

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

**IMPLEMENTACIÓN DE UN TOOLBOX DE MATLAB PARA EL
PROCESAMIENTO DE SEÑALES SÍSMICAS A PARTIR DE LAS
BASES DE DATOS DEL INSTITUTO GEOFÍSICO DE LA ESCUELA
POLITÉCNICA NACIONAL.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERA EN ELECTRÓNICA Y TELECOMUNICACIONES**

CONDOR NAVAS JENNIFFER PATRICIA

**DIRECTOR: Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO
CODIRECTOR: Ph.D. ANA MARÍA ZAMBRANO VIZUETE**

Quito, diciembre 2021

AVAL

Certificamos que el presente trabajo fue desarrollado por Jenniffer Patricia Condor Navas, bajo nuestra supervisión.

Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO
DIRECTOR DEL TRABAJO DE TITULACIÓN

Ph.D. ANA MARÍA ZAMBRANO VIZUETE
CODIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo, JENNIFFER PATRICIA CONDOR NAVAS, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración dejo constancia de que la Escuela Politécnica Nacional podrá hacer uso del presente trabajo según los términos estipulados en la Ley, Reglamentos y Normas vigentes.

JENNIFFER PATRICIA CONDOR NAVAS

DEDICATORIA

Este trabajo esta dedicado a:

A Dios, a mis padres, Hector Condor y Marlene Navas, a mis hermanos quienes han estado junto a mí durante todo este proceso brindándome cariño, apoyo y la confianza para afrontar cada reto.

A mis amigos, a Darío quienes me han apoyado y brindado su cariño, amor y confianza en todo momento.

Finalmente dedico este trabajo a todas las personas que estuvieron junto a mí, durante todo este proceso y continúan apoyándome en mis nuevos retos.

Jennifer Condor

AGRADECIMIENTO

Un agradecimiento especial a mis padres y hermanos por su apoyo incondicional, en todo momento en este recorrido.

Al Ph.D. Felipe Grijalva quien ha sido una guía en todo momento, brindándome su tiempo, conocimiento y el apoyo necesario para la elaboración de este proyecto. A los docentes de la Escuela Politécnica Nacional que con su conocimiento ayudaron en mi preparación profesional.

Al instituto Geofísico en especial al Ph.D. Mario Ruiz, por su apoyo con la base de datos usada en este proyecto. A la Escuela Politécnica Nacional por todos los conocimientos y las experiencias que me ha brindado a lo largo de todo este tiempo.

Gracias.

Jennifer

ÍNDICE DE CONTENIDO

AVAL	I
DECLARACIÓN DE AUTORIA.....	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO	V
RESUMEN	VII
ABSTRACT	VIII
1. INTRODUCCIÓN.....	1
1.1. OBJETIVO GENERAL.....	2
1.2. OBJETIVOS ESPECÍFICOS.....	2
1.3. ALCANCE.....	2
1.4. MARCO TEÓRICO	3
1.4.1. MICROSISMOS VOLCÁNICOS	4
1.4.1.1. Tipos de eventos sísmicos.....	4
1.4.2. STANDARD FOR THE EXCHANGE OF EARTHQUAKE DATA (SEED) ...	7
1.4.2.1. MiniSEED	8
1.4.3. ALGORITMO STA/LTA	9
1.4.3.1. Parámetros de detección	10
1.4.4. PREPROCESAMIENTO DIGITAL DE SEÑALES	11
1.4.5. SELECCIÓN DE CARACTERÍSTICAS	12
1.4.5.1. Características en tiempo.....	12
1.4.5.2. Características en frecuencia.....	13
1.4.5.3. Características en tiempo-frecuencia	14
1.4.6. SOFTWARE UTILIZADO	16
1.4.6.1. Matlab.....	16
1.4.7. t-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING	16
1.4.8. ESPECTROGRAMA.....	17

2. METODOLOGÍA	18
2.1. BASE DE DATOS IGEPN	19
2.2. ADQUISICIÓN Y LECTURA DE ARCHIVOS .mseed	21
2.3. COORDENADAS EN TIEMPO IGEPN	23
2.4. BÚSQUEDA Y EXTRACCIÓN DE EVENTOS	25
2.4.1. DETECTOR STA/LTA	26
2.4.2. DESCARTE DE EVENTOS	29
2.5. PREPROCESAMIENTO DE LA SEÑAL	33
2.5.1. RECORTE DE LA SEÑAL	33
2.5.2. NORMALIZACIÓN DE LA FRECUENCIA DE MUESTREO	33
2.5.3. FILTRADO	33
2.5.4. NORMALIZACIÓN DE LA SEÑAL	35
2.6. EXTRACCIÓN DE CARACTERÍSTICAS	35
3. RESULTADOS Y DISCUSIÓN	48
3.1. RESULTADOS	48
3.1.1. PRUEBAS DE FUNCIONALIDAD	49
3.2. DISCUSIÓN	60
4. CONCLUSIONES Y RECOMENDACIONES	64
4.1. CONCLUSIONES	64
4.2. RECOMENDACIONES	65
5. REFERENCIAS BIBLIOGRÁFICAS	66
ANEXOS	69

RESUMEN

El Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), realiza un monitoreo continuo de los volcanes en tiempo real mediante sismógrafos ubicados en estaciones sísmicas de manera estratégica alrededor de la red de volcanes del Ecuador. Generando bases de datos con los registros grabados y también con archivos en formato de tipo .xlsx, los cuales contienen un etiquetado manual de los eventos sísmicos detectados en las grabaciones. Estas bases de datos generadas a través de los años guardan información relevante que puede ayudar al mejor entendimiento de estos eventos sísmicos. Debido al gran volumen de datos y al formato que maneja el instituto, el cual es mini Standard for the Exchange of Earthquake Data (mSEED), el revisar cada registro de manera manual requiere de mucho tiempo y esfuerzo, además, se necesita hacer la conversión de cada uno de estos registros a formatos que permitan realizar aprendizaje automático, puesto que esta es una herramienta esencial en el análisis de señales sísmicas.

Para poder analizar todas estas bases de manera rápida y verificar el correcto etiquetado de estos eventos sísmicos, se implementa una herramienta digital llamada Toolbox en el software especializado MATLAB®, la cual permite realizar aprendizaje automático. Mediante la creación del Toolbox para procesamiento de señales sísmicas, se obtendrá como primer resultado un conjunto de datos que muestra de manera ordenada y clasificada los eventos extraídos de los registros para cada estación y canal detectado en un volcán y su respectiva confirmación o descarte como evento sísmico. Adicionalmente, los eventos confirmados pasarán por funciones de preprocesamiento como recorte, filtrado y normalización para finalmente extraer de estos 13 características en el dominio de tiempo, 21 en el dominio de la frecuencia y 50 en el dominio de la escala en total 84 características propuestas en el paper “ESeismic: Towards an Ecuadorian volcano seismic repository” [1], las cuales son útiles para el aprendizaje automático y nos permiten ver si los eventos corresponden al grupo en el que se los clasificó manualmente.

PALABRAS CLAVE: Volcanes, eventos sísmicos, mSEED, toolbox, Matlab, estación sísmica, canal, procesamiento de señales.

ABSTRACT

The Geophysical Institute of the National Polytechnic School (IGEPN), performs a continuous monitoring of volcanoes in real time through seismographs located in seismic stations strategically around the network of volcanoes in Ecuador. It generates databases with the recorded records and also with files in .xlsx format containing manual labeling of the seismic events detected in the recordings. These databases generated over the years contain relevant information that can help to better understand these seismic events. Due to the large volume of data and the format handled by the institute, which is the mini Standard for the Exchange of Earthquake Data (mSEED), reviewing each record manually requires a lot of time and effort, in addition, it is necessary to convert each of these records to formats that allow automatic learning, since this is an essential tool in the analysis of seismic signals.

In order to analyze all these bases quickly and verify the correct labeling of these seismic events, a digital tool called Toolbox is implemented in the specialized software MATLAB®, which allows automatic learning. By creating the Toolbox for seismic signal processing, the first result will be obtained in an orderly and classified manner the events extracted from the records for each station and channel detected in a volcano and their respective confirmation or discard as a seismic event. Additionally, the confirmed events will go through preprocessing functions such as trimming, filtering and normalization to finally extract from these 13 features in the time domain, 21 in the frequency domain and 50 in the scale domain, a total of 84 features proposed in the paper "ESeismic: Towards an Ecuadorian volcano seismic repository" [1], which are useful for machine learning and allow us to see if the events correspond to the group in which they were classified manually.

KEYWORDS: Volcanoes, seismic events, mSEED, toolbox, Matlab, station, channel, signal processing.

1. INTRODUCCIÓN

El estudio de eventos sísmicos es relevante para países en los cuales los volcanes representan una amenaza constante, debido a su alta actividad volcánica. Al obtener de manera rápida registros de las señales sísmicas y con una clasificación correcta de estos eventos, aumenta la posibilidad de realizar predicciones oportunas de posibles erupciones volcánicas que ponen en riesgo la seguridad de personas que residen cerca de estos lugares. Los volcanes activos presentan principalmente los siguientes tipos de señales:

VT (Volcano-tectonic): Los eventos de tipo VT tienen una duración típica más corta que 30 segundos y su contenido espectral está aproximadamente sobre los 6 Hz con espectros amplios hasta los 15 Hz [2].

LP (Long-period): Eventos de períodos largos con una duración típica de 90 segundos, poseen una baja frecuencia por lo que su espectro está contenido principalmente entre 2 y 5 Hz [2].

TR (Tremors): Los eventos tipo TR poseen una amplitud sostenida duradera de minutos a días teniendo un rango de frecuencia de 1 a 9 Hz [3].

HYB (Hybrid earthquakes): Los terremotos híbridos comparten características de eventos tipo VT y LP, presentan señales de alta frecuencia con un ancho de banda espectral amplio por encima de los 10 Hz [4].

Las señales sísmicas generadas por las estaciones son recolectadas en un centro de monitoreo en donde estos y otros tipos de eventos son detectados manualmente por expertos geólogos y vulcanólogos, quienes al revisar cada señal una por una, consumen gran cantidad de tiempo y recursos.

Las señales obtenidas por los sismógrafos son almacenadas en bases de datos, en un formato estándar internacional para el intercambio de datos sismológicos digitales llamado Standard for the Exchange of Earthquake Data (SEED), el cual es utilizado por la comunidad de investigación de terremotos. Al proporcionar este estándar, es posible hacer que la transmisión, recepción y procesamiento de datos de terremotos sea más fácil y precisa [5].

Sin embargo, SEED al estar diseñado principalmente para el archivo e intercambio de datos, no fue concebido para realizar un procesamiento de la señal volcánica, por lo cual se realiza un cambio de formato teniendo como opciones diferentes tipos de softwares, siendo uno de estos MATLAB [6].

Matlab es un software que nos permite procesar los datos extraídos del MSEED para realizar aprendizaje automático, el cual con el creciente número de estaciones volcánicas que registran continuamente señales, se ha vuelto una herramienta esencial para el análisis automático de señales sísmicas. Esta herramienta facilita la detección, clasificación de eventos y también puede ser usada para automatizar el análisis de datos volcán sísmicos [3, 4]

El Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), utiliza el formato MSEED por lo general para almacenar la información obtenida de cada volcán monitoreado, y es necesario realizar una conversión de formato de MSEED a Matlab para que los datos sean procesados con técnicas de aprendizaje automático.

Además de esto el IGEPN maneja una base de datos la cual contiene las coordenadas en tiempo de los eventos sísmicos, dichas coordenadas han sido obtenidas de forma manual por los expertos del instituto geofísico.

Por otro lado, existen muchas tareas recurrentes de preprocesamiento y extracción de características de estas señales. Específicamente en el paper “ESeismic: Towards an Ecuadorian volcano seismic repository” se describen una serie de características y tareas de preprocesamiento cuyo uso es recurrente [1]

Debido a esto se plantea un toolbox para Matlab que permita:

1. Extraer los eventos sísmicos a partir de los archivos MSEED y guardarlos en el formato de Matlab.
2. Implementar funciones típicas usadas en el preprocesamiento de señales sísmicas.
3. Extraer las características descritas en [1] de los eventos sísmicos.

1.1. OBJETIVO GENERAL

Implementar un Toolbox de Matlab para el procesamiento de señales sísmicas a partir de las bases de datos del Instituto Geofísico de la Escuela Politécnica Nacional.

1.2. OBJETIVOS ESPECÍFICOS

- Investigar los formatos de datos, tareas de preprocesamiento y extracción de características comúnmente usadas en el Instituto Geofísico para procesar señales sísmicas.
- Realizar la extracción del evento sísmico del archivo MSEED a MATLAB.
- Implementar funciones para el preprocesamiento de las señales sísmicas.
- Extraer características de tiempo, frecuencia y escala de los eventos detectados.

1.3. ALCANCE

Se desarrollará un toolbox en MATLAB que realizará las siguientes funciones:

1. Extracción del evento a partir del archivo MSEED: Los archivos en formato MSEED serán convertidos a MATLAB, mediante la utilización de la función disponible en [7]. Posteriormente se extraerá de estos los eventos sísmicos utilizando la base de datos que contiene las coordenadas en tiempo y etiqueta del evento.

2. Funciones de preprocesamiento:

- Recorte de las señales.
- Filtrado de las señales.
- Normalización de la frecuencia de muestreo.

3. Codificar funciones de extracción de características: Se implementará el código para extraer las 13 características en tiempo, 21 características en frecuencia y 50 de dominio de escala descritas en [1]. Las características que se obtendrán se muestran en la Figura 1.1.

Time-domain		Frequency-domain		Scale-domain	
ID	Feature name	ID	Feature name	ID	Feature name
f1	Mean	f29	Power	f57	Percentage of energy for D1
f2	Standard deviation	f30	Density of peaks above RMS ^d	f58	Percentage of energy for D2
f3	Variance	f31	2nd highest peak value	f59	Percentage of energy for D3
f4	Entropy ^a	f32	Freq. of 2nd highest peak	f60	Percentage of energy for D4
f5	Kurtosis	f33	3rd highest peak value	f61	Percentage of energy for D5
f6	Multiscale entropy (MSE) ^b	f34	Freq. of 3rd highest peak	f62	Percentage of energy for D6
f7	Time to reach the maximum peak	Scale-domain		f63	A6 RMS in time-domain
f8	RMS value	f35	A6 Max. peak in freq.-domain	f64	D1 RMS in time-domain
f9	Peak-to-peak value	f36	D1 Max. peak in freq.-domain	f65	D2 RMS in time-domain
f10	Peak-to-RMS ratio ^c	f37	D2 Max. peak in freq.-domain	f66	D3 RMS in time-domain
f11	Energy	f38	D3 Max. peak in freq.-domain	f67	D4 RMS in time-domain
f12	Zero-crossing rate	f39	D4 Max. peak in freq.-domain	f68	D5 RMS in time-domain
f13	Density of peaks above RMS ^d	f40	D5 Max. peak in freq.-domain	f69	D6 RMS in time-domain
Frequency-domain		f41	D6 Max. peak in freq.-domain	f70	A6 Peak-to-peak in time-domain
f14	Frequency of maximum peak	f42	A6 Freq. of max. peak	f71	D1 Peak-to-peak in time-domain
f15	Bandwidth of 90% energy ^e	f43	D2 Freq. of max. peak	f72	D2 Peak-to-peak in time-domain
f16	Entropy ^a	f44	D3 Freq. of max. peak	f73	D3 Peak-to-peak in time-domain
f17	Mean	f45	D4 Freq. of max. peak	f74	D4 Peak-to-peak in time-domain
f18	Standard deviation	f46	D5 Freq. of max. peak	f75	D5 Peak-to-peak in time-domain
f19	Variance	f47	D6 Freq. of max. peak	f76	D6 Peak-to-peak in time-domain
f20	Energy	f48	A6 Mean in freq.-domain	f77	A6 Peak-to-RMS ratio in time-domain ^f
f21	Kurtosis	f49	D1 Mean in freq.-domain	f78	D1 Peak-to-RMS ratio in time-domain
f22	Multiscale entropy ^b	f50	D2 Mean in freq.-domain	f79	D2 Peak-to-RMS ratio in time-domain
f23	Maximum peak in 10–20 Hz band	f51	D3 Mean in freq.-domain	f80	D3 Peak-to-RMS ratio in time-domain
f24	Freq. of max. peak in 10–20 Hz Band	f52	D4 Mean in freq.-domain	f81	D4 Peak-to-RMS ratio in time-domain
f25	Maximum peak in 20–30 Hz band	f53	D5 Mean in freq.-domain	f82	D5 Peak-to-RMS ratio in time-domain
f26	Freq. of max. Peak in 20–30 Hz Band	f54	D6 Mean in freq.-domain	f83	D6 Peak-to-RMS ratio in time-domain
f27	RMS value	f55	Mean energy of components ^g	f84	Mean energy of wavelet coefficients
f28	Peak-to-RMS ratio ^c	f56	Percentage of energy for A6 ^h		

^a Nonnormalized Shannon Entropy (Coffman and Wickerhauser, 1992) calculated using $E(x) = -\sum_i x_i^2 \log_2(x_i^2)$.
^b MSE at scale 1. MSE measures the degree of complexity of time series signals through Sample Entropy (Costa et al., 2005).
^c Ratio of largest absolute (i.e. L-Infinity norm) to root mean squared value.
^d Ratio of the number of peaks whose amplitudes are higher than the RMS value to the length of the signal.
^e Minimum frequency at which the cumulative energy is at least 90% of the total energy.
^f Mean energy of A6, D1, D2, D3, D4, D5 and D6 wavelet's signal components.
^g Energy fraction of wavelet's signal component with respect to total energy.

Figura 1.1: Resumen de características.Reproducido de [1].

Finalmente se tendrá un producto final demostrable que son las funciones .m del toolbox.

1.4. MARCO TEÓRICO

En la sección 1.4.1 proporciona una descripción de los eventos sísmicos y los tipos de eventos que son registrados manualmente en las bases de datos obtenidas del monitoreo continuo de los volcanes activos. Luego, en 1.4.2 se menciona el formato estándar SEED (Standard for the Exchange of Earthquake Data), en el cual se almacena la información que se obtiene de los sismógrafos en los centros de monitoreo. Mientras tanto, en las secciones 1.4.4 y 1.4.5 se presenta el procesamiento de señales y la selección de característica. En la sección 1.4.6 se presenta la herramienta utilizada para este trabajo y por último en las secciones 1.4.7 y se habla sobre el tSNE y el espectrograma.

1.4.1. MICROSISMOS VOLCÁNICOS

Los volcanes producen señales sísmicas debido a las perturbaciones que ocurren en el interior de estos, estas pueden estar racionadas con el transporte de fluidos como puede ser el magma, el agua y el gas, al interactuar estos fluidos con la roca sólida y también pueden presentarse por resonancias en las grietas o por desplazamiento del terreno.

Hay diferentes tipos de ondas sísmicas entre las principales tenemos las ondas de compresión y secundarias.

- **Ondas P:** Son el primer tipo de ondas de compresión, estas son conocidas como ondas primarias, es la primera en ser detectada por una estación sísmica. Es la más rápida de todas las ondas, su movimiento se da a través de la roca sólida y líquidos. [8].
- **Ondas S:** Estas ondas se conocen como ondas secundarias y son el segundo tipo de ondas que se desplazan a través de la tierra, su velocidad es menor que las ondas P. Su movimiento se da a través de la roca sólida. [8].
- **Ondas Superficiales:** Estas viajan solo a través de la corteza de la tierra, llega después de las ondas S y P y es fácil distinguirlas en un registro sismográfico. Las ondas superficiales son las responsables casi totalmente de los daños y la destrucción asociada a un terremoto. [9].

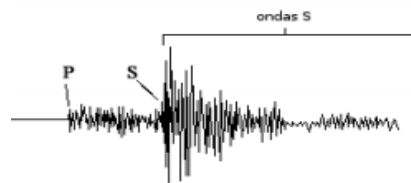


Figura 1.2: Ondas P, S y Superficiales vistas en un registro sismográfico [9].

El monitoreo continuo de los volcanes permite conocer los fenómenos que se producen en este para entender de mejor manera el comportamiento o los cambios en su actividad sísmica, y así poder dar alertas tempranas para la prevención de riesgos contra posibles erupciones que pongan en peligro zonas poblacionales.

1.4.1.1. Tipos de eventos sísmicos

Los volcanes activos presentan una gran cantidad de eventos sísmicos, los cuales son clasificados en base a la forma de su señal y su contenido espectral. Teniendo los siguientes eventos como los más recurrentes a presentarse.

Eventos Volcano Tectónicos (VT): Estos eventos están relacionados con el debilitamiento generado por la reacción de la roca a los cambios de tensión inducidos por el movimiento del magma, en el interior del volcán [3]. Son eventos de alta frecuencia con componentes de frecuencia mayores a 5Hz, y frecuencias dominantes entre 5 y 15 Hz [10].

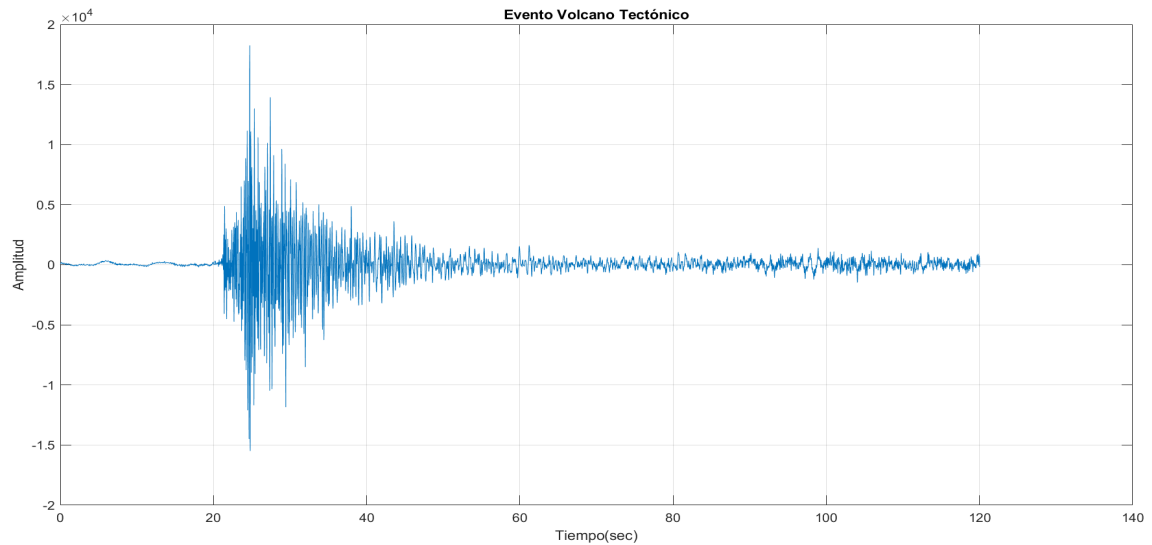


Figura 1.3: Ejemplo de un evento tipo VT.

Eventos de Largo Periodo (LP): Estos eventos están relacionados con la interacción de los conductos o grietas y los flujos de fluidos como el magma, presentándose cuando se tiene perturbaciones de presión en el fluido. Son eventos de baja frecuencia con un tiempo de duración por debajo de los 90s y con un espectro de frecuencia entre los 2 y 5Hz. Los eventos LP son indicadores importantes del estado del volcán [4, 10–12].

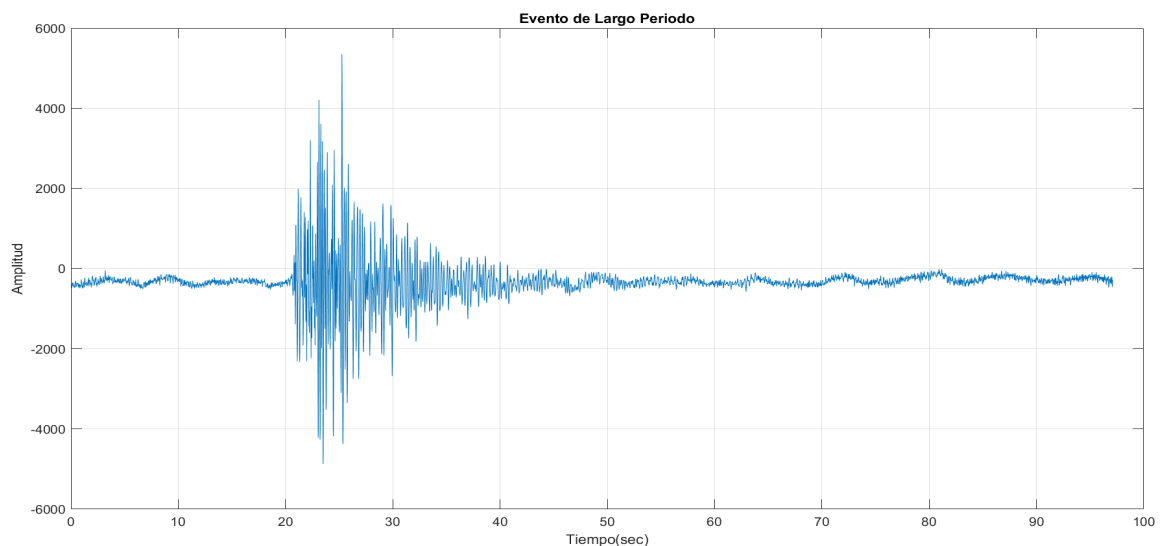


Figura 1.4: Ejemplo de un evento tipo LP.

Tremors (TR): Estos eventos en algunas regiones volcánicas pueden presentarse como la suma temporal de eventos LP, para estos eventos las perturbaciones de presión son continuas en tiempo e integran una zona más extensa, también pueden presentarse por explosiones sostenidas y con la desgasificación del magma. Este evento presenta una amplitud constante durante un largo tiempo, teniendo en cuenta el espectro de frecuencias podemos tener subgrupos de este evento [3, 13].

Se puede encontrar Tremors contenidos en altas frecuencias $>6\text{Hz}$, en frecuencias intermedias en un rango de 1 a 6Hz y para bajas frecuencias alrededor de los 0.5Hz [13].

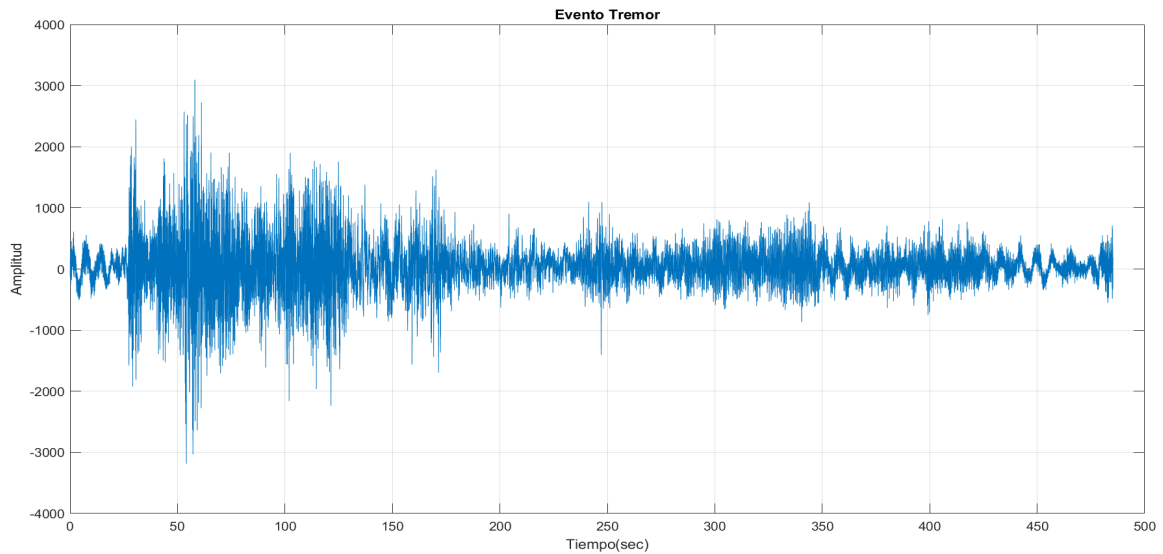


Figura 1.5: Ejemplo de un evento tipo Tremor.

Explosiones (EXP): Estos eventos, junto a los tremors, son las señales más características de un volcán que se encuentra en proceso eruptivo, se las puede reconocer por el incremento brusco de la amplitud de la señal y las frecuencias. Suelen presentarse debido a procesos de desgasificación al igual que en los tremors [13]. Las frecuencias se encuentran entre 0 y 10 Hz , estos eventos pueden alcanzar hasta 20Hz [10].

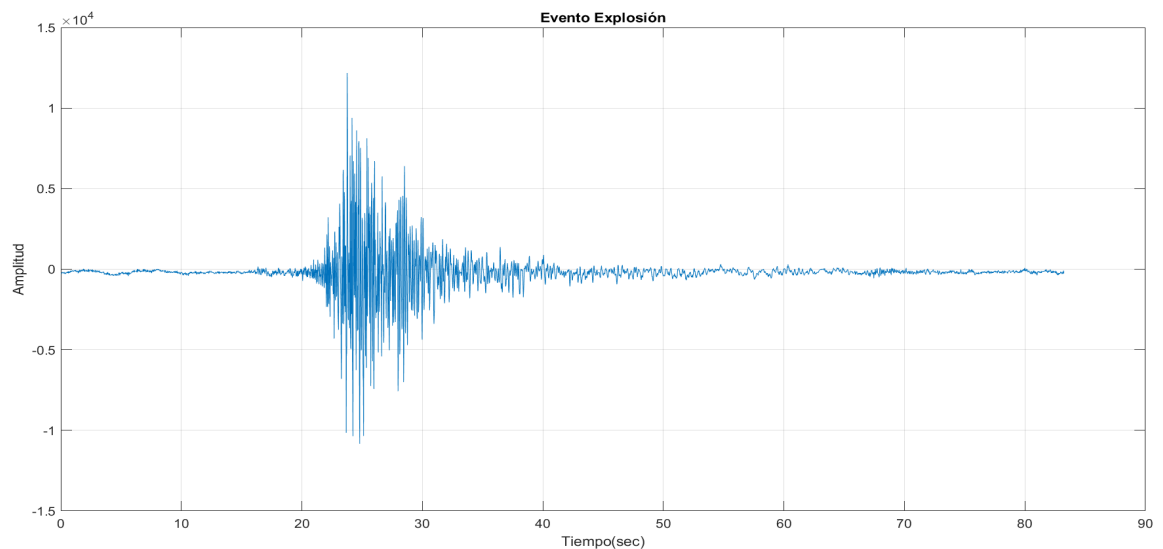


Figura 1.6: Ejemplo de un evento tipo EXPL.

Eventos de muy largo periodo (VLP): Estos eventos resultan de las fluctuaciones de presión debido al transporte inestable de masa en el sistema de descenso sub superficial. Tiene un rango de periodo de 0.2 a 2 s, estos eventos son de muy baja frecuencia el rango de fre-

cuencias se encuentra entre 0.01 y 0.5Hz [10, 14].

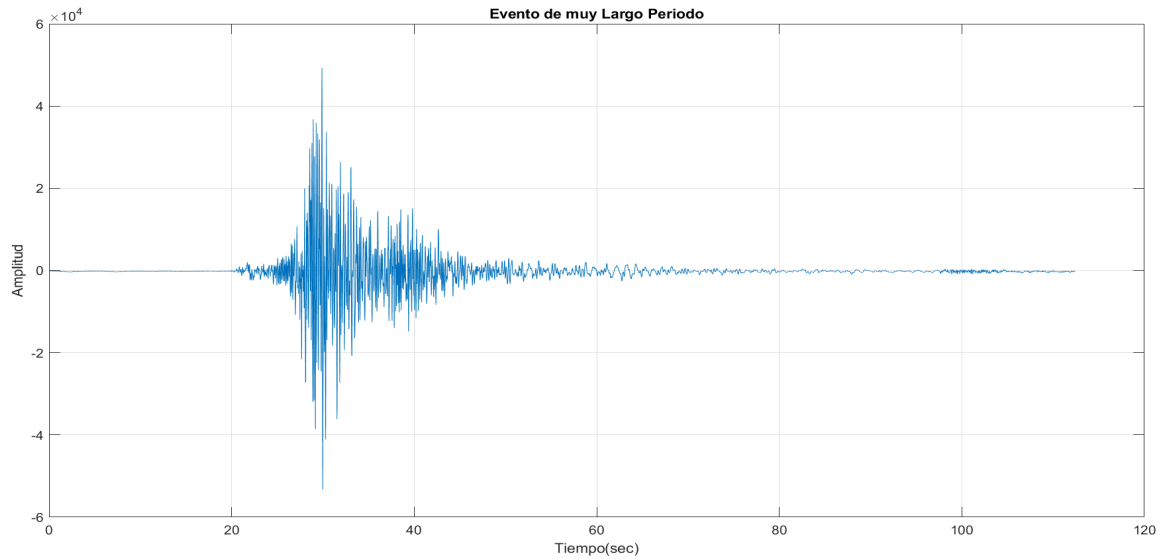


Figura 1.7: Ejemplo de un evento tipo VLP.

Eventos híbridos (HYB): Estos eventos aparecen por fragmentación de rocas y movimientos de fluidos al interior del volcán, se caracterizan por presentar una combinación entre los eventos Volcano Tectónico (VT) y de Largo Período (LP). Estos eventos sobrepasan los 10Hz [10, 15].

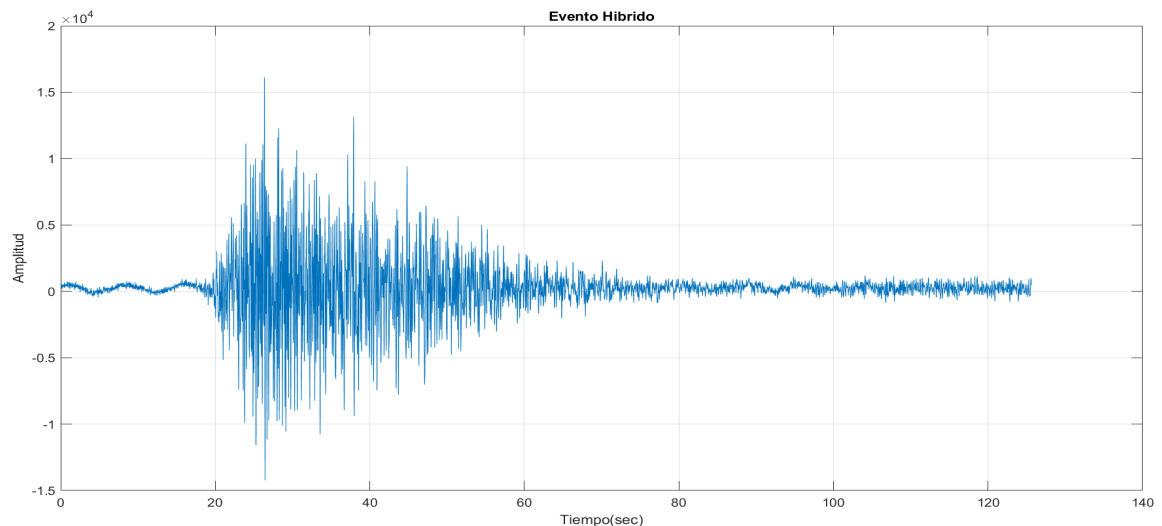


Figura 1.8: Ejemplo de un evento tipo HYB tomado de la base de datos usada.

1.4.2. STANDARD FOR THE EXCHANGE OF EARTHQUAKE DATA (SEED)

El estándar internacional SEED es un formato que se utiliza para el intercambio de datos sísmológicos digitales, fue diseñado para ser usado por la comunidad de investigación de terremotos, principalmente para el intercambio entre instituciones que manejan datos no procesados sobre el movimiento de la tierra. Los datos digitales son medidos en un punto

en el espacio y en intervalos de tiempo iguales. Este formato contiene valores temporales, metadatos o ambos. La parte que contiene solo valores temporales se conoce como mini-seed y en esta se encuentra los valores en intervalos de tiempo de las señales registradas por el sismógrafo [5, 16].

El formato SEED se puede utilizar para transferir datos desde una estación de procesado a un centro de recolección de datos, luego a un centro de gestión de datos y, finalmente, a un usuario final. Además, los centros de recopilación y gestión de datos están utilizando características del formato para almacenamiento y recuperación de datos [5].

Este formato utiliza cuatro campos para el encabezamiento de los datos (ver figura 1.9), los cuales nos permiten identificar y definir sus características.

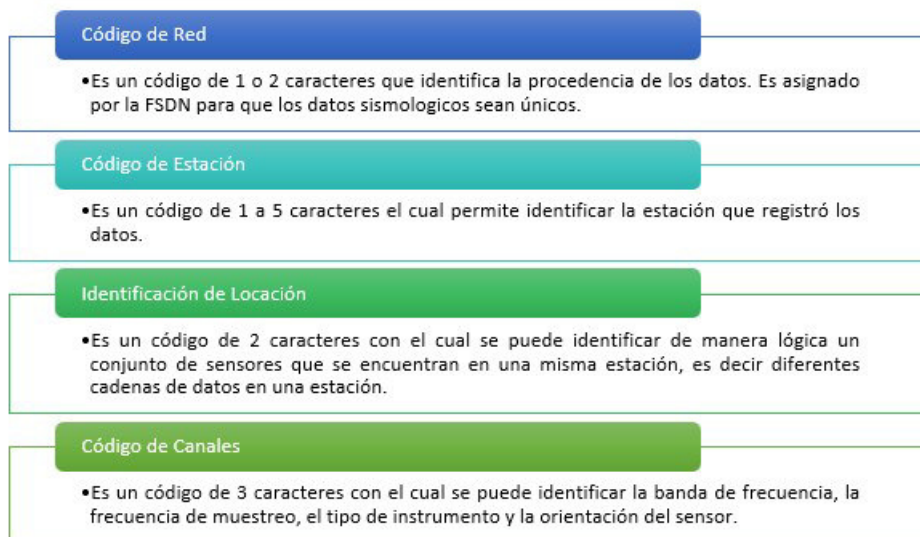


Figura 1.9: Campos de encabezado del formato SEED [17].

1.4.2.1. MiniSEED

MiniSEED es un subgrupo del estándar SEED, es usado para el intercambio solamente de series de tiempo, es más fácil de transmitir y se complementa con el datalessSEED, el cual contiene las coordenadas geográficas y la información de respuesta del instrumento. El volumen dataless puede tener un historial completo y detallado de una o varias estaciones [5, 17]. El formato SEED y MSEED es un conjunto de paquetes de datos llamados bloques que se utilizan para distribuir en un mismo formato los datos sísmicos, a diferencia de SEED que tiene bloques enumerados con códigos de 1 al 2000 en MSEED solo se usan los bloques 1000 y 1001. MSEED se compone de una cabecera fija de 48 bytes, seguido de uno o 2 bloques, el bloque que siempre va es el 1000 y opcionalmente puede ir el bloque 1001 y al final de todos los datos. En la tabla 1.10 se puede observar la cabecera del formato MSEED, en la cual ASCII y BYTE son tipos de datos de 8 bits, WORD es un tipo de dato de 16 bits, LONG de 32 bits, cuando tenemos la U esta se refiere a que no contiene signo y BTIME está

compuesto de 7 campos que son: año en formato tipo UWORD, día en formato tipo UWORD, horas en formato tipo UBYTE, minutos en formato tipo UBYTE, segundos en formato tipo UBYTE, alineación en formato tipo UBYTE y 0.0001 segundos en formato UWORD. [18]

	Campo	Tipo	Posición (bytes)	Longitud (bytes)
1	Número de Secuencia	ASCII	1-6	6
2	Indicador de calidad de datos	ASCII	7	1
3	Byte Reservado	ASCII	8	1
4	Código de Estación	ASCII	9-13	5
5	Indetificador de Localización	ASCII	14-15	2
6	Identificador de Canal	ASCII	16-18	3
7	Código de Red	ASCII	19-20	2
8	Tiempo de Inicio de Registro	BTIME	21-30	10
9	Número de Muestras	UWORD	31-32	2
10	Tasa de muestreo	WORD	33-34	2
11	Multiplicador de Tasa de Muestreo	WORD	35-36	2
12	Banderas de Actividad	UBYTE	37	1
13	Banderas de Entrada/Salida	UBYTE	38	1
14	Banderas de Calidad de Datos	UBYTE	39	1
15	Número de Blockettes Siguietes	UBYTE	40	1
16	Corrección de Tiempo (diezmilésimas de segundo)	LONG	41-44	4
17	Offset del Inicio de Datos	UWORD	45-46	2
18	Offset del Inicio del Primer Blockette	UWORD	47-48	2

Figura 1.10: Campos de cabecera del formato MSEED. Extraído de [19]

El Instituto Geofísico de la Escuela Politécnica Nacional usa el formato MiniSEED o MSEED para el almacenamiento de información de cada volcán, pero ya que este formato no fue diseñado para el procesamiento, se ayuda de diferentes softwares para procesar los datos sísmicos.

1.4.3. ALGORITMO STA/LTA

STA/LTA, es un algoritmo de disparo que se usa para la detección de señales sísmicas típicas como terremotos, explosiones subterráneas, etc. El cual cuando detecta un supuesto evento sísmico, inicia el registro y almacenamiento de la señal y se detiene cuando el algoritmo de activación declara un fin. Este algoritmo se puede traducir como razón promedio de corto plazo versus largo plazo, en el cual STA es la amplitud media en un intervalo temporal corto (t_{sta}) y LTA amplitud media en un intervalo temporal largo (t_{lta}), en comparación con un umbral de activación y desactivación. Hoy en día el algoritmo STA/LTA, es uno de los más usados y populares que otros algoritmos de disparo, debido a que otros algoritmos han demostrado ser poco manejables al tener sofisticados ajustes de los parámetros sísmicos y estar sujetos a errores en la práctica. [8, 20].

Para determinar la relación STA/LTA se define primero N_s como el número de muestras de la ventana corta, y N_l número de muestras de la ventana larga. F_s es la frecuencia de muestreo de la señal.

$$N_s = F_s * t_{sta} \quad (1.1)$$

$$N_l = F_s * t_{lta} \quad (1.2)$$

Definiendo el cálculo de STA/LTA como 1.6, en donde X_i se considera como cada valor del registro en función del tiempo.

$$STA = \frac{1}{N_s} \sum_{j=1-N_s}^i x_j^2 \quad (1.3)$$

$$LTA = \frac{1}{N_l} \sum_{j=1-N_l}^i x_j^2 \quad (1.4)$$

sí,

$$j \leq 0, x_j = \frac{x_1 + x_2}{2} \quad (1.5)$$

Finalmente, la razón STA/LTA esta dada por:

$$Radio_i = \frac{STA}{LTA} \quad (1.6)$$

1.4.3.1. Parámetros de detección

- **STA y LTA**

STA es un parámetro que mide el valor promedio instantáneo de una señal sísmica, si este valor es muy corto, el promedio de la señal sísmica no funcionará correctamente. La duración del STA es importante con respecto a los falsos disparos ya que al disminuir la duración de la ventana STA, el disparo se vuelve más sensible al ruido sísmico de tipo pico provocado por el hombre, la longitud de esta ventana debe ser menor a la duración del evento de menor duración esperado [21].

LTA es un parámetro que mide el valor promedio del ruido sísmico de amplitud media, este debe durar más que unos pocos periodos de fluctuaciones de ruido sísmico típicamente irregulares, es decir debe ser mayor a los periodos de ruido irregulares. Si la ventana LTA es de corta duración el umbral de activación debe ser de mayor amplitud, lo cual resulta beneficioso para evitar que pequeños movimientos como el del paso de un vehículo se han considerados eventos sísmicos [20, 21].

- **Trigger y dettrigger**

La relación STA/LTA, se compara continuamente con un parámetro conocido como nivel de umbral de activación (trigger), el cual depende del ruido presente en donde se encuentra el sensor y también de la sensibilidad que se requiera en determinados sistemas. Si la relación STA/LTA supera este umbral se declara la existencia de un evento y por tanto se almacenan los datos hasta que la relación cae por debajo del parámetro umbral de desactivación (dettrigger). [21].

El parámetro de dettrigger debe ser siempre menor al valor del trigger, este parámetro nos indica que el evento se ha terminado. Si el valor es muy alto se puede correr el

riesgo que no se capture la totalidad de un evento y si es muy bajo puede que se de un almacenamiento de gran longitud.

- **PEM y PET**

Los parámetros tiempo pre evento (PEM) y valores post evento (PET), son valores que agregan cierta cantidad de datos antes y después de alcanzar los valores trigger y dettrigger respectivamente. Al final, la señal se guarda con estos tiempos añadidos antes y después. Con un valor de PEM adecuado se pueden incluir las ondas P más un poco de ruidos sísmico anterior al evento, las cuales deben incluirse en el registro sísmico [20, 22].

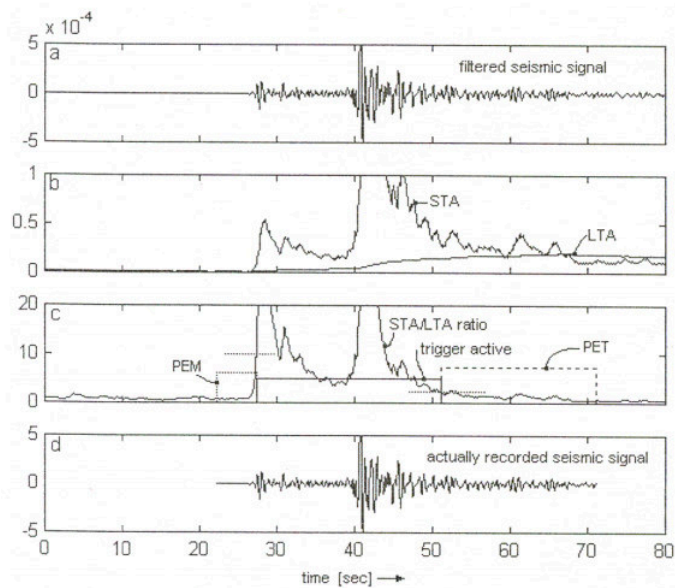


Figura 1.11: Función y variables de algoritmo STA/LTA [20].

Cada aplicación requiere un estudio del sitio sísmico, ya que solo con la experiencia práctica se puede determinar las configuraciones de disparo óptimas.

1.4.4. PREPROCESAMIENTO DIGITAL DE SEÑALES

El preprocesamiento de las señales ayuda a eliminar perturbaciones en la señal como el ruido, este es un proceso por el que las señales pasan antes de la extracción de características.

- **Filtrado:** Los filtros implementados en estas señales son de tipo no causal con distorsión de fase cero, el filtrado se usa para eliminar efectos instrumentales. Estos filtros difuminan los efectos hacia atrás en el tiempo, es decir, que su salida no depende únicamente de las entradas pasadas sino también de las futuras, y es seguro utilizar este tipo de filtro si las frecuencias de corte están lejos de las frecuencias de los eventos [1, 23]. La banda de la mayoría de eventos microsísmicos están en el rango de 0.3Hz a 15 Hz, estos valores dados por el IGEPN, aunque hay eventos como VLP

que se encuentran en un rango mas bajo 0.01 y 0.5Hz. Se debe tener cuidado con estos valores bajos ya que en 0.2 Hz existe un pico predominante relacionado con los microsismos marinos [1].

- **Normalización:** En este proceso se logra que las señales conserven sus características en frecuencia y en forma pero sus valores amplitud se transforman, cuando se normaliza al máximo se encuentran en un rango definido que esta entre el valor de [-1,1] que se obtiene dividiendo los valores para el valor absoluto máximo de la señal.

$$S_n = \frac{S_v}{\max|S_v|} \quad (1.7)$$

Donde S_n representa la señal normalizada, S_v la señal de cada evento volcánico. Existen otros tipos de normalización de los cuales se hablara en el siguiente capitulo. En la sección 2.5.

1.4.5. SELECCIÓN DE CARACTERÍSTICAS

Las características que se utilizan son las expuestas en el paper "Towards an Ecuadorian volcano seismic repository" [1], teniendo así las siguientes características para tiempo y frecuencia.

1.4.5.1. Características en tiempo

Para la obtención de las características en tiempo se utilizan características de naturaleza estadística. Para esto se utiliza funciones matemáticas en el dominio del tiempo.

- **Media:** La media se define como el promedio del valor de una señal. Siendo N el tamaño de la muestra.

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i \quad (1.8)$$

- **Desviación Estándar:** Indica cuan dispersos se encuentran los datos respecto a la media μ .

$$\delta = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2} \quad (1.9)$$

- **Varianza:** La varianza se obtiene elevando al cuadrado la desviación estándar.

$$\delta^2 = \frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2 \quad (1.10)$$

- **Entropía de Shanon no normalizada:** Se utiliza para intentar reducir o eliminar la incertidumbre en la información [24].

$$E(x) = - \sum_i x_i^2 \log_e(x_i^2) \quad (1.11)$$

- **Curtosis:** Mide la distribución de datos alrededor de una zona central de distribución [24].

$$k = \frac{1}{n} \sum_{i=1}^n \left[\frac{x_i - \mu}{\delta} \right]^4 \quad (1.12)$$

- **RMS:** El valor RMS se calcula como la raíz cuadrada de la media del cuadrado de los valores, expresa la magnitud de una cantidad variable. El nivel de raíz-media-cuadrada de un vector es x_{RMS} [25].

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N |x_n|^2} \quad (1.13)$$

- **Relación pico a RMS:** Es la relación entre el valor absoluto más grande el cual es la norma de L-infinito con el valor RMS. [24].

$$R_{p-rms} = \frac{\|X\|_{\infty}}{\sqrt{\frac{1}{N} \sum_{n=1}^N |x_n|^2}} \quad (1.14)$$

1.4.5.2. Características en frecuencia

Estas características se obtienen con la estimación de densidad espectral de potencia la cual puede ser obtenida de 3 formas utilizando mediante el periodograma el método de Busg y el método de Welch.

- **Periodograma:** El periodograma, es el estimador espectral mas simple, el periodograma de una secuencia consiste simplemente en su transformada discreta de Fourier. Para una señal X_n , muestreada el periodograma se define como:

$$\hat{P}(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} X_n e^{-j2\pi f \Delta t n} \right|^2 \quad (1.15)$$

Donde,

$$\frac{-1}{2 \Delta t} < f \leq \frac{-1}{2 \Delta t} \quad (1.16)$$

Δt , es el intervalo de muestreo y se multiplica por 2 para que se conserve la potencia total [26]. En Matlab existe la función periodogram la cual permite incluir una ventana a la señal de entrada con la cual se calcula el PSD, el uso de ventanas suaviza el inicio y fin de las ondas y es recomendado para el análisis de segmentos de la señal, este uso de enventanado se conoce como periodograma modificado y hace que la variación se reduzca, pero no desaparece totalmente [27].

- **Metodo de Welch:** Este método se utiliza para reducir la varianza del periodograma, la señal es dividida en segmentos por lo que para cada segmento se realiza el enventanado y se promedian las PSDs estimadas. Las funciones más usadas para enventanar toman los valores centrales de la secuencia que a los valores extremos, por lo que se tiene una pérdida de información y para compensar esto se introduce un solapamiento entre secuencias, esto protege contra la pérdida de información. [27, 28].
- **Método de Burg:** Burg plantea que los modelos autoregresivos (AR) deben ser tales que minimicen la suma de energías de error de predicción en adelante y retardo, definido en la ecuación 1.17 .

$$E_p = \sum_{n=p}^{N-1} |e_{pn}|^2 + \sum_{n=p}^{N-1} |b_{pn}|^2 \quad (1.17)$$

Siendo p el orden del predictor, e_{pn} el error de predicción en adelante y b_{pn} error de predicción en retardo. Sujeto a la restricción de que los parámetros satisfagan la recursión de Levinson, esto lo plantea para lograr un filtro estable. Para mayor comprensión de sus propiedades y sus condiciones de estabilidad, se puede revisar [29, 30].

El método Burg estima los coeficientes de reflexión y utiliza los coeficientes de reflexión para estimar los parámetros de AR recursivamente. Puede encontrar las relaciones de filtro de recursividad y celosía que describen la actualización de los errores de predicción hacia delante y hacia atrás. [31]

1.4.5.3. Características en tiempo-frecuencia

Para las señales que requerían ser representadas en diferentes escalas se utiliza la transformada de wavelet.

- **Transformada de Wavelet:** Esta transformada se puede utilizar para analizar en tiempo, series que contienen energía no estacionaria en muchas frecuencias diferentes. Una vez que se ha escogido la función Wavelet es preciso elegir un conjunto de escalas, si se trata de una Wavelet ortogonal se limita a un conjunto discreto de escalas y en el caso que no sea ortogonal se puede usar un conjunto arbitrario de escalas para construir una imagen más completa, es práctico poner las escalas como potencias fraccionarias de dos [32].

La transformada discreta de Wavelete (DWT), se compara con bancos de filtro discretos múltiples, es esencialmente comparable a un filtro uniforme, mediante un filtro pasa altos y pasa bajos se remueve componentes de altas o bajas frecuencias respectivamente, esto quiere decir que la señal se descompone en dos partes, por lo que una parte o región contiene las frecuencias bajas desde 0Hz hasta la mitad del valor mayor de la frecuencia de la señal y en la siguiente región se tiene desde la mitad de las frecuencias hasta el valor mayor de la señal, posteriormente cada región

se divide nuevamente de la misma manera creando niveles con cada división de la señal se muestra en la figura 1.12, donde S, es la señal, cA son los coeficientes de aproximación y cD coeficientes detallados.

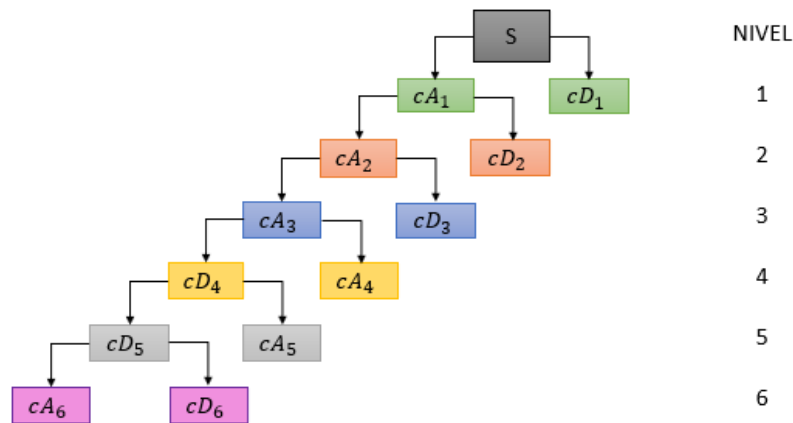


Figura 1.12: Árbol de descomposición Wavelet.

En los coeficientes detallados se encuentran las frecuencias altas y en los coeficientes de aproximación las frecuencias bajas. El tipo de Wavelet que se usa en este estudio son las de Daubechies, ya que la Wavelet madre de tipo 10 (db10), con un nivel máximo de descomposición entre 5 y 6, son las más adecuadas para trabajar con señales sísmicas [24].

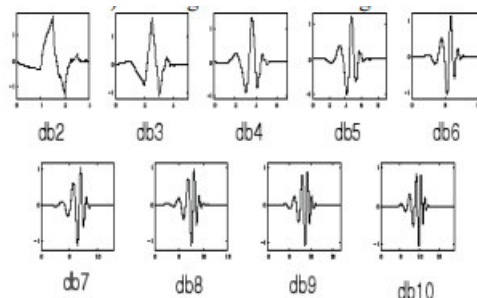


Figura 1.13: Familia de Wavelets Daubechies desde N hasta 10. [33].

Estas Wavelets son la función base y actúan como una función ventana, una Wavelet expandida es mejor para resolver componentes de frecuencia bajas de las señales con mal tiempo de resolución y una Wavelet encogida es mejor para resolver componentes de frecuencia altas de las señales con buen tiempo de resolución. La Wavelets Daubechies, son representadas como dbN, N representa los momentos de fuga, o el orden del Wavelet base, como se puede apreciar en la figura 1.13, mientras el N es mayor la Wavelet se torna mas suave.

1.4.6. SOFTWARE UTILIZADO

1.4.6.1. Matlab

Matlab, es un software en el cual se puede usar un lenguaje de programación de alto nivel desarrollado por MathWorks, usado para análisis de datos, desarrollo de algoritmos y creación de sistemas usado por ingenieros y científicos en todo el mundo. Tiene un entorno amigable con el usuario, se puede visualizar los datos mediante gráficas y cuenta con herramientas para crear diagramas personalizados e interfaces para C/C++, Java, .NET, Python, SQL, Hadoop y Microsoft Excel. Además, esta herramienta cuenta con toolboxes complementarios para diferentes áreas científicas y de ingeniería, como para el procesamiento de señales o imagen, entre otros. Este trabajo se realiza con la versión 2019b de Matlab, que presenta el siguiente entorno de trabajo 1.14.

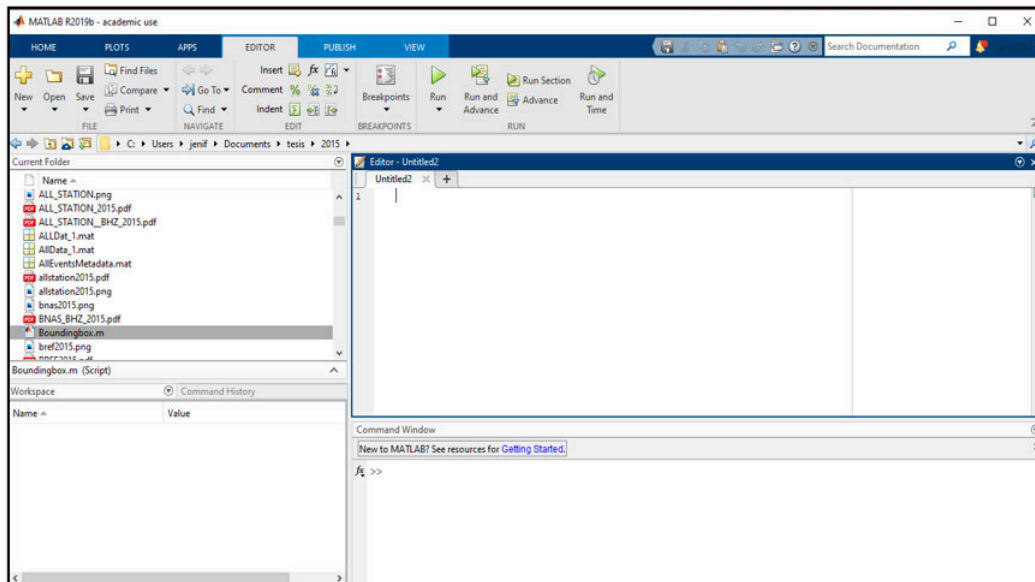


Figura 1.14: Entorno de trabajo Matlab.

Esta herramienta brinda funciones propias con las que se puede realizar el procesamiento de señales como el filtrado y la normalización, además poder medir, generar y analizar estas señales en el dominio del tiempo y la frecuencia, algunas de estas funciones se utilizan para los algoritmos implementados dentro del toolbox desarrollado en el presente trabajo.

1.4.7. t-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING

El tSNE es un algoritmo de reducción de dimensionalidad, se utiliza para la visualización de conjuntos de datos que presentan alta dimensionalidad. Los algoritmos de reducción de dimensionalidad identifican información redundante entre las características, para de esta manera comprimir los datos en un subespacio más pequeño [24]. Este algoritmo construye una distribución de probabilidad sobre parejas de muestras, de forma que muestras iguales o semejantes reciben una alta probabilidad de ser escogidas. La probabilidad se basa en la distancia entre puntos.

Como resultado del uso de tSNE, se obtiene mapas de datos que pueden ser de 2 o 3 dimensiones que pueden ser mostrados en una gráfica, el algoritmo no es lineal y se adapta a los datos subyacentes, realizando diferentes transformaciones en diferentes áreas. [24, 34].

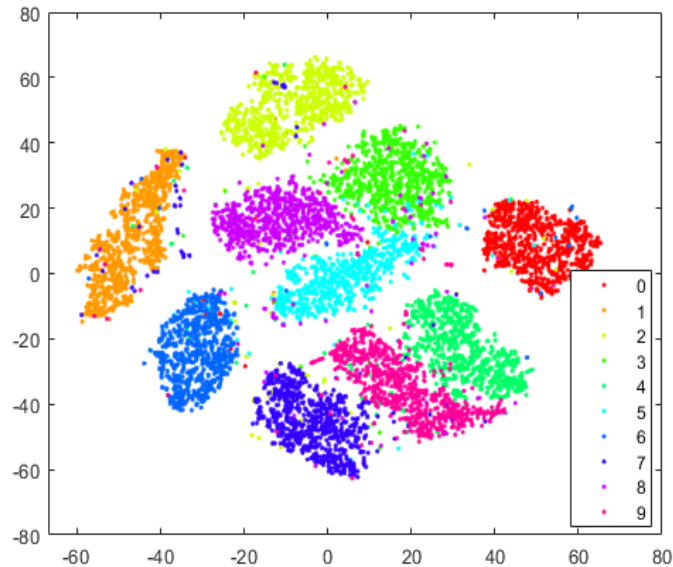


Figura 1.15: Ejemplo de tSNE. Obtenido de [35]

1.4.8. ESPECTROGRAMA

El espectrograma es una representación de forma visual del contenido de frecuencia de una señal, la intensidad de las frecuencias se muestra mediante niveles de colores a lo largo del tiempo. Se obtiene de la aplicación de la transformada de Fourier a la señal. Dependiendo de la ventana que se utilice para el análisis de la transformada de Fourier sera la resolución que tenga el espectrograma, si esta es muy grande se vera de manera más detallada pero el tiempo de cálculo será mayor y en el caso contrario con una ventana muy pequeña no será tan distinguible.

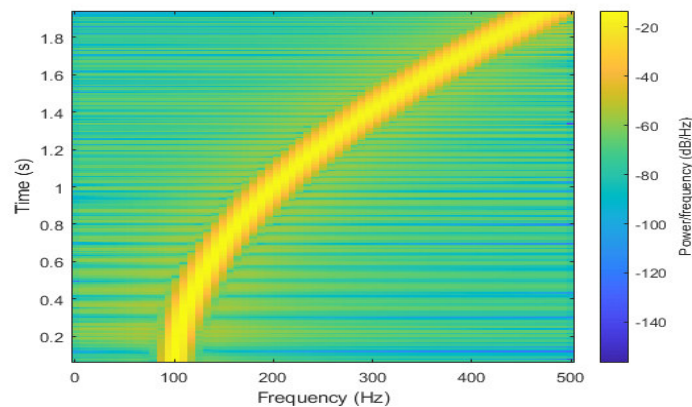


Figura 1.16: Ejemplo de espectrograma. Obtenido de [36]

2. METODOLOGÍA

En este capítulo se describe la implementación del toolbox. Para esto como primer paso se debe tener los datos de entrada en formato .mat, en la sección 2.1 se habla sobre las bases de datos que maneja el Instituto Geofísico de la Escuela Politécnica Nacional, teniendo así archivos .mseed que contienen la información del monitoreo diario de un volcán y también archivos en formato .xlsx, los cuales contienen información sobre los eventos sísmicos clasificados de manera visual por el instituto. En la sección 2.2 se indica cómo se realiza la lectura y adquisición de los archivos y con estos se procede a crear una metadata. En la sección 2.3 se habla sobre la extracción de los eventos contenidos en los archivos en formato .xlsx, los cuales contienen información sobre los eventos sísmicos clasificados de manera visual por el instituto, estos son los eventos que se buscan y extraen de la metadata como se explica en la sección 2.4 y en esta sección también se habla sobre la aplicación del detector STA/LTA con el cual se realiza una corrección de puntos de inicio y fin, después los eventos pasan por una depuración de eventos con el cual se tienen los eventos que se confirmaron o descartaron por el algoritmo. Terminados todos estos pasos se obtiene un archivo .mat el cual cuenta con la información respectiva de cada evento o eventos en cada estación y canal en el que fuese detectado. Una vez obtenido este archivo .mat en la sección 2.5 se habla sobre el preprocesamiento de la señales de los eventos confirmados teniendo así funciones de recorte, filtrado, normalización de la frecuencia de muestreo y normalización de la señal. Como último paso, en la sección 2.6 se trata sobre la extracción de características de tiempo, frecuencia y escala para cada evento confirmado. La metodología general se muestra en la siguiente Figura 2.1.

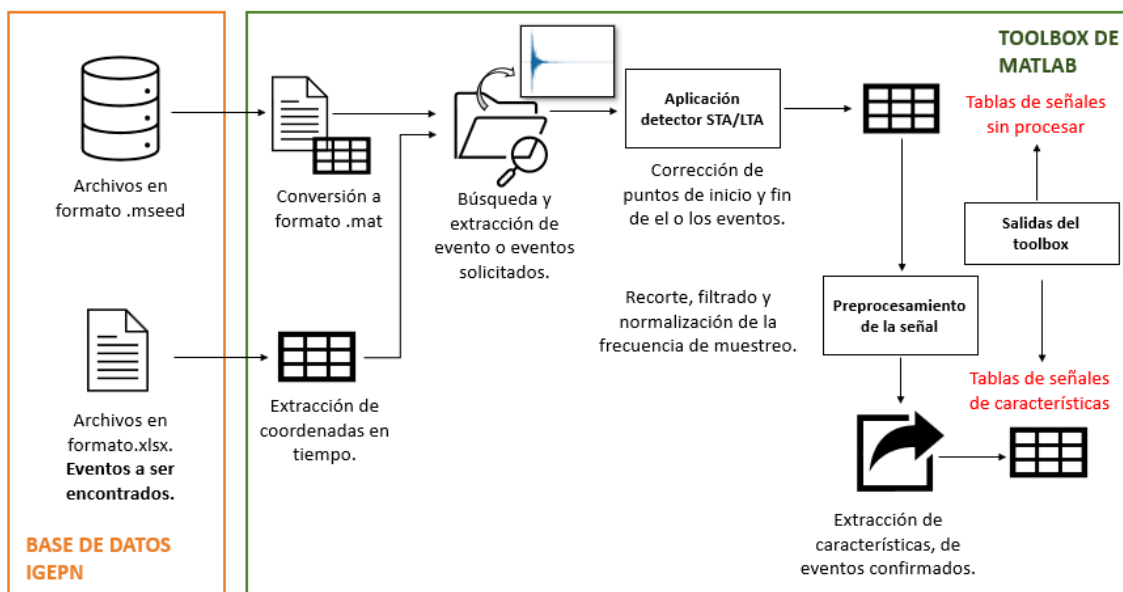


Figura 2.1: Funcionamiento de toolbox de matlab para procesamiento de señales sísmicas

En la tabla 2.1 se muestran las funciones implementadas para el funcionamiento del toolbox

y de las cuales se habla a profundidad en las siguientes secciones de este capítulo. En el Anexo A, se muestra una tabla con las funciones de entrada y salida para el toolbox.

Tabla 2.1: Descripción de las funciones implementadas en el toolbox de matlab para procesamiento de señales sísmicas.

Funciones	Descripción
seed_to_mat_conversion	Abre el directorio que contiene los archivos mseed lee estos archivos y devuelve los archivos .mat con un nombre único y almacenados en un directorio.
import_from_excel	Crea una tabla con los eventos registrados por el IGEPN.
remove_duplicates	Retira eventos que puedan estar duplicados en los registros del IGEPN.
search_and_extract	Busca en los archivos .mat, los eventos detectados por el IGEPN, una vez localizados extrae su data, puntos de inicio y fin, frecuencia de muestreo, tipo y duración para crea una tabla organizada por canales y estaciones.
trim_signal	Recorta la señal en los puntos de inicio y fin de los eventos.
remove_nan	Retira los valores de tipo NaN de cada señal.
normalize_sampling_rate	Normaliza todas las señales a una misma frecuencia de muestreo.
filter_signal	Filtra la señal, usando un filtro pasa banda de tipo butter o fir1 .
normalize_signal	Normalizar la señal con 3 diferentes tipos de métodos.
calculate_features	Calcula características en tiempo, frecuencia y escala de la señales.
signals_with_characteristics	Crea una tabla que contiene los eventos con sus respectivas características.

2.1. BASE DE DATOS IGEPN

Las bases de datos que se utiliza para el funcionamiento del toolbox son las obtenidas por el IGEPN. Este instituto se encarga del monitoreo de la actividad volcánica del Ecuador. Las bases de datos se componen de señales grabadas por estaciones sísmicas, las cuales pueden ser de banda ancha y de período corto, estas estaciones se encuentran desplegadas

en los flancos de los volcanes a ser monitoreados. Para las pruebas del toolbox, se utiliza los eventos registrados por el instituto en el año 2015, de diferentes estaciones que se encuentran localizadas en el volcán Cotopaxi, como se muestra en la figura 2.2, un ejemplo del despliegue de las estaciones.

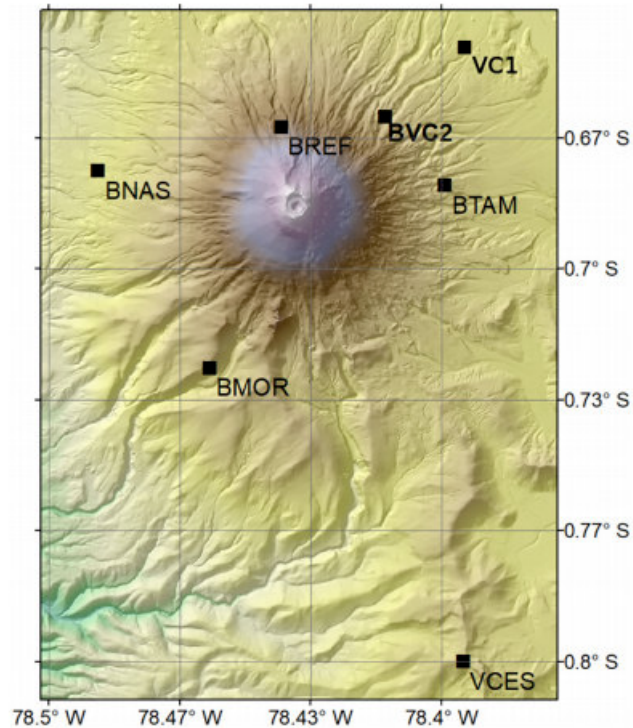


Figura 2.2: Estaciones desplegadas en el volcán Cotopaxi.

[1]

El IGEPN, aparte de manejar datos en el formato .mseed, trabaja con archivos en formato .xlsx los cuales contienen la información de cada evento que ha sido etiquetado de manera visual por el instituto, estos archivos proporcionan los siguientes parámetros: fecha y hora, fecha, tipo, diferencia de tiempo de arribo T(S-P), amplitud, coda, periodo, magnitud, energía y estación de donde se tomó esta información, como se menciona anteriormente, estos eventos fueron detectados de manera manual por lo que pueden presentar errores, por esta razón estos eventos pasan a ser extraídos para ser confirmados o descartados.

FECHA	TIPO	T(S-P)	Amp (cm)	CODA	Periodo (seg)	MAG	ENERG	ESTACIÓN
23/12/2000	HB	1,1	9,2	25	0,2	1,96	5,44E+14	
24/12/2000	HB	1,7	9,3	20	0,2	1,78	2,95E+14	
1/1/2001	HB	5	4,3	25	0,4	2,07	7,95E+14	
3/1/2001	HB		2,9	35	0,4	2,22	1,34E+15	
3/1/2001	HB	5,2	4,6	25	0,3	2,07	8,10E+14	
4/1/2001	HB	5	4,0	40	0,3	2,47	3,25E+15	
6/1/2001	HB		28,0	100	0,4	3,13	3,13E+16	
8/1/2001	HB	1,1	8,0	40	0,2	2,37	2,23E+15	
8/1/2001	HB	5	5,0	30	0,5	2,23	1,37E+15	

Figura 2.3: Datos en formato .xlsx con eventos de tipo HB detectados en el año 2015 en el volcán Cotopaxi por el IGEPN.

2.2. ADQUISICIÓN Y LECTURA DE ARCHIVOS .mseed

Los datos entregados por el IGEPN en formato .mseed se encuentran en un mismo directorio el cual contiene carpetas con nombres de la estación y el canal en el que se tomó los datos. Cada archivo cuenta con un nombre único, teniendo una descripción como la siguiente: EC.BNAS.BHZ.D.2015.106, en el cual se puede ver que cada campo brinda información sobre el archivo. EC representa el código de la red ecuatoriana, BNAS el nombre de la estación, BHZ.D el canal, 2015 el año y 106 es el día en el cual se tomó la grabación.

Para leer estos archivos se necesita una conversión al formato de Matlab, esto se lo realiza con la función `seed_to_mat_conversion`. Esta función toma el directorio en el que se encuentran los archivos de tipo .mseed, los convierte a .mat usando la función `ReadMSEEDFast`, la cual lee el archivo .mseed y devuelve la estructura del registro [7]. Luego de leer los datos en Matlab, la estructura dada por la función `ReadMSEEDFast` se rompe teniendo así para cada archivo un total de 16 variables, las cuales se presentan en la siguiente tabla 2.2.

Tabla 2.2: Variables del registro resultantes de la conversión de formato.

Variable	Descripción
network	Código de caracteres, utilizado para identificar la red ecuatoriana (EC).
station	Código de caracteres, utilizado para identificar la estación que registra los datos.
location	Código de caracteres, utilizado para identificar de forma exclusiva diferentes flujos de datos en una misma estación.
channel	Código de caracteres, utilizado para identificar el canal que registra los datos, la orientación del sensor.
dataquality	Indica el nivel de control de calidad de los datos que se ha aplicado, seteado con ("").
type	Indica el tipo de bloque, seteado con ("").
Continúa en la siguiente página	

Tabla 2.2: Variables del registro resultantes de la conversión de formato.

Variable	Descripción
startTime	Indica el tiempo de Inicio de Registro.
endTime	Indica el tiempo de Fin del Registro.
sampleRate	Indica la tasa de muestreo.
sampleCount	Indica la cuenta de muestreo.
numberOfSamples	Indica el número de muestras.
sampleType	Indica el tipo de dato del archivo, tipo de dato entero 'i', tipo de dato float 'f' y tipo de dato double 'd'.
data	Datos del registro.
dateTime	Datos del tiempo del registro de tipo double, esta estructura contiene los siguientes campos: <ul style="list-style-type: none"> • year: Año del registro. • month: Mes del registro. • day: Día del registro. • hour: Hora de inicio del registro. • minute: Minuto de inicio del registro. • second: Segundo de inicio del registro.
dateTimeString	Tiempo del registro en formato string.
matlabTimeVector	Vector de marcas de tiempo de Matlab para cada muestra de datos (generado por la función datenum).

Una vez leídos y extraídos los archivos, se almacenan con los mismos nombres que tienen en los registros .mseed. Para el caso en que se tengan estructuras mayores a 1, es decir que haya más de un registro por archivo, se procede a almacenar con los mismos nombres, pero añadiendo al final un subíndice indicando que es subconjunto del mismo archivo. Estos casos pueden darse debido a que los registros diarios pueden verse interrumpidos por cortes de energía y quedar pausados a cierta hora, teniendo así varios registros del mismo día a diferentes horas. También se debe considerar que el tamaño de número de muestras sea igual al de los datos para verificar que los archivos sean correctos para su almacenamiento. Al final, estos archivos son almacenados en un único directorio cuyo nombre es una variable de entrada en el toolbox (`dir_matlab`). A continuación, en la figura 2.4, se observa un ejemplo de un registro tomado el día 07 del 2015 por la estación BREF.

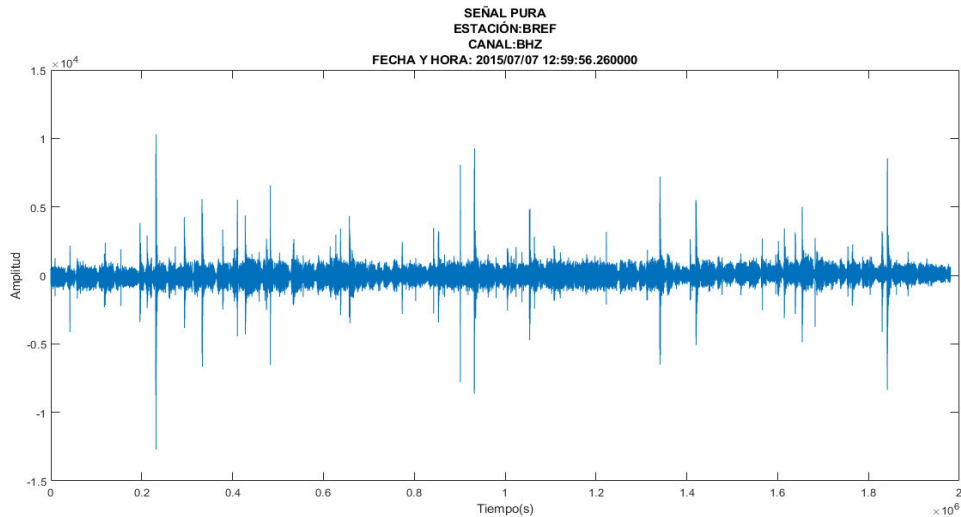


Figura 2.4: Señal sísmica extraída.

Cuando se tiene algún archivo repetido, el toolbox emite una advertencia con la siguiente descripción *'File already exists.'* y sobre escribe el archivo repetido. Si el archivo `.mseed` este dañado y no pueda ser abierto se muestra el siguiente mensaje de error *'Error occured during opening file ...'*.

Una vez terminado este proceso se crea una tabla con los archivos del directorio (`dir_matlab`). Como se muestra en la Figura 2.5. La cual contiene la metadata de cada archivo `.mseed` a excepción de los datos de la señal. En lugar de almacenar toda la data, se crea un campo en la tabla con el nombre de cada archivo, al cual se ingresa para extraer solo la data perteneciente al evento sísmico requerido. Esta metadata se crea para el proceso de búsqueda y localización de los eventos sísmicos que se explica en la sección 2.4.

1	2	3	4	5	6	7	8	9	10	11	12
file_name	start_date	number_of_samples	sample_rate	end_date	station	channel	network	location	dataquality	type	sampleType
'EC.BNAS..BDF.D.2015.098.mat'	08-Apr-2015 15:03:19	1610433	50	09-Apr-2015 00:00:07	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.099.mat'	09-Apr-2015 00:00:07	4319832	50	10-Apr-2015 00:00:03	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.100.mat'	10-Apr-2015 00:00:04	4319874	50	11-Apr-2015 00:00:01	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.101.mat'	11-Apr-2015 00:00:02	4320138	50	12-Apr-2015 00:00:04	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.102.mat'	12-Apr-2015 00:00:04	4319865	50	13-Apr-2015 00:00:01	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.103.mat'	13-Apr-2015 00:00:02	4322706	50	14-Apr-2015 00:00:56	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.104.mat'	14-Apr-2015 00:00:06	4320016	50	15-Apr-2015 00:00:06	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.105.mat'	15-Apr-2015 00:00:06	4289218	50	15-Apr-2015 23:49:50	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.106.mat'	16-Apr-2015 00:00:00	4320213	50	17-Apr-2015 00:00:04	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.107.mat'	17-Apr-2015 13:25:47	1902828	50	18-Apr-2015 00:00:03	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.108.mat'	18-Apr-2015 00:00:04	4291706	50	18-Apr-2015 23:50:38	BNAS	BDF	EC	''	''	''	'i'
'EC.BNAS..BDF.D.2015.109.mat'	19-Apr-2015 00:00:08	4287575	50	19-Apr-2015 23:49:19	BNAS	BDF	EC	''	''	''	'i'

Figura 2.5: Metadata de archivos `.mseed` del volcán Cotopaxi grabados en 2015 por el IGEPN.

2.3. COORDENADAS EN TIEMPO IGEPN

Como se habló anteriormente al inicio de este capítulo en la sección 2.1, el instituto maneja una base de datos en formato Excel en la cual se encuentran los eventos que han sido tomados manualmente. Estas coordenadas en tiempo nos brindan la información y la posición

en la que se encuentra un presunto evento sísmico. Son extraídas del archivo .xlsx a Matlab con la función `import_from_excel`, de manera que se obtiene una tabla igual a la contenida en el formato Excel. Para realizar la lectura de estos datos se utiliza la función de Matlab `readtable`, la cual permite la lectura de archivos en varios formatos, incluido el formato .xlsx. Se debe tomar en cuenta que todos estos archivos deben manejar un mismo orden en su estructura siendo esta la siguiente: `StartTime`, `Date`, `Type`, `TSP`, `Amplitude`, `Coda`, `Period`, `Magnitud`, `Energy` y `Station`, esto debido a que el toolbox lee los datos y crea el archivo .mat manteniendo este orden. Como la que se muestra en la figura 2.6. Posteriormente a la lectura y conversión de este archivo de formato .xlsx a .mat se toma algunos de estos datos para la localización y verificación del evento siendo estos:

- `StartTime`: Fecha y hora del inicio del evento.
- `Type`: Tipo de evento.
- `Coda`: Duración del evento en segundos.

1	2	3	4	5	6	7	8	9	10
StartTime	Date	Type	TSP	Amplitude	Coda	Period	Magnitud	Energy	Station
23-Dec-2000 22:43:00	23-Dec-2000 00:00:00	HB	1.1000	9.2000	25	0.2000	1.9569	5.4364e+14	<undefined>
24-Dec-2000 19:20:00	24-Dec-2000 00:00:00	HB	1.7000	9.3000	20	0.2000	1.7800	2.9508e+14	<undefined>
01-Jan-2001 21:03:00	01-Jan-2001 00:00:00	HB	5	4.3000	25	0.4000	2.0668	7.9457e+14	<undefined>
03-Jan-2001 16:20:00	03-Jan-2001 00:00:00	HB	NaN	2.9000	35	0.4000	2.2181	1.3403e+15	<undefined>
03-Jan-2001 20:51:00	03-Jan-2001 00:00:00	HB	5.2000	4.6000	25	0.3000	2.0724	8.1019e+14	<undefined>
04-Jan-2001 09:14:00	04-Jan-2001 00:00:00	HB	5	4	40	0.3000	2.4750	3.2546e+15	<undefined>
06-Jan-2001 04:37:00	06-Jan-2001 00:00:00	HB	NaN	28	100	0.4000	3.1300	3.1261e+16	<undefined>
08-Jan-2001 03:31:00	08-Jan-2001 00:00:00	HB	1.1000	8	40	0.1500	2.3651	2.2267e+15	<undefined>
08-Jan-2001 21:03:00	08-Jan-2001 00:00:00	HB	5	5	30	0.5000	2.2251	1.3730e+15	<undefined>
21-Feb-2001 09:52:00	21-Feb-2001 00:00:00	HB	NaN	6.5000	45	0.2000	2.4364	2.8486e+15	<undefined>

Figura 2.6: Tabla extraída de el formato.xlsx en Matlab con eventos de tipo HB detectados en el año 2015 en el volcán Cotopaxi por el IGEPN.

Esta tabla extraída del archivo .xlsx entra a la función `remove_duplicates`, en la cual se remueven los eventos que se encuentren duplicados. Este proceso se realiza comparando los valores de las primeras 6 columnas y encontrando los índices de los valores únicos, como se observa en el segmento de código 2.1. Al final, se ordena de mayor a menor las fechas del evento con la función de matlab `sort`.

Segmento de código 2.1: Remover Duplicados

```

1 function t_from_excel=remove_duplicates(t_from_excel)
2 w=t_from_excel(:,1); %Read the first colum.
3 p=t_from_excel(:,6); %Read the first 6 colums.
4 l=[p w];
5 [a,idxuni]=unique(l(:,1:2),'rows'); %idx unique events.
6 s=sort(idxuni); %sorts the vector elements in ...
   ascending .
7 t_from_excel=t_from_excel(s,:);

```

2.4. BÚSQUEDA Y EXTRACCIÓN DE EVENTOS

En este proceso se busca localizar y extraer los eventos de manera automática, entendiendo que esto se refiere a que se extrae de manera automática de las grabaciones diarias que guarda el IGEPN, la data de los eventos que se encuentran como coordenadas de tiempo en los archivos en formato Excel. De manera que solo se ingresa el nombre del directorio donde se almacenan las grabaciones y el nombre del archivo Excel que contiene los eventos y de esta manera se obtiene los eventos en cada canal y estación en los cuales se detecta. Para realizar esto se utiliza la función `search_and_extract`. De la metadata se obtiene toda la información de los canales y estaciones existentes y con estos se va creando la tabla final con los eventos sísmicos. La información de cada evento se presenta en la tabla de la siguiente manera:

- Las primeras 4 columnas son siempre las mismas: Event, Coda, Station y Type, posterior a esto las siguientes columnas son variables ya que dependen de las estaciones que hayan sido detectadas y a continuación de éstas, se tiene 2 columnas más, las cuales son Voting_Detection e Event_Identifier.
- Dentro de cada estación se encuentra una tabla de 1xN, en la cual las primeras columnas son variables ya que dependen de los canales que se hayan detectado y a continuación, se tiene 3 columnas más, que son: Start_Point_all, End_Point_all y Sample_Rate_all.
- Por último, dentro de cada canal se encuentra una tabla que siempre es de tamaño de 1x7 la cual contiene Name_filemat, Station, Channel, Data, StartPoint, EndPoint y Sample_rate.

1	2	3	4	5	6	7	8	9	10
Event	Coda	Station	Type	BNAS	BREF	BTAM	BVC2	Voting_Detection	Event_Identifier
01-Jul-2015 01:13:44	106.1000	<undefined>	VLP	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'f64903e695654f2992b013af927dd7ff'
02-Jul-2015 04:37:48	83.5999	<undefined>	VLP	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'df8f1baa945d7e2197d6bfb58c95c7df'

1	2	3	4	5	6	7
BDF	BHE	BHN	BHZ	Start_Point_all	End_Point_all	Sample_Rate_all
1x7 table	1x7 table	1x7 table	1x7 table	1074	2938	50

1	2	3	4	5	6	7
Name_filemat	Station	Channel	Data	StartPoint	EndPoint	Sample_rate
"EC.BREF..BHZ.D.2015.183.mat"	BREF	BHZ	1x10180 double	1074	2742	50

Figura 2.7: Ejemplo de estructura de una tabla total, para eventos de tipo VLP detectados en el volcán Cotopaxi en el año 2015.

La Figura 2.7 muestra un ejemplo de la estructura del archivo .mat final, donde se observa 4 estaciones: BNAS, BREF, BTAM y BVC2. Luego al ingresar en la estación BREF, se vi-

sualiza 4 canales los cuales son: BDF, BHE, BHN y BHZ. La obtención de cada uno de los parámetros y valores con los que se va a llenar las tablas finales, son explicados a continuación en esta sección.

Del archivo que contiene las coordenadas de tiempo se toma los datos de Event, Coda, Station y Type para ser mostrados en las primeras 4 columnas de la tabla final. Con el dato de Event se realiza una búsqueda en toda la metadata usando la información de la fecha y hora de inicio del evento (startdate) y la información de la fecha y hora de fin del evento (enddate), campos contenidos en la metadata. Con la función de Matlab isbetween, se localiza los eventos y una vez detectados se procede a buscar y extraer la señal de cada uno. En caso de ausencia de los eventos se coloca un indicador NaN.

Para la extracción de la seña del evento sísmico al inicio y final de cada señal se usa una protección en caso de que los eventos estén adelantados o retrasados al Event dado por el IGEPN. Por lo que se da una ventana de 20 segundos antes y 20 segundos después al evento sísmico detectado.

Una vez obtenidos los puntos de inicio y fin del evento se extrae la señal. Cuando los puntos iniciales o finales superan el inicio o fin de cada grabación, es decir de la los datos de la señal almacenada en los archivos .mat se realiza un relleno con NaN, proceso implementado en una función anidada llamada NaNPadding. Teniendo los siguientes posibles 3 casos:

- Relleno de NaN al inicio de la señal.
- Relleno de NaN al final de la señal.
- Relleno de NaN al inicio y al final de la señal.

Aparte de la extracción de la señal, se obtiene también la información de estación y canal de cada evento sísmico detectado datos que van llenando los campos de la tabla final. Para llenar los campos StartPoint y EndPoint, se realiza una corrección de puntos de inicio y fin puesto que los eventos que se encuentran en el archivo .xlsx son obtenidos de manera manual y con el uso de un detector se puede tener de manera más precisa el momento en que empieza a ocurrir el evento. Para esto se utiliza el detector STA/LTA descrito en el capítulo anterior en la sección 1.4.3 , sin embargo, previo al paso por el detector, la señal necesita pasar por un preprocesamiento, todo esto es mencionado a continuación.

2.4.1. DETECTOR STA/LTA

Dentro de la función search_and_extract, se utiliza un algoritmo de disparo STA/LTA con el fin de revisar cada evento dado por el IGEPN. De esa forma se puede ver si la señal que se extrae con las cordenadas de tiempo, contiene un evento o no, y de esta manera obtener los puntos de inicio y fin del evento.

Para la aplicación del algoritmo de disparo STA/LTA, se debe eliminar la componente DC para evitar un mal funcionamiento [20]. Debido a esto los datos de la señal primero son filtrados por un filtro butterworth pasa bandas de orden 5 con una banda de paso entre 0.3 y

12Hz, valores que fueron dados por el IGEPN debido a que las señales típicas de un volcán se encuentran por lo general en este rango de frecuencia. Luego pasan por un filtrado digital `filtfilt` ya que reduce el ruido en la señal y se conserva el mismo tiempo que la señal original ya que es de fase cero. Por último, los datos son normalizados utilizando una normalización al máximo, para esto se retira la media de la señal y se lo divide para el máximo valor absoluto de la señal.

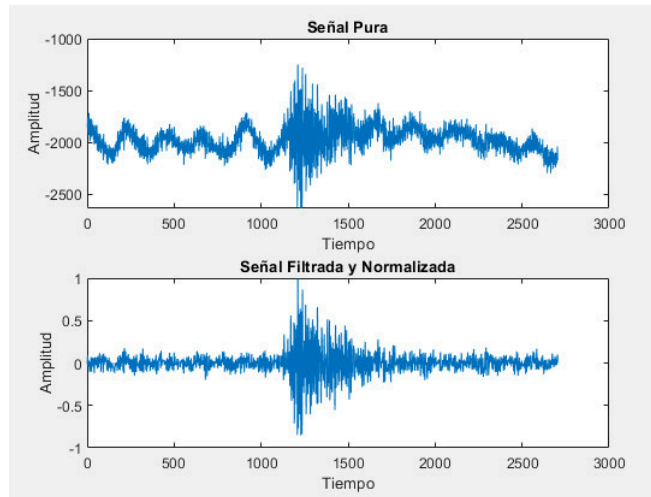


Figura 2.8: Señal de un evento tipo HB, filtrada y normalizada.

Una vez realizado estos dos procesos la señal resultante ingresa a la función `sta_lta`, en la cual se obtiene los puntos de inicio y fin de cada evento detectado. El detector para realizar la detección de estos puntos utiliza los siguientes parámetros de detección: `sta`, `lta`, `th_on`, `th_off`, `skip_int`, `min_dur`, `pem` y `pet`, descritos previamente junto al algoritmo STA/LTA en la sección 1.4.3, de este trabajo.

Para la implementación del detector se utiliza como referencia el algoritmo `sta_lta2`, descrito en [37]. Los valores de entrada del detector STA/LTA varían para cada parámetro dependiendo el tipo de evento que se desee hallar. Se tiene valores típicos que pueden ser usados para cada parámetro, los mismos que se encuentran descritos en la tabla 2.3 [20].

Tabla 2.3: Valores típicos para el algoritmo STA/LTA.

Parámetros	Valores
<code>sta</code>	Una duración de 1 o 2 segundos.
<code>lta</code>	Una duración de 60 segundos es un valor inicial común pero puede ser mas corta (si se desea excluir eventos regionales emergentes) o más larga para eventos regionales distantes.
Continúa en la siguiente página	

Tabla 2.3 – Continuación de la página anterior

Parámetros	Valores
th_on	Un nivel de umbral de activación de 4 es común para un sitio sísmico promedio, en sitios menos favorables con ruido sísmico un valor alrededor de 8 o más es adecuado, en el caso de estaciones sin ruido humano se puede usar valores menores
th_off	Un nivel de umbral de desactivación de 2 a 3 es común un lugares sísmicos tranquilos, y para lugares más ruidosos, valores más altos.
skip_int	El valor de salto se lo mantendrá en 0 segundos, debido a que el toolbox revisa evento por evento por lo que en la mayoría de caso solo tenemos una señal sísmica.
pem	El valor de memoria previa al evento debe superar el tiempo S-P máximo qu se prevee en los registros sísmicos. En el peor de los casos es igual a una duración de la ventana STA.
pet	El valor de duración de memoria después del evento depende principalmente de la aplicación, si la coda es importante y el evento es regional o largo se puede seleccionar un PET largo sino se puede elegir un valor corto para eventos locales, para los regionales un valor aceptable puede ser de 30 segundos y en la regional de 60 a 90 segundos.

Una vez ingresados los parámetros en el detector se calculan los puntos de inicio y fin del evento y estos valores son los que se almacenan en los campos StartPoint y EndPoint como muestras en tiempo. En el caso de no hallar un evento se asigna el valor de NaN, a estos campos. En la siguiente figura 2.11, se muestra un evento que ha pasado por el detector STA/LTA, contenido dentro de un bonding box, el cual está creado con StartPoint y EndPoint.

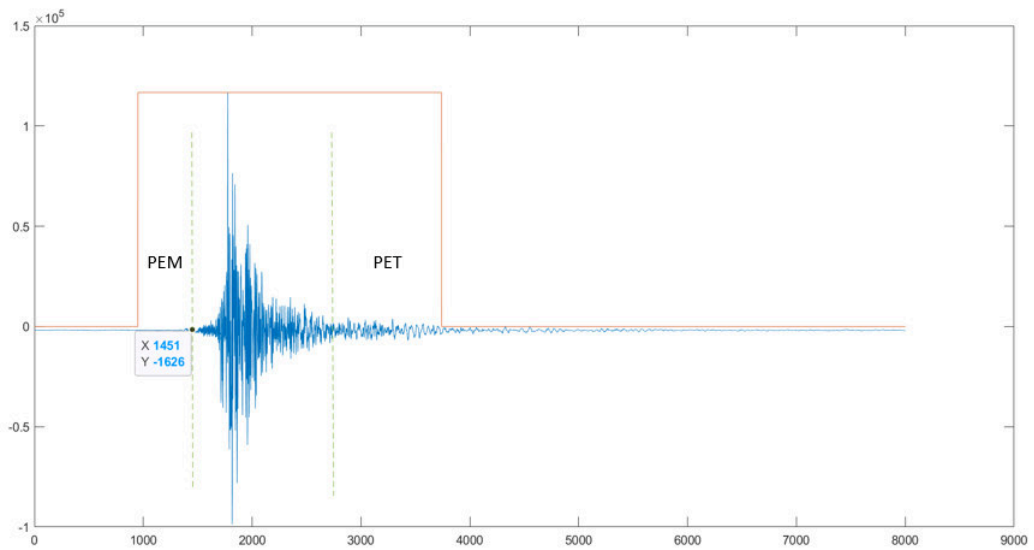


Figura 2.9: Evento tipo VLP tipo 2, después de pasar por detector STA/LTA.

Se debe aclarar que las señales pasan por este preprocesamiento debido a la implementación del detector STA/LTA, y una vez terminado este proceso en la primera tabla resultante del toolbox, los datos muestran la señal pura para que posteriormente estos datos se usen para cualquier propósito que el usuario desee. También se debe mencionar que las etapas de preprocesamiento para la extracción de características que se mencionan en la sección 2.5, incluyen algunos de estos procesos.

2.4.2. DESCARTE DE EVENTOS

Se crea un algoritmo para confirmar o descartar el evento detectado, el cual es implementado de acuerdo a las indicaciones dadas por el PhD. Mario Ruiz investigador del IGEPN.

Algorithm 1 Descarte de eventos

Input: Start _Point_all (int): Punto de inicio del evento para todas las estaciones.

Stations (categorical): Estaciones que detectaron el evento.

Output: Booleano : True or False. True: Confirmado. False: Descartado.

```
for Station: Estaciones diferentes a estación de referencia do
  if Station=='BTAM' then
    if Diferencia punto inicial respecto estación de referencia >2 segundos then
      Estación encontrada ++;
    end if
  end if
  if Station=='BNAS' then
    if Diferencia punto inicial respecto estación de referencia >2 segundos then
      Estación encontrada ++;
    end if
  end if
  if Station== otherwise then 'Add station'
  end if
end for
if Estación encontrada >1 then
  Confirmado
else
  Descartado;
end if
```

Se tiene un punto de inicio y fin que es general para cada estación los cuales se muestran en los campos Start_Point_all y End_Point_all respectivamente. Para hallar el valor del punto de inicio de cada estación (Start_Point_all) se toma del canal BHZ, por lo que es indispensable que este canal esté presente. Para el punto final de cada estación (End_Point_all), se toma el punto final mayor de todos los canales de las estaciones. Después, se usan estos valores para descartar o confirmar eventos con la función discarded_events.

Con esta función se compara la diferencia de tiempos iniciales de cada estación, y si al menos una estación respecto a otra de referencia se encuentra dentro de una ventana de 2 segundos, este evento se considera como confirmado, caso contrario el evento se considera como descartado y este resultado es visualizado en la columna Voting_detection presente en la primera tabla final. Cada evento confirmado pasa por un proceso de preprocesamiento como se observa en la sección 2.5 para poder extraer sus características.

En la figura 2.10, se muestra en el primer gráfico un evento descartado debido a que el tiempo de diferencia entre puntos inicial de 2 estaciones BREF Y BNAS es mayor a 2 segundos, en el segundo gráfico se observa un evento confirmado tiempo de diferencia entre puntos iniciales entre 2 estaciones BREF y BTAM es menor a 2 segundos.

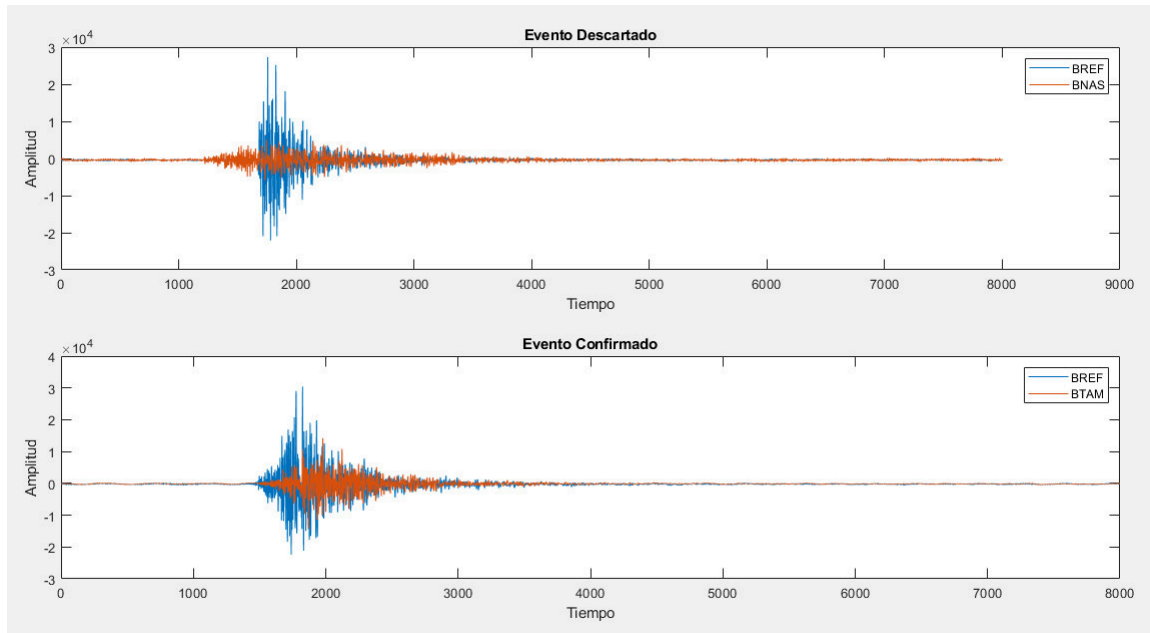


Figura 2.10: Eventos de tipo VLP tipo 2, primer gráfico muestra un evento descartado y el segundo gráfico un evento confirmado.

Al final, para terminar esta primera etapa se obtiene una tabla con las variables que se muestran a continuación en la tabla 2.4 que contiene las señales sin procesar.

Tabla 2.4: Variables del registro.

Variable	Descripción
Event	Fecha y hora de inicio del evento.
Coda	Duración del evento en segundos.
Station	Código de estación de donde los datos contenidos en el archivo Excel fueron tomados por el IGEPN .
Type	Tipo de evento detectado por el IGEPN.
Continúa en la siguiente página	

Tabla 2.4: Variables del registro.

Variable	Descripción
Name of the stations	<p>Código de las estaciones donde se realizó la búsqueda de eventos, puede ser 1 o varias columnas dependiendo el número de estaciones, dentro de este campo hay una tabla que contiene los siguientes campos:</p> <ul style="list-style-type: none"> • Name of the channels: Código de los canales donde se realiza la búsqueda de eventos, en su interior se encuentra una tabla que contiene los siguientes campos: <ul style="list-style-type: none"> • Name_filemat: Nombre del archivo donde se encuentra el evento. • Station: Estación donde se halla el evento. • Channel: Canal donde se halla el evento. • Data: Señal que se extiende 20s antes y después del evento detectado. • StartPoint: Punto inicial del evento en muestras con respecto a la variable Data. • EndPoint: Punto final del evento en muestras con respecto a la variable Data. • Sample_Rate: Frecuencia de muestreo en Hz. • Start_Point_all: Punto de inicio en muestras para todos los canales. • End_Point_all: Punto de fin en muestras para todos los canales. • Sample_Rate_all: Frecuencia de muestreo para todos los canales en Hz.
Voting_ detection	Indica si el evento es confirmado o descartado.
Event_Identifier	Muestra un Id único del evento.

Se añade al final la columna con el identificador de evento (Event_Identifier). Esto con el fin de que cada evento pueda ser encontrado solo con el identificador. Este ID para cada evento es generado con el algoritmo GetMD5 cuya función se encuentra ya implementada en Matlab [38], con el cual se obtiene un array de 128 bits y de esta manera se puede identificar posteriormente cada evento.

2.5. PREPROCESAMIENTO DE LA SEÑAL

Las señales de los eventos confirmados pasan por funciones de preprocesamiento, antes de que se extraiga sus características.

2.5.1. RECORTE DE LA SEÑAL

Las señales son recortadas en los puntos de inicio y fin de cada estación siendo estos `Start_Point_all` y `End_Point_all`, valores que son tomados de la tabla de las señales sin procesar. Antes de entrar a esta función de recorte (`trim_signal`), se confirma que existen los puntos de `Start_Point_all` y `End_Point_all` para cada evento. Esta confirmación se realiza ya que, al ser estos puntos generales para cada estación, puede haber casos en los que exista punto de inicio, pero no de fin o viceversa. Por otra parte, en el caso de que el punto final (`End_Point_all`) sea mayor a los datos de la señal se realiza el recorte en el final de la misma data de este evento.

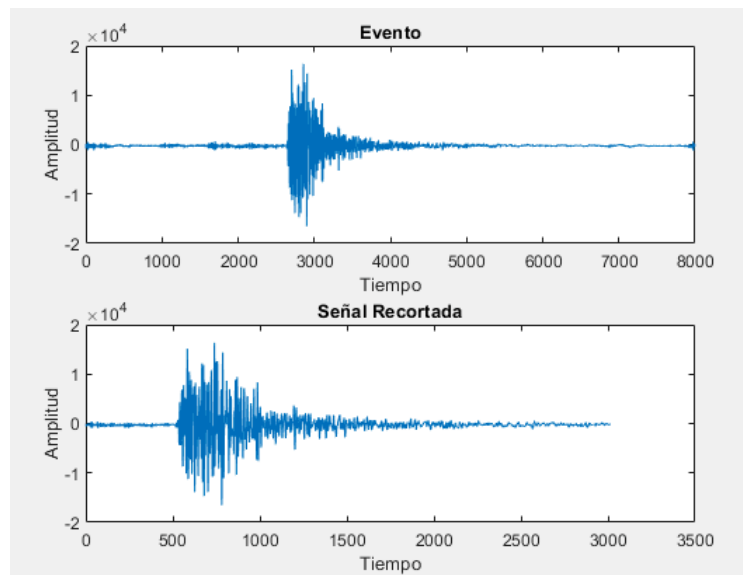


Figura 2.11: Evento de tipo VLP tipo 1, recortado en puntos de inicio y fin.

2.5.2. NORMALIZACIÓN DE LA FRECUENCIA DE MUESTREO

Luego del proceso de recorte, se procede a remover los valores `NaN` de cada señal. Una vez retirados estos valores con la función `remove_nan_trimmed`, se remuestrea la señal para trabajar con una nueva frecuencia con el fin de trabajar con una sola frecuencia de muestreo y de esta manera simplificar procesos posteriores. Para esto se utiliza la función de Matlab `resample`, contenida dentro de la función `normalize_sampling_rate`. Teniendo en cuenta que se debe normalizar siempre al máximo para evitar pérdida de información.

2.5.3. FILTRADO

Seguido estos pasos se procede a filtrar la señal para eliminar ruidos o efectos instrumentales. Luego se filtra la señal utilizando la función `signals_t_filter`, la cual utiliza un filtro an-

ticausal, los parámetros de entrada para el filtro son dados por el usuario dependiendo las señales. Teniendo en cuenta que se puede utilizar un filtro tipo butterworth o fir 1 de orden n pasa banda, los valores sugeridos por ejemplo para eventos de tipo LP, VT, HYB y TREM es en una banda de paso entre 0.5 a 15Hz, puesto que alrededor de los 0.2 Hz se encuentran el contenido espectral de los micro sismos marinos y sobre los 15Hz se encuentran los LHG eventos no volcánicos. [4, 22]. En la figura 2.12, se puede observar un filtro diseñado FIR de orden 128. Finalmente se tiene un filtrado digital de fase cero `filtfilt`, para no tener distorsión de fase en las señales. En el siguiente segmento de código se ve la función de filtrado.

Segmento de código 2.2: Función de Filtrado

```

1 function signals_t_filter=filter_signal(nan_signal,w1,w2,Fs,order,fil)
2 switch fil
3 case 'butterworth'
4 [b,g]=butter(order,[w1 w2]/(Fs/2)); % Infinite Impulse Response (IIR) ...
   filter passband
5 signal_fil=@(x) filtfilt(b,g,x);
6 signals_t_filter=cellfun(signal_fil,nan_signal,...
7 'UniformOutput',false); % Signal pre-processing
8 case 'FIR'
9 b=fir1(order,[w1/Fs w2/Fs]); % Finite Impulse Response (FIR) filter ...
   passband
10 signal_fil=@(x) filtfilt(b,1,x);
11 signals_t_filter=cellfun(signal_fil,nan_signal,...
12 'UniformOutput',false); % Signal pre-processing
13 end
14 end

```

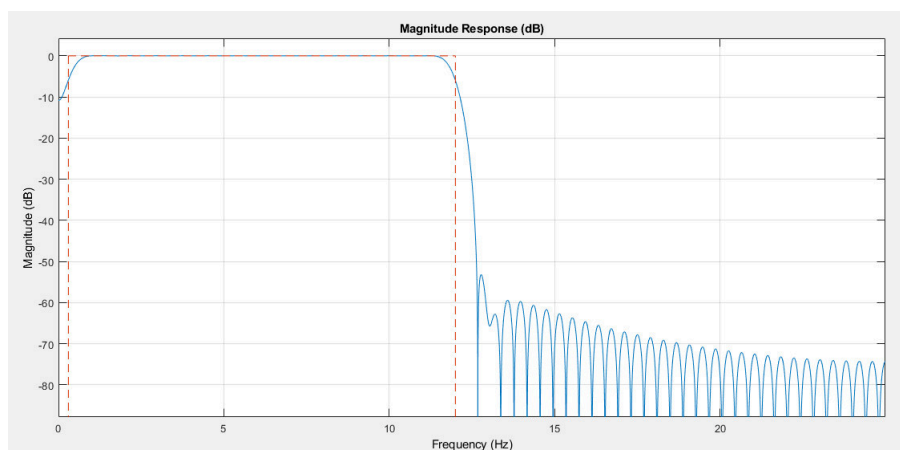


Figura 2.12: Diseño filtro.

2.5.4. NORMALIZACIÓN DE LA SEÑAL

Después que la señal ha sido filtrada pasa a ser normalizada con la función `normalize_signal`, con la cual se puede escoger entre 3 tipos de normalización:

- Normalización máxima: La señal es dividida para el valor absoluto del máximo de la señal, no se retira la media debido a que esta es una característica a ser extraída.
- Normalización con la energía: La señal es dividida para la suma del valor absoluto de la señal al cuadrado.
- Normalización con el valor RMS: La señal es dividida para el valor RMS de la señal.

2.6. EXTRACCIÓN DE CARACTERÍSTICAS

Una vez terminado el preprocesamiento se extraen 84 características, teniendo 13 características en tiempo, 21 en frecuencia y 50 en el dominio de escala, las cuales son de gran utilidad para el aprendizaje automático. Este proceso se realiza con la función `calculate_features`, la cual tiene como entrada la frecuencia de muestreo (F_s), las señales preprocesadas y también un parámetro extra con el cual se puede escoger si se desea solo las características de tiempo, de frecuencia o las de dominio de escala. En el caso que se desee obtener las 84 características este parámetro no va en la función. En la tabla 2.5 podemos ver las 13 características del dominio de tiempo.

Tabla 2.5: Características extraídas en el dominio del tiempo.

Dominio del Tiempo			
ID	Características	Descripción	Implementación
t_mean	Media	Media de la señal. Definido en la ecuación 1.8 .	Esta característica se obtiene con la utilización de la función de Matlab <code>mean(x(abs(x)≠Inf))</code> , donde x es el vector de la señal de entrada. Cuando se obtiene un valor igual a infinito se muestra un advertencia.

Continúa en la siguiente página

Tabla 2.5 – Continuación de la página anterior

ID	Características	Descripción	Implementación
t_std	Desviación Es-tándar	Variabilidad de la señal. Definido en la ecuación 1.9 .	Esta característica se obtiene con la utilización de la función de Matlab <code>std(x)</code> , donde <code>x</code> es el vector de la señal de entrada.
t_var	Varianza	Varianza de la señal. Definido en la ecuación 1.10 .	Esta característica se obtiene con la utilización de la función de Matlab <code>var(x)</code> , donde <code>x</code> es el vector de la señal de entrada.
t_entropy	Entropía	Entropía de Shannon no normalizada. Definido en la ecuación 1.11.	Esta característica se obtiene con la utilización de la función de Matlab <code>wentropy ... (x, 'shannon')</code> , donde <code>x</code> es el vector de la señal de entrada y <code>shannon</code> es la entropía especificada.
t_kurtosis	Curtosis	Medida de la simetría de la señal. Definido en la ecuación 1.12.	Esta característica se obtiene con la utilización de la función de Matlab <code>kurtosis(x)</code> , donde <code>x</code> es el vector de la señal de entrada.
t_multiscaleEntropy	Entropía multiescalar (MSE)	Mide el grado de complejidad de las señales de las series temporales a través de la Entropía de Muestra, a escala 1.	Esta característica se obtiene con la utilización de la función <code>sampen(x,1,.2)</code> , en donde <code>x</code> es el vector de la señal de entrada, 1 es la escala y 0.2 el offset.
Continúa en la siguiente página			

Tabla 2.5 – Continuación de la página anterior

ID	Características	Descripción	Implementación
t_time2peak	Tiempo para alcanzar el pico máximo	Tiempo medido en segundos.	Esta característica se obtiene con la utilización de la función <code>max(x(x≠Inf))</code> , en donde x es el vector de la señal de entrada. Luego ese valor es multiplicado por $1/F_s$ para obtener el valor en segundos. Cuando se obtiene un valor igual a infinito se muestra un advertencia.
t_rms	Valor RMS	Mide valor RMS. Definido en la ecuación 1.13.	Esta característica se obtiene con la utilización de la función de Matlab <code>rms(x)</code> , donde x es el vector de la señal de entrada.
t_peak2peak	Valor de pico a pico	Valor absoluto de la diferencia del valor máximo y mínimo de la señal.	Esta característica se obtiene con la utilización de la función de Matlab <code>peak2peak(x)</code> , donde x es el vector de la señal de entrada.
t_peak2rms	Relación pico a RMS	Relación entre la mayor magnitud absoluta (norma Linfinita) y el valor medio cuadrático. La mayor magnitud es la norma L-infinita. Definido en la ecuación 1.14.	Esta característica se obtiene con la utilización de la función de Matlab <code>peak2rms(x)</code> , donde x es el vector de la señal de entrada.

Continúa en la siguiente página

Tabla 2.5 – Continuación de la página anterior

ID	Características	Descripción	Implementación
t_energy	Energía	Suma del valor absoluto de los cuadrados de la señal.	Esta característica se obtiene con la utilización de la función <code>sum(abs(fft(x)).^2)/length(x)</code> , en donde x es el vector de la señal de entrada.
t_zcr	Tasa de paso por cero	Número de cruces por cero dividido para la longitud de la señal.	Esta característica se obtiene con la utilización de la función <code>sum(abs(diff(x>0)))/length(x)</code> , en donde x es el vector de la señal de entrada.
t_PeaksAboveRMS Density_fun	Picos por encima de la densidad RMS	Relación entre el número de picos cuyas amplitudes son superiores al valor RMS y la longitud de la señal.	Esta característica se obtiene con la utilización de la función <code>sum(findpeaks(x)>rms(x))/length(x)</code> , en donde x es el vector de la señal de entrada.

Las características en frecuencia se obtienen de la estimación de densidad espectral de potencia (PSD), para esto el usuario puede seleccionar entre 3 métodos:

- Periodogram: Se utiliza la función de Matlab `periodogram`, en la cual se tiene los siguientes parámetros de entrada:
 - La señal de entrada.
 - Una ventana que por defecto se tiene una ventana rectangular [].
 - Los números de puntos de la transformada (NFFT), cuyo valor debe ser ingresado por el usuario en la variable `parameter_PSD.NFFT`. Teniendo en cuenta que este valor debería ser múltiplo superior 2^p de la longitud de la señal.
 - La Frecuencia de muestreo, cuyo valor debe ser ingresado por el usuario en la variable `Fs`.
 - El rango de frecuencia de la estimación PSD que por defecto se usa la estimación PSD unilateral de la señal.

- Burg: Se utiliza la función de Matlab `pburg`, en la cual se tiene los siguientes parámetros de entrada:
 - La señal de entrada.
 - El orden del modelo auto regresivo, cuyo valor debe ser ingresado por el usuario en la variable `parameter_PSD.order`.
 - El valor de NFFT, cuyo valor debe ser ingresado por el usuario en la variable `parameter_PSD.NFFT`.
 - La Frecuencia de muestreo, cuyo valor debe ser ingresado por el usuario en la variable `Fs`.
- Welch: Con la función de matlab `pwelch`, en la cual se tiene los siguientes parámetros de entrada:
 - La señal de entrada.
 - Una ventana, para la cual se utiliza una ventana Hamming de tamaño del valor de NFFT, esto se realiza para mantener el mismo tamaño y simplificar posteriores procesos en los cuales se utilice aprendizaje automático.
 - Número de muestras superpuestas, especificado como un número entero positivo menor que la longitud de la ventana y cuyo valor debe ser ingresado por el usuario en la variable `parameter_PSD.noverlap`. El valor ingresado se multiplica por NFFT para mantener el mismo tamaño y simplificar posteriores procesos en los cuales se utilice aprendizaje automático.
 - El valor de NFFT, cuyo valor debe ser ingresado por el usuario en la variable `parameter_PSD.NFFT`.
 - La Frecuencia de muestreo, cuyo valor debe ser ingresado por el usuario en la variable `Fs`.

En el segmento de código 2.3, se muestra un ejemplo de los valores que el usuario debe ingresar para la extracción de las 21 características de frecuencia que se muestran en la tabla 2.6.

Segmento de código 2.3: Selección de parámetros de entrada y método de PSD.

```

1 %Parameters to obtain the characteristics in frequency domain:
2
3
4 %Parameters PSD estimate
5 parameter_PSD.method='burg';           %PSD estimate method choose ...
   'periodogram', 'welch' or 'burg'.
6 parameter_PSD.NFFT=1024;             %Number of DFT points, specified as ...
   a positive integer.
7 % If choose Burg enter the order

```



```

8 parameter_PSD.order=128;           %Order of the autoregressive model, ...
   specified as a positive
9                                     %integer.
10 % If choose Welch enter the noverlap
11 parameter_PSD.noverlap=0.5;       %Number of overlapped samples, ...
   specified as a positive
12                                     %integer smaller than the length of ...
                                       window.

```

Tabla 2.6: Características extraídas en el dominio de la frecuencia.

Dominio de la frecuencia		
ID	Características	Descripción
f_peaks_pos_1	Frecuencia del pico máximo	Muestra el primero los 3 picos mas altos de la señal.
f_90_percent_energy	Ancho de banda de 90 % de energía	Frecuencia mínima a la que la energía acumulada es al menos el 90 % de la energía total [1].
f_entropy	Entropía	Entropía de Shannon no normalizada. Definido en la ecuación 1.11 .
f_mean	Media	Media de la señal. Definido en la ecuación 1.8 .
f_std	Desviación Estandar	Variabilidad de la señal. Definido en la ecuación 1.9 .
f_var	Varianza	Varianza de la señal. Definido en la ecuación 1.10.
f_energy	Energía	Suma del valor absoluto de los cuadrados de la señal.
f_kurtosis	Curtosis	Medida de la simetría de la señal. Definido en la ecuación 1.12 .
f_multiscaleEntropy	Entropía multiescalar (MSE)	Mide el grado de complejidad de las señales de las series temporales a través de la Entropía de Muestra, a escala 1.
f_peak_1020_value	Pico máximo en la banda de 10-20 Hz	Valor de amplitud máxima en la banda de 10-20 Hz..
f_peak_1020_pos	Frecuencia del pico máximo en la banda de 10-20 Hz	Posición de pico de frecuencia máxima en la banda de 10-20 Hz.
Continúa en la siguiente página		

Tabla 2.6 – Continuación de la página anterior.

ID	Características	Descripción
f_peak_2030_value	Pico máximo en la banda de 20-30 Hz	Valor de amplitud máxima en la banda de 20-30 Hz.
f_peak_2030_pos	Frecuencia del pico máximo en la banda de 20-30 Hz	Posición de pico de frecuencia máxima en la banda de 20-30 Hz.
f_rms	Valor RMS	Mide valor RMS. Definido en la ecuación 1.13.
f_peak2rms	Relación pico a RMS	Relación entre la mayor magnitud absoluta (norma Linfinita) y el valor medio cuadrático. La mayor magnitud es la norma L-infinita. Definido en la ecuación 1.14.
f_power	Potencia	Potencia de la señal en el dominio de la frecuencia.
f_PeaksAboveRMSDensity_fun	Picos por encima de la densidad RSM	Relación entre el número de picos cuyas amplitudes son superiores al valor RMS y la longitud de la señal.
f_peaks_val_2	2º valor de pico más alto	Amplitud del segundo pico.
f_peaks_pos_2	Frecuencia del 2º pico más alto	Posición en frecuencia del segundo pico.
f_peaks_val_3	3º valor de pico más alto	Amplitud del tercer pico.
f_peaks_pos_3	Frecuencia del 3º pico más alto	Posición en frecuencia del tercer pico.

Las características en el dominio de la escala se obtuvieron con el uso de la transformada de Wavelet. Para esto se utiliza la función de Matlab *wavedec*, la cual retorna dos vectores: C que es el vector de descomposición Wavelet y L el vector que contiene el número de coeficientes por nivel. En la figura 2.13 se puede observar la organización estructural de una señal descompuesta en 3 niveles [39].

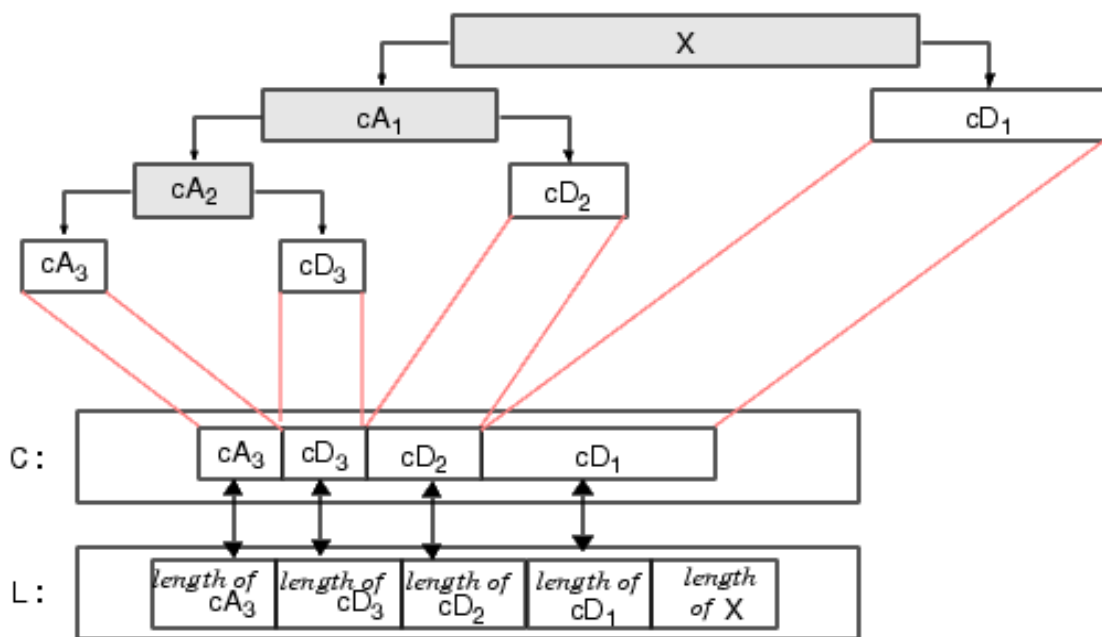


Figura 2.13: Descomposición señal de 3 niveles.

Con esta función se realiza una descomposición hasta el nivel 6 de la Wavelet Madre de tipo 10 (db10). Para la obtención de los coeficientes de aproximación (cA) y de los coeficientes detallados (cD), se utiliza la función *wrcoef*; los coeficiente utilizados para este estudio son A6, D1, D2, D3, D4, D5 y D6, ya que estos contienen la información necesaria para la extracción de características. Finalmente para hallar la energía y los porcentajes para cada componente de Wavelet se utiliza la función *wdecenergy*, en donde: $E(i)$ es la energía de la señal i -ésima, $PEC(i,1)$ es el porcentaje de energía para la aproximación del nivel de la señal i -ésima y $PECFS(i,j)$ es el porcentaje de energía para los coeficientes j de la señal i -ésima. A continuación, en la tabla 2.7 se muestran las 50 características obtenidas en el dominio de la escala.

Tabla 2.7: Características extraídas en el dominio de la escala.

Dominio de la Escala		
ID	Características	Descripción
w_f_maxval_A6	Pico máximo en el dominio de la frecuencia A6	Valor de amplitud máxima en el dominio de frecuencia de la señal A6.
w_f_maxval_D1	Pico máximo en el dominio de la frecuencia D1	Valor de amplitud máxima en el dominio de frecuencia de la señal D1.

Continúa en la siguiente página

Tabla 2.7 – Continuación de la página anterior.

ID	Características	Descripción
w_f_maxval_D2	Pico máximo en el dominio de la frecuencia D2	Valor de amplitud máxima en el dominio de frecuencia de la señal D2.
w_f_maxval_D3	Pico máximo en el dominio de la frecuencia D3	Valor de amplitud máxima en el dominio de frecuencia de la señal D3.
w_f_maxval_D4	Pico máximo en el dominio de la frecuencia D4	Valor de amplitud máxima en el dominio de frecuencia de la señal D4.
w_f_maxval_D5	Pico máximo en el dominio de la frecuencia D5	Valor de amplitud máxima en el dominio de frecuencia de la señal D5.
w_f_maxval_D6	Pico máximo en el dominio de la frecuencia D6	Valor de amplitud máxima en el dominio de frecuencia de la señal D6.
w_f_maxpos_A6	Frecuencia del pico máximo A6	Ubicación de amplitud máxima en el dominio de frecuencia de la señal A6
w_f_maxpos_D2	Frecuencia del pico máximo D2	Ubicación de amplitud máxima en el dominio de frecuencia de la señal D2.
w_f_maxpos_D3	Frecuencia del pico máximo D3	Ubicación de amplitud máxima en el dominio de frecuencia de la señal D3.
w_f_maxpos_D4	Frecuencia del pico máximo D4	Ubicación de amplitud máxima en el dominio de frecuencia de la señal D4.
w_f_maxpos_D5	Frecuencia del pico máximo D5	Ubicación de amplitud máxima en el dominio de frecuencia de la señal D5.
w_f_maxpos_D6	Frecuencia del pico máximo D6	Ubicación de amplitud máxima en el dominio de frecuencia de la señal D6.
w_f_mean_A6	Media en el dominio de la frecuencia A6	Cálculo de la media en el dominio de la frecuencia A6. Muestra un warning si detecta un valor inf.
Continúa en la siguiente página		

Tabla 2.7 – Continuación de la página anterior.

ID	Características	Descripción
w_f_mean_D1	Media en el dominio de la frecuencia D1	Cálculo de la media en el dominio de la frecuencia D1. Muestra un warning si detecta un valor inf.
w_f_mean_D2	Media en el dominio de la frecuencia D2	Cálculo de la media en el dominio de la frecuencia D2. Muestra un warning si detecta un valor inf.
w_f_mean_D3	Media en el dominio de la frecuencia D3	Cálculo de la media en el dominio de la frecuencia D3. Muestra un warning si detecta un valor inf.
w_f_mean_D4	Media en el dominio de la frecuencia D4	Cálculo de la media en el dominio de la frecuencia D4. Muestra un warning si detecta un valor inf.
w_f_mean_D5	Media en el dominio de la frecuencia D5	Cálculo de la media en el dominio de la frecuencia D5. Muestra un warning si detecta un valor inf.
w_f_mean_D6	Media en el dominio de la frecuencia D6	Cálculo de la media en el dominio de la frecuencia D6. Muestra un warning si detecta un valor inf.
w_t_meanEnergyAD	Energía media componentes	Energía media de los componentes de las wavelets A6, D1, D2, D3, D4, D5 y D6.
w_t_PEC_A6	Porcentaje de energía de A6	Fracción de energía de la componente de la señal de la wavelet A6, con respecto a la energía total.
w_t_PEC_D1	Porcentaje de energía de D1	Fracción de energía de la componente de la señal de la wavelet D1, con respecto a la energía total.
w_t_PEC_D2	Porcentaje de energía de D2	Fracción de energía de la componente de la señal de la wavelet D2, con respecto a la energía total.
w_t_PEC_D3	Porcentaje de energía de D3	Fracción de energía de la componente de la señal de la wavelet D3, con respecto a la energía total.
w_t_PEC_D4	Porcentaje de energía de D4	Fracción de energía de la componente de la señal de la wavelet D4, con respecto a la energía total.
Continúa en la siguiente página		

Tabla 2.7 – Continuación de la página anterior.

ID	Características	Descripción
w_t_PEC_D5	Porcentaje de energía de D5	Fracción de energía de la componente de la señal de la wavelet D5, con respecto a la energía total.
w_t_PEC_D6	Porcentaje de energía de D6	Fracción de energía de la componente de la señal de la wavelet D6, con respecto a la energía total.
w_t_rms_A6	RMS en tiempo de A6	Valor RMS en el dominio del tiempo de la señal A6.
w_t_rms_D1	RMS en tiempo de D1	Valor RMS en el dominio del tiempo de la señal D1.
w_t_rms_D2	RMS en tiempo de D2	Valor RMS en el dominio del tiempo de la señal D2.
w_t_rms_D3	RMS en tiempo de D3	Valor RMS en el dominio del tiempo de la señal D3.
w_t_rms_D4	RMS en tiempo de D4	Valor RMS en el dominio del tiempo de la señal D4.
w_t_rms_D5	RMS en tiempo de D5	Valor RMS en el dominio del tiempo de la señal D5.
w_t_rms_D6	RMS en tiempo de D6	Valor RMS en el dominio del tiempo de la señal D6.
w_t_peak2peak_A6	Relación pico-pico tiempo A6	Valor pico a pico en el dominio del tiempo de la señal A6.
w_t_peak2peak_D1	Relación pico-pico tiempo D1	Valor pico a pico en el dominio del tiempo de la señal D1.
w_t_peak2peak_D2	Relación pico-pico tiempo D2	Valor pico a pico en el dominio del tiempo de la señal D2.
w_t_peak2peak_D3	Relación pico-pico tiempo D3	Valor pico a pico en el dominio del tiempo de la señal D3.
w_t_peak2peak_D4	Relación pico-pico tiempo D4	Valor pico a pico en el dominio del tiempo de la señal D4.
w_t_peak2peak_D5	Relación pico-pico tiempo D5	Valor pico a pico en el dominio del tiempo de la señal D5.
w_t_peak2peak_D6	Relación pico-pico tiempo D6	Valor pico a pico en el dominio del tiempo de la señal D6.
w_t_peak2rms_A6	Relación pico-RMS tiempo A6	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal A6.

Continúa en la siguiente página

Tabla 2.7 – Continuación de la página anterior.

ID	Características	Descripción
w_t_peak2rms_D1	Relación pico-RMS tiempo D1	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D1.
w_t_peak2rms_D2	Relación pico-RMS tiempo D2	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D2.
w_t_peak2rms_D3	Relación pico-RMS tiempo D3	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D3.
w_t_peak2rms_D4	Relación pico-RMS tiempo D4	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D4.
w_t_peak2rms_D5	Relación pico-RMS tiempo D5	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D5.
w_t_peak2rms_D6	Relación pico-RMS tiempo D6	Relación entre la mayor magnitud (L-Infinity) y el valor RMS de la señal D6.
w_t_meanEnergyCoeff	Energía coeficientes wavelet	Energía del coeficiente medio calculada promediando la contribución de energía de cada coeficiente Wavelet.

Una vez calculadas las 84 características, se concluye esta segunda etapa con la creación de una tabla final con la función `signals_with_characteristics`, las variables que contiene esta tabla con las señales y sus características se muestra a continuación en la tabla 2.8.

Tabla 2.8: Variables tabla con características.

Variable	Descripción
Event_Identifier	Muestra un Id único del evento.
Station	Código de estación de donde los datos contenidos en el archivo excel fueron tomados por el IGEPN .
Channel	Canal donde se halló el evento.
Type	Tipo de evento detectado por el IGEPN.

Continúa en la siguiente página

Tabla 2.8: Variables tabla con características.

Variable	Descripción
Features	Características obtenidas de cada evento descritas en las tablas 2.7,2.6 y 2.5. Puede tener 84 columnas o variar entre 50, 21 y 13 columnas dependiendo cuales características se desee obtener.

3. RESULTADOS Y DISCUSIÓN

3.1. RESULTADOS

Para realizar las pruebas de funcionalidad del software se trabaja sobre una base de datos facilitada por el PhD. Mario Ruiz, investigador del Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), en la cual se encuentra el registro de la actividad volcánica del Cotopaxi en el año 2015, tomada en los canales BH para los diferentes ejes del sismómetro: BHZ (eje vertical), BHN (norte) y BHE (este). Además, se tiene la información del canal BDF en el cual el código DF indica que este canal contiene la medición del barómetro, el cual mide la presión por infrasonido en el interior del volcán. Estos valores son tomados en 4 estaciones vulcanológicas cuyos códigos son BNAS, BREF, BTAM, BVC2. Teniendo así un total de 4.555 archivos en formato .mseed [40].

Como se vio en la sección 2.1, esta base de datos también contiene archivos de tipo .xlsx, los cuales contienen 11 clases de eventos etiquetados de la siguiente manera: VT, LP, VLP, VLP de tipo 1, VLP de tipo 2, TREM, TREMI, EXPL, HB, TRESP y TRBA. Cada archivo contiene eventos de diferentes años, por lo que la base es filtrada para obtener solo los eventos del año 2015 y también se elimina eventos que pudieran estar duplicados. En la siguiente tabla 3.1 se muestra la cantidad de señales que se obtuvieron para el año 2015 por cada tipo.

Tabla 3.1: Número de eventos por tipo.

Tipo	Cantidad
VT	9674
LP	8692
VLP	135
VLP TIPO 1	198
VLP TIPO 2	157
TREM	6757
TREMI	4221
EXPL	110
HB	538
TRESP	169
TRBA	142

Teniendo un total de 30793 eventos por ser localizados y extraídos en cada canal y estación antes mencionados.

3.1.1. PRUEBAS DE FUNCIONALIDAD

Primero para la búsqueda y extracción de cada evento, se ingresa en el toolbox el nombre del directorio en el que se encuentran los archivos .mseed (mseed), el nombre del directorio donde se almacenan los archivos convertidos (allfilesmat), el archivo Excel que contiene el tipo de evento que se desee hallar y la estación de referencia, la cual se utiliza para el descarte o confirmación de eventos para esta base de datos la estación de referencia es BREF. Como se observa en el siguiente segmento de código.

Segmento de código 3.4: Entradas toolbox

```
1 % Inputs :
2
3 dir_mseed='mseed';           %mseed files directory
4 dir_matlab='allfilesmat';    %directory converts mseed mat
5 excel_file='HB.xlsx';       %file .xlsx with events to find
6 reference_station='BREF';    %reference station for discard events
```

Los archivos contenidos en archivos de tipo mseed son convertidos exitosamente al formato .mat con la función `seed_to_mat_conversion`, como se muestra en el ejemplo de la figura 3.1 y son almacenados en el directorio que se ha creado con el nombre `allfilesmat`.

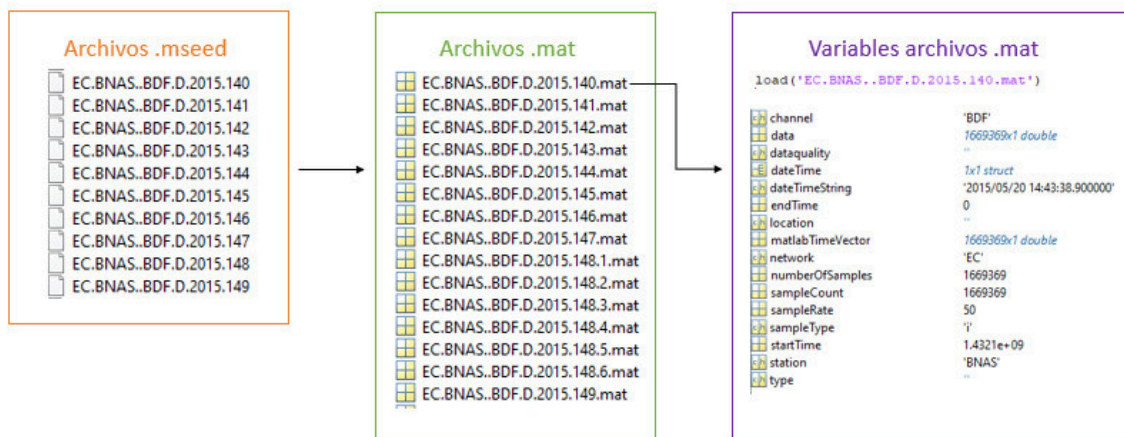


Figura 3.1: Conversión eventos .mseed a .mat.

Luego de esto con los archivos de tipo .xlsx se procede a probar la función `t_from_excel`, en donde se realiza de manera correcta la conversión del tipo .xlsx a .mat, generando así tablas para cada tipo de evento en formato .mat. Debido a que estos archivos almacenan eventos de diferentes años se realiza un filtrando de los datos para solo extraer los pertenecientes a el año 2015 y también los eventos duplicados son retirados de manera correcta. En el siguiente segmento de código se observa lo antes mencionado.

Segmento de código 3.5: Extracción de eventos de archivo .xlsx

```

1 t_from_excel=import_from_excel(excel_file); ...
   %Extracción de coordenadas en excel
2 t_from_excel=t_from_excel(t_from_excel.StartTime.Year == 2015,:); %filtrado ...
   de señales 2015
3 t_from_excel=remove_duplicates(t_from_excel); %eliminar ...
   duplicados del excel

```

Como resultado se obtuvo tablas en formato .mat, como la que se observa en el gráfico 3.2. Estas muestran de manera correcta todo el contenido del archivo Excel y además se observa que para el correcto funcionamiento se necesita una versión de Matlab superior a la 2019a, ya que en las pruebas realizadas con versiones anteriores se presenta un error al utilizar la función de Matlab readtable.

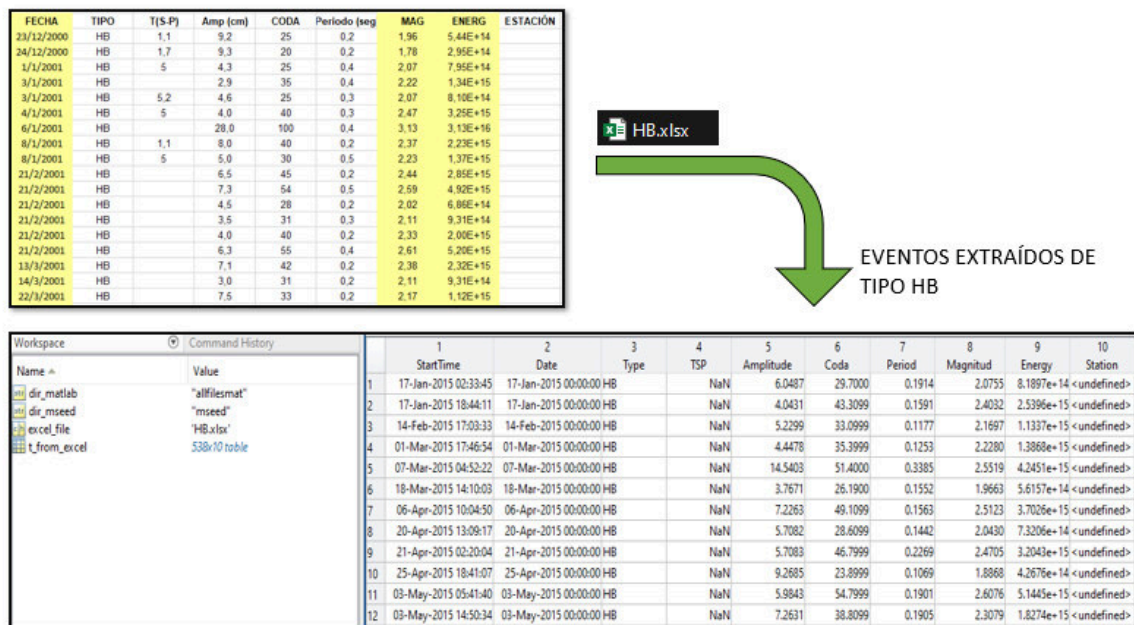


Figura 3.2: Extracción de eventos del archivo .xlsx.

Posterior a esto para realizar la búsqueda y extracción se prueba la función search_and_extract, con 30793 eventos para ser buscados en 4555 archivos .mat. Se observa que un solo evento tarda alrededor de 2 segundos en ser buscado y extraído de cada estación y canal de un total de 30 días. Por lo que es evidente que el tiempo que toma la extracción de estos eventos es menor a la que emplearía un investigador al realizar este proceso de manera manual, por otra parte, esta herramienta no solo extrae los datos también almacena información importante para posteriores desarrollos, donde se aplique aprendizaje automático. La figura 3.3, muestra un ejemplo de la extracción realizada por el toolbox de un evento.

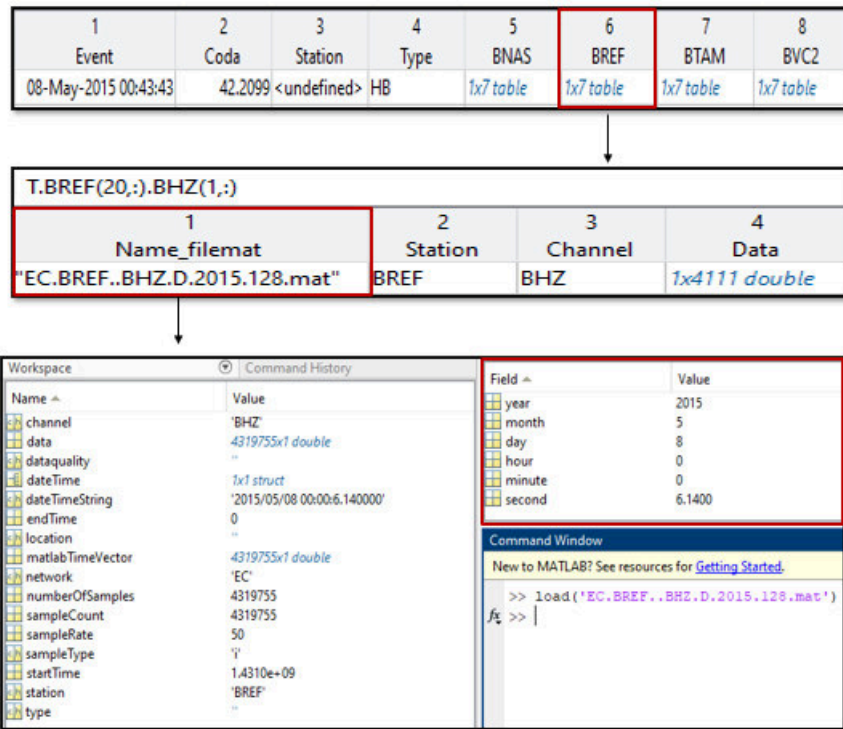


Figura 3.3: Localización y extracción de la data de un evento tipo HB para la estación BREF y canal BHZ.

Como siguiente paso con los datos obtenidos de cada evento, se utiliza un detector STA/LTA. Esto debido a que el PhD. Mario Ruíz sugirió utilizar un algoritmo STA/LTA en las señales para identificar solo las señales con alta relación señal a ruido y de esta manera poder reconocer de mejor manera los eventos sísmicos. Los parámetros de detección son escogidos dependiendo el tipo de señal, tomando como referencia los valores sugeridos en la tabla 2.3. Se debe tener en cuenta que lo mas importante es la elección de los triggers, por lo que para cada tipo de evento se toma señales aleatorias, con estas se observa su gráfica de relación STA/LTA y de acuerdo a esta se selecciona los parámetros que se ajustan mejor para ese tipo de eventos. A continuación, en la gráfica 3.4, se muestran 2 eventos tipo EXPL en la estación BREF para el canal BHZ, de 2 días diferentes en los que se aprecia que el umbral empieza en 6 y decae por debajo de 0.3. Se debe tomar en cuenta que en toda base siempre existe ruido, por lo que los valores escogidos para estos casos podrían no ajustarse a todas las señales y estas deberían ser analizadas de manera independiente.

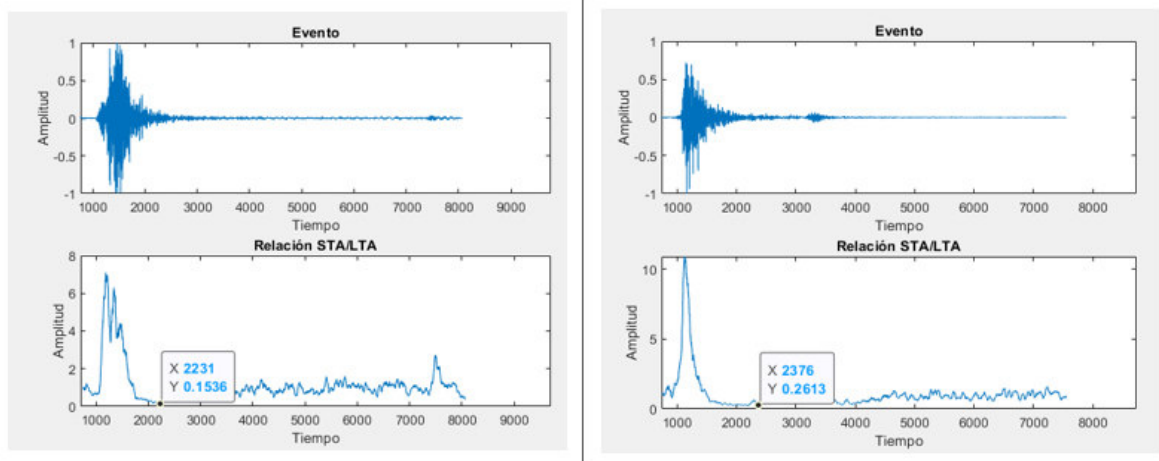


Figura 3.4: Eventos tipo EXPL con su respectiva relación sta/lta.

En la siguiente tabla 3.2, se definen los siguientes parámetros como los más óptimos para la mayoría de señales, después de experimentar con señales aleatorias para cada tipo de evento con un rango de ventana de sta de 1s y lta hasta 30s.

Tabla 3.2: Valores de parámetros para cada tipo de evento, determinados en este trabajo.

	STA	LTA	Activación	Desactivación	Salto	Duración mínima	PEM	PET
VT	1	15	2	1	0	2	5	10
LP	1	10	2	1	0	2	5	10
VLP	1	15	3	0.5	0	2	10	10
VLP TIPO 1	1	30	6	0.3	0	2	10	20
VLP TIPO 2	1	25	5	0.5	0	2	10	20
TREM	1	12	3	0.3	0	2	10	10
TREMI	1	20	3	0.3	0	2	10	10
EXPL	1	15	6	0.4	0	2	5	20
HB	1	15	3	0.5	0	2	10	10
TRESP	1	30	2	0.5	0	2	10	10
TRBA	1	30	2.5	0.5	0	2	10	10

Con estos valores se procede a probar el detector STA/LTA implementado en el toolbox y una vez seteados estos parámetros, se obtienen los puntos de inicio y fin para cada evento que detecta el disparador. En la figura 3.5 y 3.6, se puede ver un ejemplo de cómo se presentan estos puntos de inicio y fin para un evento detectado y uno no detectado, estos puntos se presentan como muestras cuando un evento a sido detectado y en el caso de no haber

detectado un evento los puntos de inicio y fin quedan con el valor de NaN, como se puede observar a continuación.

T.BREF(20,:).BHZ(1,:)						
1	2	3	4	5	6	7
Name_filemat	Station	Channel	Data	StartPoint	EndPoint	Sample_rate
"EC.BREF..BHZ.D...."	BREF	BHZ	1x4301 dou...	860	2962	50

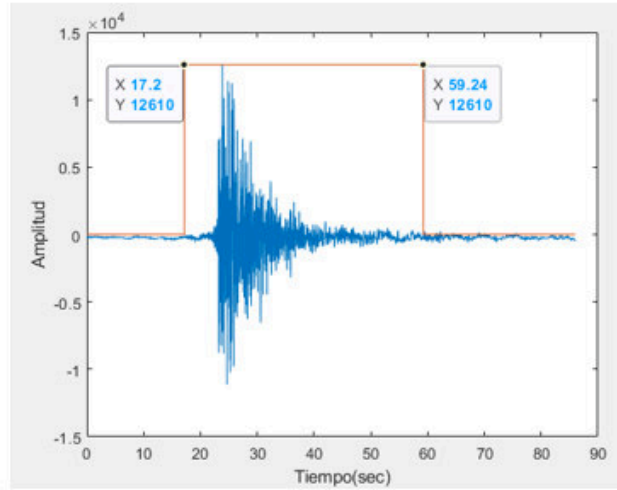


Figura 3.5: Evento tipo EXPL detectado, con su punto de inicio y fin en segundos.

T.BTAM(20,:).BHZ(1,:)						
1	2	3	4	5	6	7
Name_filemat	Station	Channel	Data	StartPoint	EndPoint	Sample_rate
"EC.BTAM..BHZ.D.2015.292.mat"	BTAM	BHZ	1x4301 dou...	NaN	NaN	50

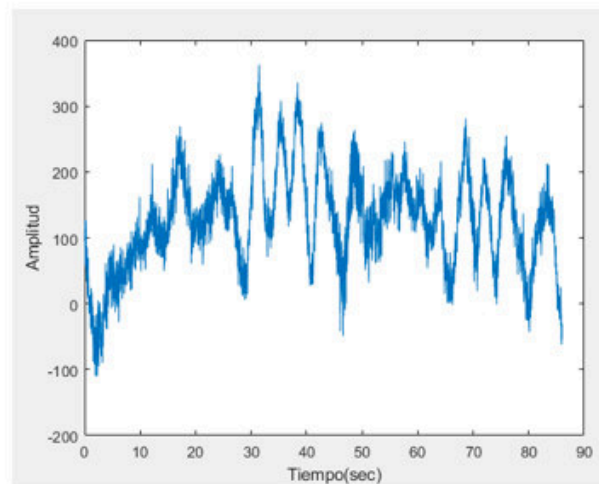


Figura 3.6: Evento tipo EXPL no detectado, con valores NaN en su punto de inicio y fin.

Después de este proceso se puede ver que varios eventos etiquetados no son detectados como eventos, del total de 30793 se consideraron como eventos con la aplicación del algoritmo STA/LTA 17286, de manera que, casi el 42 % de los eventos detectados manualmente pueden a ver sido producto del ruido de la señal, en la siguiente tabla se muestra la cantidad de eventos que se descartan para la estación BREF en el canal BHZ, que es la estación de referencia para el siguiente paso.

Tabla 3.3: Eventos detectados y descartados por detector STA/LTA.

Tipo	Eventos detectados	Eventos descartados
VT	8315	1359
LP	6535	2157
VLP	103	32
VLP TIPO 1	124	74
VLP TIPO 2	68	89
TREM	517	6240
TREMI	1044	3177
EXPL	62	48
HB	358	180
TRESP	127	42
TRBA	33	109

Con los eventos ya pasados por el detector STA/LTA, se procede a probar la confirmación o descarte de evento que realiza el toolbox. Esto se implementa según las indicaciones del PhD. Mario Ruiz del IGEPN, el cual indicó que la estación BREF se tome como la estación de referencia para las demás estaciones y que el tiempo para que un evento sea descartado es de una ventana de 2s de diferencia de arribo con respecto el resto de estaciones, de tal manera, que un evento es confirmado si existe al menos en una estación a parte de BREF. Con esto en la siguiente figura 3.7, se observa un ejemplo de como el toolbox muestra el descarte o confirmación de eventos.

	1	2	3	4	5	6	7	8	9	10
	Event	Coda	Station	Type	BNAS	BREF	BTAM	BVC2	Voting_Detection	Event_Identifier
1	17-Jan-2015 02:33:45	29.7000	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'99fa45754ebcdcd3999fb5b9cf6a259'
2	17-Jan-2015 18:44:11	43.3099	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'cf134224213e185c7854af27b31aeb40'
3	14-Feb-2015 17:03:33	33.0999	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'0d10bf099e7853fca78cec04ba2af266'
4	01-Mar-2015 17:46:54	35.3999	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'08838032a18f4c01f7ab7a0e19380392'
5	07-Mar-2015 04:52:22	51.4000	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'38c4678b3dbba0b67704502a0abad17922'
6	18-Mar-2015 14:10:03	26.1900	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'00ce3ae74d3fee43adc877d8dc629866'
7	06-Apr-2015 10:04:50	49.1099	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'e7e81075226e3fab4f19d80944e78a71'
8	20-Apr-2015 13:09:17	28.6099	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'fbd9e36495b12e361370a686bd3b95fd'
9	21-Apr-2015 02:20:04	46.7999	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'baf804a2533893585c83bd2a22cc0eff'
10	25-Apr-2015 18:41:07	23.8999	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'132b65f8ac7073bb224cd7c680f9545c'
11	03-May-2015 05:41:40	54.7999	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'84f9c647d5fc83cc290ddacda2f15eb3'
12	03-May-2015 14:50:34	38.8099	<undefined>	HB	1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'e01939143cf79532b0c72a4111f5428e'

Figura 3.7: Tabla con 12 eventos de tipo HB confirmados y descartados.

Al realizar el descarte de eventos se obtiene un total de 7788 eventos confirmados, con los cuales se realiza los procesos siguientes. En la siguiente tabla 3.4, se muestran los eventos confirmados y descartados para cada tipo de evento.

Tabla 3.4: Eventos confirmados y descartados por tipo.

Tipo	Eventos confirmados	Eventos descartados
VT	2815	6859
LP	4143	4549
VLP	71	64
VLP TIPO 1	72	126
VLP TIPO 2	44	113
TREM	118	6639
TREMI	252	3969
EXPL	28	82
HB	191	347
TRESP	48	121
TRBA	6	136

Se procede a probar las funciones de preprocesamiento del toolbox y la extracción de características de cada evento tomado por el canal BHZ de cada estación, puesto que este es el canal que tiene la mayor información de las señales. Las señales primero pasan por la función de recorte trim_signal en donde son recortadas en los puntos Start_Point_all y End_Point_all. Como se puede observar en la figura 3.8.

1	2	3	4	5
BHZ	BHE	Start_Point_all	End_Point_all	Sample_Rate_all
1x7 table	1x7 table	417	2176	50

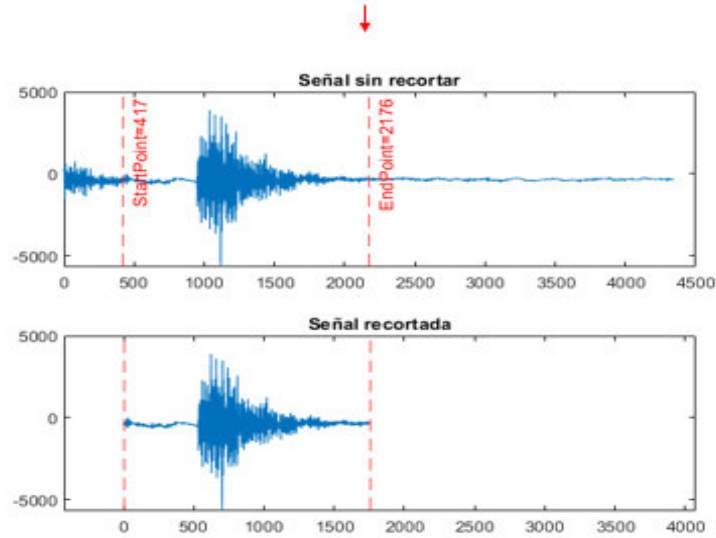


Figura 3.8: Ejemplo de señal recortada de tipo HB

El proceso de normalización no se realiza en estas señales puesto que la frecuencia de muestreo es la misma, para este caso es igual a 50 Hz. Pero se prueba el funcionamiento de la función de remuestreo cambiando la Fs de 50Hz a 100 Hz. Como se observa en la figura 3.9.

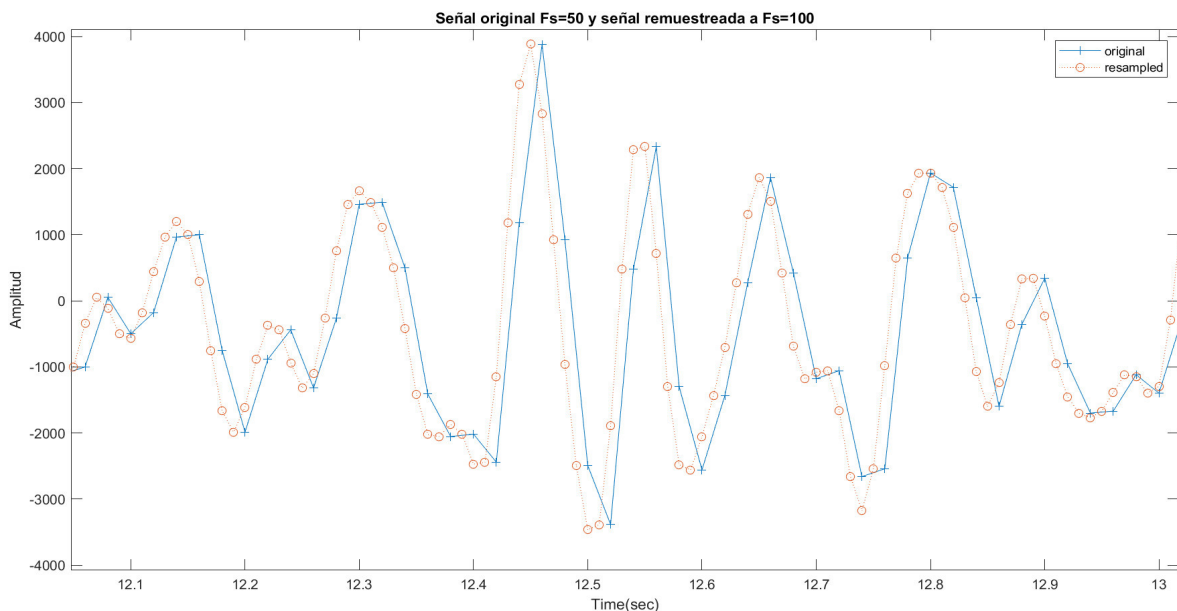


Figura 3.9: Ejemplo de señal remuestreada a 100Hz

En el proceso de filtrado, dependiendo el tipo de evento se escoge el mejor rango de frecuencias para cada señal, se prueba el funcionamiento del filtro con una señal de tipo HB con un rango de entre 0.5 y 15 Hz que es donde se encuentra el mayor espectro de frecuencia

como se observa en la figura 3.10.

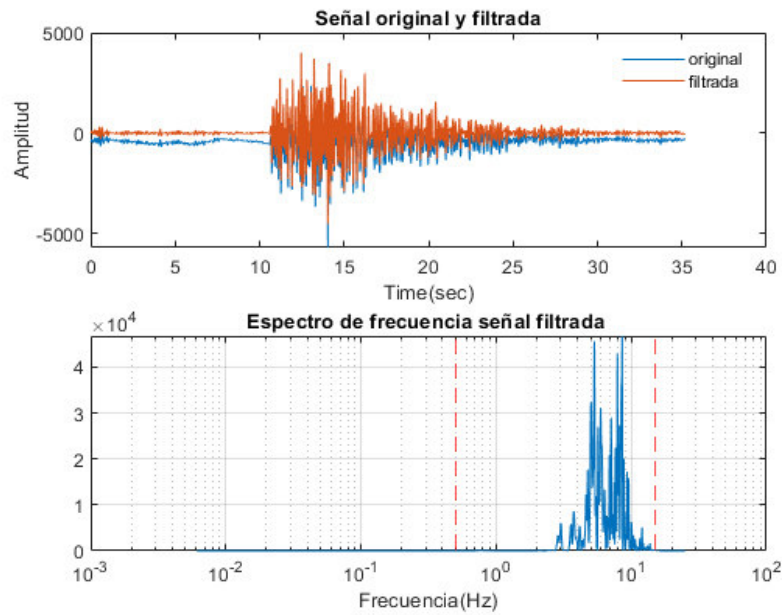


Figura 3.10: Ejemplo de señal de tipo HB filtrada.

Para terminar la etapa de preprocesamiento las señales son normalizadas con el método de normalización máxima. Se prueba el toolbox con los tres tipos de normalizaciones implementadas para ver su correcto funcionamiento, como se observa en la siguiente figura 3.11, un ejemplo de una señal de tipo HB normalizada por los 3 métodos.

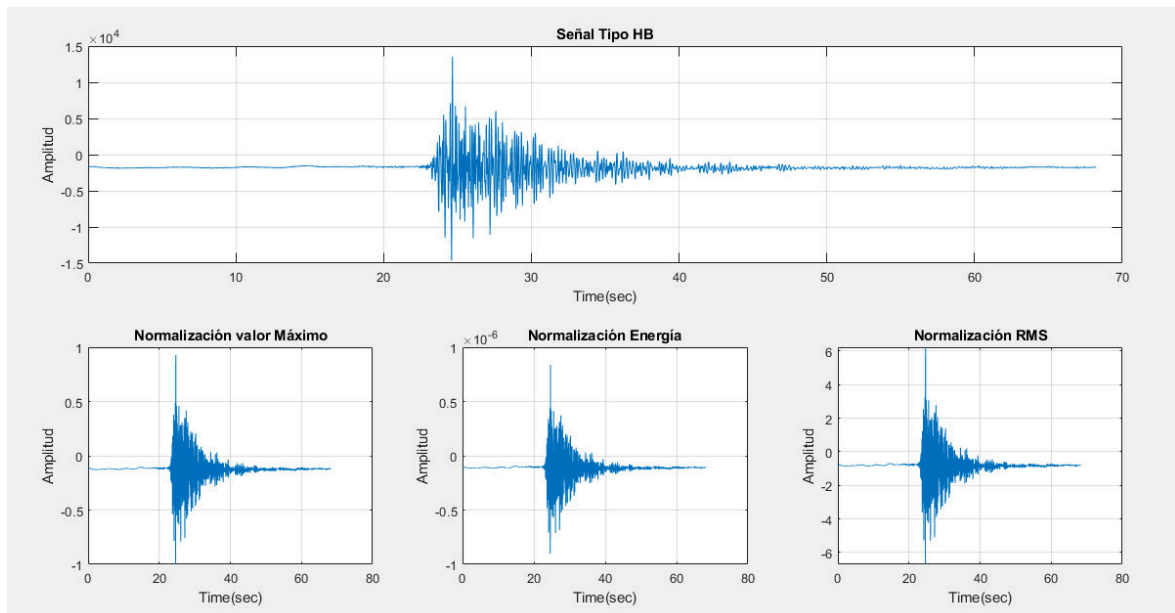


Figura 3.11: Señal tipo HB, normalizada por 3 métodos.

Una vez terminada esta etapa se procede a probar las funciones para la extracción de las 84 características, para cada evento. Primero se obtiene las 14 características de tiempo

como se aprecia en la figura 3.12, la cual muestra un ejemplo de los resultado obtenido por el toolbox para una señal de tipo HB.

1	2	3	4	5	6	7	8	9	10
Event_Identifier	Station	Channel	Type	t_mean	t_std	t_var	t_entropy	t_kurtosis	t_multiscaleEntropy
'6049ffd2ab53e6fd...	BREF	BHZ	HB	9.2715e-05	0.1462	0.0214	73.7041	11.1102	1.0674

11	12	13	14	15	16	17
t_time2peak	t_rms	t_peak2peak	t_peak2rms	t_energy	t_zcr	t_PeaksAboveRMSDensity_fun
12.4600	0.1462	1.8798	6.8398	37.6208	0.2523	0.0375

Figura 3.12: Características en el dominio de el tiempo, de una señal de tipo HB.

Para obtener las características de frecuencia se puede escoger entre tres métodos, para las señales dadas se utiliza el método del periodograma, utilizando para NFFT un valor de 1024, esto debido a que la mayor parte de señales utilizadas no sobrepasan esta longitud y experimentalmente una ventana mas o menos de ese tamaño no ha hecho la diferencia, además que la carga computacional no es tan grande. De igual forma se prueba el funcionamiento de los 3 métodos con señales aleatorias. En la figura 3.13, se puede apreciar un ejemplo de los resultados obtenidos al aplicar cualquiera de los 3 métodos a una señal de tipo HB.

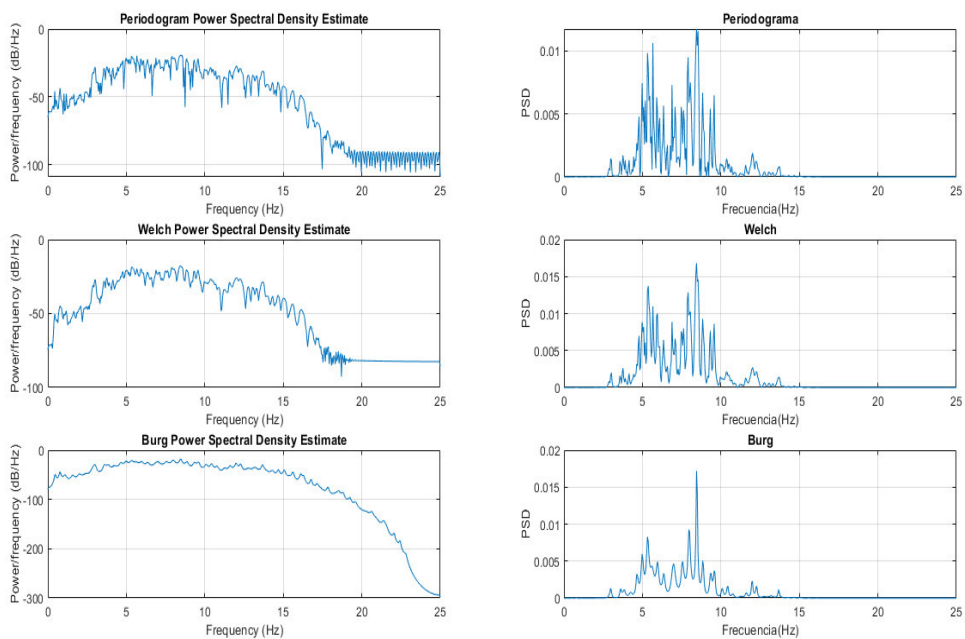


Figura 3.13: Estimación de densidad espectral obtenida por 3 métodos.

En la figura 3.14, se puede apreciar un ejemplo de los resultados obtenidos de la extracción de características de frecuencia de una señal de tipo HB.

19	20	21	22	23	24	25	26	27
f_90_percent_energy	f_entropy	f_mean	f_std	f_var	f_energy	f_kurtosis	f_multiscaleEntropy	f_peak_1020_value
9.6680	-1.6974e+07	-55.6691	27.1374	736.4366	37.6208	1.6819	1.8338	-27.2210

28	29	30	31	32	33	34
f_peak_1020_pos	f_peak_2030_value	f_peak_2030_pos	f_rms	f_peak2rms	f_power	f_PeaksAboveRMSDensity_fun
12.0117	-89.9401	20.4102	61.9197	5.5839	0.4413	0.0526

35	36	37	38
f_peaks_val_2	f_peaks_pos_2	f_peaks_val_3	f_peaks_pos_3
-19.3721	8.5449	-19.7402	5.6641

Figura 3.14: Características en el dominio de la frecuencia, de una señal de tipo HB.

Por último se prueba las funciones para obtener las características en el dominio de la escala, para esto se utiliza la transformada de Wavelet, en la figura 3.15, donde se puede apreciar las señales y los espectrogramas con un NFFT=1024, de los coeficientes de la descomposición de Wavelet hasta el nivel 6 para un evento de tipo HB, donde las principales componentes se encuentran en D1, D2 y D3 y las frecuencias más bajas se encuentran en D5, D6 y A6.

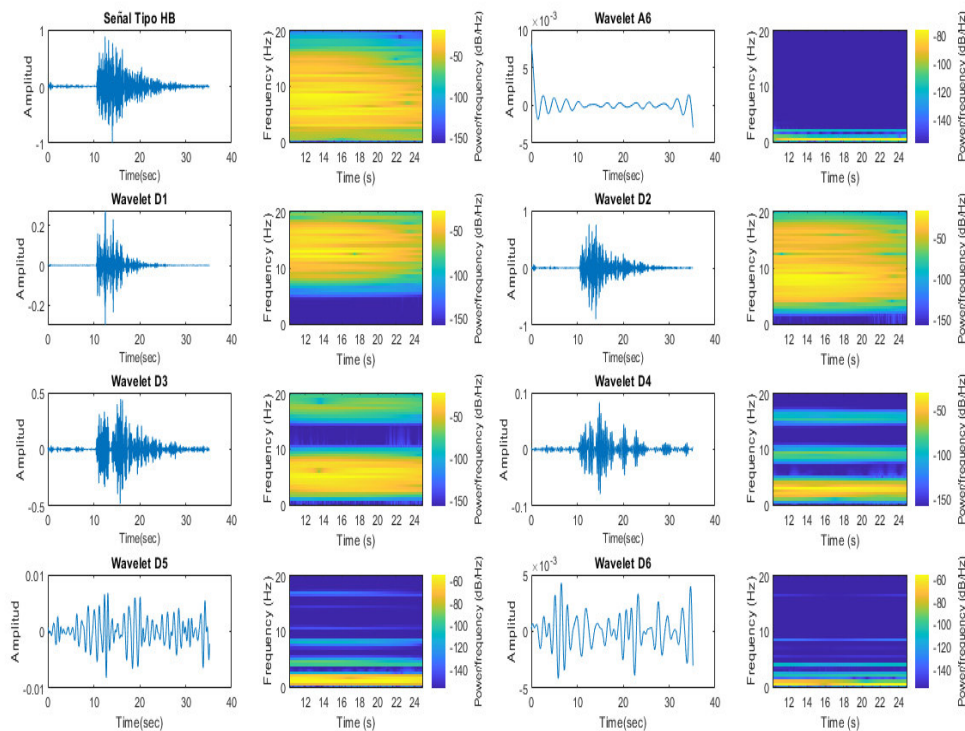


Figura 3.15: Espectrogramas de las señales resultantes de la descomposición Wavelet para un evento de tipo HB.

En la figura 3.16, se puede apreciar un ejemplo de los resultados obtenidos de la extracción de características en el dominio de la escala de una señal de tipo HB.

39	40	41	42	43	44	45
w_f_maxval_A6	w_f_maxval_D1	w_f_maxval_D2	w_f_maxval_D3	w_f_maxval_D4	w_f_maxval_D5	w_f_maxval_D6
-53.9799	-30.1844	-19.4993	-20.6940	-30.9558	-46.0361	-48.9718

46	47	48	49	50	51	52
w_f_maxpos_A6	w_f_maxpos_D2	w_f_maxpos_D3	w_f_maxpos_D4	w_f_maxpos_D5	w_f_maxpos_D6	w_f_mean_A6
0.4395	8.4473	5.3223	2.9785	0.7813	0.7813	-88.1911

53	54	55	56	57	58	59	60
w_f_mean_D1	w_f_mean_D2	w_f_mean_D3	w_f_mean_D4	w_f_mean_D5	w_f_mean_D6	w_t_meanEnergyAD	w_t_pec_A6
-79.1794	-57.9200	-72.4563	-91.1240	-99.0489	-95.9742	5.3900	0.2300

61	62	63	64	65	66	67	68	69
w_t_pec_D1	w_t_pec_D2	w_t_pec_D3	w_t_pec_D4	w_t_pec_D5	w_t_pec_D6	w_t_rms_A6	w_t_rms_D1	w_t_rms_D2
4.1146	62.1672	32.4324	1.0068	0.0320	0.0169	9.6947e-04	0.0297	0.1154

70	71	72	73	74	75	76
w_t_rms_D3	w_t_rms_D4	w_t_rms_D5	w_t_rms_D6	w_t_peak2peak_A6	w_t_peak2peak_D1	w_t_peak2peak_D2
0.0833	0.0146	0.0024	0.0014	0.0111	0.5785	1.6576

77	78	79	80	81	82
w_t_peak2peak_D3	w_t_peak2peak_D4	w_t_peak2peak_D5	w_t_peak2peak_D6	w_t_peak2rms_A6	w_t_peak2rms_D1
0.9253	0.1619	0.0149	0.0084	8.4087	10.1680

83	84	85	86	87	88
w_t_peak2rms_D2	w_t_peak2rms_D3	w_t_peak2rms_D4	w_t_peak2rms_D5	w_t_peak2rms_D6	w_t_meanEnergyCoeff
7.7558	5.7540	5.6869	3.4117	2.9994	0.0202

Figura 3.16: Características en el dominio de la escala, de una señal de tipo HB.

3.2. DISCUSIÓN

Obtenidas todas las características para cada evento, se procede a visualizar los datos en el t-SNE (t-Distributed Stochastic Neighbor Embedded), el cual se usa comúnmente para la reducción de la dimensión y permite la visualización de datos de alta dimensión. Esto se lo realiza con la finalidad de medir la calidad de las etiquetas, de tal manera que se pueda verificar si las etiquetas están correctas. Teniendo en las gráficas 3.17, 3.18, 3.19 y 3.20 la agrupación por tipo de evento para cada estación. En estas figuras se muestra un diagrama bidimensional, en donde se encuentran los 7788 eventos confirmados de la base de datos y su ubicación de acuerdo al tipo de evento al que pertenecen.

En la figura 3.17, se observa una clasterización de los eventos de tipo VT y LP de manera clara. Los eventos de tipo HB, se encuentran entre estos 2 tipos de eventos puesto que estos son una combinación de los dos tipos. Con los eventos de tipo TREMI se observa cierta confusión puesto que estos se encuentran más dispersos. También se puede ver ciertas superposiciones de eventos que potencialmente pueden ser producto de un mal etiquetado recordando que este es hecho de forma manual, de tal manera que se puede observar el ruido que existe en la base de datos.

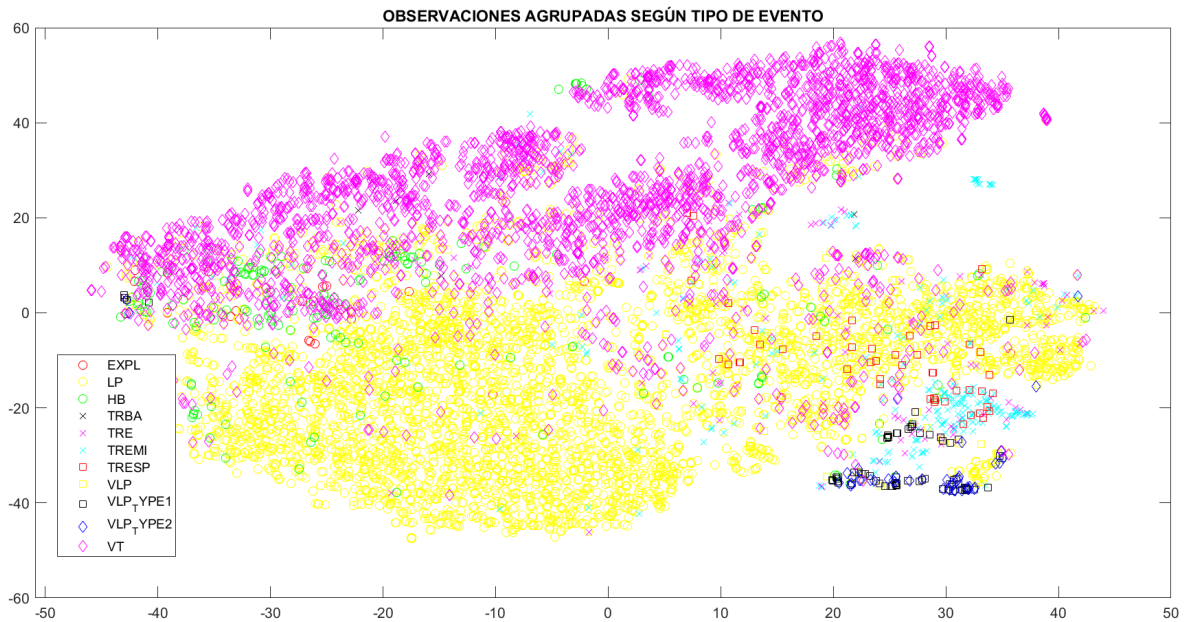


Figura 3.17: TSNE y gscatter sobre los datos de la estación BREF.

En la figura 3.18, se observa una claustrización de los eventos de tipo VT, LP, VLP TIPO 1, TRESP Y TREMI de manera clara. Los eventos de tipo HB, nuevamente se observan como entre los eventos tipo VT y LP. Igualmente se puede ver ciertas superposiciones de eventos que potencialmente pueden ser producto de un mal etiquetado recordando que este es hecho de forma manual, de tal manera que se puede observar que en esta estación también existe ruido.

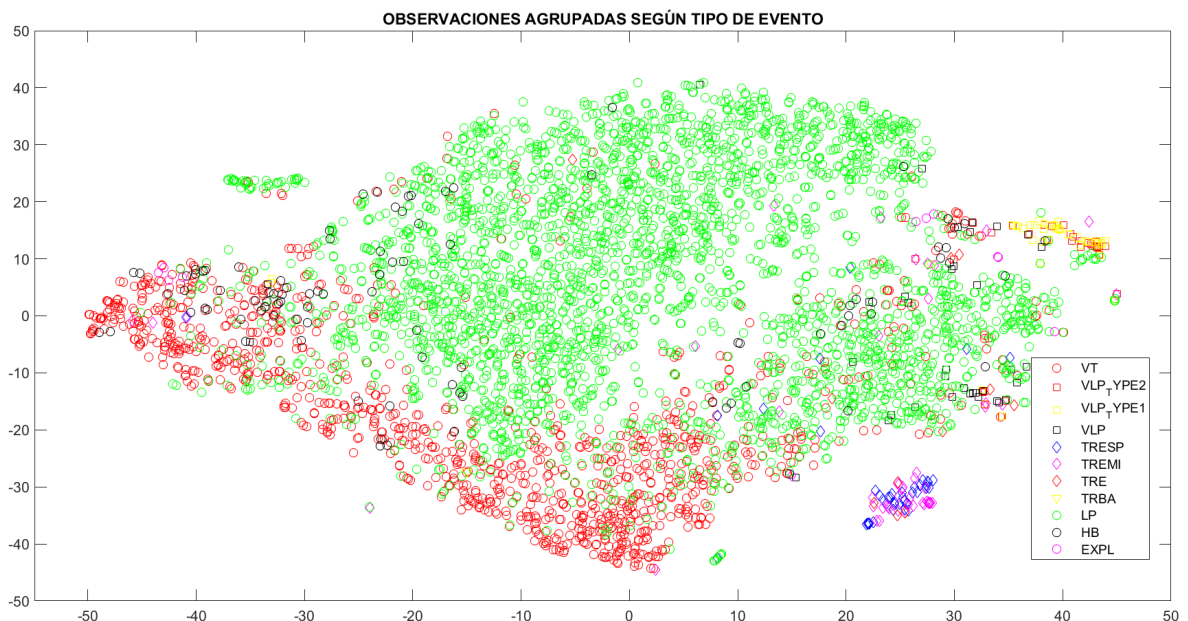


Figura 3.18: TSNE y gscatter sobre los datos de la estación BNAS.

En la figura 3.19, se observa una claustrización de los eventos de tipo VT, LP, TREMI, TRESP Y HB de manera clara. Los eventos de tipo VLP se ven un poco dispersos respecto

a los otros tipos de eventos.

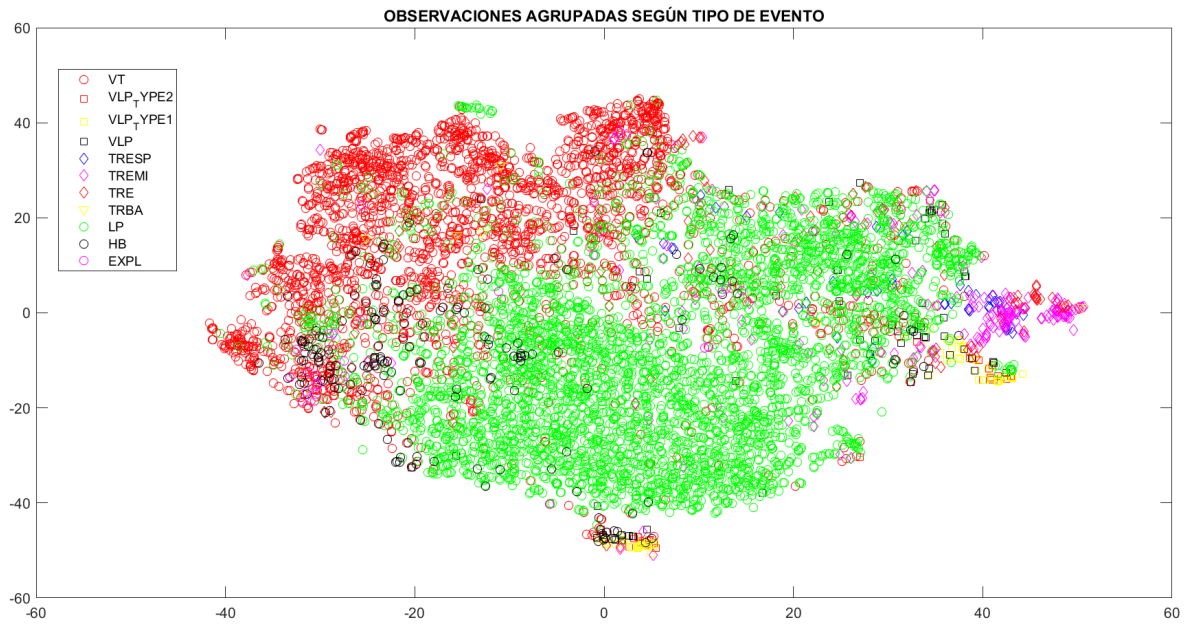


Figura 3.19: TSNE y gscatter sobre los datos de la estación BTAM.

En la figura 3.20, se observa una claustrización de los eventos de tipo VT, LP, de manera clara ya que existen mayor número de muestras de estos eventos. Los eventos de tipo VLP, VLP TIPO 1 y VLP TIPO 2 pueden verse también de manera más clara que en las otras estaciones agrupados en una pequeña sección. Igualmente se observa ciertas superposiciones de eventos.

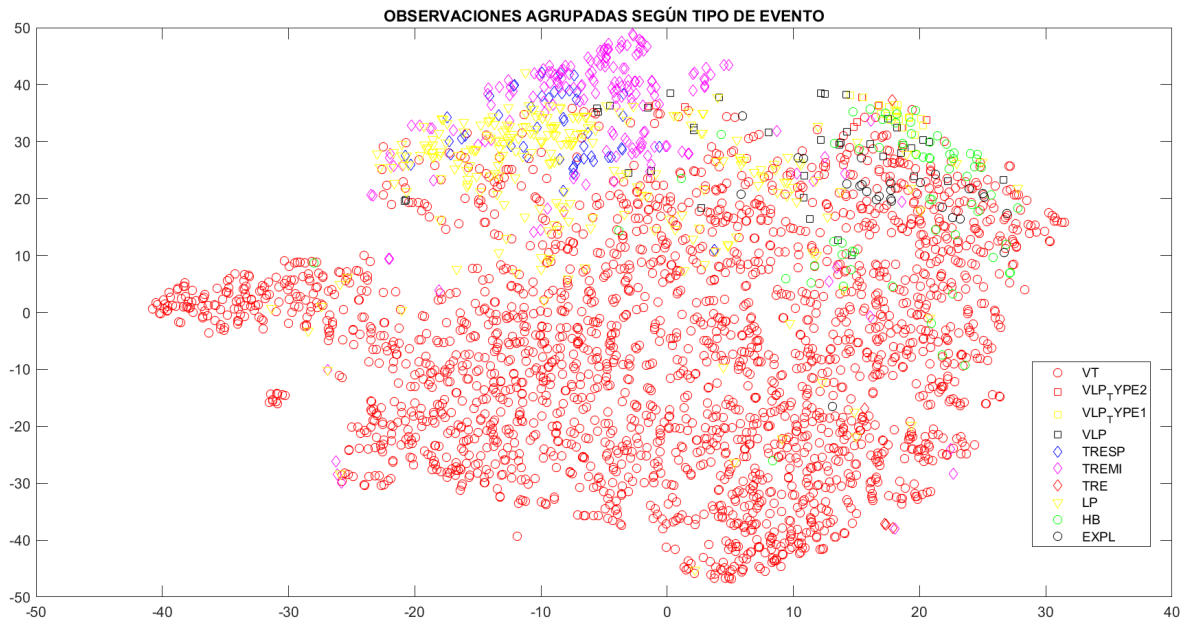


Figura 3.20: TSNE y gscatter sobre los datos de la estación BVC2.

En la figura 3.21, se muestra un diagrama bidimensional, en donde se tomaron al azar 100 muestras por cada estación, teniendo un total de 400 muestras. Se puede observar que las

muestras se ven agrupadas por estación pero no se ve una clasterización completa debido a que se tiene diferentes tipos de eventos dentro de estas muestras.

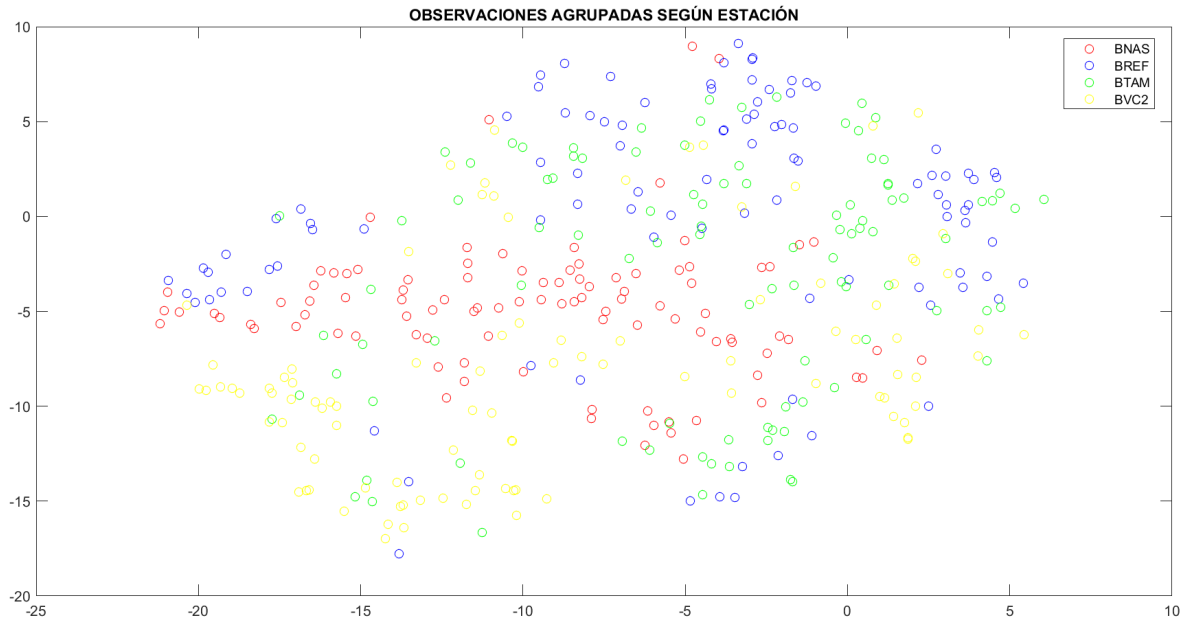


Figura 3.21: TSNE y gscatter tomando 100 muestras al azar por estación.

Por lo que se puede observar en la tabla 3.4, existe un alto porcentaje de descarte de eventos alrededor de un 70 %, esto debido a los requerimientos dados por el PhD Mario Ruiz, explicados anteriormente para la confirmación de eventos. De esta manera, se puede ver que el etiquetado manual no es eficiente si se toman en cuenta estos términos. También hay que considerar que en esta base de datos algunos registros están recortados o no existen en el archivo en formato .mseed, por lo que se pudo haber descartado varios eventos debido a la falta de registros.

El toolbox realizado y las bases de datos obtenidas mediante el uso de este han sido entregados al PhD. Mario Ruíz, quien ha certificado que el producto entregado cumple con los requerimientos solicitados y su conformidad con el mismo. Este certificado se adjunta en los anexos.

4. CONCLUSIONES Y RECOMENDACIONES

4.1. CONCLUSIONES

La conversión del formato mSEED a otros formatos como Matlab, es importante puesto que el formato mSEED se usa solo para almacenar o transferir datos sísmológicos entre instituciones de investigación de terremotos. Por lo tanto, para realizar un procesamiento de las señales sísmicas utilizando técnicas de aprendizaje automático es necesaria una conversión de los datos. Debido a esto la implementación de este proceso en el toolbox es de gran ayuda para todo aquel que desee trabajar con las bases de datos del IGEPN en un futuro.

La utilidad de este toolbox radica en la estandarización de los procesos y técnicas con las que se tratan las señales sísmicas existentes en las bases de datos del IGEPN. De tal manera que se obtiene reproducibilidad en las futuras investigaciones de eventos sísmicos y con estos estudios posteriormente poder realizar comparaciones de los resultados obtenidos.

Debido a que las señales pasan por procesos irreversibles es importante conservar los datos originales de los registros .mseed, ya que esto permite que se tenga trazabilidad e integridad de los datos. De manera que estos puedan ser utilizados en estudios posteriores aplicando técnicas de aprendizaje automático, herramienta que se ha vuelto fundamental en el análisis de señales sísmicas.

El escoger los correctos parámetros para el algoritmo de detección STA/LTA, asegura un funcionamiento correcto del toolbox en la detección de cualquier tipo de evento. Sin embargo, cabe recalcar que para una gran cantidad de señales en algunos de estos puede darse casos de detecciones erróneas, debido a que señales con múltiples choques pueden estar etiquetadas como un solo evento.

Las funciones de pre-procesamiento son importantes antes de la extracción de características puesto que estas permiten que las señales sean limpiadas y normalizadas para el correcto funcionamiento de las funciones empleadas en la extracción.

Para la extracción de características, el uso de la transformada de Wavelet en lugar de solo la transformada de Fourier hace que se puedan extraer características tanto en el dominio del tiempo como en el de la frecuencia, y esto permite que se pueda acceder de manera más eficiente a la información situada sobre la señal, mejorando el desempeño del análisis de señales sísmicas mediante el aprendizaje automático.

Las bases de datos del IGEPN contienen una gran cantidad de información y analizar cada una de estas bases de datos de manera manual requeriría de mucho tiempo y esfuerzo por parte de los investigadores, debido a esto la implementación del toolbox es de gran utilidad para la depuración de eventos sísmicos de manera rápida, teniendo claramente un beneficio para el IGEPN ya que le permite ahorrar tiempo y recursos.

Se pudo ver la necesidad de verificar el etiquetado manual realizado por el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), ya que como se mostró con la herramienta t-SNE, hay varios eventos que no fueron etiquetados correctamente, teniendo así señales que no pertenecen al tipo de evento en el cual se los clasifico lo cual puede provocar errores en posteriores investigaciones donde se empleen estas señales.

4.2. RECOMENDACIONES

Para un futuro trabajo se puede acoplar el toolbox en tiempo real lo cual permitirá que los archivo .mseed puedan ser convertidos inmediatamente sin necesidad de tener que acceder al directorio donde están almacenados. Igualmente se puede adaptar el toolbox para que se conecte automáticamente a las bases de datos del IGEPN que se van creando manualmente y de esta manera hacer que la búsqueda y extracción del evento sea inmediata.

Para estudios posteriores se propone realizar funciones estandarizadas de aprendizaje automático para la confirmación de los eventos sísmicos contenidos en las bases del IGEPN, de manera que se pueda seguir teniendo reproducibilidad en las investigaciones.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] N. Pérez, D. Benítez, F. Grijalva, R. Lara-Cueva, M. Ruiz, and J. Aguilar, “Eseismic: Towards an ecuadorian volcano seismic repository,” *Journal of Volcanology and Geothermal Research*, vol. 396, p. 106855, 2020.
- [2] P. Venegas, N. Pèrez, D. S. Benítez, R. Lara-Cueva, and M. Ruiz, “Building machine learning models for long-period and volcano-tectonic event classification,” in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. IEEE, 2019, pp. 1–6.
- [3] M. Malfante, M. Dalla Mura, J.-P. Métaxian, J. I. Mars, O. Macedo, and A. Inza, “Machine learning for volcano-seismic signals: Challenges and perspectives,” *IEEE Signal Processing Magazine*, vol. 35, no. 2, pp. 20–30, 2018.
- [4] R. A. Lara-Cueva, D. S. Benítez, E. V. Carrera, M. Ruiz, and J. L. Rojo-Álvarez, “Automatic recognition of long period events from volcano tectonic earthquakes at cotopaxi volcano,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5247–5257, 2016.
- [5] I. I. R. I. for Seismology, “Iris: Seed,” 2021. [Online]. Disponible en: <https://ds.iris.edu/ds/nodes/dmc/data/formats/seed/>
- [6] M. of the Federation of Digital Seismographic Networks, “Seedmanual,” 2012. [Online]. Disponible en: http://www.fdsn.org/pdf/SEEDManual_V2.4.pdf
- [7] MathWorks, “Readmseedfast,” Febrero 2019. [Online]. Disponible en: <https://la.mathworks.com/matlabcentral/fileexchange/46532-readmseedfast>.
- [8] Y. J. E. BRAVO, “Estudio comparativo de algoritmos para la segmentación de señales volcánicas orientado a la clasificación de sismos,” 2019.
- [9] P. E. 1er Cuat, “Análisis de señales geológicas señales sísmicas,” Master’s thesis, Junio 2011.
- [10] G. L. Arias Méndez, “Estudio de las señales sísmicas de muy largo período del volcán cotopaxi,” B.S. thesis, Quito, 2016., 2016.
- [11] S. Petrosino, P. Cusano, M. La Rocca, D. Galluzzo, J. Orozco-Rojas, M. Bretón, J. Ibáñez, and E. Del Pezzo, “Source location of long period seismicity at volcàn de colima, méxico,” *Bulletin of volcanology*, vol. 73, no. 7, pp. 887–898, 2011.
- [12] H. Kumagai and B. A. Chouet, “Acoustic properties of a crack containing magmatic or hydrothermal fluids,” *Journal of Geophysical Research: Solid Earth*, vol. 105, no. B11, pp. 25 493–25 512, 2000.
- [13] J. Ibáñez and E. Carmona, “Sismicidad volcánica,” *Curso Internacional de Volcanología y Geofísica Volcánica, Serie Casa de los Volcanes*, vol. 7, pp. 269–282, 2000.

- [14] B. A. Chouet, "Long-period volcano seismicity: its source and use in eruption forecasting," *Nature*, vol. 380, no. 6572, pp. 309–316, 1996.
- [15] I. Londoño and R. Lara, "Implementación de un sistema de monitorización de señales sísmicas del volcán cotopaxi empleando una red de sensores inalámbricos," *Escuela Politécnica del Ejército. Proyecto de grado*, 2011.
- [16] AbrirArchivos, ".seed extensión de archivo." [Online]. Disponible en: <https://abrirarchivos.info/extension/seed>
- [17] V. P. BROSSARD, "Título: Análisis del ruido sísmico en el servicio sismológico nacional cubano," Ph.D. dissertation, Universidad de Oriente, 2018.
- [18] M. Mariotti and T. Utheim, "A white paper about miniseed for liss and data compression using steim1 and steim2," *Dept. of Earth Science, University of Bergen*, 2006.
- [19] K. S. C. BENAVIDES, "Facultad de ingeniería carrera de electrónica y telecomunicaciones," Ph.D. dissertation, Universidad de Cuenca, 2015.
- [20] A. Trnkoczy, "Understanding and parameter setting of sta/lta trigger algorithm," in *New Manual of Seismological Observatory Practice (NMSOP)*. Deutsches GeoForschungs-Zentrum GFZ, 2009, pp. 1–20.
- [21] L. Gallo Rivera, "Implementación de un sistema para la evaluación del efecto de sitio del suelo del edificio 350," 2017.
- [22] A. S. PÉREZ QUISAGUANO, "Detección automática de eventos sísmicos en el volcán cotopaxi mediante técnicas de aprendizaje de maquinas," 2018.
- [23] G. A. Rousselet, "Does filtering preclude us from studying erp time-courses?" *Frontiers in psychology*, vol. 3, p. 131, 2012.
- [24] C. A. Brusil Cruz, "Análisis comparativo entre aprendizaje supervisado y aprendizaje semi-supervisado para la clasificación de señales sísmicas vulcanológicas del volcán cotopaxi," B.S. thesis, Quito, 2020., 2020.
- [25] mathworks, "rms." [Online]. Disponible en: <https://la.mathworks.com/help/signal/ref/rms.html>
- [26] —, "periodogram." [Online]. Disponible en: <https://la.mathworks.com/help/signal/ref/periodogram.html#btt5utg>
- [27] P. Gomis Román, "Estimación espectral de señales biomédicas. métodos clásicos (fft) y paramétricos: aplicaciones prácticas con matlab," 2013.
- [28] mathworks, "pwelch." [Online]. Disponible en: <https://la.mathworks.com/help/signal/ref/pwelch.html>

- [29] M. Bayas Paredes, “Modelos autorregresivos de análisis espectral,” B.S. thesis, QUITO/EPN/1984, 1984.
- [30] S. M. Kay and S. L. Marple, “Spectrum analysis—a modern perspective,” *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.
- [31] Matlab, “arburg.” [Online]. Disponible en: <https://la.mathworks.com/help/signal/ref/arburg.html>
- [32] C. Torrence and G. P. Compo, “A practical guide to wavelet analysis,” *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
- [33] A. Dixit and S. Majumdar, “Comparative analysis of coiflet and daubechies wavelets using global threshold for image denoising,” *International Journal of Advances in Engineering & Technology*, vol. 6, no. 5, p. 2247, 2013.
- [34] M. Wattenberg, F. Viégas, and I. Johnson, “How to use t-sne effectively,” *Distill*, vol. 1, no. 10, p. e2, 2016.
- [35] Matlab, “tsne settings.” [Online]. Disponible en: <https://la.mathworks.com/help/stats/tsne-settings.html>
- [36] —, “spectrogram.” [Online]. Disponible en: https://la.mathworks.com/help/signal/ref/spectrogram.html#bultpz_-2
- [37] D. Ketner, “stalta2,” 2016. [Online]. Disponible en: https://github.com/daneketner/SEMS/blob/master/Detection/sta_lta_2.m
- [38] J. Simon, “Getmd5,” 2019. [Online]. Disponible en: <https://la.mathworks.com/matlabcentral/fileexchange/25921-getmd5>
- [39] Matlab, “wavedec.” [Online]. Disponible en: <https://la.mathworks.com/help/wavelet/ref/wavedec.html#d123e106966>
- [40] I. I. R. I. for Seismology, “Iris: Seed.” [Online]. Disponible en: <https://ds.iris.edu/ds/nodes/dmc/data/formats/seed-channel-naming/>

ANEXOS

ANEXO A. Tabla con funciones y variables implementadas en el toolbox.

ANEXO B. Certificado institución auspiciante.

Los siguientes anexos se encuentran en formato digital:

Toolbox para procesamiento de señales sísmicas. (Formato digital)

Manual de usuario. (Formato Digital)

ANEXO A

TABLA CON LAS FUNCIONES Y VARIABLES IMPLEMENTADAS EN EL TOOLBOX

Tabla 6.1: Descripción de las funciones implementadas en el toolbox de matlab para procesamiento de señales sísmicas.

Funciones	Descripción	Variables de entrada y salida
seed_to_mat_conversion	Abre el directorio que contiene los archivos mseed lee estos archivos y devuelve los archivos .mat con un nombre único y almacenados en un directorio.	<p>Entradas: La función recibe dos variables de tipo string.</p> <ul style="list-style-type: none"> • dir_mseed: Nombre del directorio donde se encuentran los archivos de tipo mseed. • dir_matlab: Nombre del directorio donde se guardaran los archivos de tipo mat. <p>Salidas: La función retorna un archivo .mat, que contiene argumentos de tipo string, int y double.</p>
import_from_excel	Crea una tabla con los eventos registrados por el IGEPN.	<p>Entradas: La función recibe un archivo de tipo .xlsx.</p> <ul style="list-style-type: none"> • excel_file (char): Archivo que contiene los eventos registrados por el IGEPN. <p>Salidas: La función retorna un arreglo en formato de tabla que contiene los eventos.</p>
remove_duplicates	Retira eventos que puedan estar duplicados en los registros del IGEPN.	<p>Entradas: La función recibe un arreglo de tipo tabla.</p> <ul style="list-style-type: none"> • t_from_excel: Tabla con eventos registrados por el IGEPN. <p>Salidas: La función retorna un arreglo en formato de tabla sin eventos duplicados.</p>

Continúa en la siguiente página

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
search_and_extract	<p>Busca en los archivos .mat, los eventos detectados por el IGEPN, una vez localizados extrae su data, puntos de inicio y fin, frecuencia de muestreo, tipo y duración para crea una tabla organizada por canales y estaciones.</p>	<p>Entradas: La función recibe 3 variables un arreglo de tipo tabla, un string y un arreglo de estructura.</p> <ul style="list-style-type: none"> • t_from_excel: Tabla con eventos registrados por el IGEPN. • dir_matlab(string): Nombre del directorio donde se encuentran los archivos de tipo mat. • detection_parameter(struct): <ul style="list-style-type: none"> sta: Tamaño de ventana sta. Tipo double. lta: Tamaño de ventana lta. Tipo double. th_on: Trigger on. Tipo double. th_off: Trigger off. Tipo double. skip_int: Salto hacia adelante después del final del evento en segundos. Tipo double. min_dur: Duración mínima de un evento en segundos. Tipo double. pem: Valor de memoria pre evento en segundos. Tipo double. pet: Valor de memoria post evento en segundos. Tipo double. • reference_station(char): Estación de referencia para confirmar o descartar eventos. <p>Salidas: La función retorna un arreglo en formato de tabla con la data de eventos confirmados y los descartados.</p>
Continúa en la siguiente página		

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
trim_signal	Recorta la señal en los puntos de inicio y fin de los eventos.	<p>Entradas: La función recibe 4 variables un arreglo tipo celda con tipo de datos double, 2 variables de tipo double y un vector con índices lineales.</p> <ul style="list-style-type: none"> ● signals_t(double): Señal pura de los eventos confirmados. ● Start_Point_all(double): Punto de inicio de un evento, tomado del canal BHZ de cada estación. ● End_Point_all(double): Punto final de un evento obtenido al tomar el valor más alto de todos los canales de una estación. ● idx(vector): Índice que permite hallar los eventos en los que existe punto de inicio y de fin. <p>Salidas: La función retorna un tipo de dato double de la señal recortada.</p>
remove_nan	Retira los valores de tipo NaN de cada señal.	<p>Entradas: La función recibe una variable de tipo double.</p> <ul style="list-style-type: none"> ● signals_t_trimmed: Señal recortada en puntos de inicio y fin del evento. <p>Salidas: La función retorna un tipo de dato double de la señal sin valores de tipo NaN.</p>
Continúa en la siguiente página		

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
normalize_sampling_rate	Normaliza todas las señales a una misma frecuencia de muestreo.	<p>Entradas: La función recibe 4 variables de tipo double.</p> <ul style="list-style-type: none"> ● signals_t_filter (double): Señal filtrada. ● Sample_Rate_all(double): Frecuencia de muestreo propia de cada canal. ● fs(double): Frecuencia de muestreo a la que se desea remuestrear a la señal. ● idx(double): Índice que permite hallar los eventos en los que existe punto de inicio y de fin. <p>Salidas: La función retorna un tipo de dato double de la señal remuestreada a una misma frecuencia.</p>
Continúa en la siguiente página		

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
filter_signal	Filtra la señal, usando un filtro pasabanda de tipo butter o fir1 .	<p>Entradas: La función recibe 5 variables de tipo double y una de tipo char.</p> <ul style="list-style-type: none"> ● remove_nan_trimmed(double): Señal sin valores de tipo NaN. ● Fc_low(double): Frecuencia de corte baja. ● Fc_high(double): Frecuencia de corte alta. ● Fs(double): Frecuencia de muestreo a la que se desea remuestrear la señal. ● order(double): Orden del filtro pasabanda. ● type(char): Escoja el tipo de filtro que desee usar Se ingresa 'butterworth', para utilizar un filtro tipo IIR pasabandas. Se ingresa 'FIR', para utilizar un filtro tipo FIR pasabandas. <p>Salidas: La función retorna un tipo de dato double de la señal filtrada.</p>
Continúa en la siguiente página		

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
normalize_signal	Normalizar la señal con 3 diferentes tipos de métodos.	<p>Entradas: La función recibe 1 variable de tipo double y tipo char.</p> <ul style="list-style-type: none"> • signals_t_filter (double): Señal con la nueva frecuencia de muestreo. • type(char): Escoge el método de normalización a usar. <p>Se ingresa 'Max', para normalizar con el valor máximo.</p> <p>Se ingresa 'Energy', para normalizar con la energía.</p> <p>Se ingresa 'RMS', para normalizar con el valor RMS de la señal.</p> <p>Salidas: La función retorna un tipo de dato double de la señal normalizada.</p>
Continúa en la siguiente página		

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
<p>calculate_features</p>	<p>Calcula características en tiempo, frecuencia y escala de la señales.</p>	<p>Entradas: La función recibe 3 o 4 variables, 2 de tipo double, una de tipo estructura y una tipo char.</p> <ul style="list-style-type: none"> • signals_t_resampled_normalized (double): Señal normalizada con la nueva frecuencia de muestreo. • fs(double): Frecuencia de muestreo nueva. • parameter_PSD (estructura): Parámetros para calculo de la PSD. method (char): Se puede escoger entre tres métodos 'periodogram', 'welch' o 'burg'. NFFT (double): Numero de puntos de la TDF. order (double): Orden del modelo auto regresivo. noverlap (double): Número de muestras superpuestas. • feature(char): Para obtener las 84 características no se ingresa nada, caso contrario se ingresa lo siguiente: Se ingresa 'Time', para calcular las 13 características en tiempo. Se ingresa 'Frequency', para calcular las 21 características en frecuencia. Se ingresa 'Scale', para calcular 50 características de dominio de escala. <p>Salidas: La función retorna un arreglo en formato tabla que contiene las señales con sus características.</p>

Continúa en la siguiente página

Tabla 6.1 – Continuación de la página anterior

Funciones	Descripción	Variables de entrada y salida
signals_with_characteristics	Crea una tabla que contiene los eventos con sus respectivas características.	<p>Entradas: La función recibe 5 variables, 3 de tipo char, 1 tabla y 1 dato de tipo double.</p> <ul style="list-style-type: none"> • Event_Identifier: Identificador único para cada evento. • Type: Tipo de evento. • station_channel: Estación de cada evento. • features: Tabla con las características calculadas de cada señal. • idx: Índice que permite hallar los eventos en los que existe punto de inicio y de fin. <p>Salidas: La función retorna un arreglo en formato tabla que contiene los eventos confirmados con sus respectivas características, la estación, tipo de evento y su identificador único.</p>

ANEXO B



**ESCUELA POLITÉCNICA NACIONAL
"CAMPUS POLITÉCNICO JOSÉ RUBÉN ORELLANA RICAURTE"
INSTITUTO GEOFÍSICO**

Quito, 6 de septiembre de 2021

Certificado de Culminación de la Institución Auspiciante

Yo, **Mario Calixto Ruiz Romero**, en mi calidad de Docente Investigador del Instituto Geofísico (IG) de la Escuela Politécnica Nacional, unidad auspiciante del proyecto de titulación

"IMPLEMENTACIÓN DE UN TOOLBOX DE MATLAB PARA EL PROCESAMIENTO DE SEÑALES SÍSMICAS A PARTIR DE LAS BASES DE DATOS DEL INSTITUTO GEOFÍSICO DE LA ESCUELA POLITÉCNICA NACIONAL",

elaborado por la señorita **Jennifer Patricia Condor Navas**, estudiante de la carrera de **Ingeniería en Electrónica y Telecomunicaciones**, y dirigido por el **Ph.D Felipe Grijalva**, docente del Departamento de Electrónica, Telecomunicaciones y Redes de Información, certifico que el producto final demostrable del mencionado trabajo de titulación ha sido culminado al 100% y está acorde a las requerimientos del Instituto Geofísico.

**MARIO
CALIXTO RUIZ
ROMERO**

Firmado digitalmente por MARIO
CALIXTO RUIZ ROMERO
Nombre de reconocimiento (DN): c=EC,
o=SECURITY DATA S.A. 1, ou=ENTIDAD
DE CERTIFICACION DE INFORMACION,
serialNumber=161120112007, cn=MARIO
CALIXTO RUIZ ROMERO
Fecha: 2021.09.06 08:24:28 -05'00'

Ph.D. Mario Calixto Ruiz Romero
Investigador del Instituto Geofísico de la Escuela Politécnica Nacional

ORDEN DE EMPASTADO