros y velocidad de 110 m/s. Esto implica que es posible que los mapas de periodos dominantes requieran de una resolución con mucho más detalle, por ejemplo, un mapa de periodos dominantes para cada delegación.

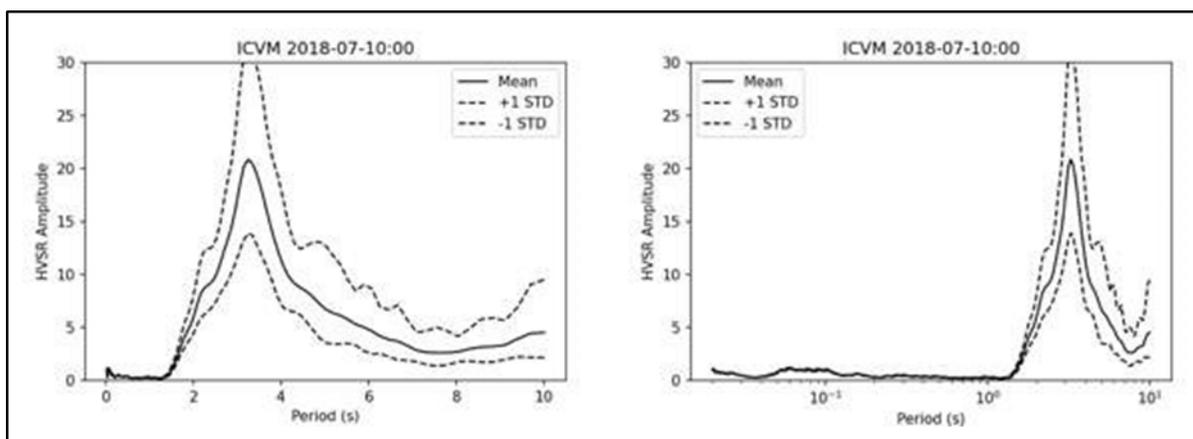


Figura 18 ICVM_2018-07-10/00.



Figura 19. Estación ICVM. La estación se encuentra adentro de una escuela, muy cerca de los edificios.

Estación VRVM (Venustiano Carranza)

Esta estación es similar a la estación Iztacalco, se encuentra cerca de las instalaciones del Instituto Mexicano del Seguro Social desplantadas en un solo piso, sus espectros HVSR son bien definidos, ya que no parece haber variaciones considerables, la curva sólo cambia en amplitud, pero nunca en periodo.

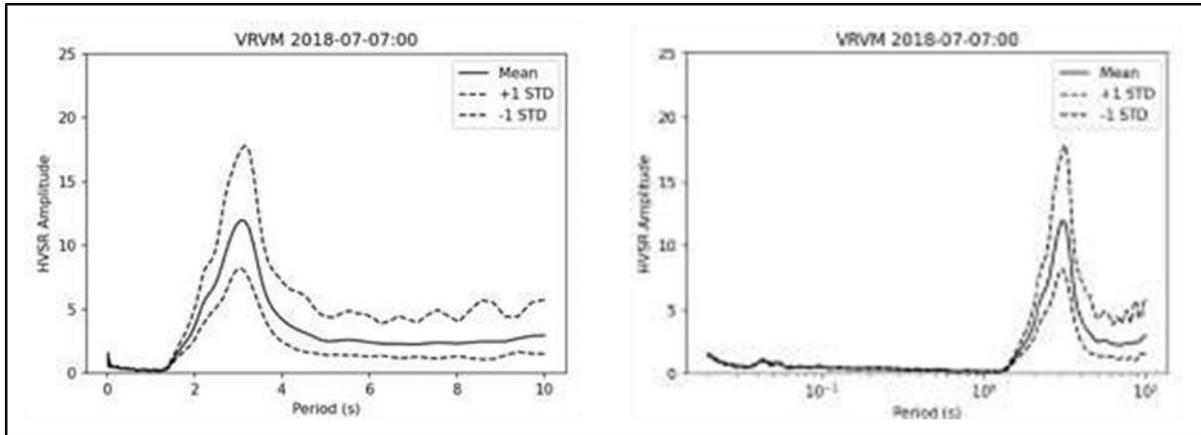


Figura 20. VRVM_2018-07-07/00.

Esta es la única estación donde los resultados realmente representan la información del actualizado mapa de periodos dominantes (Lermo et al., 2020), es importante destacar que esta estación está cerca del Aeropuerto Internacional de la Ciudad de México, por lo tanto, tenemos una fuente continua de ruido ambiental en varias frecuencias, además es una zona relativamente homogénea en el sentido geológico y urbano.



Figura 21. Estación VRVM. La estación se encuentra en un asilo cerca de un IMSS y entre estructuras de un nivel.

Resultados de H/V en la Zona de Transición

A pesar de que la nueva nomenclatura no define la Zona de Transición, es importante resaltar que estos sitios corresponden a las zonas más complejas. En general se observan que los periodos no están bien definidos, es decir, las amplitudes máximas son difíciles de identificar. A veces aparecen amplitudes máximas a periodos inesperados, es decir, que las amplitudes máximas no siempre se encuentran en el periodo dominante de vibración. En la parte de discusión se explican las posibles causas de estos importantes cambios.

Estación APVM (Azcapotzalco)

No se visitó en persona esta estación, pero se obtuvieron imágenes de la página del Servicio Sismológico Nacional (SSN, 2022), para observar que había a su alrededor. Esta estación se encuentra en las instalaciones de protección civil. Se observa una estructura de un piso dentro de un espacio amplio. El pico principal no está bien definido y representa periodos cercanos a 1 segundo. También existen diversos máximos en modos superiores.

Pensamos que esta estación tiene una litología compleja, el sitio puede sufrir los efectos de cambios hidrológicos, climáticos y antropogénicos, la amplitud de H/V es en general baja. Esta zona probablemente ha sido influenciada por diferentes efectos de tipo litológico y sedimentológico, en algunos momentos el espectro representa formas completamente diferentes cuando se observan a tiempos variables.

Es importante destacar que la zona de transición aparentemente refleja una zona de alta peligrosidad sísmica, esto se debe a que los periodos dominantes asignados se encuentran ligeramente debajo de un segundo, Buendía Sánchez y Reinoso Angulo (2019) realizaron un análisis de las estructuras colapsadas en la CDMX donde relacionan el periodo elástico de la estructura con el periodo fundamental del suelo, la gran mayoría de las estructuras que colapsó se encontraban en suelos con periodos entre 0.5 y 1.5 segundos. Por lo que se sugiere que se estudie esta estación más a detalle debido a que es difícil asignarle con claridad un valor al periodo fundamental.

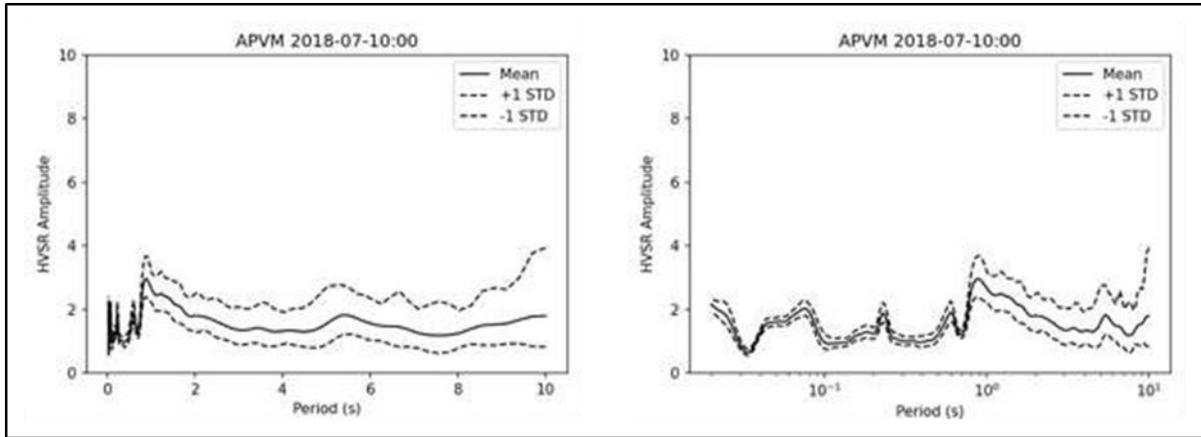


Figura 22. APVM_2018-07-10/00.

Estación BJVM (Benito Juárez)

La estación se encuentra en una casa de cultura de la Ciudad de México. Aparentemente es un sitio con un periodo bien definido a los 0.5 segundos. Sin embargo, en nuestra visita obtuvimos información relevante. Este sitio se encuentra sobre una placa de concreto que pertenecía a una antigua fábrica textil. Esa losa nunca se removió y se encuentra enterrada a pocos metros.

Observamos que esta estación no presentó variación anual de ningún tipo, pero en cambio era evidente que existían otros periodos característicos que no se pueden distinguir, se piensa que son los que realmente representan el sitio. Un punto interesante es que logramos identificar que los sitios que se encuentran “contaminados” por estructuras ocultas, son los sitios que no presentan cambios estacionales. Sus periodos son constantes a lo largo del año. En cambio, los periodos que realmente representan las características dinámicas del sitio, son lo que encontramos en los espectros que tienen ligeras variaciones en periodos y amplitudes.

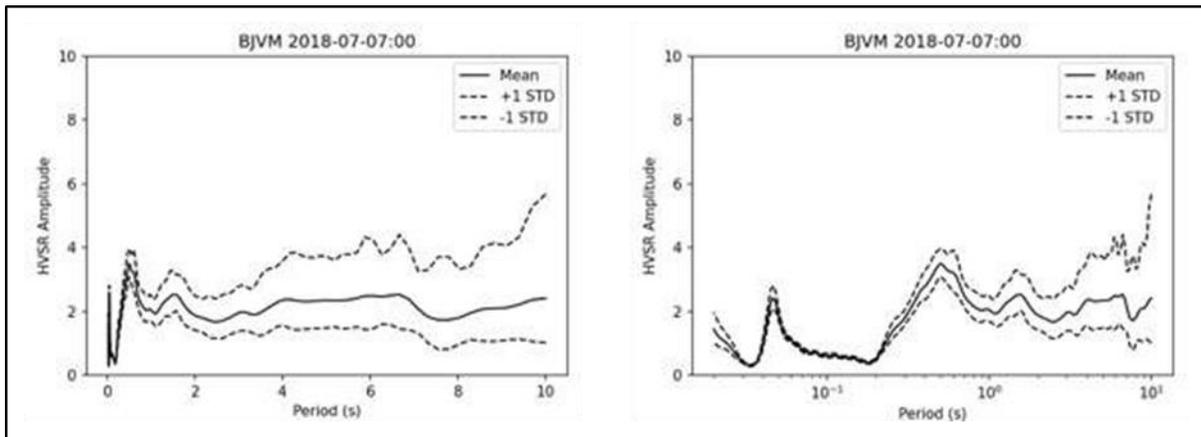


Figura 23. BJVM_2018-07-07/00.



Figura 24. Estación BJVM. La estación se encuentra adentro de una casa de cultura entre dos edificios.

Estación THVM (Tláhuac)

La estación Tláhuac no se visitó pero es una estación que presenta un pico a los 5 segundos. Aparentemente este valor es incorrecto debido a su ubicación. Sin embargo, se cree que es un sitio que puede presentar periodos largos. Primeramente, el pico máximo tiene una amplitud de 4 en los espectros HVSR, esta amplitud es cuatro veces menor que Venustiano Carranza, Iztacalco y Cuauhtémoc. Se encuentra en un sitio donde las rocas de la formación Tarango y las Sierras Mayores del Cuaternario Superior han cubierto completamente otros sedimentos aluviales de grandes planicies. Es probable que estos sedimentos se encuentren entre 800 y 1000 metros de profundidad. Si hacemos una aproximación por medio de la relación entre los parámetros de profundidad, velocidad y periodo dominante (CFE, 2008), las rocas de 800 m/s de esta zona estimarían periodos dominantes de 5 segundos.

Cuando comparamos con el mapa de periodos dominantes parece ser que esta estación está completamente fuera del lugar, es decir, el periodo fundamental está en 5 segundos, en cambio el periodo dominante del mapa de la Figura 3 debería corresponder a 0.5 segundos. Algunos investigadores han observado este tipo de inconsistencias, por ejemplo Jaramillo et al. (2012) observó periodos muy largos en zonas relativamente competentes. Creemos que es importante comparar la amplitud de los espectros H/V para poder diferenciar los periodos fundamentales de los periodos característicos que ocurren cuando se excitan capas relativamente duras, pero con bajas velocidades que generan amplitudes menores que no representan la excitación del antiguo Vaso de Texcoco. Es decir, si existe un periodo fundamental pero que no necesariamente está relacionado al periodo dominante.

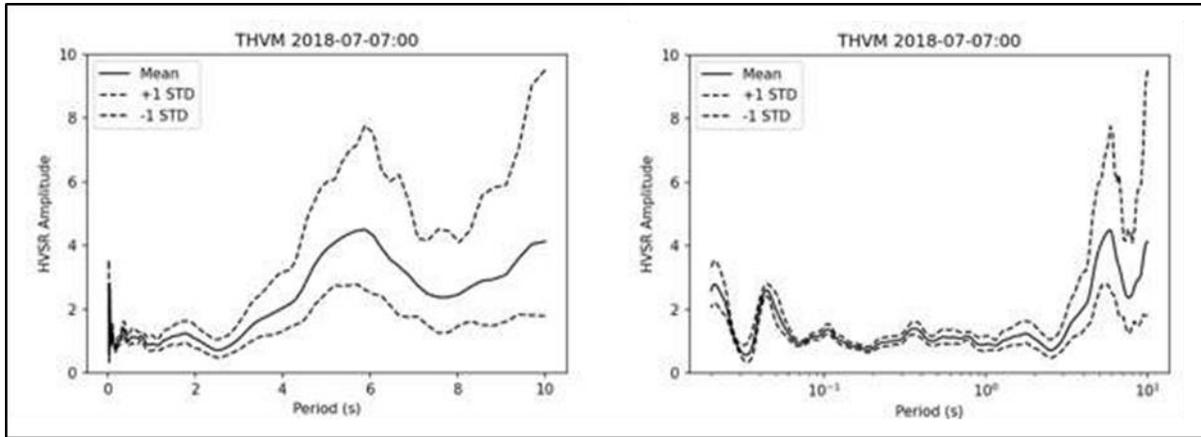


Figura 25. THVM_2018-07-07/00.

Estación TXVM (Texcoco)

La estación Texcoco predice correctamente los valores esperados en los mapas de periodos dominantes (Figura 3). Se puede usar como un punto de referencia para estudios posteriores, sobre todo, para entender las zonas de altos gradientes hidrológicos.

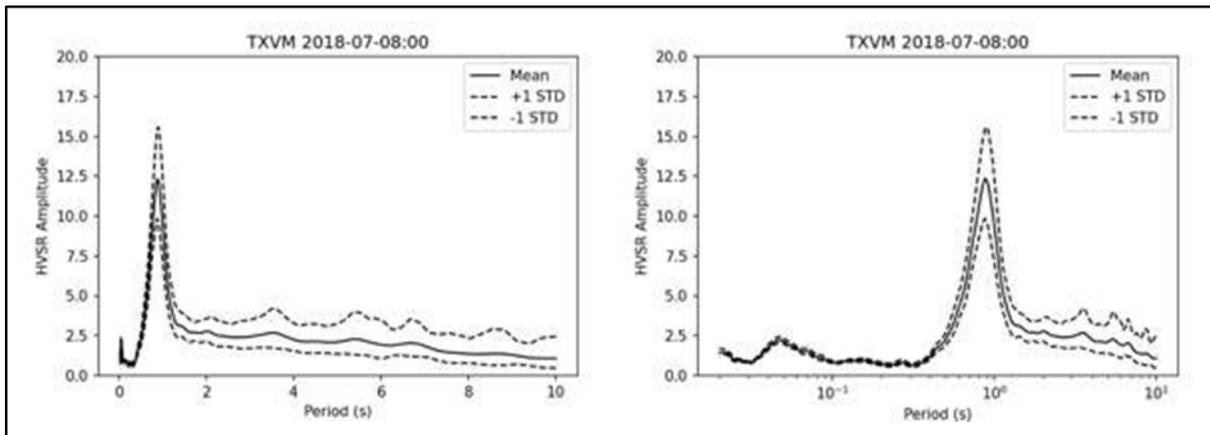


Figura 26. TXVM_2018-07-08/00.

Estación XCVM (Xochimilco)

Esta estación presenta cambios interesantes, su periodo dominante coincide con el fundamental. Sin embargo, presenta elementos similares a la estación Tláhuac, debido a su entorno geológico y proximidad, es probable que entre 700 y 900 mts. de profundidad se

encuentren sedimentos de abanicos aluviales, es decir, espesores ligeramente menores pero que pueden estar controlados por el mismo sistema profundo.



Figura 27. Estación XCVM. La estación se encuentra cerca de un árbol grueso y alto, atrás hay un salón de clases y a un lado se encuentra un complejo departamental.

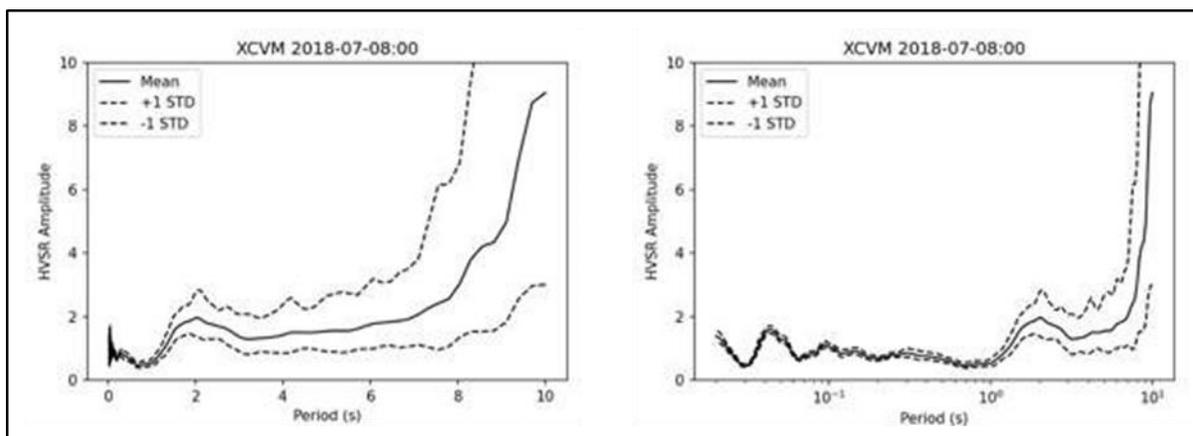


Figura 28. XCVM_2018-07-08/00.

Resultados de H/V en la Zona de Lomas

La zona de Lomas es la zona de mayor dificultad de estudio. Esta zona debe reflejar espectros planos. Teóricamente, los espectros H/V se representan en forma de una línea constante cuando no existe amplificación debido a que tenemos una roca competente sin interfases de sedimentos blandos. Pero nuestras observaciones muestran todo lo contrario, estos espectros resultan variar fuertemente con el tiempo.

Estación AMVM (Amecameca)

Esta estación se encuentra en la Zona de Lomas, con un periodo dominante estimado de 0.4 segundos, como podemos apreciar en la Figura 29 el periodo fundamental se observa a los 0.2 segundos, lo cual implica una variación menor.

Por ser un sitio con basamento rocoso, esperaríamos una curva plana en todos los periodos, pero solo ocurre de 0.6 a 10 segundos, al inicio aparecen tres picos, el primero parece que su máximo está en 0.2 segundos. Sin embargo, la estructura es más compleja, porque hay otros periodos que hacen interferencia, por lo que se divide en tres (intervalo de 0.2 a 0.4 segundos), en el mes de junio el pico del centro llega a superar el periodo dominante a lo largo del día (amplitud máxima en 0.3 segundos), estos periodos corresponden a estructuras de 50 metros aproximadamente.

Debido a que los cambios ocurren anualmente pensamos que se trata de variaciones de origen antropogénico, el dominante (0.2 segundos) no parece mostrar variación significativa durante el año, el segundo pico tiene su máximo en 0.08 segundos, pero en enero de 2020 también llega a superar la amplitud del periodo dominante durante el día, a principios de abril parece disiparse y a finales de julio vuelve a incrementar. Esta variación probablemente se encuentra controlada por los cambios cíclicos del nivel freático.

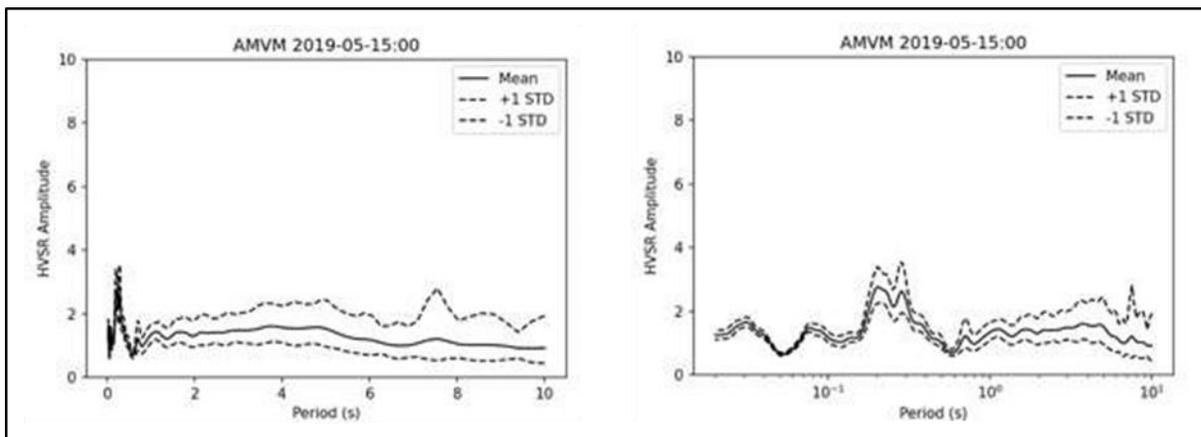


Figura 29. AMVM_2019-05-15/01.

Estación AOVM (Álvaro Obregón)

Teóricamente cualquier espectro H/V en la Zona de Lomas debe ser plano para todas las frecuencias, esto se debe a que no deberían existir depósitos de sedimentos blandos, por lo tanto, esta relación se debe mantener constante a lo largo del año. Sin embargo, encontramos que esto no es cierto, en el caso de la estación Álvaro Obregón se observan efectos de todo tipo, primero un periodo fundamental claro a los 1.6 segundos, la cual es variable en todos los sentidos, esto es temporal y espectral, es decir, aumenta ligeramente en modos superiores en un lapso de varios meses.

Por otro lado, los periodos menores también tienen variabilidad a lo largo del año, todas estas variaciones son claramente estacionales y se disipan a lo largo de algunos meses, pero aparecen nuevamente al siguiente año.

Una consecuencia interesante de estas observaciones se refiere a la definición de lo que es una estación de referencia, es decir, teóricamente se supone que cualquier estación en la Zona de Lomas o de roca dura debe tener características estacionarias, pero no es así. Creemos que es necesario hacer una revisión cuidadosa en los sitios de roca dura o en la Zona de Lomas.

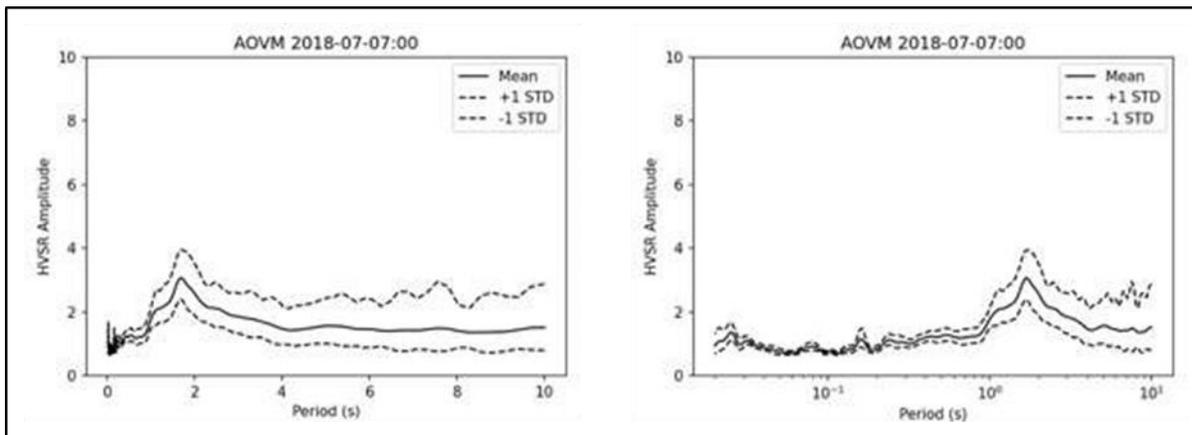


Figura 30. AOVM_2018-07-07/00.

Estación ATVM (Atacomulco)

Atacomulco se encuentra muy lejos del Valle de México y está localizado en otro tipo de valle aluvial, en este se puede observar un periodo fundamental de 0.7 segundos, pero con una amplitud considerable. Es importante destacar que a pesar de que los estudios de periodos dominantes se han realizado desde hace varios años, solo se han hecho en sitios muy específicos, especialmente en la Ciudad de México. Sin embargo, los núcleos de población se han esparcido a lo largo de toda la República, incluyendo sitios como Atacomulco, en donde se debería realizar una zonificación más a detalle.

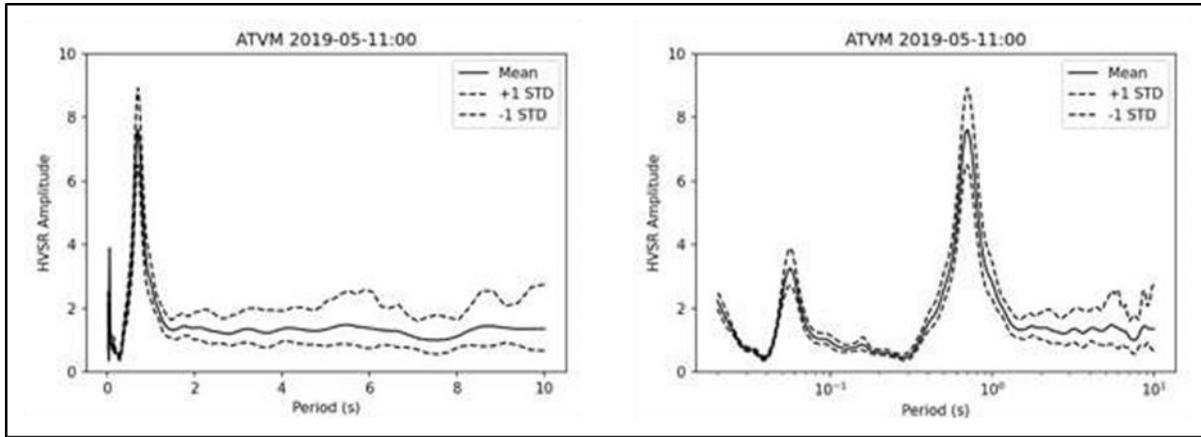


Figura 31. ATVM_2019-05-11/00.

Estación AZVM (Atizapán)

Atizapán es una estación que es muy difícil de analizar porque a pesar de que se encuentra en Zona de Lomas y que refleja relativamente un espectro plano, las amplitudes máximas están invertidas. Si analizáramos una función de transferencia de un sitio donde la frecuencia dominante siempre es la de menor valor, es decir, a medida que las frecuencias aumentan, las amplitudes disminuyen. En este caso, ocurre lo contrario, por lo tanto, interpretamos que el periodo característico de los sedimentos de menor espesor son los que controlan la vibración de este sitio.

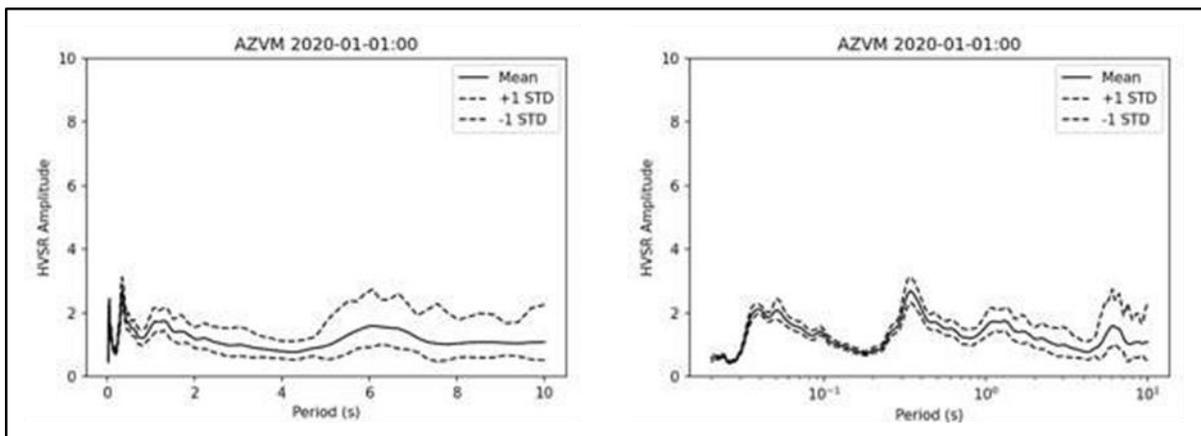


Figura 32. AZVM_2020-01-01/00.

Estación CJVM (Cuajimalpa)

En este caso, la primera curva no varía en amplitud, pero la meseta máxima se amplía y se reduce a lo largo del día, pasa de entre 0.6 a 1.8 a 0.6 a 2 segundos. No conocemos las causas de porqué se observa una meseta a ciertas horas y luego desaparece, además de que hay poca información del periodo que se analizó. Una posibilidad podría ser condiciones muy locales, por ejemplo, el cauce del río que se encuentra a pocos metros de la estación.

Además, los cambios que se observan en los periodos más bajos, como que tiende a aumentar en ciertas ocasiones, por ejemplo el 09 de julio de 2019 a la 01:00am pasó de 4 a 7 en amplitud, volvió a su amplitud normal en cuestión de horas, pueden estar relacionados con la altura del techo que vemos en la Figura 34, por cambios de velocidad en el viento o por la actividad hidrológica del arroyo cerca que se encuentra a pocos metros de distancia y a una profundidad aproximada de 5 metros, con una velocidad de corte de 40 m/s que corresponde a los sedimentos no consolidados. Se pueden observar las variaciones de la saturación del nivel freático, es decir, los cambios del nivel de agua.

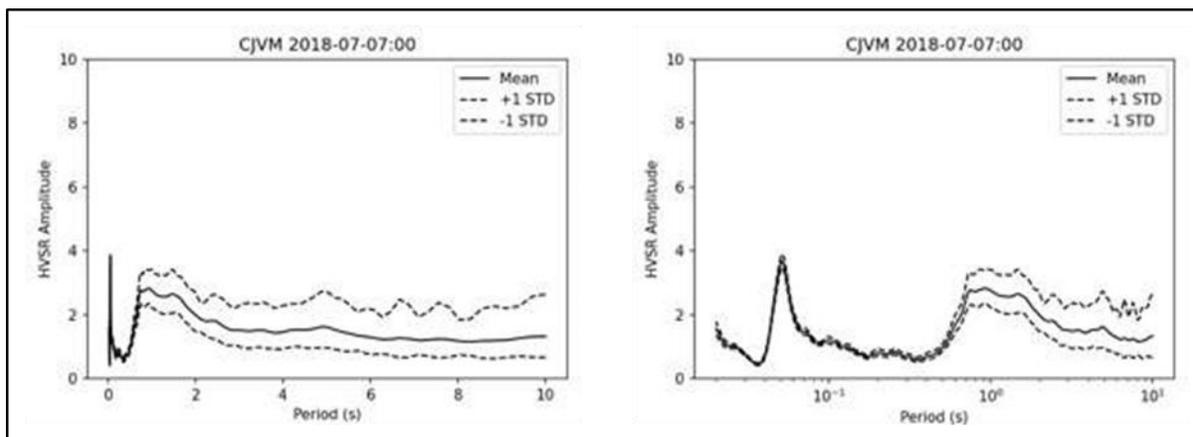


Figura 33. CJVM_2018-07-07/00.



Figura 34. Estación CJVM. La estación se encuentra en una escuela, al lado de estructuras altas y cerca de un río.

Estación INVM (ININ)

Esta estación debería presentar una curva plana parecida al tramo de 1 a 10 seg., sin embargo, podemos observar 3 picos que varían en amplitud, pero su forma y periodos se mantienen estables en el tiempo.

La zona plana tampoco presenta variaciones considerables, sin embargo, se puede inferir que la señal real debería ser plana en su totalidad por la zona en que se encuentra, pero debido a que está dentro de un Instituto de Investigaciones Nucleares, los tres picos pueden considerarse como una señal externa que detecta el instrumento y recubre la señal real.

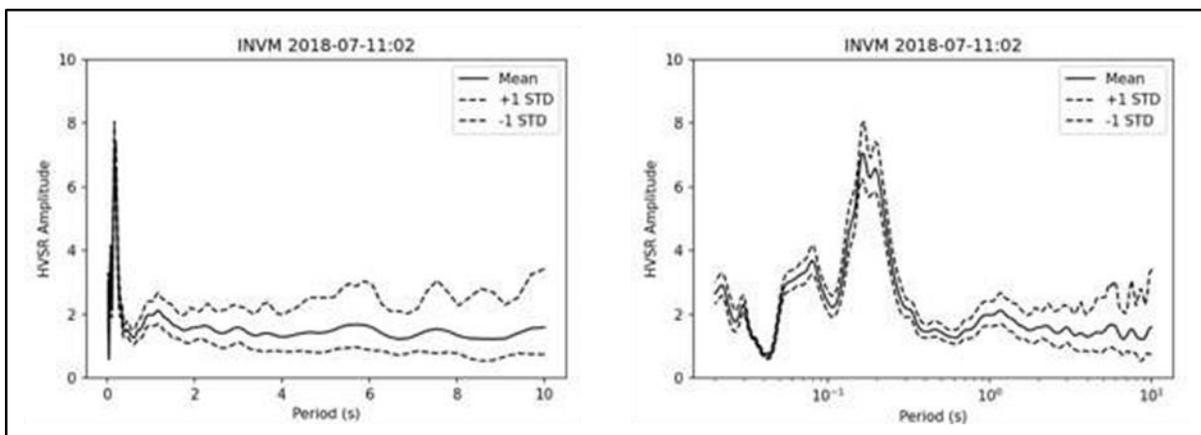


Figura 35. INVM_2018-07-11/02.

Estación MAVM (Malinalco)

Esta estación también refleja que no se manifiestan cambios hidrológicos todo el tiempo en todas las estaciones sísmicas, nuevamente es un caso similar al de Atlacomulco, donde el sitio se encuentra muy lejano al Valle de México y presenta algún tipo de efecto de sitio.

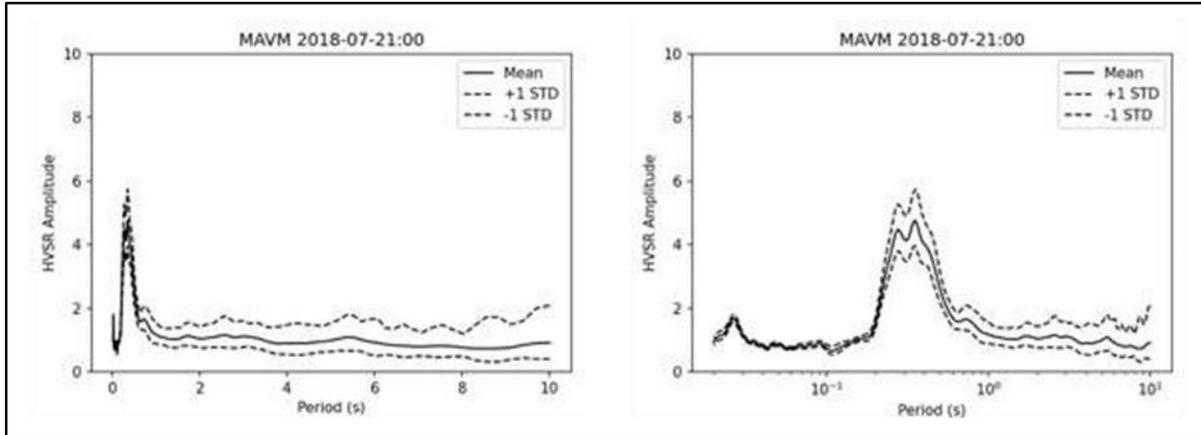


Figura 36. MAVM_2018-07-21/00.

Estación MCVM (Magdalena Contreras)

Es probable que debajo de este sitio tengamos dos acuíferos a diferentes profundidades, el primer acuífero se encuentra no confinado y representa el pico de la derecha porque tiene variaciones estacionales muy definidas, en cambio, el que tiene la frecuencia dominante se encuentra en un acuífero confinado, por lo tanto, no cambia de amplitud a lo largo de las estaciones del año. Esta interpretación se puede correlacionar con los mapas geológicos del Valle de México donde se han caracterizado hasta 4 acuíferos diferentes y muestran en algunos casos que se encuentran debajo de las rocas volcánicas más jóvenes.

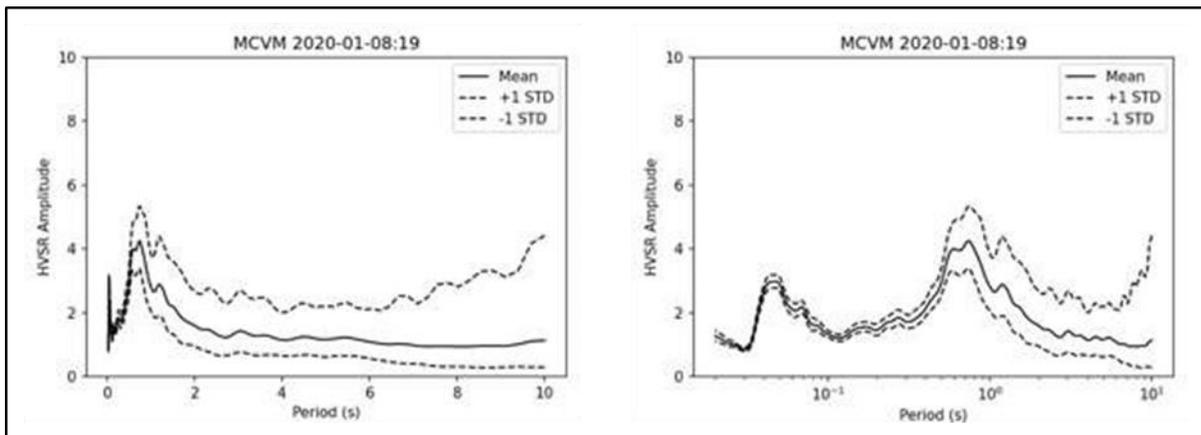


Figura 37. MCVM_2020-01-08/19.

Estación MHVM (Miguel Hidalgo)

Esta estación se encuentra dentro del Panteón Civil de Dolores, el cual está dentro del cerro Chapultepec que pertenece a rocas volcánicas del periodo Terciario, por la zona en la que se encuentra se espera una curva más plana y uniforme, sin embargo, se pueden apreciar cuatro picos, de los cuales los tres no presentan cambios en amplitud o periodo, se mantienen constantes a lo largo del tiempo, el primero si llega a superar la amplitud del segundo. Por lo anterior, podemos decir que es difícil identificar cual es el periodo dominante, ya que el periodo fundamental no coincide con el que se especifica en la Figura 3.

La estación Miguel Hidalgo ocupa una mención especial en este trabajo debido a que, los cambios en los modos superiores son claramente estacionales y se encuentran controlados por los flujos subterráneos de todo el Valle de México. Debido a esto, se realizó un análisis cuidadoso para entender su relación con los cambios de agua del subsuelo, los cuales se presentan en el capítulo 5. Comparando la Figura 11 con la Figura 38 podemos ver la aparición y desaparición de modos superiores muy cerca de los 0.05 segundos. Que corresponden a una profundidad aproximada de 8 metros, que es precisamente donde se localiza el nivel freático. Esta zona es interesante porque representa los cambios de masas de agua de todo el Valle de México, tal vez porque funciona como una zona de recarga para los acuíferos del Valle de México.

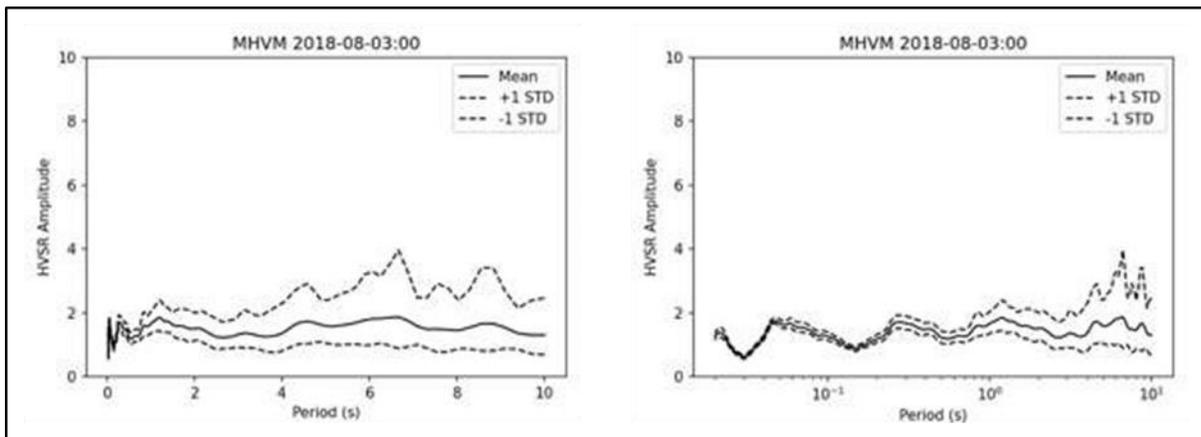


Figura 38. MHVM_2018-08-03/00



Figura 39. Estación MHVM. La estación se encuentra adentro de un panteón, en una zona despejada, al lado tiene un edificio de un nivel.

Estación MPVM (Milpa Alta)

Esta estación refleja actividad antropogénica en todos los sentidos, probablemente relacionado a cambios en la actividad cultural, por ejemplo, en época de siembra o de eventos importantes como días festivos o cambios de estaciones que es cuando cambian las actividades agrícolas, además a periodos bajos es probable que existan algunas relaciones con el patrón hidrológico de la zona.

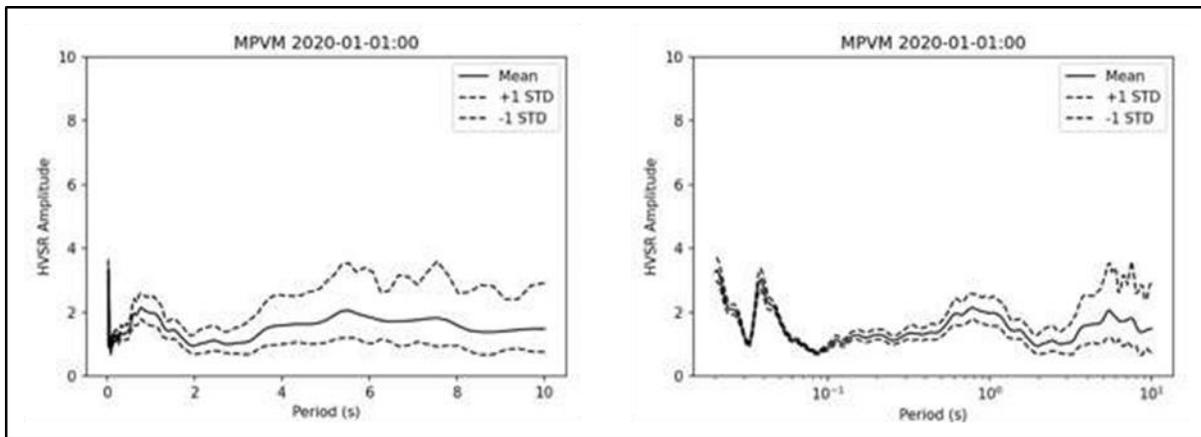


Figura 40. MPVM_2020-01-01/00.

Estación MZVM (Mezontepec)

Esta estación tiene muy poca información para poder asociar los cambios en el espectro a la litología del lugar. Su periodo fundamental no corresponde con el mapa de periodos dominantes, por lo que se vuelve a hacer hincapié en un análisis a detalle de las estaciones ubicadas en la Zona de Lomas.

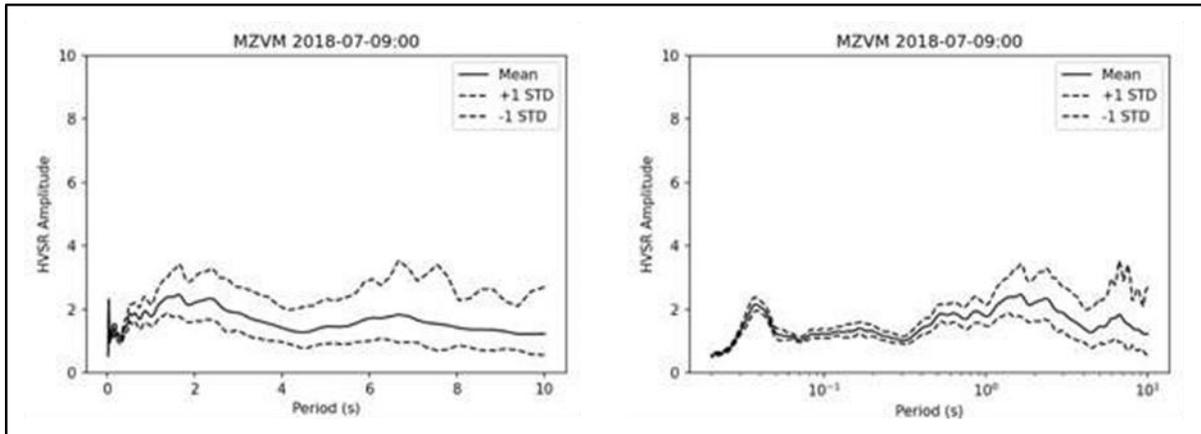


Figura 41. MZVM_2018-07-09/00.

Estación PBVM (Peñón de los Baños)

En esta estación podemos observar una curva plana y dos picos con mayor amplitud en periodos bajos, sin embargo, éstos parecen cambiar con el paso de los días, pero manteniendo una secuencia, primero se forman dos con la misma amplitud, después el de la derecha aumenta su amplitud y el de la izquierda disminuye, por último, aparece uno con una amplitud considerable, esto se repite durante los 3 años. En cambio, el resto del espectro no presenta variación, se mantiene estable en periodo y amplitud. Por la ubicación en la que se encuentra esta estación, esas variaciones en los periodos bajos pueden deberse a la interferencia que provoca una torre de radar perteneciente al aeropuerto que se encuentra a un lado de la caseta como podemos observar en la Figura 43. En este sitio no fue posible encontrar el periodo dominante reportado en el mapa de la Figura 3, probablemente se debe al tipo de roca volcánica en el cual se encuentra la estación de Peñón de los Baños. En los mapas geológicos del Valle de México se observa que el Peñón de los Baños corresponde a edificios volcánicos que no se encuentran asociados al resto de la litología del Valle de México.

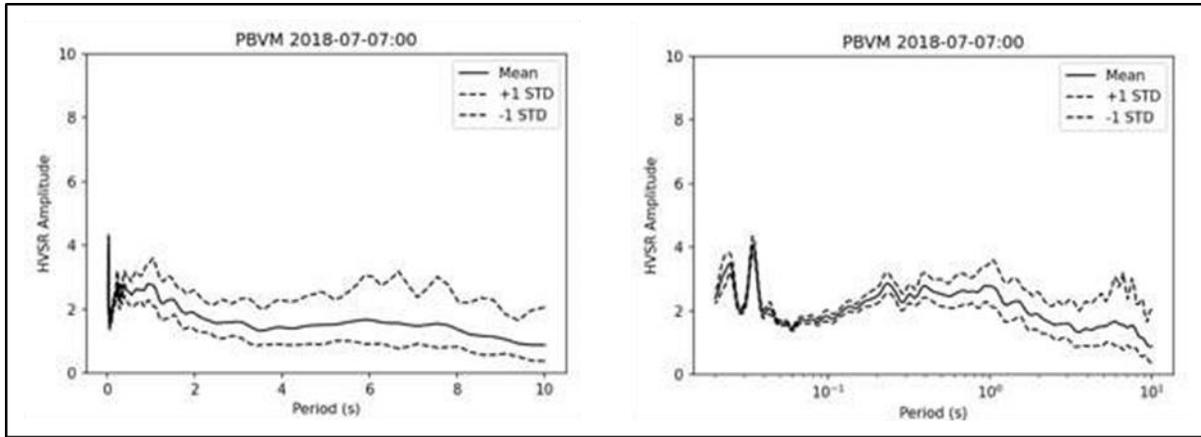


Figura 42. PBVM_2018-07-07/00.

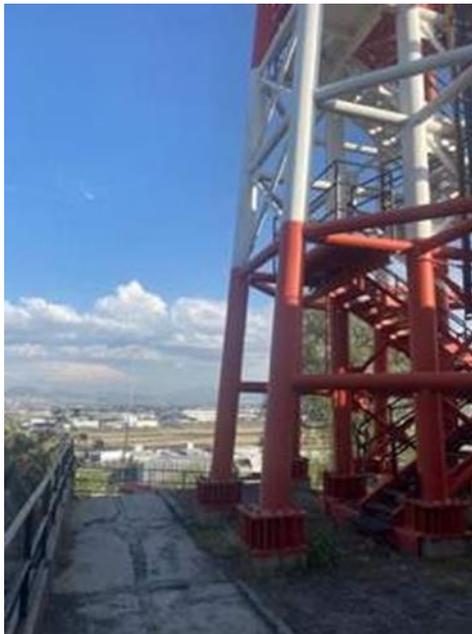


Figura 43. Estación PBVM. La estación se encuentra cerca de una torre de radar.

Estación PTVM (Pico Tres Padres)

Hay muy poca información registrada en esta estación, por lo que se pudieron analizar todas las imágenes. Las primeras dos imágenes son horas del mismo día de enero donde podemos ver una curva relativamente plana, que se mantiene estable ese día, de ahí se brinca al 14 de abril donde se sigue manteniendo estable las primeras horas (de 2 a 13hrs), a las 14 hrs hay interferencia en la señal y la recuperamos hasta las 17 hrs, la curva sigue sin presentar variaciones, brinca al 21 de mayo a las 0 hrs en donde la variación está en los periodos bajos, donde el primer pico parece tener un pequeño aumento en su amplitud pero solo durante 3

hrs ya que a las 4 hrs el pico se redujo a la misma amplitud que el resto de la curva, aumenta conforme pasan las horas hasta las 8 hrs donde empezó a disminuir nuevamente. A las 14 hrs del mismo día la señal vuelve a registrar una interferencia, se recupera la señal hasta las 18 hrs, aún sin variación hasta la última imagen que es a las 21 hrs.

Debido a la escasa información, este sitio es similar a una medición clásica donde no se hace un análisis estacional, en este caso no es posible saber si el aparente periodo fundamental que ocurre a los 1.5 seg. realmente se trata de un efecto permanente o estacional. Esto probablemente corresponde a observaciones similares donde otros investigadores han encontrado periodos dominantes de más de un segundo en Zona de Lomas.

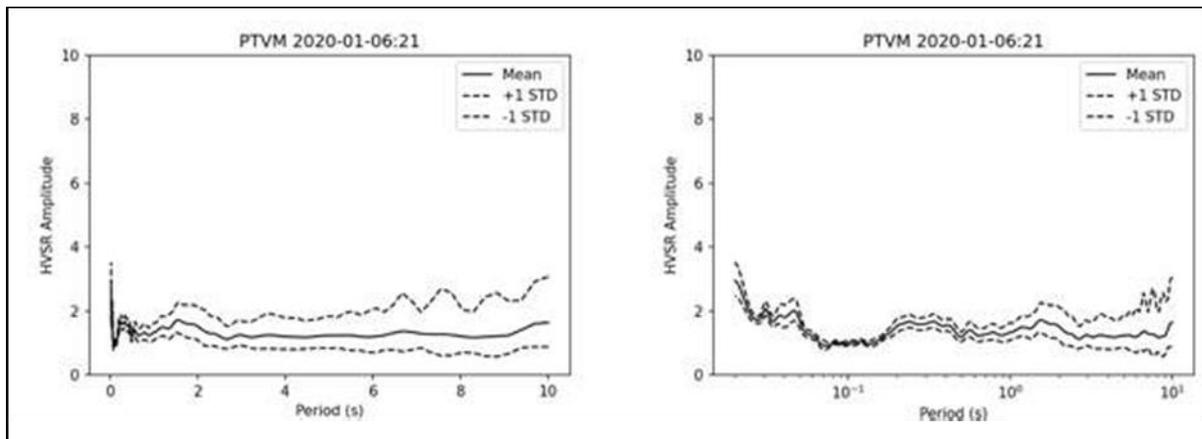


Figura 44. PTVM_2020-01-06/21.

Estación TLVM (Tlalpan)

No hay mucha información consecutiva para esta estación, pero las variaciones no parecen ser significativas, podría decirse que no existen ya que no se aprecian aumentos de amplitud, en ocasiones una disminución de la misma pero no se tendría una razón muy certera debido a que ocurre en pocos días y no sé sabe si el resto de los días pasó lo mismo. Una situación similar a lo que vimos en Pico Tres Padres.

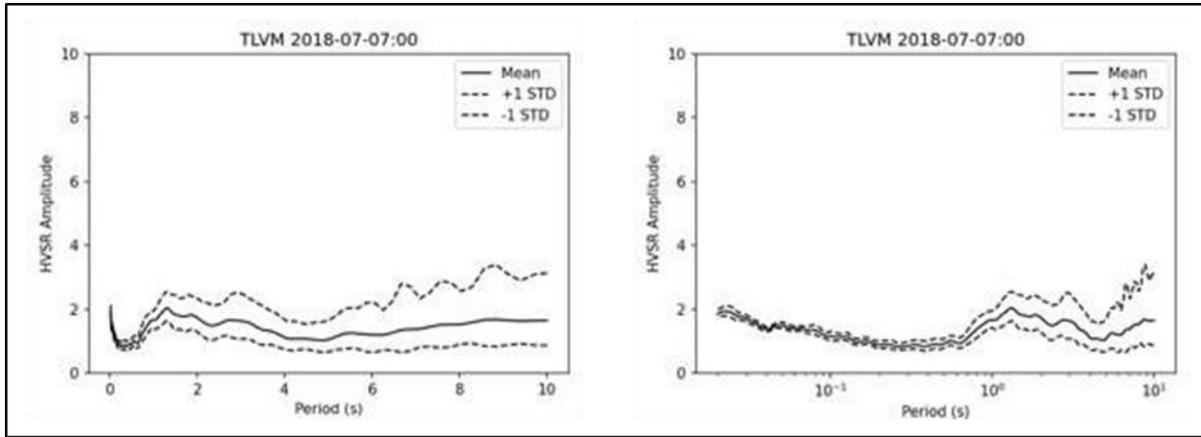


Figura 45. TLVM_2018-07-07/00.

Estación TOVM (Toluca)

En esta estación podemos observar una curva plana, muy limpia que no presenta variación en amplitud o periodo después de 0.1 seg., ya que antes tenemos un pico que, si bien no varía considerablemente, su forma puede cambiar en ocasiones, suele disiparse, pero no aumentar en amplitud. Creemos que Toluca es el mejor sitio de referencia en todo el centro de México, porque realmente representa un sitio donde todo el espectro es constante y plano a lo largo de todo el año, incluso a los cambios ambientales e hidrológicos, tal vez esto se deba al tipo de roca impermeable y a las condiciones litológicas del sitio.

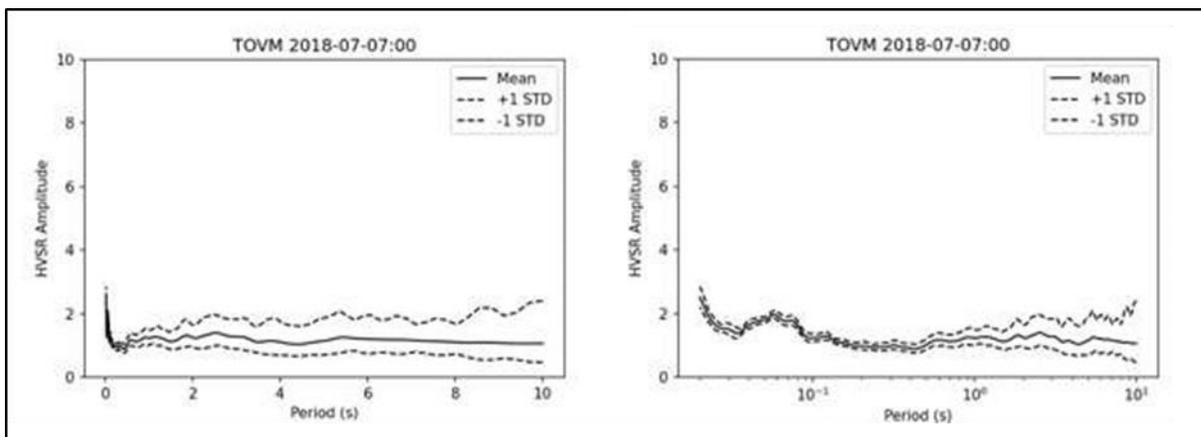


Figura 46. TOVM_2018-07-07/00.

Estación VTVM (Valle de Teotihuacán)

Esta estación probablemente es de las pocas donde se ha logrado obtener un periodo dominante en la Zona de Lomas adecuado al mapa de la Figura 3. Esta situación también demuestra la enorme dificultad de analizar los sitios con periodos menores a 1 seg.

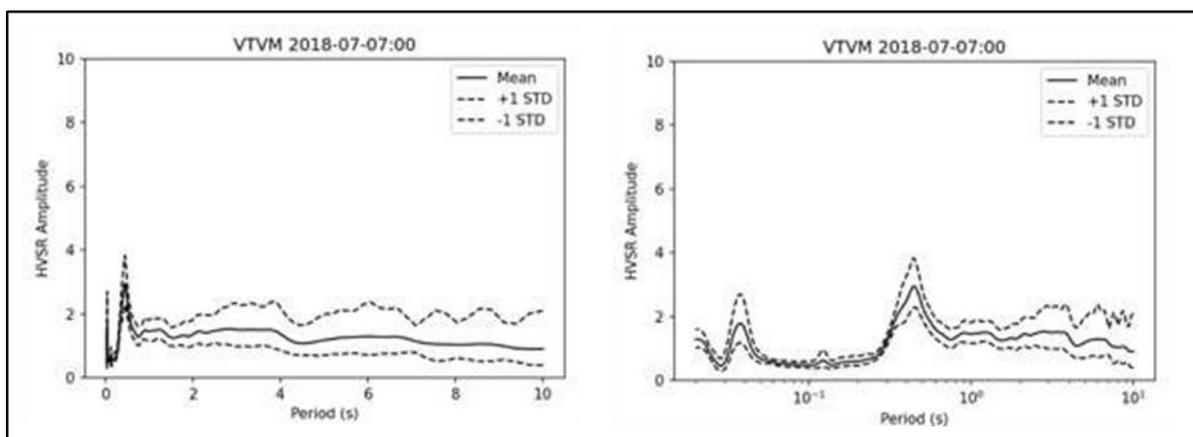


Figura 47. VTVM_2018-07-07/00.



Figura 48. Estación VTVM. La estación se encuentra en un espacio despejado cerca de un árbol y a pocos metros de una sala de ordeño.

Estación ZUVM (Zumpango)

Esta estación se encuentra cerca de un valle aluvial alejado del Valle de México, por lo tanto, debe analizarse fuera del contexto de los mapas de periodos dominantes exclusivos para los periodos fundamentales asociados al Lago de Texcoco y al valle central.

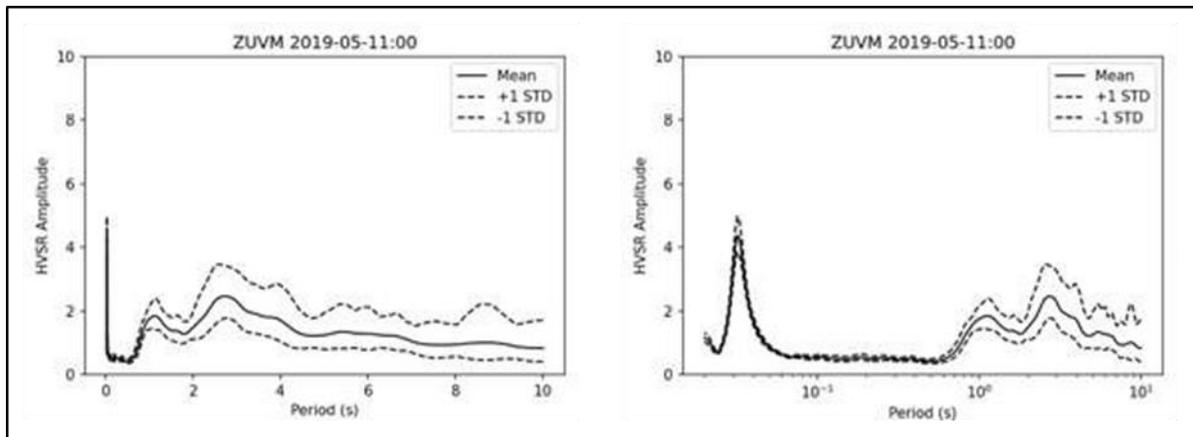


Figura 49. ZUVM_2019-05-11/00.

APÉNDICE 3

Fotogramas de campo

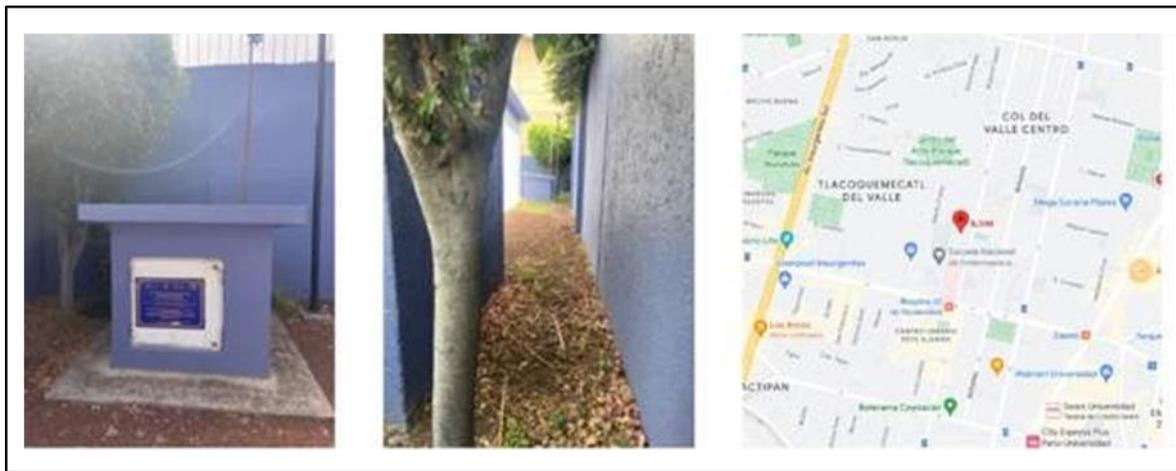


Figura 50. Estación BJVM (Benito Juárez). Ubicada en Zona II (Zona de Transición).



Figura 51. Estación COVM (Coyoacán). Ubicada en Zona II (Zona de Transición). No registró datos.

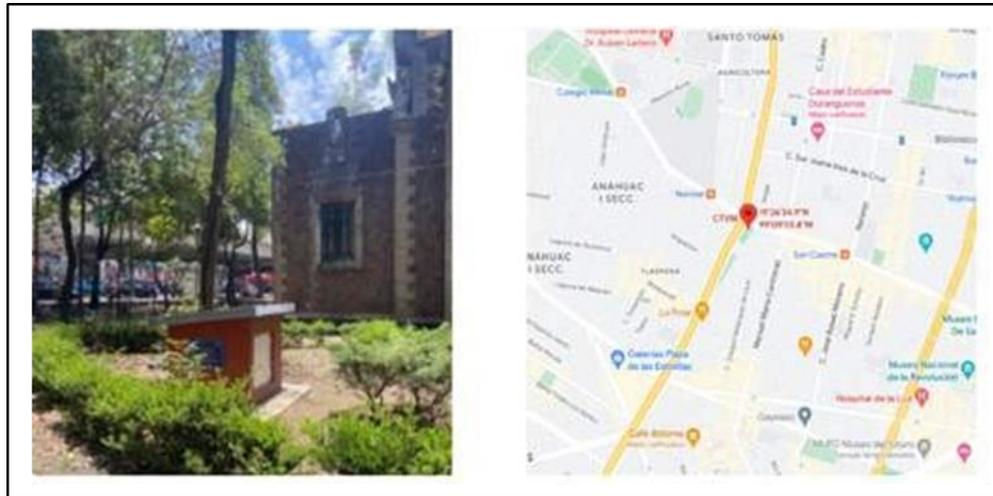


Figura 52. Estación CTVM (Cuauhtémoc). Ubicada en Zona III (Zona del Lago).

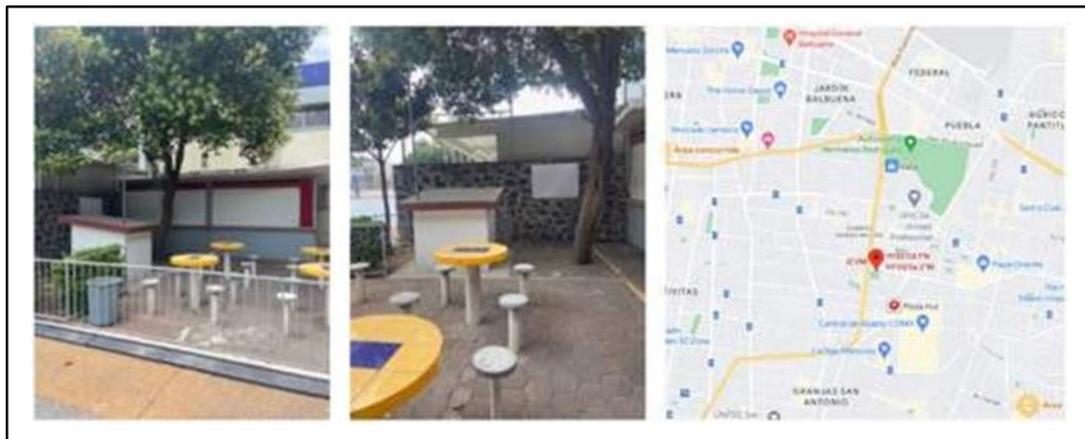


Figura 53. Estación ICVM (Iztacalco). Ubicada en Zona III (Zona del Lago).

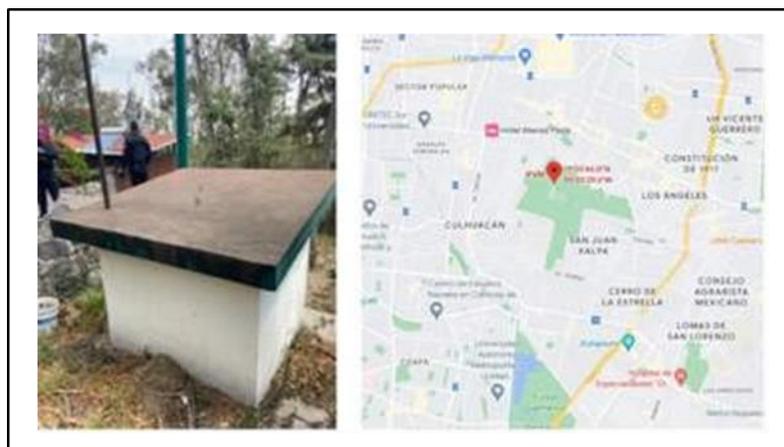


Figura 54. Estación IPVM (Iztapalapa). Ubicada en Zona I (Zona de Lomas). No registró datos.



Figura 55. Estación MHVM (Miguel Hidalgo). Ubicada en Zona I (Zona de Lomas).

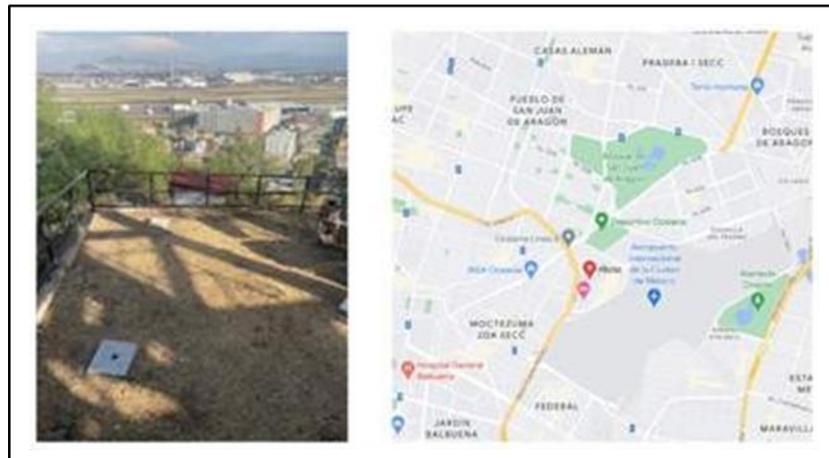


Figura 56. Estación PBVM (Peñón de los Baños). Ubicada en Zona III (Zona del Lago).



Figura 57. Estación XCVM (Xochimilco). Ubicada en Zona II (Zona de transición).



Figura 58. Estación CJVM (Cuajimalpa). Ubicada en Zona I (Zona de Lomas).

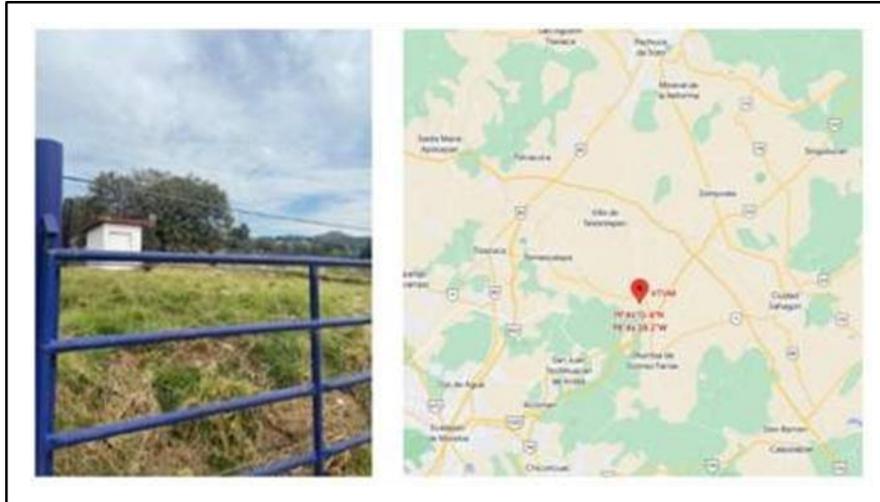


Figura 59. Estación VTVM (Valle de Teotihuacán). Ubicada en Zona I (Zona de Lomas).