

**Creación de bases de datos sintéticas de eventos sísmicos producidos en el volcán  
Llaima con Bootstrap.**

Caisaluisa Hurtado, Henry Fabricio

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en  
Telecomunicaciones

Ing. Lara Cueva, Román Alcides PhD.

25 de agosto de 2023

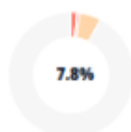


### O5\_Caisaluisa\_Fabricio.pdf

#### Scan details

Scan time: August 25th, 2023 at 18:56 UTC      Total Pages: 60      Total Words: 14953

#### Plagiarism Detection



Types of plagiarism		Words
Identical	1.5%	221
Minor Changes	0.7%	109
Paraphrased	5.6%	830
Omitted Words	0%	0

#### AI Content Detection



Text coverage  
 ● AI text  
 ○ Human text

#### Plagiarism Results: (57)

##### Your File

1.6%

MARIANITA

No introduction available.

##### T-ESPE-019175.pdf

1.2%

<https://repositorio.espe.edu.ec/bitstream/21000/28429/1/t-e...>

Ariel Villacís

1 Generación de señales sintéticas de eventos sismo-volcánicos del volcán Cotopaxi mediante el método de Bootstrap. Pérez Moreno, Brend...



ROMAN ALCIDES LARA CUEVA

**Ing. Román Alcides Lara Cueva, PhD.**  
**Director**



**Departamento de Eléctrica, Electrónica y Telecomunicaciones**  
**Carrera de Telecomunicaciones**

**Certificación**

Certifico que el trabajo de integración curricular: "**Creación de bases de datos sintéticas de eventos sísmicos producidos en el volcán Llaima con Bootstrap**" fue realizado por el señor **Caisaluisa Hurtado, Henry Fabricio**, el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 25 de agosto de 2023



**Ing. Román Alcides Lara Cueva, PhD.**

C. C: 1713988218



**Departamento de Eléctrica, Electrónica y Telecomunicaciones**

**Carrera de Telecomunicaciones**

**Responsabilidad de Autoría**

Yo, **Caisaluisa Hurtado, Henry Fabricio**, con cédula de ciudadanía n° 1722552443, declaro que el contenido, ideas y criterios del trabajo de integración curricular: **Creación de bases de datos sintéticas de eventos sísmicos producidos en el volcán Llaima con Bootstrap** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

**Sangolquí, 25 de agosto de 2023**

**Caisaluisa Hurtado, Henry Fabricio**

C.C.: 1722552443



**Departamento de Eléctrica, Electrónica y Telecomunicaciones**  
**Carrera de Telecomunicaciones**

**Autorización de Publicación**

Yo **Caisaluisa Hurtado, Henry Fabricio**, con cédula de ciudadanía n° 1722552443, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **Creación de bases de datos sintéticas de eventos sísmicos producidos en el volcán Llaima con Bootstrap** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

**Sangolquí, 25 de agosto de 2023**

**Caisaluisa Hurtado, Henry Fabricio**

C.C.: 1722552443

### **Dedicatoria**

El presente trabajo está dedicado a las personas más importantes en mi vida, mi madre Esmeralda y mi hermano Juan Carlos, los cuales nunca han dudado ni una sola vez de lo lejos que puedo llegar, siendo mi apoyo moral en todo momento de esta larga travesía.

A mi mascota Cassie que ha estado conmigo cada mala noche motivándome con un maullido y durmiendo sobre mis cuadernos porque ya no pudo mantenerse despierta.

Quiero expresarles mi agradecimiento, admiración y gratitud, porque gracias ha ellos me he podido levantar de las adversidades.

***Henry Fabricio Caisaluisa Hurtado***

## **Agradecimiento**

Quiero agradecer a cada persona que ha formado parte de mi vida en esta Universidad, agradecer a los docentes por compartirnos su conocimiento, anécdotas y experiencia en el transcurso de estos años. Espero encontrarlos en un futuro, siendo un excelente profesional. Agradezco de todo corazón a mi familia por ayudarme a culminar mis metas y estar conmigo en el final de este largo camino.

***Henry Fabricio Caisaluisa Hurtado***

## Índice de Contenido

Resumen.....	16
Abstract.....	17
Capítulo I.....	18
Introducción .....	18
Antecedentes.....	20
Justificación e Importancia.....	22
Objetivo .....	23
Actividades.....	23
Capítulo II.....	24
Materiales y Métodos .....	24
Materiales .....	24
Identificación de propiedades y variables.....	24
Volcán Llama .....	27
Características geológicas del volcán Llama .....	28
Reunión de la base de datos representativa .....	29
Densidad Espectral de Potencia .....	29
Eventos Microsísmicos.....	30
Método Bootstrap.....	35



Ventajas del Método.....	35
Desventajas del Método .....	36
Preprocesamiento.....	36
Filtrado .....	37
Estandarizado .....	38
Normalizado .....	39
Transformada de Fourier .....	40
Análisis Matemático de la Transformada de Fourier .....	41
Análisis Matemático de la Transformada Inversa de Fourier .....	43
Capítulo III.....	45
Diseño e Implementación .....	45
Caracterización de microsismos .....	45
Transformada de Fourier de las señales.....	45
Selección de fases.....	50
Normalizar las Magnitudes.....	55
Método de Bootstrap a las Magnitudes .....	55
Igualar número de Fases solicitadas.....	71
Unión Magnitud sintética y Fase real .....	72
Transformada Inversa de Fourier.....	72

	10
Recuperación del desfase.....	73
Desfasar el peor de los casos .....	75
Desfase por media .....	77
Diseño Aplicativo de Interfaz Gráfica .....	79
Uso de la Aplicación .....	79
Capítulo IV .....	83
Pruebas y Resultados .....	83
Resultados señales sintéticas.....	84
Evento LP.....	85
Evento TC .....	86
Evento TR .....	87
Evento VT .....	89
Evaluación de desempeño por expertos del Instituto Geofísico de la Escuela Politécnica Nacional .....	90
Resultados del clasificador .....	98
Capítulo V .....	106
Conclusiones y Trabajos Futuros .....	106
Conclusiones .....	106
Trabajos Futuros.....	107
Referencias .....	109

Apéndices .....113

## Índice de Tablas

<b>Tabla 1</b> Especificaciones del computador.....	24
<b>Tabla 2</b> Distribución de Base de datos de eventos microsísmicos.....	34
<b>Tabla 3</b> Identificación de datos ideales para seleccionar fases lineales.....	52
<b>Tabla 4</b> Cantidad de fases ideales, después de realizar el proceso de selección .....	55
<b>Tabla 5</b> Parámetros mínimos de Bootstrap.....	85
<b>Tabla 6</b> Distribución de las señales sintéticas y reales en el formulario enviado al IGEPN .....	92
<b>Tabla 7</b> Resultados obtenidos a partir del formulario respondido por los expertos del IGEPN en relación a las señales sintéticas y reales.....	93
<b>Tabla 8</b> Resultados de las señales sintéticas en términos del porcentaje de engaño .....	95
<b>Tabla 9</b> Porcentaje total de engaño a los expertos del IGEPN según el tipo de evento microsísmico.....	96
<b>Tabla 10</b> Tipos de Eventos que Generan Confusión entre los Expertos del IGEPN.....	97
<b>Tabla 11</b> Porcentaje total de confusión en los eventos .....	98
<b>Tabla 12</b> Métricas de evaluación del rendimiento del clasificador DT .....	101
<b>Tabla 13</b> Métricas de evaluación del rendimiento del clasificador k-NN.....	103
<b>Tabla 14</b> Métricas de evaluación del rendimiento del clasificador SVM .....	104
<b>Tabla 15</b> Métricas de evaluación del sistema de votación .....	105

## Índice de Figuras

<b>Figura 1</b> Diagrama de bloque de la metodología propuesta .....	26
<b>Figura 2</b> Foto del volcán Llaima visto desde río Blanco.....	28
<b>Figura 3</b> Evento microsísmico Volcano tectónico .....	30
<b>Figura 4</b> Evento microsísmico Largo Período.....	31
<b>Figura 5</b> Evento microsísmico Tremor.....	32
<b>Figura 6</b> Evento microsísmico Tectónico.....	33
<b>Figura 7</b> Diagrama de flujo de etapa de preprocesamiento de la base de datos.....	36
<b>Figura 8</b> Señales microsísmicas antes y después de pasar por la etapa de preprocesamiento .....	40
<b>Figura 9</b> Representación temporal y en frecuencia de señal LP.....	46
<b>Figura 10</b> Representación temporal y en frecuencia de señal TC.....	47
<b>Figura 11</b> Representación temporal y en frecuencia de señal TR.....	48
<b>Figura 12</b> Representación temporal y en frecuencia de señal VT.....	48
<b>Figura 13</b> Fases totales de los 1488 eventos microsísmicos TC .....	50
<b>Figura 14</b> Fases totales de los 304 eventos microsísmicos VT .....	51
<b>Figura 15</b> Fase seleccionada para generar una pendiente ideal perteneciente al evento TR ..	52
<b>Figura 16</b> Zona delimitada alrededor de la fase ideal para el evento TC .....	54
<b>Figura 17</b> Ejemplo de proceso Bootstrap.....	57
<b>Figura 18</b> Magnitud sintética LP generada mediante el método de Bootstrap .....	58
<b>Figura 19</b> Representación de las Magnitudes reales e intervalo de confianza del evento LP ..	59
<b>Figura 20</b> Representación de las Magnitudes sintéticas del evento LP .....	60
<b>Figura 21</b> Valor medio e intervalo de confianza de las magnitudes sintéticas del evento LP ...	61
<b>Figura 22</b> Representación de las Magnitudes reales e intervalo de confianza del evento TC..	63
<b>Figura 23</b> Representación de las Magnitudes sintéticas del evento TC.....	64

<b>Figura 24</b>	Valor medio e intervalo de confianza de las magnitudes sintéticas del evento TC...	64
<b>Figura 25</b>	Representación de las Magnitudes reales e intervalo de confianza del evento TR..	65
<b>Figura 26</b>	Representación de las Magnitudes sintéticas del evento TR.....	66
<b>Figura 27</b>	Valor medio e intervalo de confianza de las magnitudes sintéticas del evento TR...	67
<b>Figura 28</b>	Representación de las Magnitudes reales e intervalo de confianza del evento VT ..	68
<b>Figura 29</b>	Representación de las Magnitudes sintéticas del evento VT .....	69
<b>Figura 30</b>	Valor medio e intervalo de confianza de las magnitudes sintéticas del evento VT...	70
<b>Figura 31</b>	Señal sintética correspondiente al evento LP con desfase .....	73
<b>Figura 32</b>	Señal sintética correspondiente al evento VT con desfase.....	74
<b>Figura 33</b>	Señal sintética correspondiente al evento TR.....	75
<b>Figura 34</b>	Recuperar el desfase de la señal sintética mediante proceso en los límites de los rangos.....	75
<b>Figura 35</b>	Propuesta de recuperar el desfase al considerar el peor de los casos .....	76
<b>Figura 36</b>	Media de las fases ideales LP utilizada para recuperar el desfase .....	77
<b>Figura 37</b>	Desfase recuperado con éxito en los distintos tipos de eventos microsísmicos.....	78
<b>Figura 38</b>	Diagrama de flujo de la aplicación.....	79
<b>Figura 39</b>	Interfaz gráfica: Pantalla de inicio.....	80
<b>Figura 40</b>	Interfaz gráfica: Pantalla de sistema de generación de señales sintéticas.....	81
<b>Figura 41</b>	Base de datos sintética generada por el método de Bootstrap .....	82
<b>Figura 42</b>	Eliminación del pico en 0.2 Hz mediante el preprocesamiento de las señales.....	83
<b>Figura 43</b>	Señal original microsísmica de tipo LP, su PSD y forma temporal.....	85
<b>Figura 44</b>	Señal sintética microsísmica de tipo LP, su PSD y forma temporal .....	85
<b>Figura 45</b>	Señal original microsísmica de tipo TC, su PSD y forma temporal .....	86
<b>Figura 46</b>	Señal sintética microsísmica de tipo TC, su PSD y forma temporal.....	87
<b>Figura 47</b>	Señal original microsísmica de tipo TR, su PSD y forma temporal .....	87

<b>Figura 48</b> Señal sintética microsísmica de tipo TR, su PSD y forma temporal.....	88
<b>Figura 49</b> Señal original microsísmica de tipo VT, su PSD y forma temporal .....	89
<b>Figura 50</b> Señal sintética microsísmica de tipo VT, su PSD y forma temporal.....	89
<b>Figura 51</b> Formulario enviado a los expertos del IGEPN .....	91
<b>Figura 52</b> Ejemplo de pregunta incluida en el formulario enviado, señal N.019.....	92
<b>Figura 53</b> Resultados de la clasificación de las 10000 señales sintéticas al utilizar DT.....	100
<b>Figura 54</b> Resultados de la clasificación de las 10000 señales sintéticas al utilizar k-NN.....	102
<b>Figura 55</b> Resultados de la clasificación de las 10000 señales sintéticas al utilizar SVM.....	103
<b>Figura 56</b> Resultados del sistema de votación de los tres diferentes clasificadores .....	104

## Resumen

El Ecuador se encuentra situado en una región altamente propensa a actividad sísmica debido a su ubicación dentro del cinturón de Fuego del Pacífico, en este contexto resulta crucial establecer un sistema de monitorización efectivo con el propósito de prevenir erupciones y mitigar sus impactos. Sin embargo, las bases de datos de volcanes suelen ser limitadas en cuanto a la cantidad de datos disponibles para este fin. Una alternativa factible es desarrollar bases de datos sintéticas, las cuales, a partir de la cantidad de datos existentes, permitan generar nuevos datos que conserven las mismas características de los datos originales.

Este estudio explora el método de muestreo aleatorio con la técnica de Bootstrap junto a la transformada de Fourier, para generar datos sintéticos de los eventos microsísmicos; Largo Período (LP), Volcano Tectónicos (VT), Tremor (TR) y Tectónicos (TC) producidos en el volcán Llaima, a partir de una base de datos registrados por las estaciones sismológicas y provistos por el observatorio "Observatorio Volcanológico de los Andes del Sur". El objetivo es generar señales sintéticas que repliquen con precisión las características distintivas de cada evento, para asegurar la preservación de las características temporales y espectrales, se emplearon magnitudes sintéticas en conjunto con fases reales selectas.

Finalmente se obtienen señales sintéticas que fueron evaluadas por expertos del Instituto Geofísico de la Escuela Politécnica Nacional con resultados de: 94.45% en LP, 0.00% en TC, 55.53% en TR y 71.44% en VT. Este porcentaje ilustra la habilidad de las señales sintéticas para pasar desapercibidas como señales reales. De igual manera se evaluaron mediante el uso de los clasificadores desarrollados en paralelo en el grupo de investigación, con resultados del 99.90% de exactitud en cada una de las señales sintéticas.

*Palabras clave:* Bootstrap, Llaima, Datos sintéticos, Muestreo aleatorio, Transformada de Fourier.



## Abstract

Ecuador is located in a highly seismic-prone region due to its position within the Pacific Ring of Fire. In this context, establishing an effective monitoring system is crucial to prevent eruptions and mitigate their impacts. However, volcano databases often have limited data available for this purpose. A feasible alternative is to develop synthetic databases which, based on the existing data, can generate new data that retain the same characteristics as the original data.

This study explores the method of random sampling using the Bootstrap technique along with the Fourier transform to generate synthetic data for microseismic events: Long Period (LP), Volcano Tectonic (VT), Tremor (TR), and Tectonic (TC) events produced at the Llaima volcano. This is done using a database recorded by seismic stations and provided by the "Observatorio Volcanológico de los Andes del Sur". The objective is to generate synthetic signals that accurately replicate the distinctive characteristics of each event. To ensure the preservation of temporal and spectral features, synthetic magnitudes were used in conjunction with selected real phases.

Finally, synthetic signals were obtained and evaluated by experts from the Geophysical Institute of the National Polytechnic School. The results were as follows: 94.45% accuracy for LP, 0.00% for TC, 55.53% for TR, and 71.44% for VT. This percentage illustrates the ability of synthetic signals to go unnoticed as real signals. Similarly, they were evaluated using classifiers developed in parallel within the research group, yielding results of 99.90% accuracy for each of the synthetic signals.

*Keywords:* Bootstrap, Llaima, Synthetic data, Random sampling, Fourier transform.

## Capítulo I

### Introducción

La cordillera de los Andes es una cadena montañosa ubicada en América del Sur comprendida entre los 11° de latitud N y los 55° de latitud S (Mendoza, 2014), que atraviesa Argentina, Bolivia, Chile, Colombia, Ecuador, Perú y parte de Venezuela. Con una longitud aproximada de 7500 km, se establece como la cordillera más extensa. El volcán Llaima se destaca como uno de los volcanes más activos en el cono sur de Sudamérica y se encuentra ubicado dentro de la zona Volcánica de los Andes del Sur en las coordenadas (38.7° S y 71.7°W) (Navarrete, 2017), además es considerado el segundo volcán más peligroso de Chile debido a que junto con el volcán Villarrica, registran más de la mitad de las erupciones registradas en el país.

Debido a la frecuente actividad volcánica registrada en los últimos años (Franco, 2019) en el volcán Llaima, éste se presenta como un objeto de estudio altamente relevante para su investigación. El volcán ha presentado constante erupción de material volcánico, acompañado de emisión de gases y actividad sísmica cercana al área, esto aporta una gran cantidad de información acerca de los procesos geológicos y vulcanológicos. La institución estatal (J. P. Canário et al., 2020) encargada de monitorizar y tomar las decisiones pertinentes con el volcán es el Observatorio Vulcanológico de los Andes Sur (OVDAS). El análisis de dichos eventos ayuda a prevenir desastres que puedan causar grandes daños a infraestructuras cercanas y pérdidas de vidas humanas.

Un Instituto Geofísico es una organización científica dedicada al estudio de los procesos físicos y geológicos tales como; la estructura interna del planeta, geodinámica, actividad sísmica, climatológica y meteorológica. Los institutos geofísicos son encargados de realizar investigaciones y experimentos con el fin de comprender dichos fenómenos, realizan trabajos

de monitorización, recopila datos y también realizan vigilancia para poder prevenir desastres naturales. El Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) es el principal centro de investigación del Ecuador encargado de la monitorización y vigilancia de los peligros sísmicos y volcánicos del país.

En los últimos años, se ha producido un notable desarrollo en el ámbito de la Inteligencia Artificial (AI, del inglés *Artificial Intelligence*). En la actualidad, se emplean algoritmos de aprendizaje automático para llevar a cabo tareas relacionadas con la clasificación de imágenes o la transcripción de voz. Estos sistemas se fundamentan en modelos matemáticos entrenados con grandes cantidades de datos para mejorar su precisión. Se observa un avance en los sistemas de reconocimiento automático que prescinden de la participación humana en la ejecución de tareas particulares. La distinción de estos sistemas radica en su empleo de algoritmos de aprendizaje automático y aprendizaje profundo (DL, del inglés *Deep Learning*), los cuales realizan las tareas de manera autónoma. La necesidad de utilizar *Data Augmentation*, se presenta al trabajar con DL, ya que es una técnica empleada para incrementar la cantidad de datos de entrenamiento al crear nuevas instancias a partir de las ya existentes (Ali et al., 2023).

El método de Bootstrap (Belío Miranda, 2020) es una técnica estadística utilizada para crear réplicas de una base de muestra, a partir de las cuales se caracteriza la distribución del estimador de interés al utilizar una distribución empírica, la cual evita la necesidad de realizar complicados cálculos analíticos. En el contexto de las bases de datos sintéticas, el método de Bootstrap se utiliza para generar una nueva base de datos totalmente sintética que sean representativos de los datos reales. El método implica tomar muestras aleatorias con reemplazo de la base de datos original, a partir de estas muestras se generan los nuevos datos. Luego, se evalúa la calidad de estos datos y, en caso de ser necesario, se ajustan los parámetros del método para mejorar su calidad.

## Antecedentes

Los métodos de remuestreo son técnicas indispensables para la generación de datos sintéticos, en el trabajo de investigación de (Ledesma, 2008) se hace uso del método de Bootstrap al ser una opción recomendada para la creación de datos sintéticos, debido a que no requiere de una alta complejidad matemática, sino simplemente de un computador capaz de simular un proceso de muestreo aleatorio de la base de datos. El método de Bootstrap es una técnica que permite resolver problemas relacionados con la estimación de intervalos de confianza mediante la idea de selección aleatoria de datos.

El uso de la transformada de Fourier representa un método en el estudio de la frecuencia para señales discretas aperiódicas como se menciona en el trabajo de (Cabrera Riobamba & Rosero Narvaez, 2019), por esto es una herramienta fundamental para la obtención del contenido espectral de las señales microsísmicas. El estudio en el dominio tiempo – frecuencia aporta una mejor percepción a la representación de la señal, la cual permite observar el comportamiento de la frecuencia respecto al tiempo y facilita el análisis de los diferentes eventos microsísmicos.

La monitorización permanente del volcán Cotopaxi permite predecir eventos futuros para poder accionar a tiempo frente a las catástrofes, el trabajo de investigación de (Zapata, 2022), aplica la técnica de Bootstrap para generar eventos artificiales debido a la escasa base de datos utilizada y la necesidad de obtener miles de datos para poder entrenar correctamente algoritmos de Machine Learning. La técnica de Bootstrap fue evaluada al utilizar la inspección visual y la métrica de evaluación (FD, del inglés *Frechet Distance*) obtiene buenos resultados en la representación tiempo – frecuencia. Conjuntamente las señales fueron sintetizadas por espectros empíricos de Fourier en amplitud y fase, las señales resultantes se compararon con

las señales sintéticas las cuales logran semejanza en la distribución de energía de la señal (Camacho, 2022).

En el trabajo de investigación de (Aguinaga Navarreta, 2018) utiliza varios métodos para el análisis de señales sísmicas del volcán Cotopaxi en Ecuador. Se emplean técnicas de análisis espectral, como métodos paramétricos y no paramétricos, para estimar la densidad espectral de potencia de los eventos microsísmicos. Además, se aplica la técnica de Bootstrap para mejorar el análisis y establecer intervalos de confianza para cada tipo de evento. Por último, se utilizan algoritmos de clasificación, como máquinas de vectores de soporte y árboles de decisión, para clasificar los eventos en base a las características seleccionadas

En diferentes regiones del mundo, surgen múltiples inquietudes a causa de terremotos y erupciones volcánicas. Entre estas preocupaciones, figura la situación que enfrenta Ecuador que se encuentra situado en el Cinturón de Fuego del Pacífico. En los últimos años se ha presentado actividad por parte del volcán Cotopaxi por lo que este se encuentra en constante vigilancia y monitorización. En el trabajo de investigación de (Pérez, 2022), se procede a generar bases de datos sintéticas equilibradas de señales microsísmicas provenientes del volcán Cotopaxi. Esto se logra mediante la combinación del método Bootstrap y la transformada de Fourier. Una vez aplicada la transformada de Fourier a las señales de cada evento, se lleva a cabo una separación en términos de magnitud y fase. Al analizar tanto sus componentes de magnitud y fase, se determina que la fase contiene la información de la envolvente y, por tanto, la forma de la señal.

Las magnitudes, son pasadas por filtros para seleccionar aquellas que se adapten a los parámetros establecidos antes de aplicar el método de Bootstrap para obtener las señales sintéticas en magnitud. El mismo proceso se aplica con la respectiva fase al obtener “fases madres” provenientes de la media y mediana de todas las fases sintéticas. De esta manera, se

generan datos sintéticos que conservan las características de las señales originales. La finalidad es entregar al IGEPN un sistema de reconocimiento automático desarrollada en el software de MATLAB® de un detector y clasificador de eventos de tipo Largo Periodo (LP, del inglés *Long Period*) y Volcano Tectónico (VT, del inglés *Volcano Tectonic*), desarrollado con redes neuronales convolucionales mediante los espectrogramas obtenidos.

### **Justificación e Importancia**

La motivación que fundamenta este trabajo de integración curricular es la preservación de vidas en un posible escenario de erupciones volcánicas. En este contexto, nuestro enfoque se dirige a establecer procesos sólidos de detección, clasificación y, en un sentido más amplio, generación de señales sintéticas. A través de estas iniciativas, buscamos brindar a las comunidades herramientas que les permitan anticipar y responder de manera efectiva, lo cual contribuye a la mitigación de los riesgos asociados a estos fenómenos naturales.

Uno de los mayores problemas que se presentan al entrenar un modelo de clasificador con Machine Learning son las bases de datos desbalanceadas (He & Garcia, 2009), destacan que, si una clase tiene mucho mayor peso que otra, esto puede incurrir en una reducción de la capacidad predictiva al momento de generar la clase de interés la cual normalmente se encuentra asociada a la clase minoritaria.

La generación de señales sintéticas utiliza el método de Bootstrap y se presenta como una herramienta efectiva para completar datos existentes. Esto es especialmente útil si los datos originales son limitados. Al agregar una señal sintética a la base de datos, se enriquece la información disponible y se obtiene una visión más completa de los fenómenos que se pueden presentar (Ledesma, 2008).

Existen diferentes técnicas de balanceo de datos, en este caso se va a utilizar la técnica Bootstrap para resolver el desequilibrio de la base, las limitaciones de la base se presentan

debido que, al trabajar con datos de señales reales, estos son de difícil adquisición como para tener una base de forma masiva. La motivación del presente trabajo se basa en disponer una cantidad suficiente de datos sintéticos para entrenar modelos de detección y clasificación al utilizar técnicas de DL. Esto se lleva a cabo en paralelo en otros trabajos de investigación, con el objetivo de no depender exclusivamente de los datos reales, los cuales suelen ser limitados y difíciles de obtener.

Por lo cual, el propósito del presente trabajo de investigación es generar suficientes datos sintéticos de los cuatro diferentes tipos de eventos microsísmicos del volcán Llaima proporcionado por Observatorio Vulcanológico de los Andes Sur con el objetivo de apoyar las labores del IGEPN. El IGEPN utiliza estas bases de datos sintéticas para mejorar la monitorización, comprensión y predicción de posibles riesgos volcánicos.

### **Objetivo**

- Crear bases de datos sintéticas de eventos microsísmicos producidos en el volcán Llaima con Bootstrap.

### **Actividades**

- Identificación de las propiedades y variables del problema a tener en cuenta.
- Reunión de la base de datos representativa, con datos provistos por el observatorio “Observatorio Vulcanológico de los Andes del Sur”.
- Creación de una estructura de datos (temporales, espaciales, otros.) con un soporte común.
- Caracterización de microsismos.
- Generación de bases sintéticas con la técnica de remuestreo Bootstrapping.
- Pruebas y evaluación de desempeño.

## Capítulo II

### Materiales y Métodos

#### Materiales

El presente trabajo de integración curricular se desarrolla en el software de MATLAB® en su versión R2022a, una herramienta ampliamente reconocida y utilizada a lo largo de la carrera académica.

La implementación del método de Bootstrap, al no necesitar una complejidad computacional elevada, se realiza en una computadora con las especificaciones detalladas en la Tabla 1 que se presenta a continuación.

**Tabla 1**

*Especificaciones del computador*

Especificaciones	
<b>Procesador</b>	Intel® Core™ i7 8th Generación 2.00 GHz
<b>Memoria RAM</b>	12 GB DDR4
<b>Gráficos</b>	Gráficos Integrados Intel UHD Graphics 620
<b>Almacenamiento</b>	Disco SSD M.2, 500 GB Disco HDD, 500 GB
<b>Sistema Operativo</b>	Windows 10 Home

#### Identificación de propiedades y variables

En el primer proceso de análisis correspondiente a la Actividad 1, en esta se lleva a cabo la identificación de las propiedades, variables y problemas relacionados con el trabajo de



integración curricular, basándose en trabajos previos realizados en el volcán Cotopaxi, algunos puntos que se identificaron fueron:

- Características de los eventos microsísmicos, se identificaron las propiedades relevantes de los eventos, como magnitud, ubicación del volcán, duración y tipo de evento microsísmico.
- Se determinaron los parámetros necesarios para generar señales sintéticas al utilizar el método de Bootstrap en MATLAB®, como el tamaño de muestra, el número de iteraciones y los criterios de remuestreo.
- Se realizó un análisis de la influencia de las variables en los resultados, como el modelo de ruido, la selección de estaciones sísmicas, la distribución espacial de los eventos, entre otros, pueden afectar los resultados y la representatividad de las señales sintéticas.
- Se evaluó la calidad de las señales sintéticas generadas, comparándolas con registros sísmicos reales del volcán Cotopaxi en términos de similitud, respuesta en frecuencia y coherencia temporal.

Mediante el diagrama de la Figura 1 se presentan algunos aspectos que son necesarios conocer para entender la metodología aplicada. La cual empieza con la carga de base de datos del volcán Llaima con las 3592 señales otorgadas por el Observatorio Volcanológico de los Andes del Sur. Estas señales son preprocesadas por filtrado, estandarizado y normalizado. Una vez las señales están listas se clasifica en los 1310 eventos LP, 1488 Tectónicos (TC, del inglés *Tectonic*), 490 Tremor (TR, del inglés *Tremor*) y 304 VT.

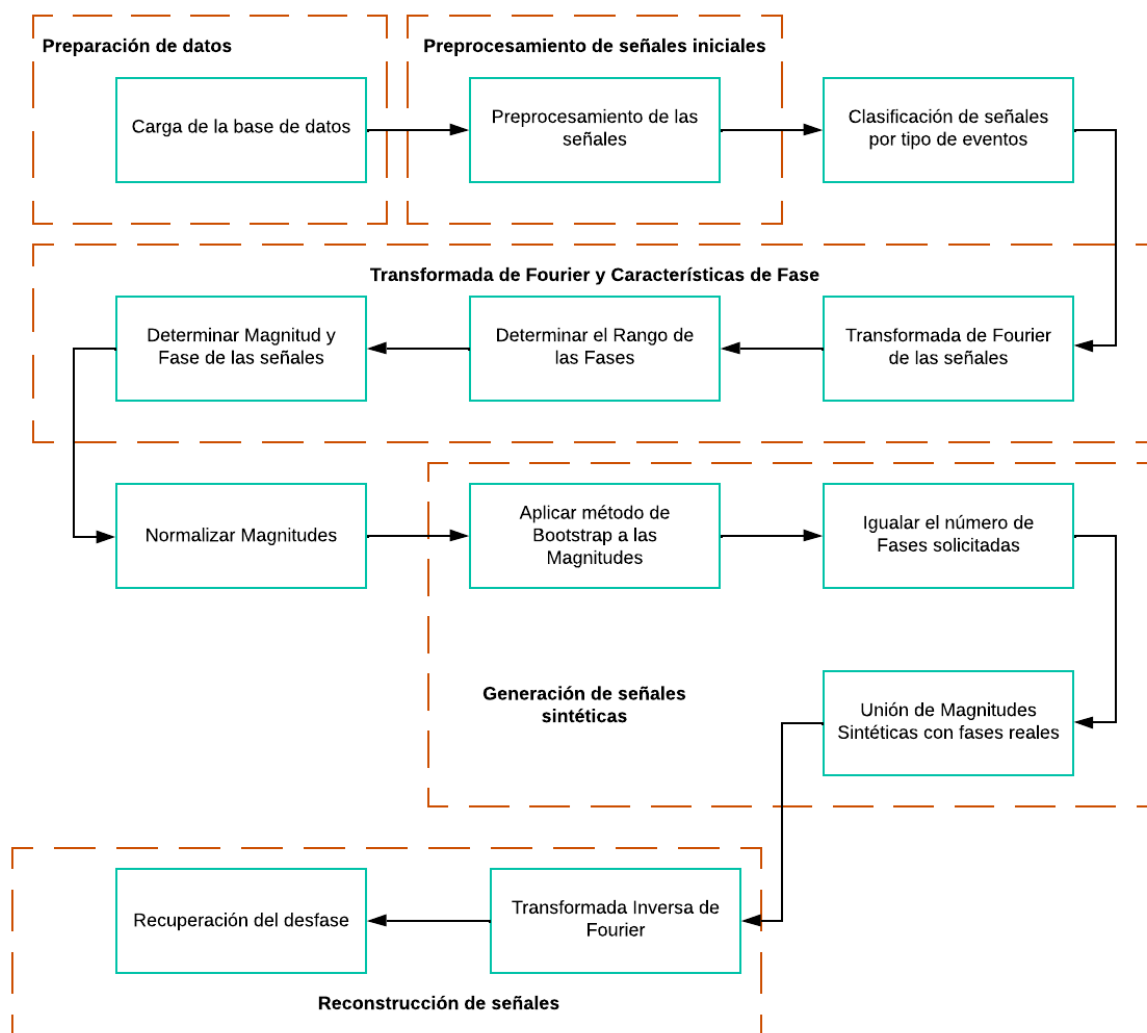
Una vez clasificado en los diferentes tipos de eventos, se aplica la Transformada de Fourier para determinar la magnitud y las fases representativas de cada uno. Se determina el

rango que contenga las fases más lineales para cada evento y sean almacenados en un vector, mientras que las magnitudes reales son almacenadas en otro para posteriormente ser normalizadas.

Las señales, una vez normalizadas en una escala común, son sometidas al método de Bootstrap para generar magnitudes sintéticas adicionales. Estas nuevas magnitudes se combinan con fases aleatorias previamente seleccionadas y, posteriormente, se realiza la Transformada Inversa de Fourier para generar las señales sintéticas en función del tiempo. Este proceso permite obtener señales sintéticas que reproducen las características de las señales originales, pero con variaciones que reflejan la incertidumbre y la aleatoriedad del proceso. Un efecto de la aleatoriedad produce un desfase en las señales sintéticas, las cuales deben pasar por un proceso para recuperar dicho desfase.

### **Figura 1**

*Diagrama de bloque de la metodología propuesta*



*Nota.* La Figura 1 ilustra visualmente la metodología propuesta a través de un diagrama, en el cual se resaltan mediante bloques los diversos tipos de procesos realizados desde la preparación de los datos hasta la reconstrucción y obtención de las señales sintéticas.

## Volcán Llaima

El volcán Llaima considerado el segundo volcán más peligroso de Chile, se encuentra ubicado en la IX región de la Araucanía, con una altitud de 3.195 m (Carmona, 2017). El volcán es considerado uno de los que posee mayor actividad en el país y en América del Sur, pues presenta 23 eventos mayores de volcanismo, el último más importante en 1994, y continúa en

actividad durante el siglo XX. La última erupción de gran magnitud registrada ocurrió el 1 de enero de 2008 a las 18:20 hora Chile. La erupción fue clasificada como una erupción explosiva y se caracterizó por una intensa actividad volcánica, emisiones de gases, expulsión de material piroclástico y flujos de lava, además la erupción generó una columna de erupción de varios kilómetros de altura lo que ha causado importantes daños en las zonas cercanas al volcán.

## **Figura 2**

*Foto del volcán Llaima visto desde río Blanco*



*Nota.* La Figura representa el volcán Llaima observado desde el río Blanco, en el camino hacia Curacautín. Tomado de *Volcan Llaima (3125 M.)*, por P.Cox, 2020, AndesHandBook.

### ***Características geológicas del volcán Llaima***

Es un volcán compuesto, mixto y de escudo, que posee una caldera cubierta de 40 conos adventicios de escoria como se observa en la Figura 2, presenta en un costado una chimenea secundaria en los 2.900 m. Posee una pendiente moderada. El Llaima se encuentra activo desde el pleistoceno superior, produce normalmente lava y piroclastos andesítico-basáltico. El volcán ha presentado numerosas erupciones violentas de estilo vulcaniano y estramboliano, las cuales han dejado capas de 50 m de arenisca y tobas volcánicas, así como flujos de lava basáltica rugosa. En su historia registra 37 erupciones a contar desde el año

1952, en las cuales las de 1927 y 1957 se consideran las más violentas, convirtiéndolo en uno de los volcanes más activos y peligrosos de Chile (Schuster & Millar, 2020).

### **Reunión de la base de datos representativa**

Para la Actividad 2 en el presente trabajo se hace uso de la base de datos previamente construida y provista por el Observatorio OVDAS del volcán Llaima compuesto por 3592 señales separadas por clase y filtradas para determinar su evento sísmico. Esta estación de monitorización almacena diferentes tipos de microsismos presentes en el volcán recolectadas de “LAV”, que es una de las 7 estaciones sísmicas del volcán, adquiridas entre los años 2010 a 2016, muestreadas a 100 Hz y filtradas con un filtro tipo Butterworth de orden 10 entre el rango de [1 -10] Hz en términos de la componente Z-vertical (J. Canário, 2021), dicha componente contiene toda la información que se necesita para clasificar los diferentes tipos de eventos registrados; Eventos LP, VT, TR y TC (J. P. Canário et al., 2020).

En la ecuación 1 se presenta la matriz de la base de datos,  $X$ , que tiene una longitud,  $M$ , correspondiente a las 3592 señales otorgadas por el Observatorio el operador OVDAS:

$$X = \{x_1^T, x_2^T, \dots, x_M^T\}^T. \quad (1)$$

### **Densidad Espectral de Potencia**

La Densidad Espectral de Potencia (PSD, del inglés *Power Spectral Density*), ofrece una representación detallada de cómo se distribuye la energía de la señal en el espectro de frecuencia. Este análisis es útil para identificar las componentes frecuenciales dominantes presentes en cada tipo de evento.

### ***Eventos Microsísmicos***

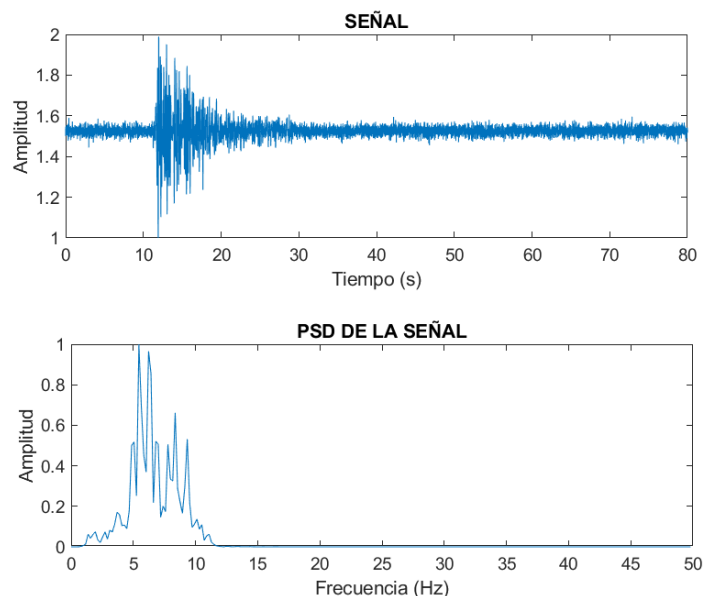
En una actividad volcánica se presentan una enorme variedad de señales microsísmicas, estas son clasificadas al tomar en cuenta su apariencia en el sismograma en sus aspectos de cambio de frecuencia, duración, entre otros. En un principio (Minakami, 1974) clasificó estos eventos en 4 clases de señales, tipo A o de alta frecuencia, tipo B o de baja frecuencia, explosiones y tremor volcánico.

### **Volcano Tectónicos**

Los eventos VT son el resultado de las fracturas de las paredes volcánicas debido a la presión generada por el ascenso del magma.(Cabrera & Rosero, 2019) El estudio de los eventos VT son importantes para detectar y monitorizar la actividad volcánica en curso, así como para identificar zonas de falla activas y áreas propensas a riesgos sísmicos y volcánicos. Estos eventos son caracterizados por tener una duración variable, que puede oscilar desde pocos segundos hasta minutos. Exhiben una forma de onda que muestra un inicio abrupto y una extinción rápida, como se puede observar en la Figura 3.

### **Figura 3**

*Evento microsísmico Volcano tectónico*



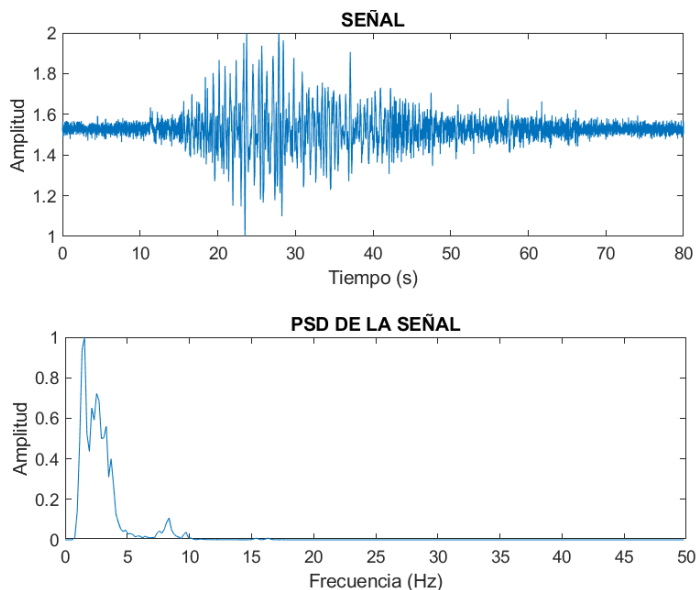
*Nota.* La Figura 3 representa un evento microsísmico Volcano Tectónico del Volcán Llaima. El evento microsísmico fue extraído de la base de datos sísmicos del Observatorio OVDAS. Se observa su comportamiento en tiempo y en frecuencia mediante la PSD.

### **Largo Período**

Los eventos LP están asociados a la resonancia de fluidos volcánicos como resultado de fluidos multifase de gases y líquidos que se desplazan a través de fracturas y grietas saturadas por estos fluidos. También se registran casos de eventos LP durante la fragmentación del magma en las paredes internas de los ductos. Por lo general estos eventos suelen originarse a una profundidad inferior a 1km en las proximidades de la superficie y presentan frecuencias inferiores a los 3 Hz (UNA, 2023).

### **Figura 4**

*Evento microsísmico Largo Período*



*Nota.* La Figura representa un evento microsísmico Largo Período del Volcán Llaima localizado en Chile. El evento sísmico fue extraído de la base de datos sísmicos del Observatorio OVDAS. Se visualiza su comportamiento en tiempo y en frecuencia.

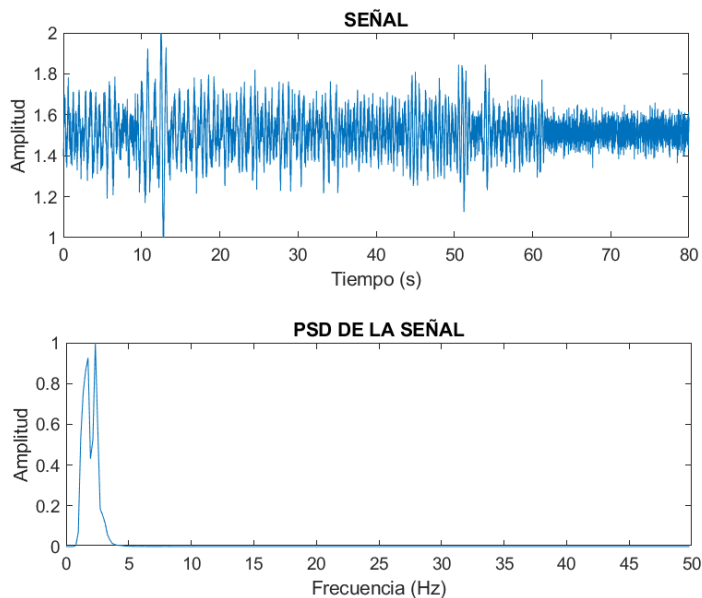
### **Tremores**

Los eventos TR son un fenómeno sismológico que se caracteriza por la presencia constante de movimientos vibratorios de baja frecuencia y amplitud. La duración de un evento TR puede variar desde breves periodos de tiempo hasta extenderse durante varios meses (Cadena, 2021). En el trabajo de investigación de (Cortés Moreno, 2015) se menciona que debido a su larga duración también son denominados como ruido volcánico y que no existe una teoría aceptada en cuanto a su procedencia, estos eventos sugieren que su origen se debe a procesos como la liberación de gases, resonancia de cavidades y cambios en la presión y temperatura en sistemas hidrotermales. También se menciona que son señales principalmente bajas entre [0.5 - 10] Hz.

### **Figura 5**

*Evento microsísmico Tremor*





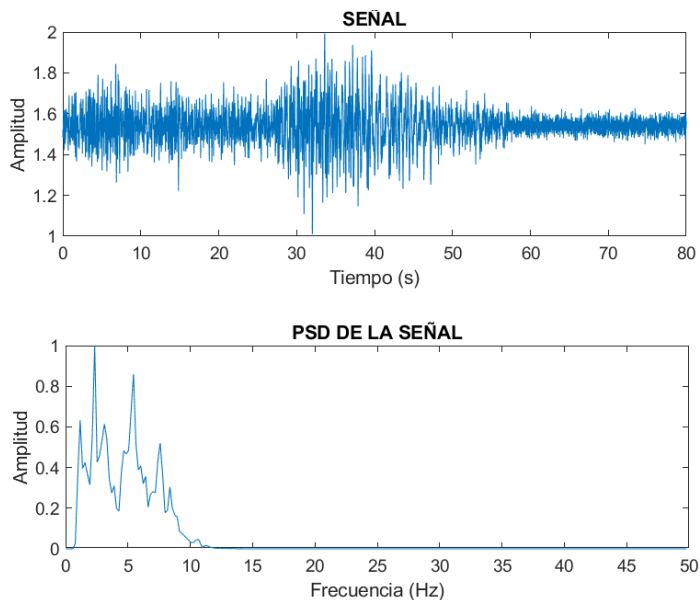
*Nota.* La Figura representa un evento microsísmico Tremor del Volcán Llaima localizado en Chile. El evento sísmico fue extraído de la base de datos sísmicos del Observatorio OVDAS. Se visualiza su comportamiento en tiempo y en frecuencia.

### **Tectónico**

El tipo de evento TC es un tipo de evento que no se encuentra relacionado con la actividad volcánica, se produce por la dinámica de las fallas geológicas. En el artículo (Curilem, Canário, & Franco, 2018) se menciona la siguiente información; que los eventos pueden ser, locales, regionales o distantes, referentes a la ubicación del epicentro. Mientras el evento se encuentre a una mayor distancia, menor será su patrón de frecuencia. Además, depende de la proximidad geográfica entre el epicentro y la estación, puede ocasionar que los eventos TC sean confundidos con eventos LP o VT. El contenido espectral de un evento TC es caracterizado por poseer un comienzo impulsivo y una caída exponencial, muy parecido a los LP, a diferencia que las señales TC llevan más energía.

### **Figura 6**

*Evento microsísmico Tectónico*



*Nota.* La Figura representa un evento microsísmico Tectónico del Volcán Llaima localizado en Chile. El evento sísmico fue extraído de la base de datos sísmicos del Observatorio OVDAS. Se visualiza su comportamiento en tiempo y en frecuencia.

La Tabla 2 que se presenta a continuación recopila la cantidad de señales microsísmicas según su tipo de evento de la base de datos proporcionada por el observatorio OVDAS, la cual se va a utilizar en el análisis.

**Tabla 2**

*Distribución de Base de datos de eventos microsísmicos*

<b>Tipo de Evento</b>	<b>Número de señales</b>
<b>LP</b>	1310
<b>TC</b>	1488
<b>TR</b>	490
<b>VT</b>	304

*Nota.* Base de datos del volcán Llaima proporcionada por el Observatorio Volcanológico de los Andes del Sur.

La Tabla 2 representa la creación de la estructura de datos con un soporte común, tal como se hace mención en la Actividad 3.

### **Método Bootstrap**

El autor (Belío Miranda, 2020) reafirma que el método Bootstrap, también conocido como método de remuestreo, es un método que consiste en generar muestras de datos mediante remuestreo directo o a través de un modelo ajustado, para así crear réplicas de conjuntos de datos. Estas réplicas permiten evaluar la variabilidad de los estimadores de interés sin la necesidad de realizar análisis complejos.

### ***Ventajas del Método***

Las ventajas registradas en el trabajo de (Gil, 2018) sobre este método son enlistadas a continuación:

- El método de remuestreo posibilita una aproximación de las propiedades de la distribución al muestrear un estadístico específico.
- En el caso que el modelo teórico sea validado, el método de Bootstrap puede proporcionar estimadores más precisos en intervalos más estrechos y confiables en comparación con enfoques no paramétricos.
- El método permite simular el proceso de obtención de nuevas muestras al utilizar un conjunto de datos único. Esto nos brinda la oportunidad de, por ejemplo, validar modelos mediante la generación de múltiples muestras.

### ***Desventajas del Método***

Según el autor (Hesterberg, 2015) este método también presenta las siguientes limitaciones:

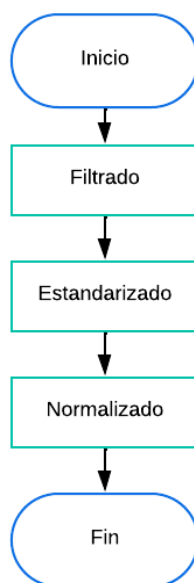
- Primero es necesario seleccionar un modelo teórico que pueda ajustarse al proceso generador real de los datos. Si el ajuste es deficiente, los resultados obtenidos a través del método Bootstrap pueden carecer de utilidad práctica.
- En situaciones que los datos presentan una alta correlación entre ellos, el método de remuestreo puede ser menos eficaz o menos preciso en sus resultados.
- El método de remuestreo puede requerir un alto costo computacional, especialmente al trabajar con bases de datos extensos o al realizar un gran número de iteraciones.

### **Preprocesamiento**

Las señales entregadas por el Observatorio OVDAS necesitan pasar por un preprocesamiento antes de poder utilizarlas con el propósito de asegurar la fiabilidad, coherencia y adecuación de los datos, se procede a eliminar cualquier forma de ruido e interferencia presentes en la señal original, a continuación, se presenta un diagrama de flujo, el cuál ayuda a comprender de forma gráfica y sistemática el desarrollo de los códigos en MATLAB®.

### **Figura 7**

*Diagrama de flujo de etapa de preprocesamiento de la base de datos*



*Nota.* La Figura 7 representa un diagrama de flujo con el código que se utiliza en la etapa de preprocesamiento de las señales provenientes de la base de datos original del volcán Llaima.

### **Filtrado**

Las señales de la base de datos originales deben pasar por un proceso de filtrado en la etapa de preprocesamiento para reducir o eliminar ruidos no deseados. Otra razón importante es para determinar las frecuencias específicas de interés.

Para el filtrado se hace uso de funciones anónimas de MATLAB®, las cuales son una forma flexible de crear funciones sin la necesidad de crear archivos separados para la función. Las funciones son definidas en la propia ventana de comandos y se encuentran disponibles hasta que se limpie el área de trabajo. Se utiliza la función `'filter_fun = @(x) miFiltro(x,fs,0)'` presente en el Anexo 1, la cual admite dos parámetros de entrada; la señal a la cual se va a aplicar el filtro y la frecuencia de muestreo. En este caso la frecuencia a la cual fue muestreada la señal corresponde a 100 Hz como se menciona en el trabajo de (Curilem, Canário, Franco, et al., 2018).

El filtro utilizado es un filtro pasa banda Filtro de Respuesta al Impulso Finito (FIR, del inglés *Finite Impulse Response*), lo que significa que su respuesta en frecuencia se caracteriza por tener una duración finita en el tiempo. Los valores de 0,70 y 49,50 Hz corresponden a las frecuencias de corte del filtro en Hz. Un filtro FIR de orden 256 indica que se utiliza 257 coeficientes para definir la respuesta del filtro en el dominio del tiempo. La respuesta en frecuencia del filtro se calcula en 1024 puntos para tener una representación más detallada de cómo el filtro afecta a diferentes frecuencias.

En cuanto al proceso de filtrado, el filtro pasa banda permite que las frecuencias dentro del rango pasen relativamente sin cambios, mientras que reduce la amplitud de las frecuencias fuera de ese rango. La función `filtfilt` se emplea para asegurar que la señal filtrada no presente un desplazamiento en el tiempo, ya que el filtro se aplica en ambas direcciones. Este enfoque bidireccional es particularmente útil para conservar la fase de la señal, lo que es esencial en muchas aplicaciones de análisis de señales.

Este método de Filtrado,  $H$ , se representa matemáticamente al aplicar a cada una de las señales de la base de datos mediante el operador  $h_M\{X_M\}$ . El resultado es almacenado en la matriz, donde  $h_1$  representa cada señal filtrada, cómo:

$$H = \{h_1^T, h_2^T, \dots, h_M^T\}^T. \quad (2)$$

### **Estandarizado**

El proceso de estandarización se utiliza sobre la base de datos para eliminar cualquier sesgo o desplazamiento sistemático presente en las señales y tiene como resultado las señales centradas en torno a cero.

El proceso matemático de la estandarización,  $Z_S$ , está dada por la estandarización z-score, el cual representa que tal alejado se encuentra un valor de la media en términos de la desviación estándar:

$$Z_{SM} = \frac{h_M - \mu_M}{\sigma_M}, \quad (3)$$

Dónde  $\mu_M$  representa la media, y  $\sigma_M$  su desviación estándar.

Cada valor estandarizado  $Z_{Si}$  es almacenado en la matriz  $Z_S$ , cómo:

$$Z_S = \{z_{S1}^T, z_{S2}^T, \dots, z_{SM}^T\}^T. \quad (4)$$

### **Normalizado**

El proceso de normalización se realiza en la base de datos para poder compararlas en una escala común y facilitar su análisis.

Se hace uso de la función anónima 'norm\_max\_fun = @(x) x/max(abs(x))', presente en el Anexo 1, esta función normaliza la señal y divide cada señal de la base de datos por el valor absoluto máximo presente en la señal, ajusta la señal a un rango específico.

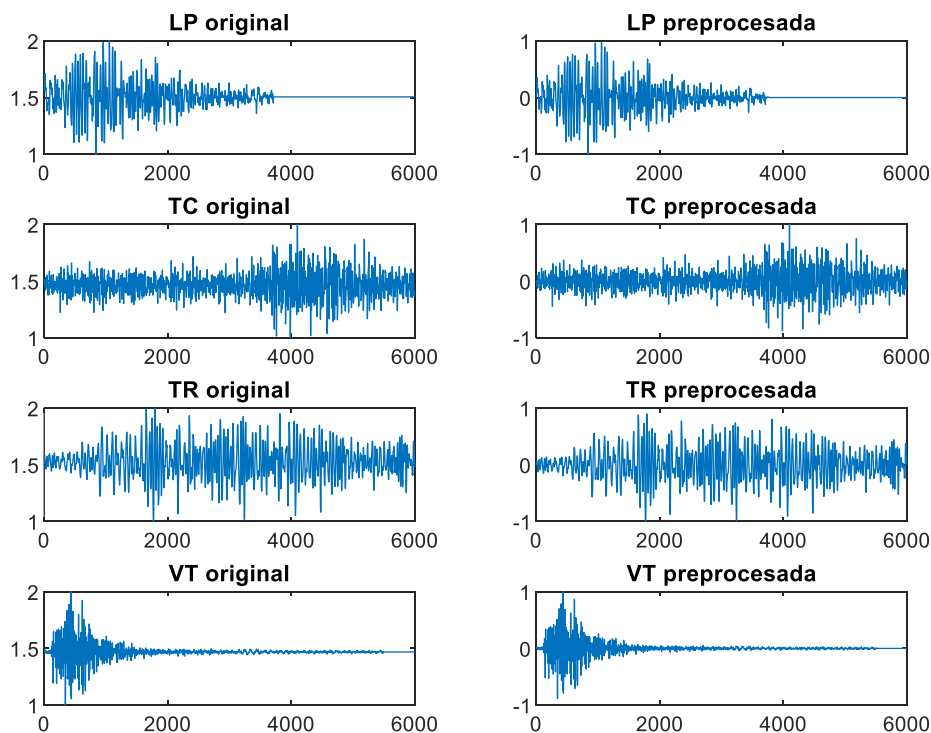
La representación matemática de la base de datos una vez estandarizados, se da en la matriz,  $Z_1$ , la cual normaliza las señales entre los valores de [-1 a 1] dividiendo cada señal para el valor absoluto del máximo valor  $z_S$ , cómo:

$$Z_{1M} = \frac{Z_{SM}}{\max|Z_{SM}|},$$

$$Z_1 = \{z_{11}^T, z_{12}^T, \dots, z_{1M}^T\}^T. \quad (5)$$

## Figura 8

*Señales microsísmicas antes y después de pasar por la etapa de preprocesamiento*



Nota. En la Figura se presentan las señales preprocesadas, las cuales han pasado por las etapas de filtrado, estandarización y normalización.

En la Figura 8 se nota la eliminación de un desfase de 1,50 en las señales, y se exhiben estas mismas señales normalizadas en un rango entre 1 y -1.

## Transformada de Fourier

La transformada de Fourier es una técnica utilizada para analizar y extraer información en frecuencia de una señal. Esta técnica se aplica en diversos ámbitos, especialmente en el procesamiento de señales, permite transformar una señal en el dominio del tiempo al dominio de la frecuencia (Duoandikoetxea, 2003).



### ***Análisis Matemático de la Transformada de Fourier***

En este trabajo de investigación, se emplea la Transformada de Fourier en lugar de la PSD para crear señales sintéticas microsísmicas del volcán Llaima. Esto se debe a que la Transformada de Fourier ofrece una eficiencia computacional superior, la cual permite el procesamiento de grandes volúmenes de datos en un lapso razonable. Por otro lado, el cálculo de la PSD implica la Transformada de Fourier de la autocorrelación de la señal, lo que resulta más exigente en términos de recursos computacionales.

Además, destaca la facilidad de representación en el dominio de la frecuencia. La Transformada de Fourier proporciona una representación directa de la señal en dicho dominio, permite un análisis y manipulación más conveniente de las componentes de la señal. En contraste, la PSD se enfoca en medir la distribución de energía de la señal en el dominio de la frecuencia, sin ofrecer una representación directa de la señal misma.

Esta elección se respalda en la flexibilidad que brinda la Transformada de Fourier para manipular señales, como filtrado, ajuste de amplitud y desplazamientos de frecuencia de manera sencilla. Estas características son fundamentales para la generación de señales sintéticas. En contraparte, la PSD no es recomendada para este propósito, ya que se emplea principalmente para el análisis de la distribución de energía de la señal y no proporciona las mismas capacidades de manipulación (Stoica & Moses, 2005).

Las series de Fourier son utilizadas para representar funciones definidas en un intervalo de la recta o, de forma equivalente, funciones periódicas en la recta. Sin embargo, al necesitar representar funciones definidas en toda la recta y no periódicas, se recurre a la transformada de Fourier. En este contexto, trabajar con números complejos se presenta como una opción

conveniente. La serie de Fourier en su forma de números complejos se representa con la siguiente fórmula:

$$\sum_{n=-\infty}^{\infty} \left( \frac{1}{2l} \int_{-l}^l f(t) e^{-\pi i n t / l} dt \right) e^{\pi i n x / l}, \quad (6)$$

Al ser:

$$\xi_n = n/2l, \quad (7)$$

y

$$h(\xi) = \int_{-l}^l f(t) e^{-2\pi i \xi t} dt, \quad (8)$$

Se reemplaza la ecuación (7) y (8) en (6),

$$\sum_{n=-\infty}^{\infty} (\xi_n - \xi_{n-1}) h(\xi_n) e^{2\pi i \xi_n x}. \quad (9)$$

La ecuación (9) exhibe el aspecto de una suma de Riemann, se tiene de esta la siguiente igualdad,

$$f(x) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt \right] e^{2\pi i \xi x} d\xi, \quad (10)$$

La ecuación (10) se conoce como transformada de Fourier de  $f$ . Mientras  $f$  sea una función integrable y definida dentro de los Reales. Su transformada de Fourier también estará definida dentro de los Reales y representada por  $\hat{f}$  dada por la siguiente expresión:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx. \quad (11)$$

Al igual que las series de Fourier, la transformada de Fourier descompone una función  $f$  en componentes; pero en este caso, en lugar de presentar solo frecuencias discretas al formar una sucesión, se obtiene un rango continuo de frecuencias en todos los números reales. En el caso de cada frecuencia  $\xi$ , le corresponderá un coeficiente de frecuencia  $\hat{f}(\xi)$  el cual es un número complejo. Su módulo representa la amplitud y su fase representa el argumento (Duoandikoetxea, 2003).

Se representa matricialmente la transformada de Fourier,  $F$ , sobre las señales normalizadas, mediante el operador  $f\{X_{z1}\}$ , obteniendo una matriz con cada  $f_i$  expresado en el dominio de la Frecuencia:

$$F = \{f_1^T, f_2^T, \dots, f_M^T\}^T. \quad (12)$$

### **Análisis Matemático de la Transformada Inversa de Fourier**

A partir de la transformada de Fourier, es posible recuperar la señal  $f(t)$  a través de lo que se denomina la transformada inversa de Fourier.

Si  $F(\omega)$  representa la transformada de Fourier de  $f(t)$ ;

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt. \quad (13)$$

Entonces  $f(t)$  viene dado por:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega. \quad (14)$$

Así pues, la transformada de Fourier realiza la transformación de una señal en el dominio de la frecuencia  $\omega$  al dominio temporal  $t$  (Cruz, 2022).

La transformada inversa de Fourier representada matricialmente,  $IF$ , se utiliza en la matriz correspondiente a señales sintéticas,  $X_B$ , con el propósito de regresar las señales al dominio temporal.

## Capítulo III

### Diseño e Implementación

#### Caracterización de microsismos

Para caracterizar los microsismos mencionados en la Actividad 4, se analiza los cuatro tipos de eventos; LP, TC, TR y VT. Además, se efectúa un análisis detallado de la forma de cada evento en el dominio del tiempo, y se examina su magnitud y fase en el dominio de la frecuencia.

#### Transformada de Fourier de las señales

Se emplea la transformada de Fourier como una herramienta fundamental para el análisis de diversos tipos de eventos microsísmicos en el dominio de la frecuencia. La transformada de Fourier ofrece la capacidad de examinar tanto la magnitud como la fase de las señales microsísmicas de interés.

La magnitud microsísmica, representa la energía liberada por un evento, es un parámetro clave de la identificación de cada tipo de evento microsísmico. En paralelo, el análisis de la fase microsísmica es otro aspecto esencial que aborda esta investigación. La fase proporciona información valiosa sobre el comportamiento y la forma de las señales.

Se realiza la transformada de Fourier a todas las señales al utilizar un tamaño de 8192 puntos. Esta elección se debe a la necesidad de una alta resolución de 8k, lo que proporciona una mayor claridad para capturar detalles precisos en las señales y mejorar la calidad de las señales sintéticas generadas. Gracias a esta transformada, se logra obtener tanto la magnitud como la fase de cada señal microsísmica.

La representación matricial de la magnitud,  $M_S$ , y de la fase,  $P_S$ , está dada por:

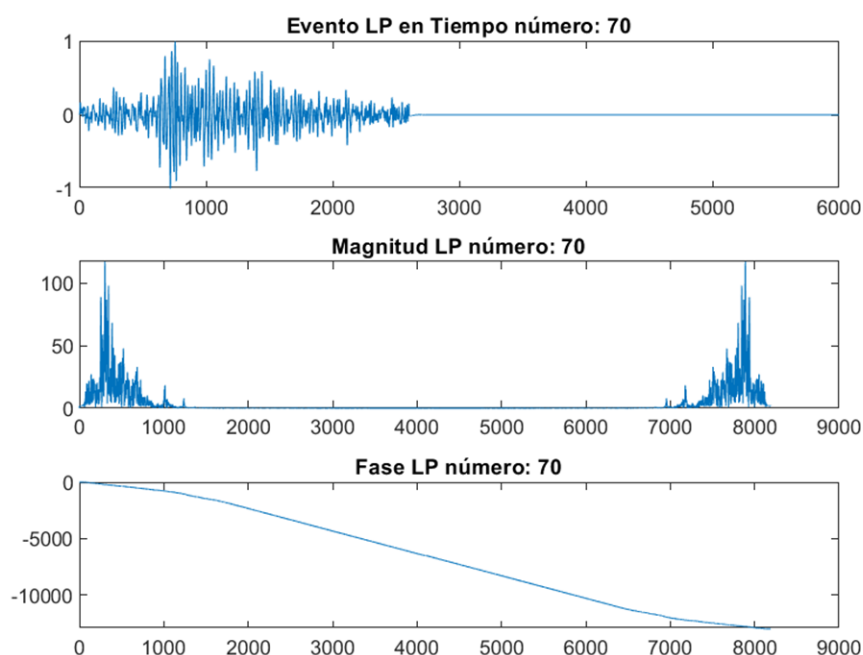
$$M_s = \{m_{s1}^T, m_{s2}^T, \dots, m_{sN}^T\}^T, \quad (15)$$

$$P_s = \{p_{s1}^T, p_{s2}^T, \dots, p_{sN}^T\}^T, \quad (16)$$

Dónde  $N$  representa el número de eventos de estudio, este varía dependiendo la cantidad de tipos de eventos conforme a la Tabla 2.

### Figura 9

*Representación temporal y en frecuencia de señal LP*



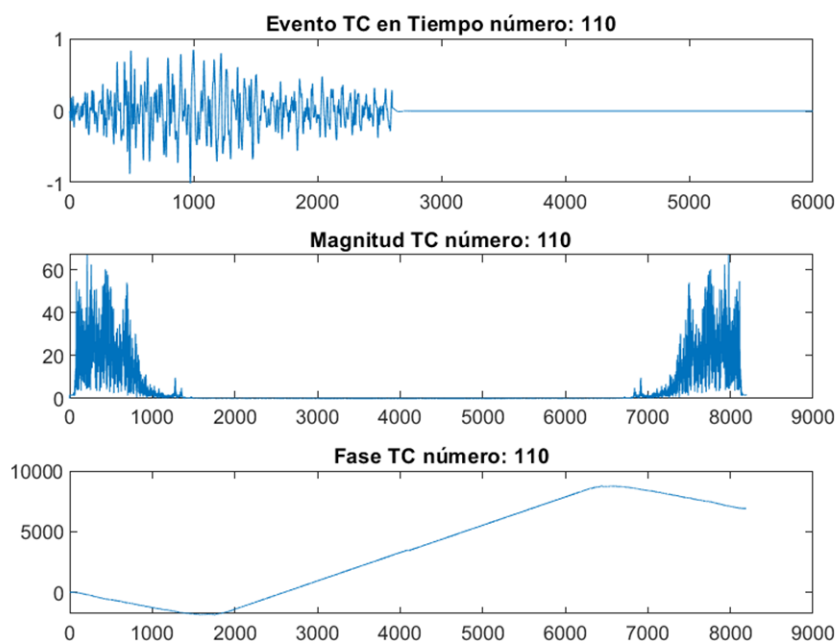
*Nota.* En el análisis de los eventos LP, se identifica claramente la forma característica de estos eventos en el dominio del tiempo. Además, se lleva a cabo un análisis exhaustivo del comportamiento de la magnitud y fase en el dominio de la frecuencia.

La Figura 9 se observa la forma distintiva en el tiempo de un evento LP, a diferencia de otros tipos de eventos microsísmicos, los eventos LP muestran un inicio gradual seguido de un período de actividad constante antes de disminuir gradualmente. La forma de onda de un

evento LP puede ser más prolongada y exhibir una mayor variabilidad en comparación con otros tipos de eventos (Cabrera & Rosero, 2019).

### Figura 10

*Representación temporal y en frecuencia de señal TC*

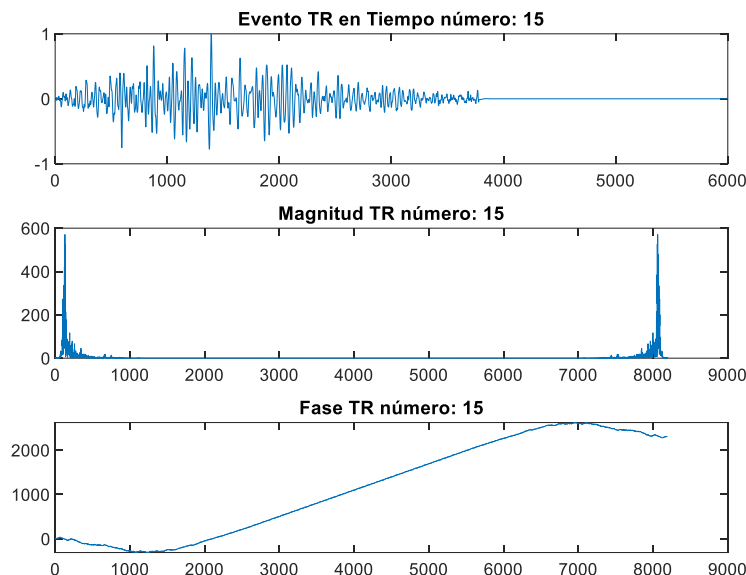


*Nota.* Se observa la forma característica de un evento TC la cual se distingue por su carácter oscilatorio y sostenido en el tiempo, también se analiza su comportamiento en frecuencia, para determinar la forma de su magnitud y fase.

A diferencia de los eventos microsísmicos de impulso breve, los eventos TC muestran una actividad continua y prolongada en el tiempo, sin un inicio o final claro. Estos eventos presentan una serie de oscilaciones rítmicas con una amplitud y frecuencia relativamente constantes a lo largo de su duración observada en la Figura 10. La forma de onda de un evento TC puede exhibir una variabilidad en la amplitud y frecuencia de las oscilaciones, lo que indica la dinámica interna del fenómeno (J. P. Canário et al., 2020).

## Figura 11

### *Representación temporal y en frecuencia de señal TR*



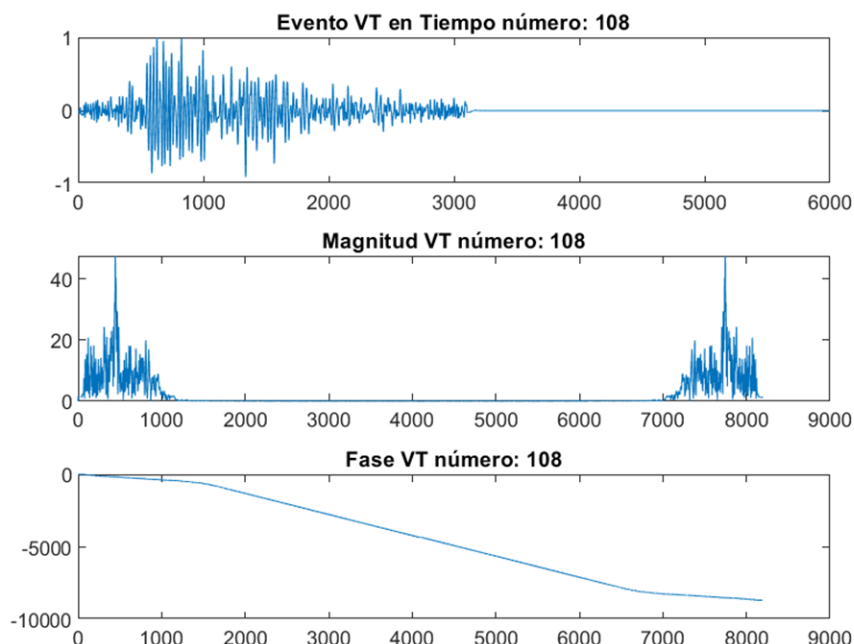
*Nota.* Se representa la forma característica de un evento TR; en el dominio del tiempo se caracteriza por su patrón repetitivo y periódico de oscilaciones, también se analiza su comportamiento en frecuencia, para determinar la forma de su magnitud y fase.

Estos eventos exhiben una serie de pulsos oscilatorios que se repiten a intervalos regulares a lo largo de su duración como se logra apreciar en la Figura 11. Cada pulso puede tener una forma de onda similar, con una amplitud y frecuencia relativamente constante. La forma de onda de un evento TR puede mostrar una regularidad en la periodicidad y una coherencia en las oscilaciones a lo largo del tiempo. Sin embargo, es posible que también exista cierta variabilidad en la amplitud y frecuencia de los pulsos entre eventos TR diferentes (Cadena, 2021).

## Figura 12

### *Representación temporal y en frecuencia de señal VT*





*Nota.* Se representa la forma característica de un evento VT; en el dominio del tiempo se caracteriza por su naturaleza impulsiva y de corta duración, también se analiza su comportamiento en frecuencia, para determinar la forma de su magnitud y fase.

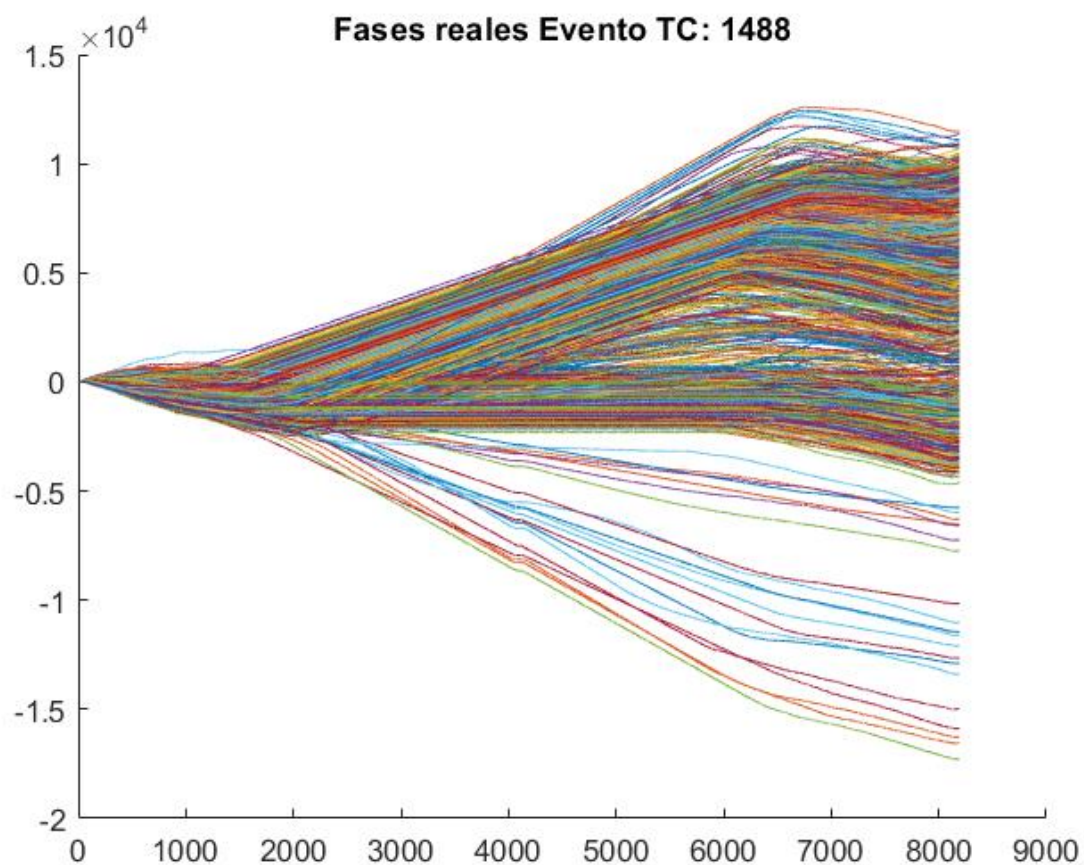
La Figura 12 representa un evento VT, estos eventos muestran una forma de onda que se inicia abruptamente y se extingue rápidamente, generalmente con un pico de amplitud pronunciado. A diferencia de los eventos microsísmicos continuos como el tremor, los eventos VT suelen ser de corta duración y están asociados con la actividad tectónica volcánica; como fracturas de rocas o movimiento de fluidos en el sistema volcánico. La forma de onda de un evento VT puede variar en términos de amplitud, duración y características espectrales, depende de la energía liberada y las características del evento en particular (Cabrera & Rosero, 2019).

### Selección de fases

Durante el proceso de selección de fases, se analizan cada una de las señales correspondiente a cada tipo de evento, en la Figura 13 se observa las fases totales de las 1488 señales correspondientes al evento TC.

### Figura 13

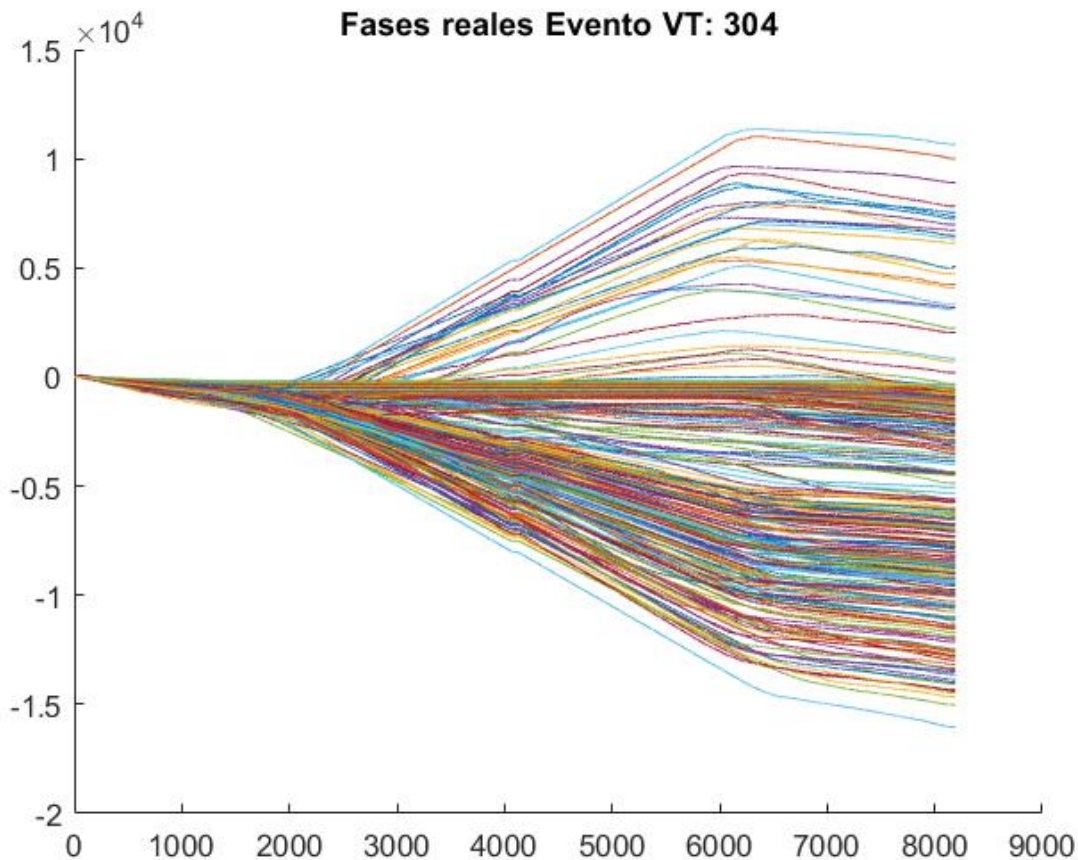
*Fases totales de los 1488 eventos microsísmicos TC*



*Nota.* Se observa en la Figura todas las 1488 fases correspondientes a las señales del evento microsísmico TC, con el objetivo de identificar patrones o correlaciones que pudieran existir entre ellas.

**Figura 14**

*Fases totales de los 304 eventos microsísmicos VT*



*Nota.* Al observar las distintas clases de eventos, se percibe una particularidad en los eventos VT. En comparación con los otros tipos de eventos, se observa una tendencia hacia una mayor concentración de fases negativas en su patrón.

En el trabajo de (Pérez, 2022) se concluye que las fases que exhiben una forma linealizada son las que mejor conservan las características distintivas de cada evento, estas preservan así su forma distintiva en el dominio del tiempo.

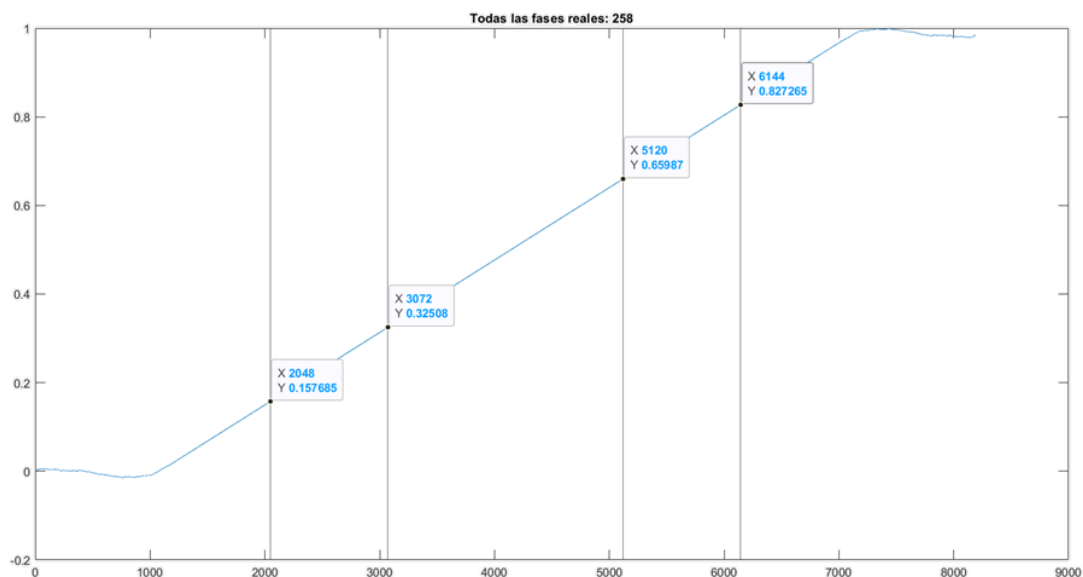
Una vez se identifican los diversos patrones asociados a cada tipo de evento; como se evidencia en las Figuras 13 y 14, se procede a crear un modelo de pendiente ideal en el rango

de frecuencias que presenta una mayor concentración de estas fases. Para ello, se consideró un intervalo de confianza tanto a la derecha como a la izquierda de la pendiente correspondiente a cada tipo de evento.

En este proceso, los datos correspondientes a las fases se normalizaron para facilitar su tratamiento. Se utiliza un modelo de fase para generar una pendiente ideal como se muestra en la Figura 15.

### Figura 15

*Fase seleccionada para generar una pendiente ideal perteneciente al evento TR*



*Nota.* A partir del conjunto de fases con mayor concentración pertenecientes al evento TR, se obtiene una pendiente linealizada. De esta pendiente se extraen cuatro puntos clave ubicados en los puntos  $3M/8$ ,  $5M/8$ ,  $M/4$  y  $3M/4$ . Donde M representa los 8192 puntos de la transformada de Fourier.

### Tabla 3

*Identificación de datos ideales para seleccionar fases lineales*

Posición	Amplitud LP	Amplitud TC	Amplitud TR	Amplitud VT
<b>2048</b>	-0.13	0.13	0.16	-0.14
<b>3072</b>	-0.25	0.32	0.33	-0.32
<b>5120</b>	-0.75	0.68	0.66	-0.68
<b>6144</b>	-0.88	0.86	0.83	-0.86

*Nota.* Propuesta de datos para generar una pendiente ideal en cada tipo de evento microsísmico que han sido tomados de los cuatro diferentes puntos seleccionados. Basado en fases lineales como se observa en la Figura 15.

Tras haber identificado los datos necesarios para crear una fase ideal de cada tipo de evento, tal como se evidencia en la Tabla 3, se procede a emplear la fórmula, específica para la pendiente con el fin de obtener la ecuación representativa de la fase ideal.

La ecuación que se utiliza para determinar la pendiente de una fase ideal es la siguiente:

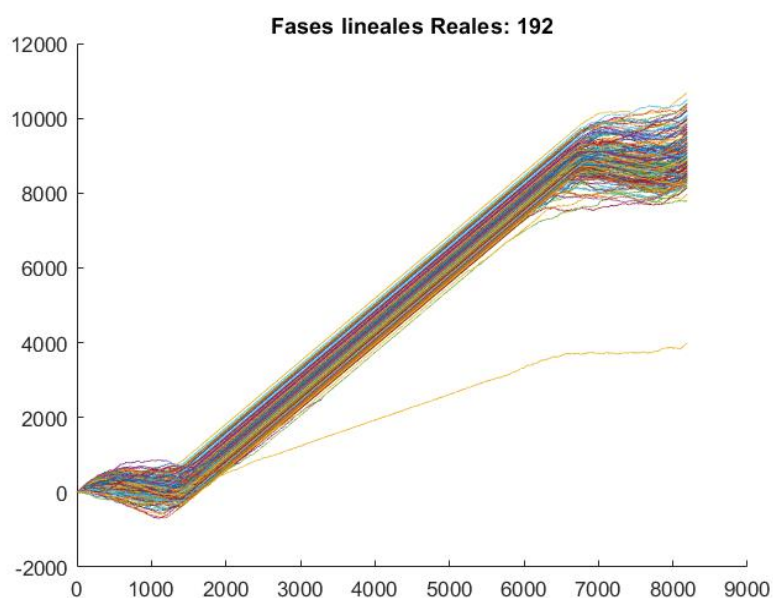
$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (17)$$

Después de reemplazar los datos de la Tabla 3 en la ecuación 17, se obtienen ecuaciones representativas para cada tipo de evento, correspondientes a fases ideales. Al aplicar un intervalo de confianza del 20% tanto a la derecha como a la izquierda, se delimita un conjunto de fases linealizadas. De acuerdo con la investigación de (Pérez, 2022), este conjunto de fases contiene aquellas que mejor capturan las características representativas de cada tipo de evento y, por lo tanto, son idóneas para ser replicadas mediante el método de Bootstrap.

En el estudio de investigación realizado por (Pérez, 2022), se explica que el uso del intervalo de confianza del 20% se debe a que los valores de las pendientes, obtenidos al reemplazar los valores de la Tabla 3, se encuentran dentro de dicho intervalo. Como resultado, las fases no se ven afectadas en su linealidad debido a la presencia de ruido, y las señales que caen dentro de este rango exhiben de manera más precisa las características inherentes a las señales.

### Figura 16

*Zona delimitada alrededor de la fase ideal para el evento TC*



*Nota.* Al delimitar la zona que contiene las fases más lineales alrededor del 20% de la recta de la fase ideal, se logran extraer 192 fases que conservan mejor las características representativas del evento TC.

Al realizar este procedimiento para cada uno de los cuatro tipos de eventos, se lleva a cabo un proceso de selección para obtener lo que llamamos "Fases ideales". Los resultados de este proceso se muestran en la Tabla 4.

**Tabla 4**

*Cantidad de fases ideales, después de realizar el proceso de selección*

<b>Tipo de evento</b>	<b>Cantidad de Fases ideales</b>
<b>Largo Período</b>	157
<b>Tectónicos</b>	192
<b>Tremor</b>	106
<b>Volcano Tectónicos</b>	138

*Nota.* Al completar el proceso de selección, se identifica la cantidad de fases que cumplen con el modelo de ecuación de cada fase ideal aplicado un intervalo de confianza del 20% tanto a la derecha como a la izquierda.

### **Normalizar las Magnitudes**

Para asegurar un procedimiento efectivo en la generación de señales sintéticas mediante el método de Bootstrap, se lleva a cabo la normalización de todas las magnitudes de los eventos correspondientes, representado matemáticamente en la ecuación (3). La normalización de las magnitudes garantiza una comparación equitativa entre los datos y facilita un proceso coherente y confiable en la generación de las señales sintéticas.

### **Método de Bootstrap a las Magnitudes**

La aplicación del método de Bootstrap implica realizar un remuestreo de las magnitudes originales previamente normalizadas para crear múltiples conjuntos de señales sintéticas. La metodología empleada en este trabajo se basa en el enfoque presentado por (Rocco S. & Zio, 2005), quienes explican el algoritmo de Bootstrap para estimar la variación estándar de un parámetro. Esta técnica se describe de la siguiente manera:

- Definir un valor  $B$  que representa el número de muestras de Bootstrap independientes. En este caso, dicho valor se refiere a la longitud de las muestras de remuestreo. Cada una de estas muestras consta de  $N$  valores de datos extraídos con reemplazo de la muestra original. Para la explicación del método lo indica como un vector  $Z_{1N}$ . Específicamente en nuestra situación, dicho valor de  $N$  corresponde a los 8192 puntos de la Transformada de Fourier.
- Calcular la estadística de interés para cada muestra de Bootstrap para constituir el conjunto de salida  $X_B$ .
- Estimar la desviación estándar al utilizar la desviación estándar de las  $B$  repeticiones.

El último punto se explica de la siguiente manera: una vez se ha calculado la estadística de interés para cada muestra Bootstrap, se toma la desviación estándar de todas las estadísticas y se utiliza como estimación de la desviación estándar de la población. Este último paso se lleva a cabo porque el método de Bootstrap busca aproximar la distribución muestral de la estadística de interés.

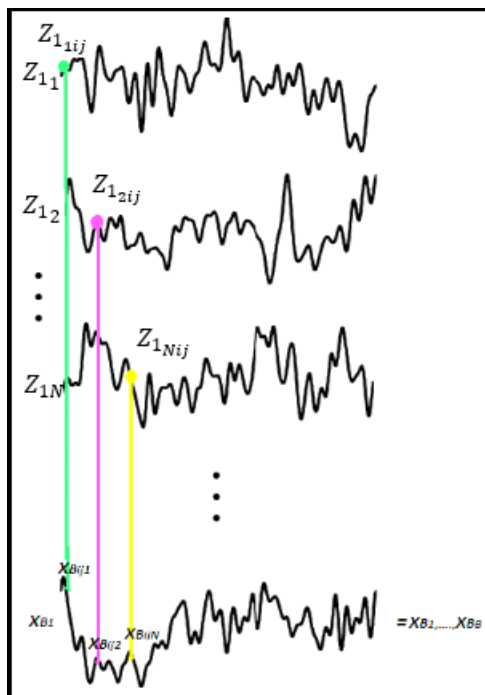
Un aspecto destacado señalado por (Rocco S. & Zio, 2005), es la recomendación de considerar valores de  $B \geq 1000$  en caso de que se pretenda estimar intervalos de confianza. Esto también sirve para limitar el tiempo de procesamiento computacional necesario para el cálculo de las réplicas en caso de que se pretenda estimar intervalos de confianza. Esto también sirve para limitar el tiempo de procesamiento computacional necesario para el cálculo de las réplicas.

La visualización de dicho proceso puede ser ejemplificada con claridad en el trabajo de (Aguinaga Navarreta, 2018), donde se presenta de forma gráfica la técnica de Bootstrap en la Figura 17.



## Figura 17

*Ejemplo de proceso Bootstrap*



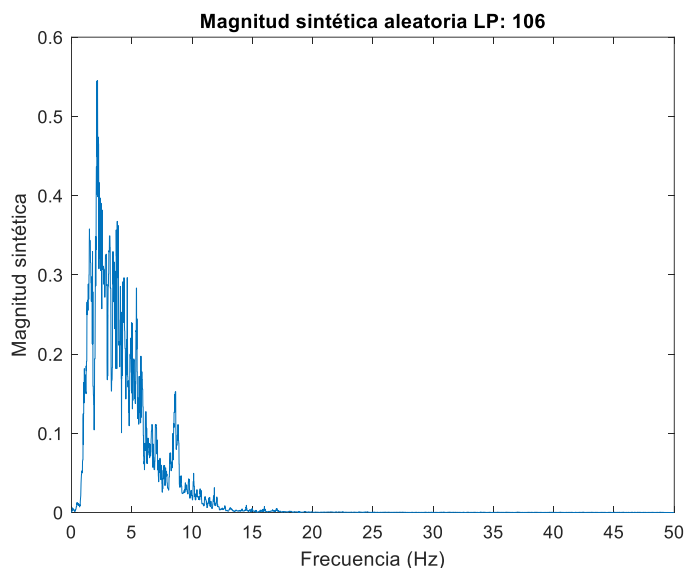
*Nota.* La Figura presenta de manera gráfica el proceso de generación de señales sintéticas al utilizar la técnica de Bootstrap. Tomado de (Aguinaga Navarreta, 2018).

La Figura 17 ilustra la metodología utilizada en este trabajo, donde se parte de una magnitud perteneciente a un evento  $Z_{11}$  y se toma un punto  $Z_{1ij}$  para conformar el punto  $x_{B1ij1}$  de la señal  $X_{B1}$  correspondiente a la magnitud sintética, se toma otro punto  $Z_{1Nij}$  que pertenece al evento  $Z_{1N}$ , el cual forma otro punto  $x_{B1ijN}$  de señal sintética  $X_{B1}$ , este proceso se repite de forma aleatoria desde las magnitudes  $Z_{11}$  hasta  $Z_{1N}$ , el proceso toma puntos en orden hasta completar la primera réplica o magnitud sintética. Una vez que el proceso se repite las  $B$  veces, da como resultado nuestro conjunto de  $B$  magnitudes sintéticas. Se representa matemáticamente la matriz que contiene las señales sintéticas,  $X_B$ , cómo:

$$X_B = \{x_{B1}^T, x_{B2}^T, \dots, x_{BN}^T\}^T. \quad (18)$$

### Figura 18

*Magnitud sintética LP generada mediante el método de Bootstrap*



*Nota.* Magnitud sintética de una señal microsísmica LP antes de aplicar el sistema de espejo.

Solo se representan sus primeros 4096 puntos.

Una vez que se han obtenido las magnitudes sintéticas, las cuales se encuentran en un estado normalizado, se procede a combinar cada magnitud con una fase aleatoria. Para lograr esto, primero se realiza un proceso mediante el cual se recupera el valor original de la magnitud: se elige al azar una magnitud real y se multiplica por la magnitud normalizada obtenida. De esta manera, se genera una magnitud sintética con características similares a las magnitudes reales.

En las magnitudes sintéticas generadas, al igual que en las magnitudes reales, se está definiendo el intervalo de confianza (IC). Este intervalo representa un rango numérico en el cual

se espera que el valor real de un parámetro poblacional desconocido se encuentre. En nuestro contexto, este parámetro se relaciona con las señales sintéticas el cual se representa de manera gráfica en las siguientes Figuras. Con un nivel de confianza del 95%, se evalúa si las magnitudes sintéticas cumplen con estar dentro de este rango. Además, se calcula la media de las señales para llevar a cabo un análisis de similitud con las magnitudes reales.

Para determinar la representación matemática se parte de la ecuación (3) de estandarización:

$$Z_{SM} = \frac{h_M - \mu_M}{\sigma_M} = \frac{h_M - \mu}{\sigma/\sqrt{n}}, \quad (19)$$

Se reemplaza tanto  $\mu_M$  por  $\mu$ , cómo  $\sigma_M$  por  $\sigma/\sqrt{n}$ , debido al teorema del límite central.

Se resuelve  $\mu$  en términos de  $Z_S$ , dando como resultado:

$$\mu = h_M \pm Z_{SM} \sigma/\sqrt{n}, \quad (20)$$

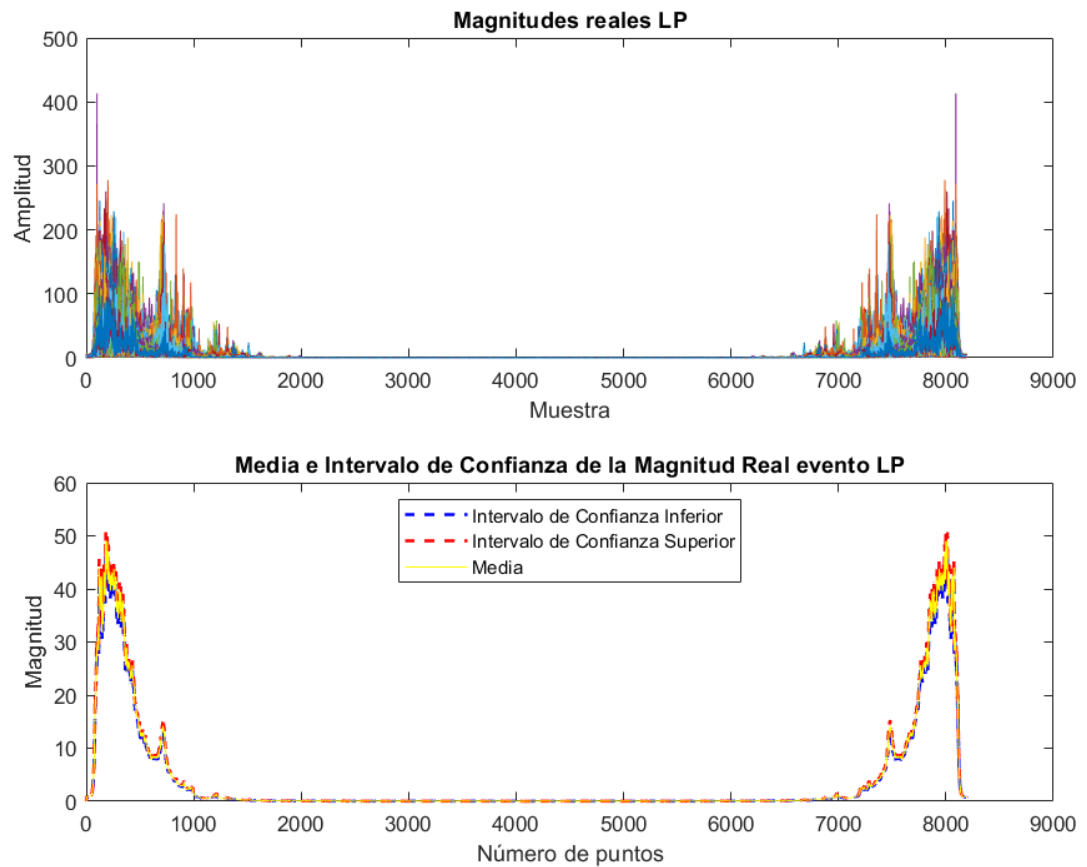
El mismo teorema del límite central, establece que la distribución muestral tiende a ser normal, y esta es simétrica. Si se establece en estos términos y en función de las señales sintéticas, se obtiene:

$$X_B - Z_\alpha \left( \frac{\sigma}{\sqrt{n}} \right) \leq \mu \leq X_B + Z_\alpha \left( \frac{\sigma}{\sqrt{n}} \right). \quad (21)$$

La fórmula (21) establece el intervalo de confianza para las magnitudes sintéticas. Mientras que el valor reemplazado de  $Z_\alpha$  representa el nivel de confianza, dado que  $\alpha$  es la probabilidad de que el rango no contenga a la magnitud sintética (Rice University, 1999).

## Figura 19

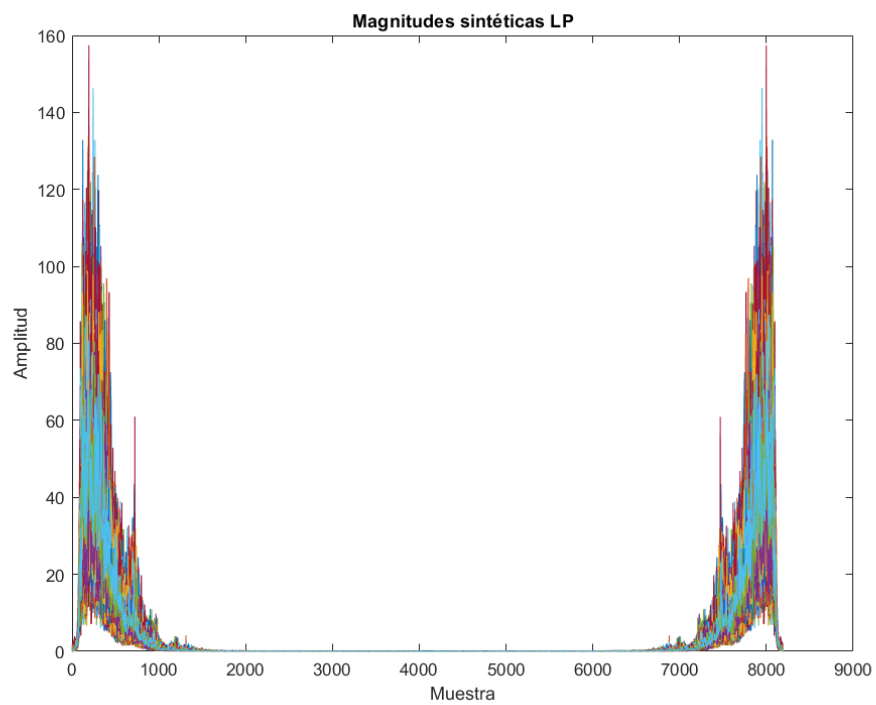
*Representación de las Magnitudes reales e intervalo de confianza del evento LP*



*Nota.* La Figura tiene como objetivo visualizar la variabilidad y la incertidumbre asociada a estas magnitudes. Este intervalo proporciona una estimación de la dispersión de las magnitudes observadas del evento LP.

### Figura 20

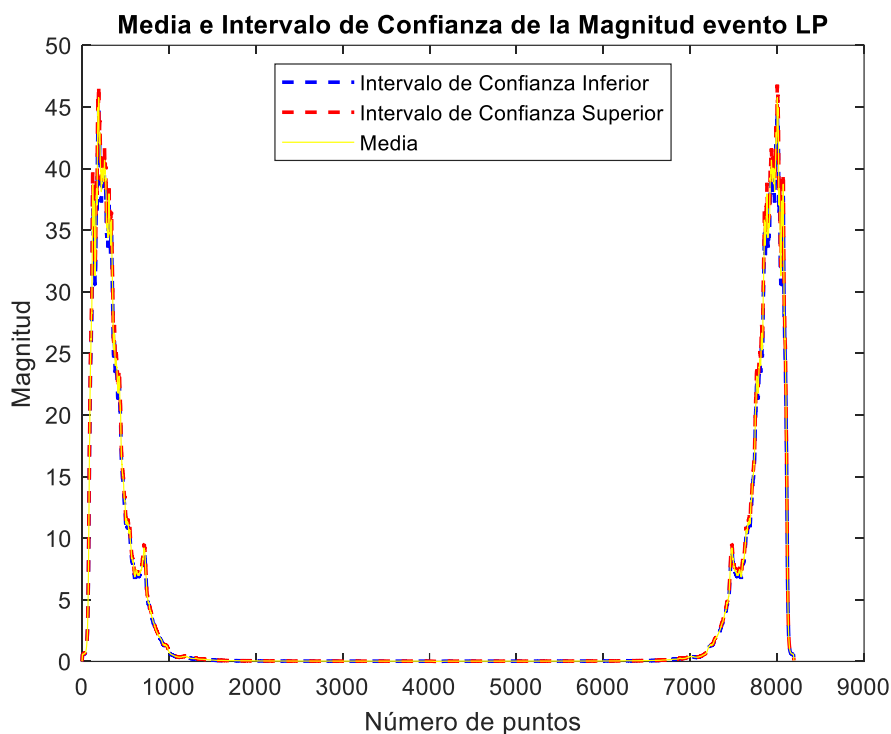
*Representación de las Magnitudes sintéticas del evento LP*



*Nota.* La Figura muestra 1000 magnitudes sintéticas correspondientes al evento LP.

### **Figura 21**

*Valor medio e intervalo de confianza de las magnitudes sintéticas del evento LP*



*Nota.* En la Figura se observa que la media de las magnitudes se encuentra dentro del intervalo de confianza del 95%.

En la Figura 21, se calcula y presenta gráficamente el intervalo de confianza de la media para un conjunto de datos de eventos microsísmicos LP. Primero se define un nivel de confianza deseado del 95%. Luego, se calcula la media y la desviación estándar de las filas de datos correspondientes a estos eventos. Esta información se utiliza para determinar el intervalo de confianza al calcular los límites inferior y superior basados en el valor crítico de la distribución t de Student y el tamaño de la muestra. Posteriormente, genera un gráfico que representa los intervalos de confianza inferior y superior mediante líneas punteadas azules y rojas, respectivamente. También dibuja una línea amarilla que indica las medias de los eventos LP.

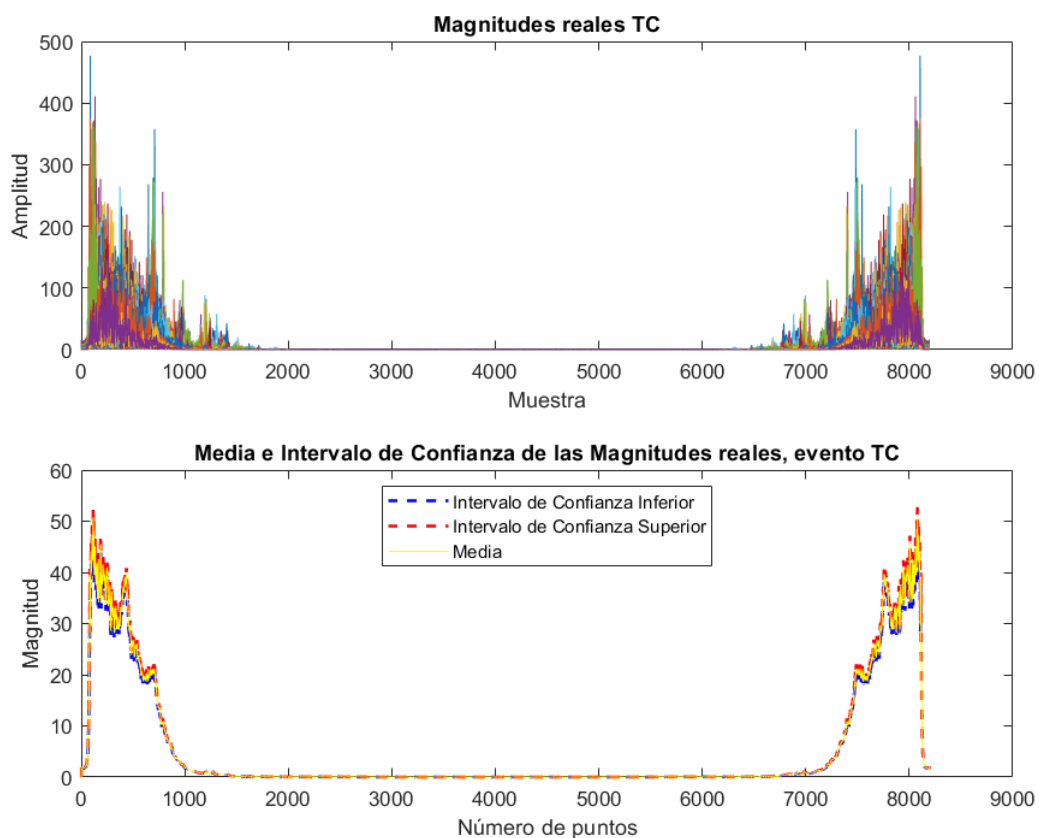
La Figura 19 exhibe la variabilidad presente entre las magnitudes reales y las magnitudes sintéticas, al alcanzar un pico de 400 de amplitud en las magnitudes reales y de

140 de amplitud en las sintéticas tal como se visualiza en la Figura 20. A pesar de esta variación, es relevante señalar que tanto la amplitud promedio como la forma muestran concordancia, lo que respalda que el método de Bootstrap logró capturar de manera efectiva las características distintivas del evento.

El proceso mencionado se repite para los eventos microsísmicos restantes, lo que produce la obtención de las siguientes Figuras.

### Figura 22

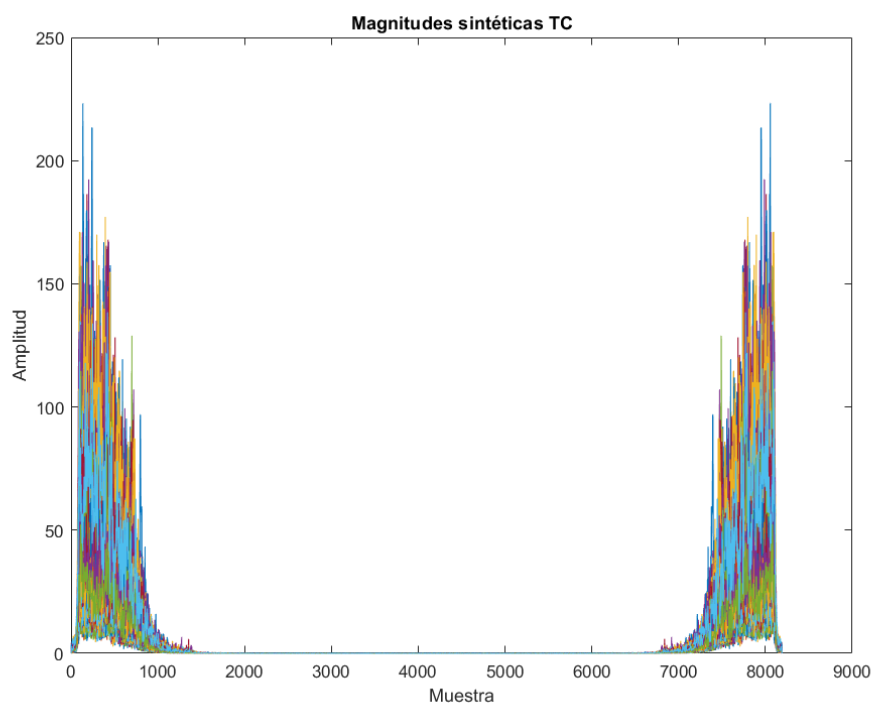
*Representación de las Magnitudes reales e intervalo de confianza del evento TC*



*Nota.* La Figura tiene como objetivo visualizar la variabilidad y la incertidumbre asociada a estas magnitudes. Este intervalo proporciona una estimación de la dispersión de las magnitudes observadas del evento TC.

### Figura 23

*Representación de las Magnitudes sintéticas del evento TC*

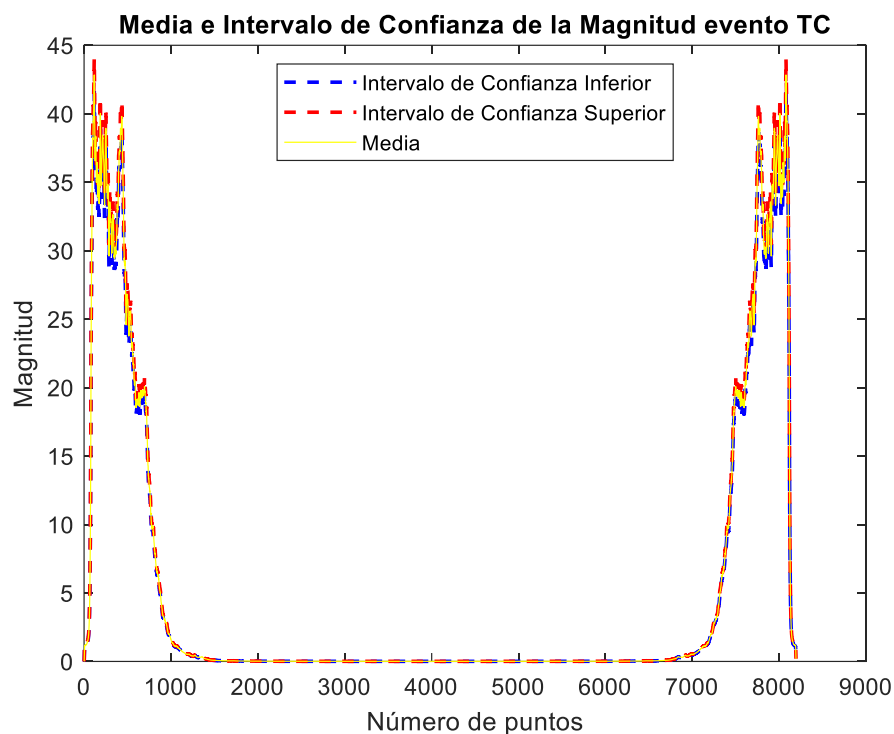


*Nota.* La Figura muestra 1000 magnitudes sintéticas correspondientes al evento TC.

### Figura 24

*Valor medio e intervalo de confianza de las magnitudes sintéticas del evento TC*



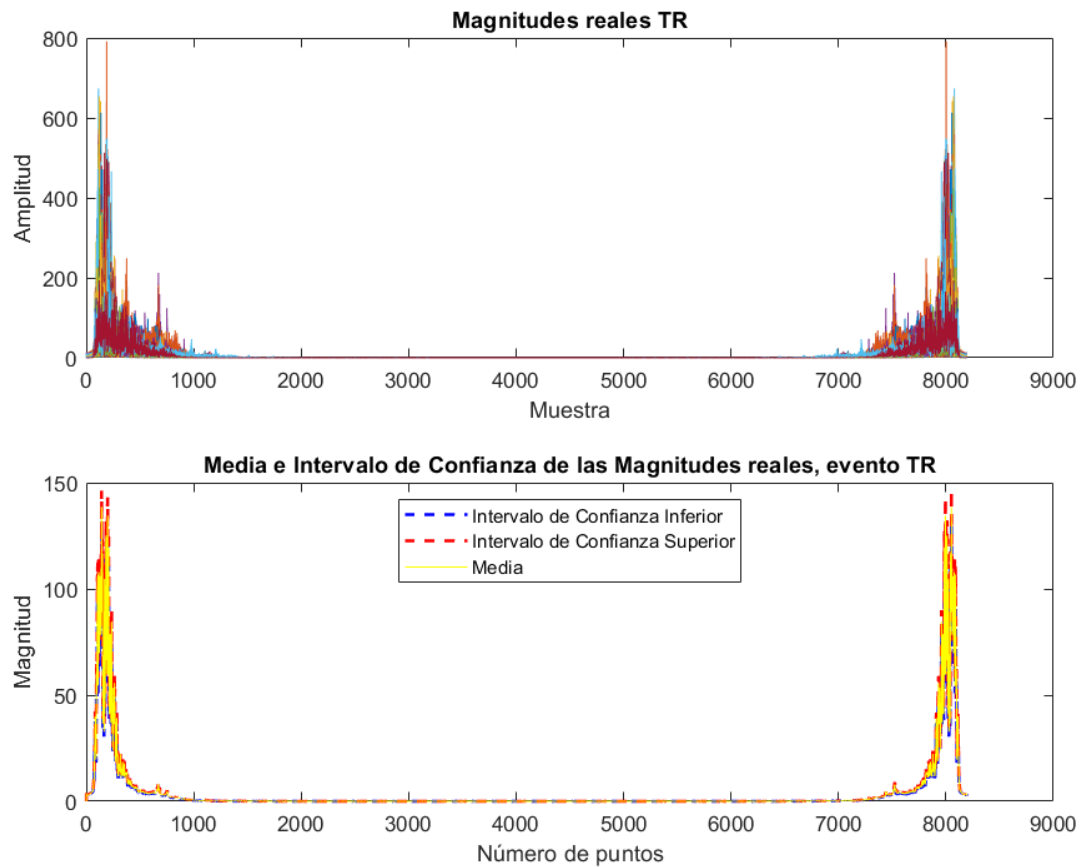


*Nota.* En la Figura se observa que la media de las magnitudes se encuentra dentro del intervalo de confianza del 95%.

La Figura 22 exhibe la variabilidad presente entre las magnitudes reales y las magnitudes sintéticas, al alcanzar un pico de aproximadamente 500 de amplitud en las magnitudes reales y de aproximadamente 220 de amplitud en las sintéticas tal como se visualiza en la Figura 23. A pesar de esta variación, es relevante señalar que tanto la amplitud promedio como la forma muestran concordancia, lo que respalda que el método de Bootstrap logró capturar de manera efectiva las características distintivas del evento TC.

### Figura 25

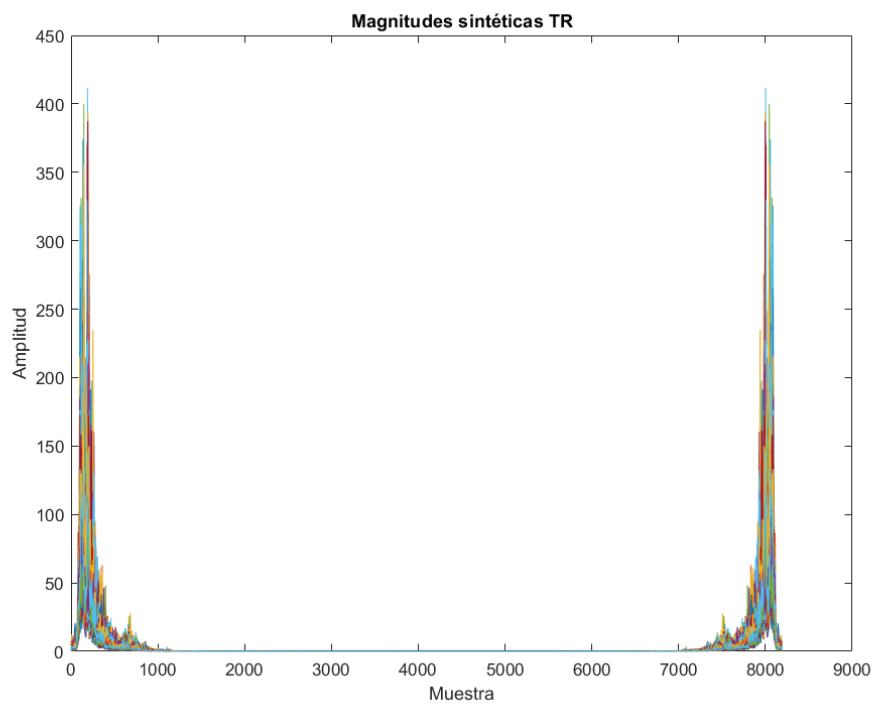
*Representación de las Magnitudes reales e intervalo de confianza del evento TR*



*Nota.* La Figura tiene como objetivo visualizar la variabilidad y la incertidumbre asociada a estas magnitudes. Este intervalo proporciona una estimación de la dispersión de las magnitudes observadas del evento TR.

### Figura 26

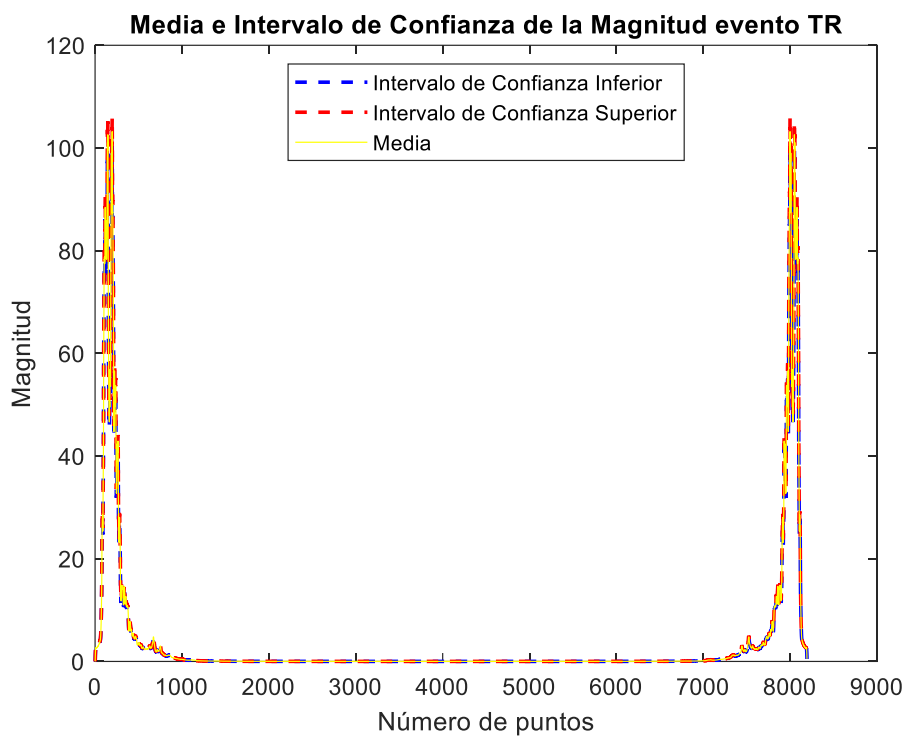
*Representación de las Magnitudes sintéticas del evento TR*



*Nota.* La Figura muestra 1000 magnitudes sintéticas correspondientes al evento TR.

### **Figura 27**

*Valor medio e intervalo de confianza de las magnitudes sintéticas del evento TR*

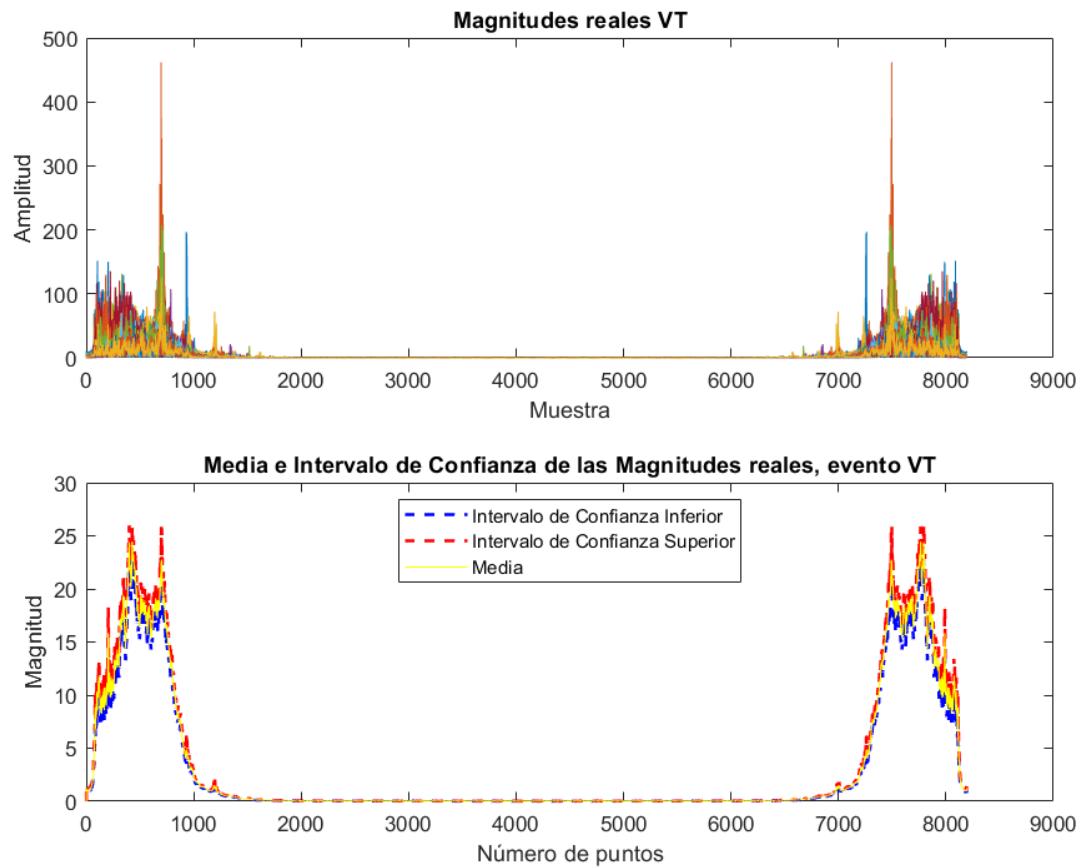


*Nota.* En la Figura se observa que la media de las magnitudes se encuentra dentro del intervalo de confianza del 95%.

La Figura 25 exhibe la variabilidad presente entre las magnitudes reales y las magnitudes sintéticas, al alcanzar un pico de aproximadamente 800 de amplitud en las magnitudes reales y de aproximadamente 400 de amplitud en las sintéticas tal como se visualiza en la Figura 26. A pesar de esta variación, es relevante señalar que tanto la amplitud promedio como la forma muestran concordancia, lo que respalda que el método de Bootstrap logró capturar de manera efectiva las características distintivas del evento TC.

### **Figura 28**

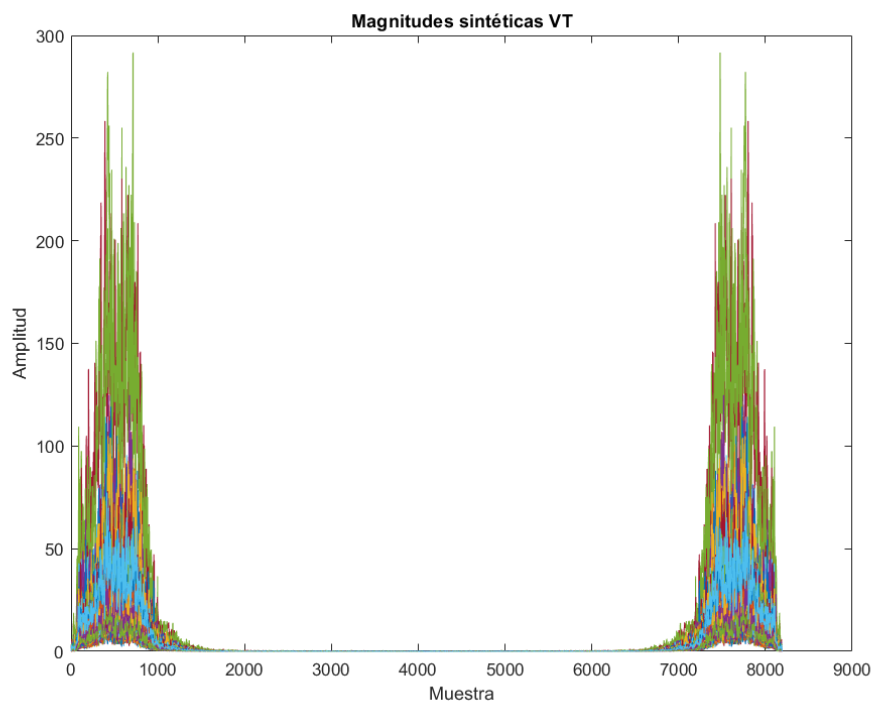
*Representación de las Magnitudes reales e intervalo de confianza del evento VT*



*Nota.* La Figura tiene como objetivo visualizar la variabilidad y la incertidumbre asociada a estas magnitudes. Este intervalo proporciona una estimación de la dispersión de las magnitudes observadas del evento VT.

### Figura 29

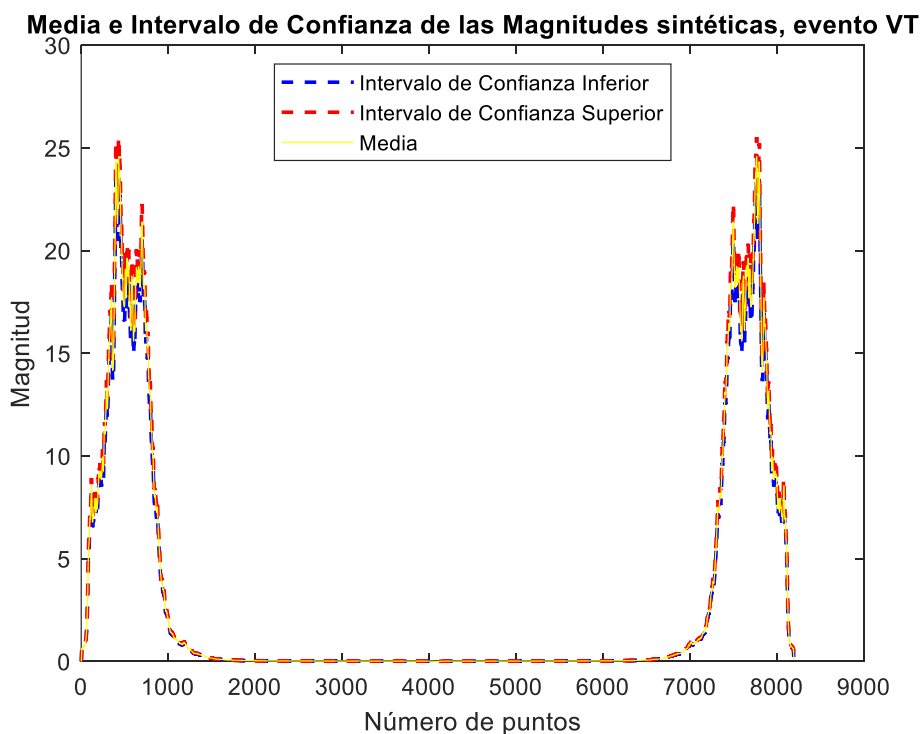
*Representación de las Magnitudes sintéticas del evento VT*



*Nota.* La Figura muestra 1000 magnitudes sintéticas correspondientes al evento VT.

### **Figura 30**

*Valor medio e intervalo de confianza de las magnitudes sintéticas del evento VT*



*Nota.* En la Figura se observa que la media de las magnitudes se encuentra dentro del intervalo de confianza del 95%.

La Figura 28 exhibe la variabilidad presente entre las magnitudes reales y las magnitudes sintéticas, al alcanzar un pico de aproximadamente 500 de amplitud en las magnitudes reales y de aproximadamente 300 de amplitud en las sintéticas tal como se visualiza en la Figura 29. A pesar de esta variación, es relevante señalar que tanto la amplitud promedio como la forma muestran concordancia, lo que respalda que el método de Bootstrap logró capturar de manera efectiva las características distintivas del evento TC.

### **Igualar número de Fases solicitadas**

Una vez se generan las "B" magnitudes sintéticas mediante el método de Bootstrap, se procede a realizar una selección de "B" Fases ideales de la base de datos almacenada. El objetivo es asegurar que el número de magnitudes y fases coincida para generar las señales

microsísmicas sintéticas. Esta selección garantiza que cada magnitud sintética se encuentre asociada a una Fase ideal específica, permite crear un conjunto coherente de señales sintéticas que preserven las características representativas de cada tipo de evento.

### **Unión Magnitud sintética y Fase real**

En la siguiente sección del código, una vez igualadas las fases ideales con las magnitudes sintéticas generadas mediante el método de Bootstrap, se lleva a cabo la combinación de cada magnitud sintética con una Fase ideal seleccionada aleatoriamente. Esta combinación se realiza mediante una solución matemática que permite representar un número complejo en términos de su magnitud y fase. El resultado de esta operación es un conjunto de "B" señales microsísmicas sintéticas en el dominio de la frecuencia.

Además, se realiza un proceso de reflejo para obtener los 8192 puntos del espectro de frecuencia que tiene simetría conjugada. Como se mencionó con anterioridad esta selección se basa en la necesidad de alcanzar una alta resolución de 8k, lo que aporta una mayor claridad para capturar de manera precisa los detalles presentes en las señales. Como resultado, se experimenta una mejora sustancial en la calidad de las señales sintéticas que son generadas.

Esto implica que las frecuencias positivas y negativas son idénticas en magnitud, pero difieren en fase. La obtención de ambos lados del espectro de frecuencia es fundamental para tener una representación completa y precisa de las señales en el dominio de Fourier, y así garantizar que las características de la señal en el dominio de la frecuencia sean correctamente preservadas en las señales sintéticas resultantes.

### **Transformada Inversa de Fourier**

Por último, se realiza la transformada Inversa de Fourier para pasar las señales al dominio del tiempo. Estas señales combinan de manera coherente las magnitudes sintéticas con Fases ideales específicas, esto asegura que cada señal sintética refleje adecuadamente



las características representativas de cada tipo de evento microsísmico ahora en el dominio temporal.

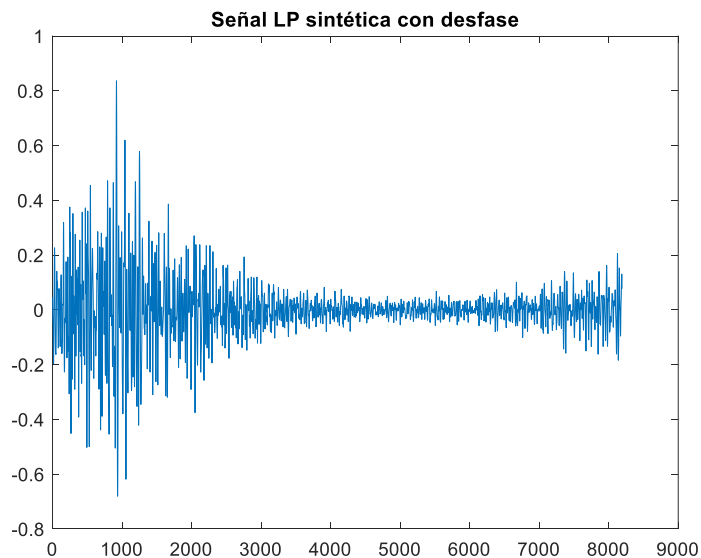
### **Recuperación del desfase**

Al transformar la señal sintética al dominio del tiempo se observa que se presenta un desplazamiento temporal, esto está relacionado con la fase compleja de la señal al utilizar fases aleatorias dentro del rango. Después de combinar magnitudes sintéticas con fases reales; es una consecuencia de la transformada inversa de Fourier y la información de desfase temporal que se encuentra en la representación compleja de la señal.

El desplazamiento temporal en las señales sintéticas surge debido al proceso de transformación del dominio de la frecuencia al dominio del tiempo mediante el uso de la transformada de Fourier inversa. Durante esta transformación, la información de fase se traduce en desplazamientos en el tiempo en la señal sintética. Esto significa que la señal sintética resultante experimenta un desplazamiento temporal debido a los desfases introducidos por las fases ideales en la combinación con las magnitudes sintéticas.

### **Figura 31**

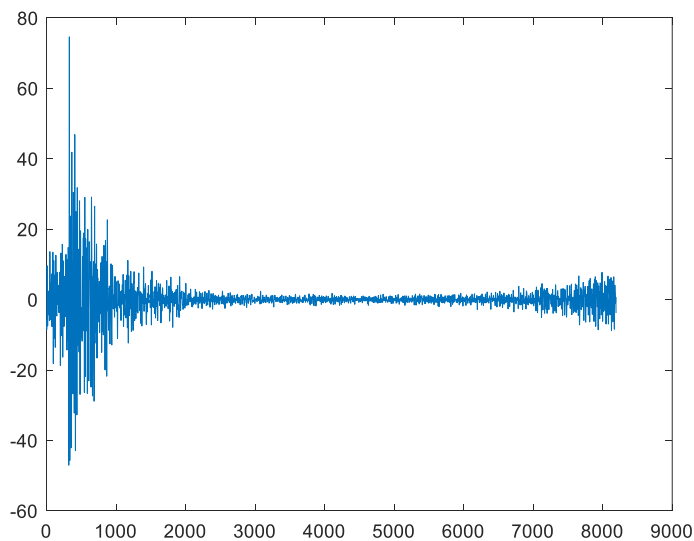
*Señal sintética correspondiente al evento LP con desfase*



*Nota.* En la Figura, es evidente la presencia de un desfase en los puntos finales de la señal sintética. Esta característica surge debido a la variabilidad inherente a la fase ideal elegida.

### Figura 32

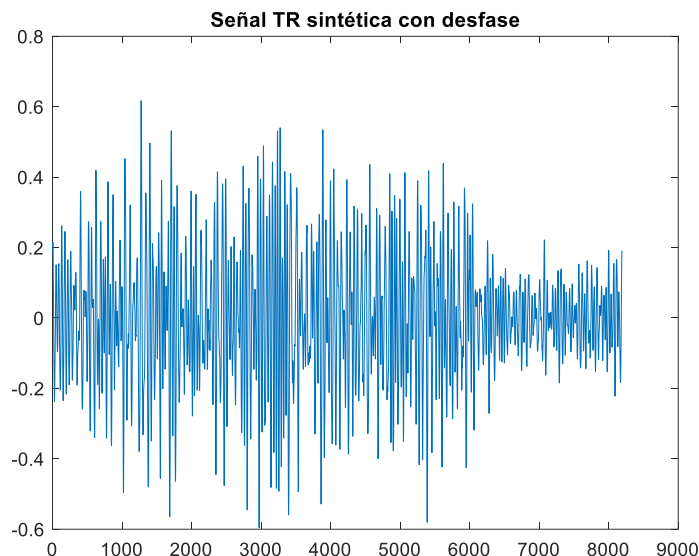
*Señal sintética correspondiente al evento VT con desfase*



*Nota.* En la Figura, es evidente la presencia de un desfase en los puntos finales de la señal sintética. Esta característica surge debido a la variabilidad inherente a la fase ideal elegida.

### Figura 33

*Señal sintética correspondiente al evento TR*



*Nota.* En la Figura, se nota con menor claridad el desfase en los puntos finales de la señal sintética debido a la irregularidad de la señal en todo el dominio del tiempo. Esta característica surge debido a la variabilidad inherente a la fase ideal elegida.

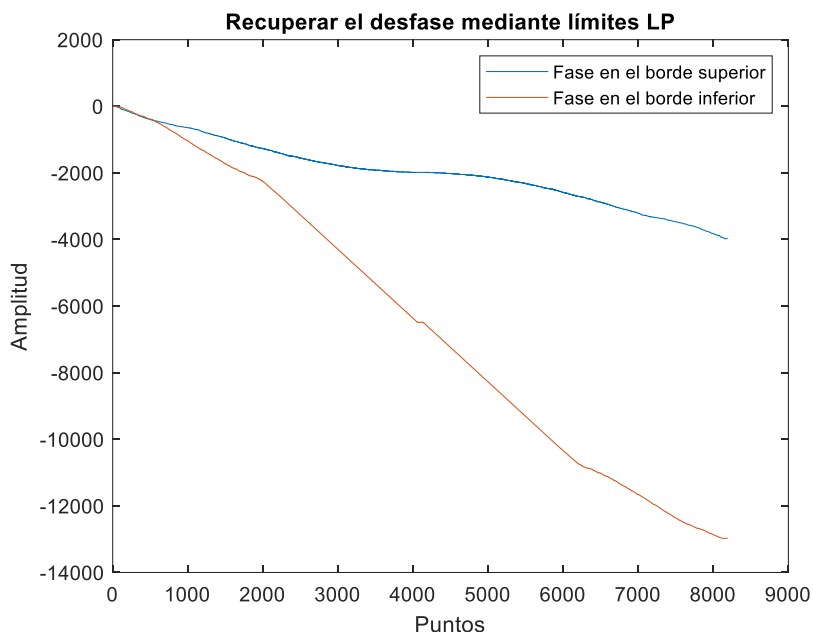
Para compensar el desfase se realizan dos métodos:

#### ***Desfasar el peor de los casos***

Se lleva a cabo un proceso para recuperar el desfase perdido debido a la selección aleatoria de la Fase ideal. Esto implica seleccionar los límites superior e inferior del rango que contenía el conjunto de Fases ideales y calcular el ángulo existente entre estos límites. Posteriormente, se aplica este nuevo ángulo complejo para desfasar la señal y así corregir el desfase inicial.

### Figura 34

*Recuperar el desfase de la señal sintética mediante proceso en los límites de los rangos*

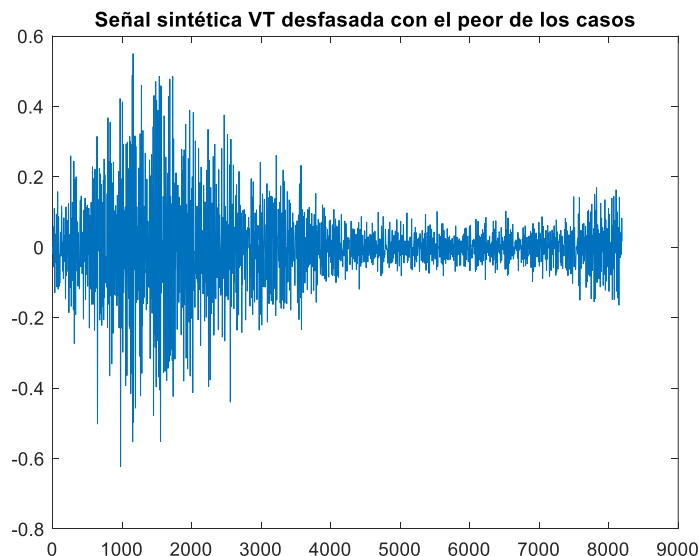


*Nota.* Se toma la fase en los bordes tanto superior, como inferior para determinar el ángulo complejo que existe entre ellas, al ser el caso más extremo a considerar.

Al aplicar este procedimiento resulta que no proporciona los resultados esperados debido a que al intentar desfasear el peor y más extremo de los casos, este desfase se aplica a cada fase aleatoria seleccionada dentro del rango. Como resultado, se produce un desfase excesivo en todas las fases, ya que se recurre al caso más extremo en lugar de ajustar el desfase de manera más precisa y específica para cada fase individualmente.

### Figura 35

*Propuesta de recuperar el desfase al considerar el peor de los casos*



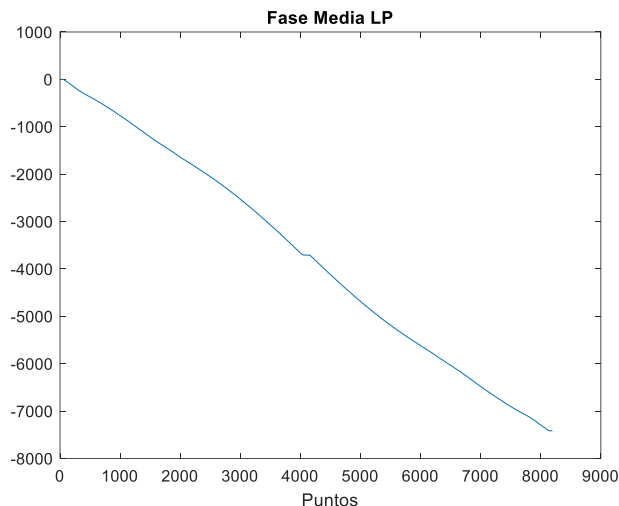
*Nota.* La Figura revela que la recuperación del desfase no se ha logrado de manera exitosa. Este resultado se origina en la consideración del caso extremo en el que se eligen las fases en los límites de sus rangos posibles.

### ***Desfase por media***

Al comprobar que desfasar el caso extremo no arroja resultados satisfactorios debido a que no se toma en cuenta el desfase individual, se opta por calcular una media de las fases complejas del conjunto de fases ideales. Esta aproximación más equilibrada y precisa permite mejorar la calidad de las señales sintéticas, ya que se tiene en cuenta la variabilidad de cada fase individual.

### **Figura 36**

*Media de las fases ideales LP utilizada para recuperar el desfase*

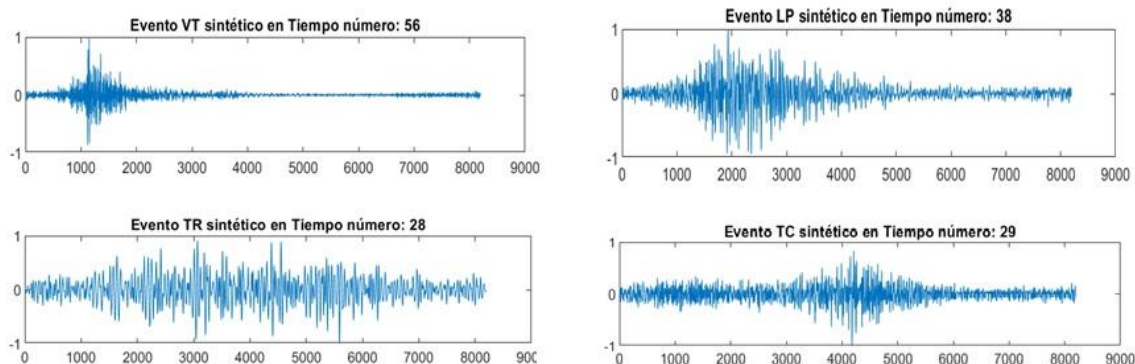


*Nota.* La Figura muestra la fase compleja utilizada para recuperar el desfase provocado por la aleatoriedad del uso de Fases ideales.

Al utilizar la media de las fases, se aplica el método matemático previamente utilizado para combinar la señal en el dominio de la frecuencia con este nuevo desfase. Al llevar a cabo este procedimiento, se logra una recuperación satisfactoria del desfase en todos los casos. La combinación de la señal con el desfase de la media de las fases, permite obtener resultados precisos en cada señal sintética, corrige de manera efectiva el desfase introducido por el uso de Fases ideales aleatorias. Este enfoque demuestra ser exitoso y garantiza la coherencia y calidad de las señales sintéticas. De esta manera, se concluye exitosamente la Actividad 4 correspondiente a la generación de bases sintéticas con la técnica de remuestreo Bootstrap.

### **Figura 37**

*Desfase recuperado con éxito en los distintos tipos de eventos microsísmicos*



*Nota.* La Figura ilustra el proceso de recuperación del desfase en las señales sintéticas mediante la aplicación de la técnica de desplazamiento basada en la media.

### **Diseño Aplicativo de Interfaz Gráfica**

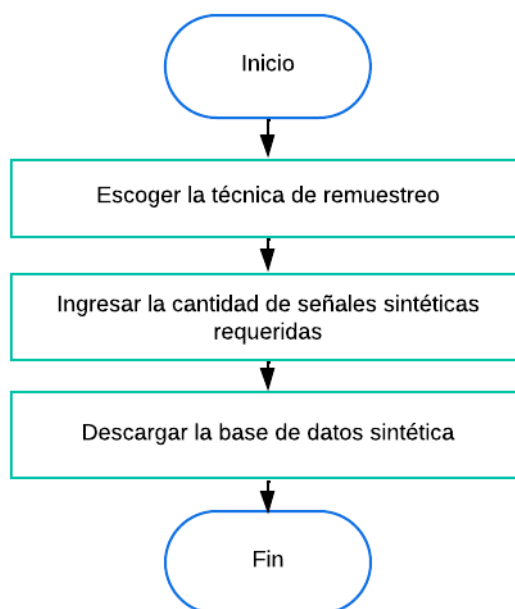
El trabajo se presenta a través de una interfaz gráfica desarrollada en el software de MATLAB®, la cual se destaca por su diseño estético y eficiente. Mediante esta interfaz los usuarios pueden seleccionar los datos de forma intuitiva y sencilla, permitiéndoles generar fácilmente cualquier cantidad de señales sintéticas microsísmicas. La herramienta brinda una experiencia amigable y efectiva para la generación de estas señales.

### ***Uso de la Aplicación***

La interfaz es diseñada de manera intuitiva, esto facilita que el usuario final pueda generar datos sintéticos sin dificultades.

### **Figura 38**

*Diagrama de flujo de la aplicación*



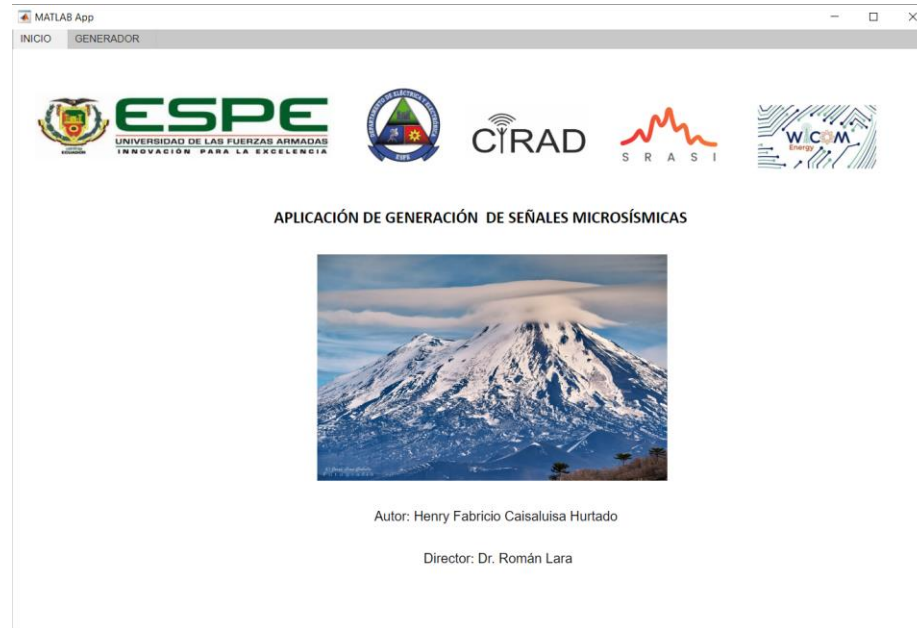
*Nota.* La Figura ilustra el diagrama de flujo del proceso de utilización de la aplicación para generar las señales sintéticas de los cuatro eventos microsísmicos.

En el primer paso, el usuario elige la técnica de remuestreo que desea utilizar, ingresa la cantidad de señales sintéticas, las cuales se distribuyen uniformemente en los cuatro tipos de eventos disponibles conforme se explica en el ícono de información de la aplicación. Finalmente, se descarga la tabla que contiene las señales sintéticas, junto con toda su información correspondiente en la carpeta que contiene la aplicación.

### **Figura 39**

*Interfaz gráfica: Pantalla de inicio*

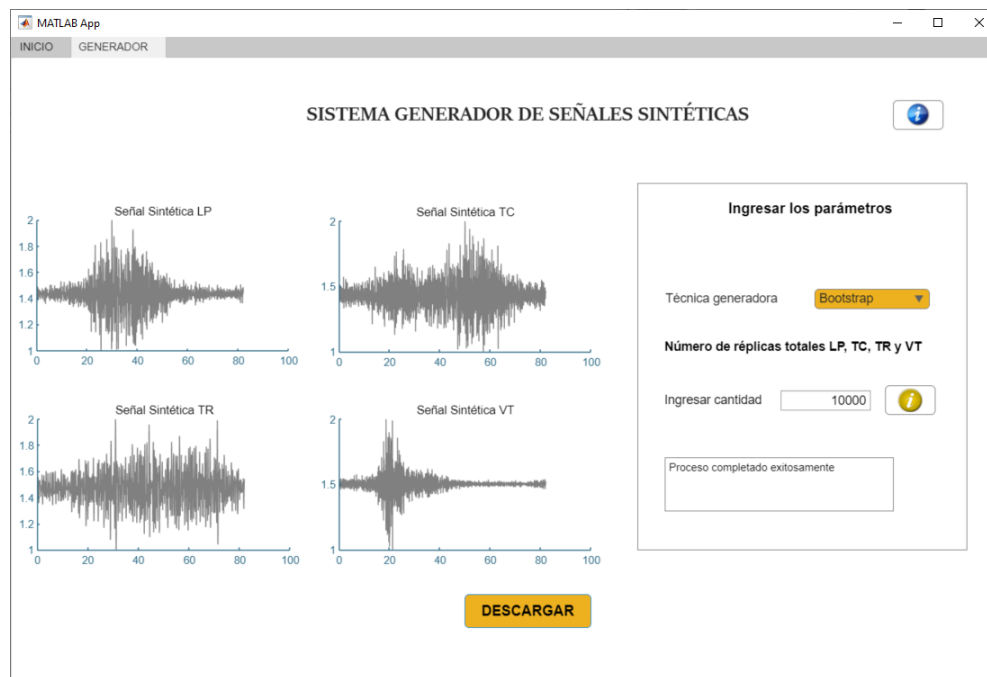




*Nota.* La Figura muestra la pantalla inicial de la aplicación, ofrece una visualización de cómo se presenta la interfaz al usuario.

### Figura 40

*Interfaz gráfica: Pantalla de sistema de generación de señales sintéticas*



*Nota.* La Figura muestra la interfaz del generador de señales sintéticas. En el lado derecho, se introducen los parámetros requeridos, mientras que en la parte izquierda se visualizan las señales sintéticas asociadas a cada evento microsísmico.

La Figura 40 muestra los parámetros ingresados, donde se aprecia que la cantidad máxima de señales sintéticas que se puede generar está limitada a diez mil. Esta cifra se considera apropiada para realizar análisis y para prevenir un posible colapso de la aplicación.

### Figura 41

*Base de datos sintética generada por el método de Bootstrap*

	1	2	3	4	5	6	7	8	9	10	11
	Network	Station	SampleRate	Component	Year	Month	Type	Duration	StartPoint	EndPoint	Data
1	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
2	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
3	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
4	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
5	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
6	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
7	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
8	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
9	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
10	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...
11	"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	'LP'	81.9200	1	8192	1x8192 dou...

*Nota.* La Figura ilustra cómo la aplicación proporciona la nueva base de datos sintética que comprende los cuatro tipos de eventos microsísmicos.

Al hacer uso de una laptop con características detalladas en la Tabla 1, el proceso de generación de la base de datos sintética, que engloba un total de diez mil señales correspondientes a los eventos microsísmicos LP, TC, TR y VT, se completa en un tiempo de 19 minutos.

## Capítulo IV

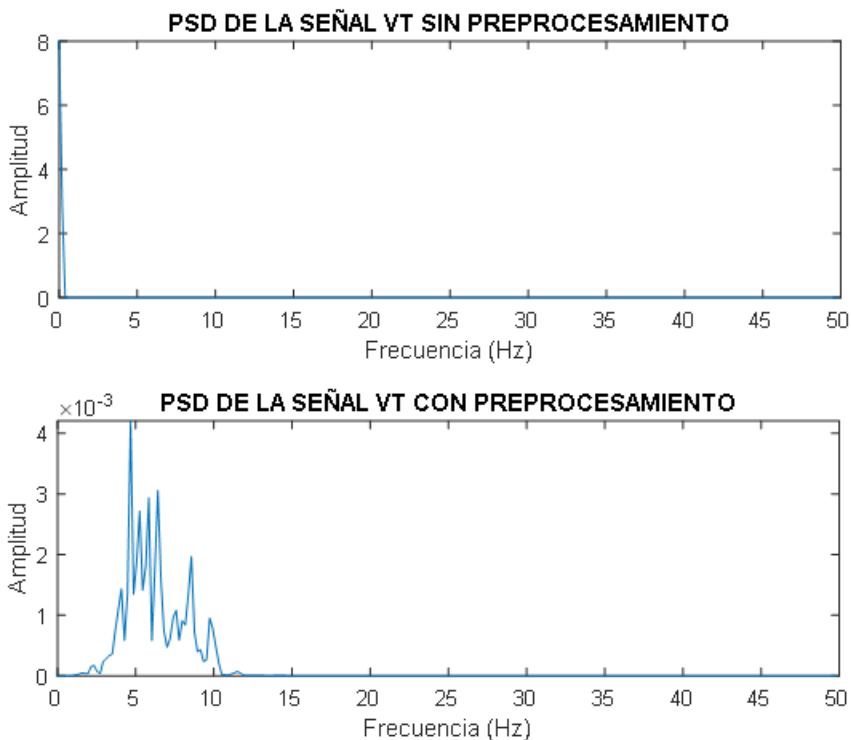
### Pruebas y Resultados

Este capítulo presenta los resultados de las señales sintéticas en conjunto con las señales reales, lo que permite observar su semejanza. Además, se ha efectuado un análisis de la PSD con el propósito de que los expertos del IGEPN puedan identificar el tipo de evento a través de su forma en el dominio del tiempo y la cantidad de energía que exhibe la PSD en el dominio de la frecuencia.

Un aspecto crucial del preprocesamiento radica en la eliminación del pico en 0.2 Hz presente en las señales originales. El libro de referencia "The Seismic Wavefield Volume I: Introduction and Theoretical Development" de B.L.N. Kennett destaca la importancia de filtrar las señales debido a la presencia de características destacadas de las ondas S en la corteza terrestre y de las ondas P en la capa oceánica. A medida que aumenta la frecuencia, la influencia de la capa oceánica se vuelve más pronunciada. Estas ondas son una característica fundamental en un componente significativo del entorno oceánico, y emergen debido a las considerables variaciones de velocidad entre las ondas, por consiguiente, es esencial suprimir dicho pico en el desplazamiento de las ondas superficiales (Kennett, 2001).

#### **Figura 42**

*Eliminación del pico en 0.2 Hz mediante el preprocesamiento de las señales*



*Nota.* La Figura revela la presencia de un pico en 0.2 Hz que es tan marcado que prácticamente reduce la señal microsísmica a una magnitud despreciable.

La Figura 42 resalta de manera evidente la relevancia de someter las señales al proceso de preprocesamiento, particularmente a través de la etapa de filtrado.

### Resultados señales sintéticas

En esta sección, se muestran los resultados de las señales sintéticas generadas para cada tipo de evento microsísmico, las cuales se presentan junto a una señal real para permitir observar las similitudes entre ellas.

La aplicación del método de Bootstrap se realizó con los siguientes parámetros mínimos:

**Tabla 5**

*Parámetros mínimos de Bootstrap*

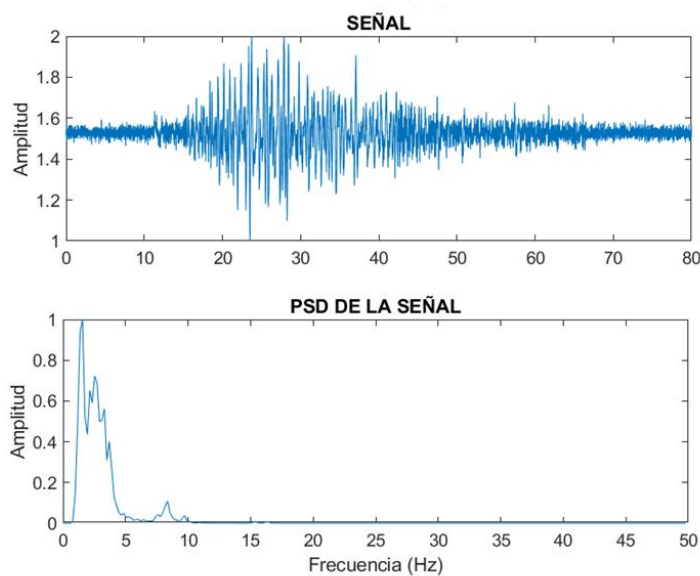
<b>Variable</b>	<b>Valor</b>
B	10000
N	8192

En la Tabla 5 se muestran los parámetros mínimos utilizados para generar las señales sintéticas. Adicional, al final las magnitudes sintéticas pasan por un filtro de mediana con tamaño de ventana 10, con el objetivo de reducir el ruido a la salida de Bootstrap.

### **Evento LP**

**Figura 43**

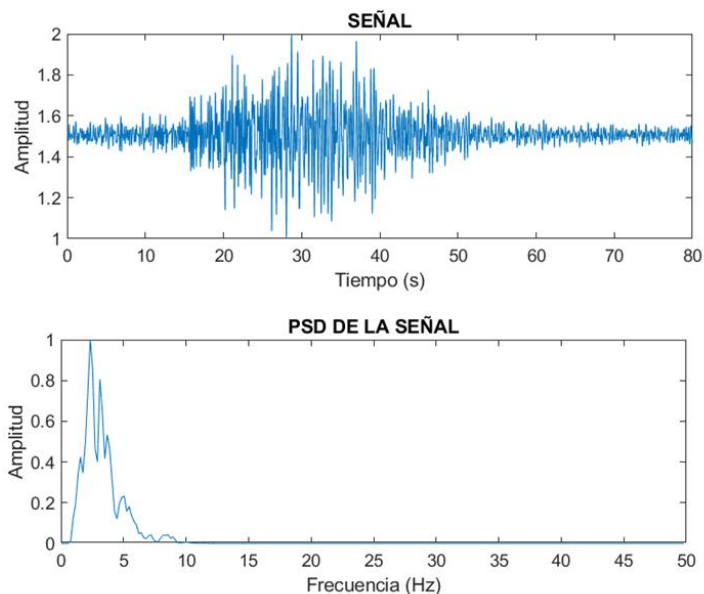
*Señal original microsísmica de tipo LP, su PSD y forma temporal*



*Nota.* Figura de una señal original tipo LP con fines comparativos.

**Figura 44**

*Señal sintética microsísmica de tipo LP, su PSD y forma temporal*

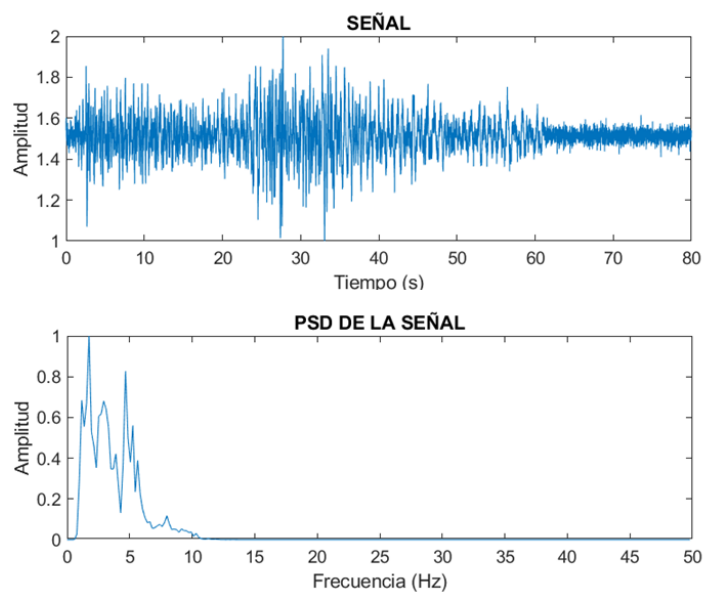


*Nota.* La Figura muestra una señal sintética LP obtenida. Este evento es un indicador importante de actividades inminentes antes de las erupciones volcánicas, se presentó la mayor concentración de energía en el rango de frecuencia de [1-5] Hz.

### **Evento TC**

#### **Figura 45**

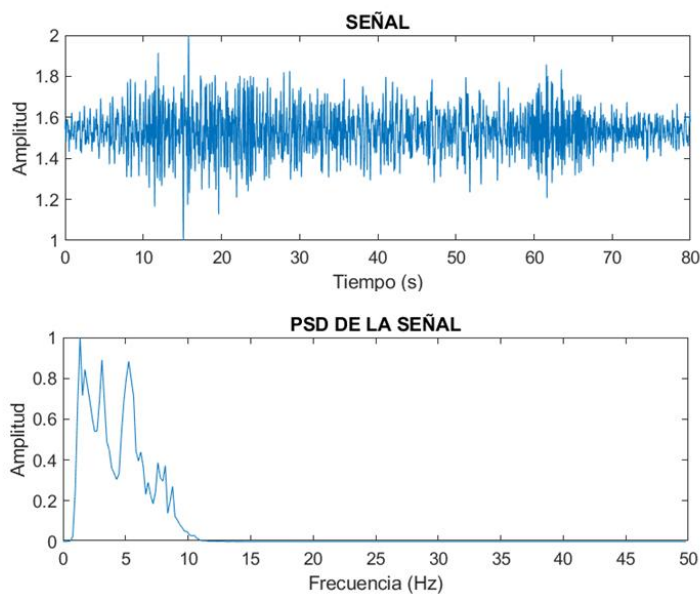
*Señal original microsísmica de tipo TC, su PSD y forma temporal*



*Nota.* Figura de una señal original tipo TC con fines comparativos

### Figura 46

*Señal sintética microsísmica de tipo TC, su PSD y forma temporal*

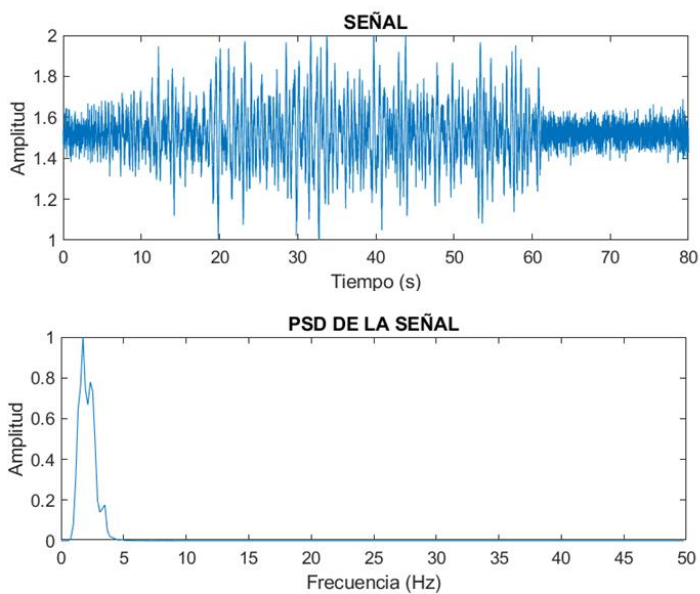


*Nota.* La Figura muestra una señal sintética tipo TC obtenida. Es el único evento que no está relacionado con la actividad volcánica, al ser este producto de fallas geológicas. Presentó la mayor concentración de energía en el rango de frecuencia de [1-10] Hz.

### Evento TR

### Figura 47

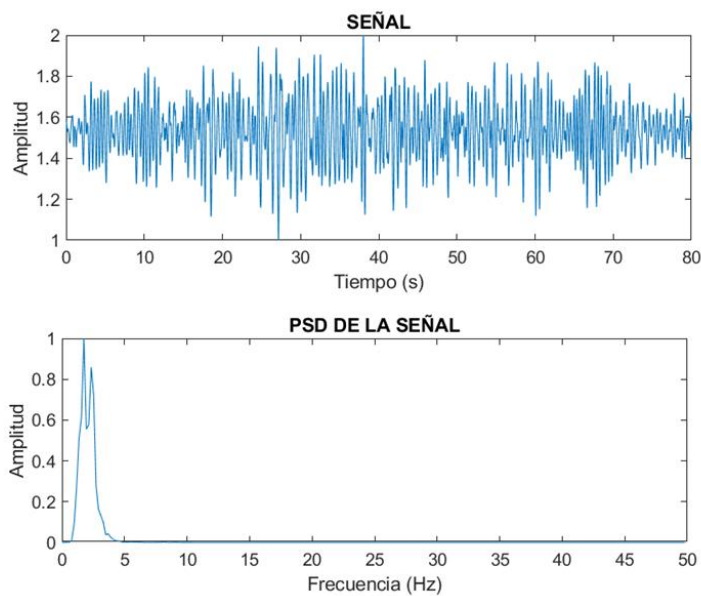
*Señal original microsísmica de tipo TR, su PSD y forma temporal*



*Nota.* Figura de una señal original tipo TR con fines comparativos.

#### Figura 48

*Señal sintética microsísmica de tipo TR, su PSD y forma temporal*



*Nota.* La Figura muestra una señal sintética tipo TR obtenida. Es un evento de gran amplitud generalmente producido por el movimiento del flujo de magma a través de las grietas en

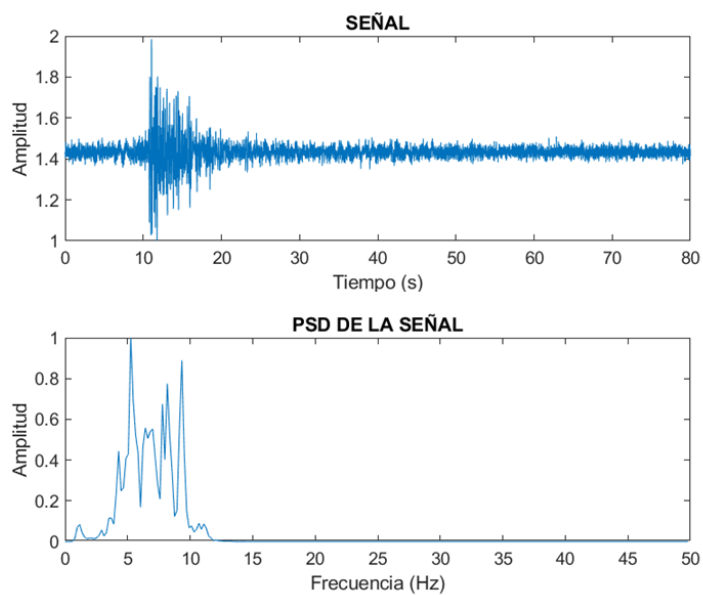


tiempos prolongados desde minutos a días. Se determinó que su mayor concentración de energía se encuentra alrededor de [2 – 3] Hz.

### **Evento VT**

#### **Figura 49**

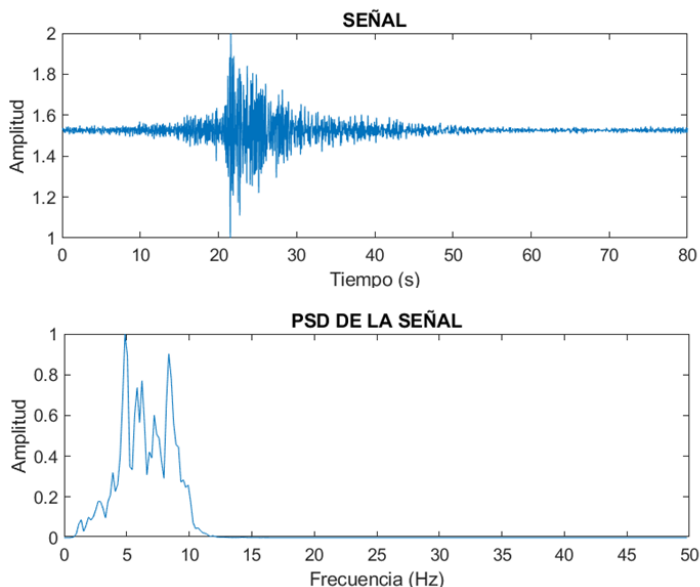
*Señal original microsísmica de tipo VT, su PSD y forma temporal*



*Nota.* Figura de una señal original tipo VT con fines comparativos.

#### **Figura 50**

*Señal sintética microsísmica de tipo VT, su PSD y forma temporal*



*Nota.* La Figura muestra una señal sintética tipo VT obtenida. Este evento está relacionado con las fallas frágiles de las rocas dentro del complejo volcánico. Presentó una duración mínima con amplio contenido espectral entre [0,1 - 10] Hz.

Se llevó a cabo dos pruebas de desempeño en concordancia con la Actividad 5, las cuales involucraron la evaluación de las señales sintéticas. Estas evaluaciones fueron realizadas por expertos del Instituto Geofísico de la Escuela Politécnica Nacional, así como a través de la utilización de clasificadores multiclase que fueron desarrollados simultáneamente en el grupo de investigación.

### **Evaluación de desempeño por expertos del Instituto Geofísico de la Escuela Politécnica Nacional**

El 19 de junio de 2023, se implementó un cuestionario, como se muestra en la Figura 51, que contenía un conjunto de 40 señales. Estas señales incluían tanto señales sintéticas como señales reales. El propósito de este proceso fue someter estas señales a la evaluación de tres expertos del IGEPN. El objetivo principal era determinar la calidad de las señales

sintéticas y si eran capaces de pasar como señales auténticas ante la evaluación de estos expertos. El formulario completo se encuentra en el Anexo 2.

### Figura 51

*Formulario enviado a los expertos del IGEPN*

The image shows a screenshot of a survey form. At the top, the title 'Señales Naturales y Sintéticas Microsísmicas' is displayed in a large, bold font. Below the title, the user's email address 'hfcaisaluisa@espe.edu.ec' is shown, along with a link to 'Cambiar de cuenta' and a cloud icon. A red asterisk indicates that the question is mandatory. The form is divided into sections, with the first section titled 'Parte 1'. The text in this section describes the survey's purpose: to evaluate experts on volcanic microseismic signals from the Llaima volcano in Chile, distinguishing between natural and synthetic signals. It specifies a 10-minute duration. The second section provides definitions for event types: LP (Long Period), VT (Volcano-Tectonic), TR (Tremor), and TC (Tectonic).

*Nota.* La Figura presenta el formulario enviado a los expertos del IGEPN. Este formulario incluyó un conjunto de 40 señales, compuesto tanto por señales sintéticas como señales reales de los eventos LP, VT, TR y TC. El objetivo de este formulario fue someter a prueba las señales sintéticas y evaluar si los expertos son capaces de distinguirlas de las señales reales.

Con el propósito de no influir en los resultados, se optó por no efectuar una distribución uniforme de las señales. La disposición de las señales sintéticas y reales se detallan en la Tabla 6.

**Tabla 6**

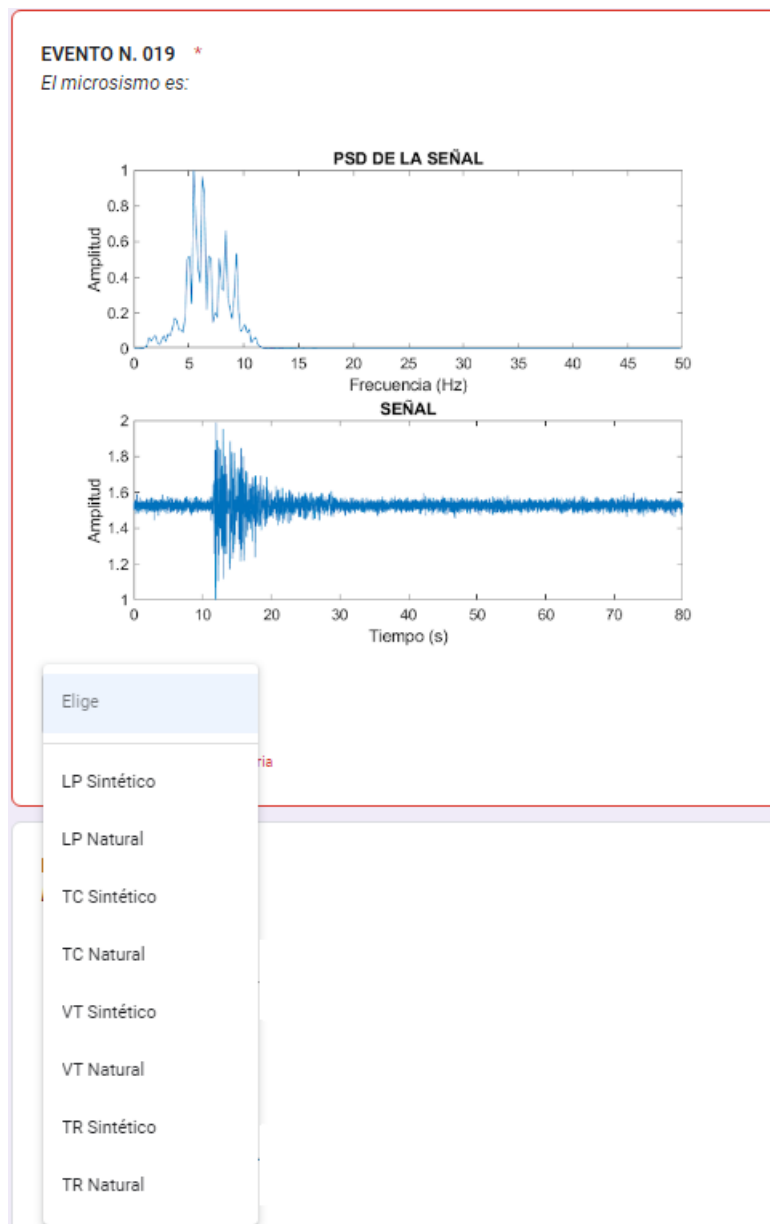
*Distribución de las señales sintéticas y reales en el formulario enviado al IGEPN*

<b>Tipo de evento</b>	<b>Cantidad sintéticos</b>	<b>Cantidad reales</b>
<b>Largo Período</b>	6	4
<b>Tectónicos</b>	4	6
<b>Tremor</b>	6	4
<b>Volcano Tectónicos</b>	7	3

*Nota.* Esta tabla muestra la distribución de las 40 señales que se incluyeron en el formulario con la finalidad de que los expertos del IGEPN evaluaran si las señales eran reales o sintéticas.

**Figura 52**

*Ejemplo de pregunta incluida en el formulario enviado, señal N.019*



*Nota.* La Figura muestra un ejemplo de una pregunta realizada en el formulario, se habilitó una pestaña desplegable que permitió elegir a los expertos del IGEPN seleccionar a que tipo de evento pertenecía la señal según su experiencia y criterio.

### **Tabla 7**

*Resultados obtenidos a partir del formulario respondido por los expertos del IGEPN en relación a las señales sintéticas y reales*

Número	Tipo de evento	Respuesta		
		Sintético (%)	Real (%)	Evento diferente (%)
1	VT Real	100.00	0,00	0.00
2	TC Real	0.00	33.33	66.70
3	LP Sintética	0.00	100.00	0.00
4	TR Real	100.00	0.00	0.00
5	LP Sintética	33.33	66.70	0.00
6	TC Real	0.00	0.00	100.00
7	TR Sintética	33.33	33.33	33.33
8	VT Sintética	0.00	100.00	0.00
9	TR Real	100.00	0.00	0.00
10	VT Sintética	66.70	33.33	0.00
11	LP Real	33.30	66.70	0.00
12	TC Sintética	0.00	0.00	100.00
13	TR Real	100.00	0.00	0.00
14	LP Sintética	0.00	100.00	0.00
15	VT Sintética	33.30	66.70	0.00
16	LP Real	66.70	33.30	0.00
17	TC Real	0.00	0.00	100.00
18	TR Sintética	0.00	100.00	0.00
19	VT Real	33.30	66.70	0.00
20	LP Sintética	0.00	100.00	0.00
21	TR Sintética	66.70	33.30	0.00
22	LP Real	33.30	66.70	0.00
23	VT Sintética	33.30	66.70	0.00
24	TC Sintética	0.00	0.00	100.00
25	TR Real	66.70	33.30	0.00
26	LP Sintética	0.00	100.00	0.00
27	TC Real	0.00	33.30	66.70
28	VT Real	66.70	33.30	0.00
29	TR Sintética	66.70	33.30	0.00
30	TC Real	0.00	33.30	66.70
31	LP Sintética	0.00	100.00	0.00
32	VT Sintética	33.30	66.70	0.00
33	TR Sintética	0.00	100.00	0.00
34	TC Sintética	0.00	0.00	100.00
35	VT Sintética	0.00	100.00	0.00
36	TC Real	0.00	0.00	100.00
37	VT Sintética	33.30	66.70	0.00
38	LP Real	100.00	0.00	0.00
39	TC Sintética	0.00	0.00	100.00
40	TR Sintética	66.70	33.30	0.00

*Nota.* La tabla presenta los resultados obtenidos del formulario entregado a los expertos del IGEPN.

La Tabla 7 compila los resultados del formulario, detalla las 40 señales y el porcentaje de su naturaleza (sintética o real) según la evaluación de los expertos. Las celdas se destacan en un tono verde claro para indicar si la señal es originalmente sintética o real, esto permite un análisis posterior. Además, se incorpora una tercera columna para identificar los eventos que no fueron clasificados correctamente dentro de su respectivo tipo.

**Tabla 8**

*Resultados de las señales sintéticas en términos del porcentaje de engaño*

Número de evento	Tipo Evento	Porcentaje de engaño (%)
3	LP	100.00
5	LP	66.70
7	TR	33.30
8	VT	100.00
10	VT	33.30
12	TC	0.00
14	LP	100.00
15	VT	66.70
18	TR	100.00
20	LP	100.00
21	TR	33.30
23	VT	66.70
24	TC	0.00
26	LP	100.00
29	TR	33.30
31	LP	100.00
32	VT	66.70
33	TR	100.00
34	TC	0.00
35	VT	100.00
37	VT	66.70
39	TC	0.00
40	TR	33.30
<b>Promedio ponderado</b>		<b>60.87%</b>

*Nota.* La Tabla 8 recoge exclusivamente las 26 señales sintéticas y el porcentaje en el que se logró engañar a los expertos del IGEPN haciéndolas pasar por señales reales, esto respalda los resultados de la Tabla 7.

La Tabla 8 exhibe los porcentajes de engaño correspondientes a cada señal sintética. En específico, se destacan las 4 señales del evento TC, las cuales muestran un porcentaje de engaño del 0.00%. Con el propósito de lograr una comprensión más precisa de los resultados obtenidos a partir de la generación de señales sintéticas, se lleva a cabo un análisis exhaustivo de cada tipo de evento. Estos resultados se presentan detalladamente en la Tabla 8 que se muestra a continuación. Además, se efectúa un promedio ponderado con el fin de comprender la tendencia general presente en los datos sintéticos. El resultado obtenido es de 60.87%, un valor que disminuye notablemente debido a la ausencia de reconocimiento de señales TC, lo cual ejerce un fuerte impacto en el promedio.

### **Tabla 9**

*Porcentaje total de engaño a los expertos del IGEPN según el tipo de evento microsísmico*

<b>Tipo de evento</b>	<b>Porcentaje total de engaño (%)</b>
LP	94.45
TC	0.00
TR	55.53
VT	71.44

*Nota.* La Tabla 9 resume los resultados completos de la encuesta llevada a cabo entre los expertos del IGEPN, muestra el porcentaje de éxito logrado para hacer pasar las señales sintéticas como reales en cada tipo de evento microsísmico.

Tras calcular el porcentaje total de engaño en las señales sintéticas microsísmicas, se procedió a analizar con qué tipos de eventos se producían estas confusiones. La información correspondiente se detalla en la siguiente Tabla 10.



**Tabla 10**

*Tipos de Eventos que Generan Confusión entre los Expertos del IGEPN*

Número de evento	Tipo Evento sintético	Evento con el que se confunde
3	LP	-
5	LP	LP sintético (33.30%)
7	TR	TR sintético (33.30%), LP real (33.30%)
8	VT	-
10	VT	VT sintético (66.70%)
12	TC	TR real (100%)
14	LP	-
15	VT	VT sintético (33.30%)
18	TR	-
20	LP	-
21	TR	TR sintético (66.70%)
23	VT	VT sintético (33.30%)
24	TC	TR real (100%)
26	LP	-
29	TR	TR sintético (66.70%)
31	LP	-
32	VT	VT sintético (33.30%)
33	TR	-
34	TC	TR real (33.30%), TR sintético (33.30%), LP sintético (33.30%)
35	VT	-
37	VT	VT sintético (33.30%)
39	TC	TR natural (33.30%), TR sintético (66.70%)
40	TR	TR sintético (66.70%)

*Nota.* La Tabla presenta el porcentaje de confusión de los diferentes tipos de eventos.

En la Tabla 10, se presentan los resultados que llevaron a la confusión entre los expertos del IGEPN, se empleó una representación con colores. Se resalta en color amarillo aquellos eventos que no fueron confundidos con ningún otro tipo, demuestran que los expertos identificaron adecuadamente que dicha señal era sintética. Este reconocimiento respalda su competencia en la materia.

**Tabla 11**

*Porcentaje total de confusión en los eventos*

Tipo de señal	Evento que causa confusión			
	TR real (%)	TR sintético (%)	LP sintético (%)	LP real (%)
TC	66.65	25.00	8.33	0.00
TR	0.00	0.00	0.00	5.55

*Nota.* En la Tabla 11 se presenta una ponderación total de los dos únicos eventos que causaron confusión a los expertos.

La confirmación de este fenómeno se observa en la Tabla 11. En una tonalidad azul, se muestran principalmente las cuatro señales sintéticas de tipo TC, las cuales fueron confundidas en un 66.65% con señales reales de tipo TR, en un 25% con señales sintéticas de tipo TR y en un 8.33% con señales sintéticas de tipo LP.

Existe un caso particular, destacado en color verde, se registra un suceso único; una señal TR fue confundida en un 5.55% con una señal real de tipo LP.

### **Resultados del clasificador**

El clasificador de señales es un trabajo de integración curricular que se desarrolló en paralelo dentro del grupo de investigación, el proceso se realizó al utilizar una extracción de 84 características de las señales. El clasificador fue implementado en el entorno de MATLAB®, al aprovechar técnicas de Machine Learning tradicional, para abordar la clasificación de eventos microsísmicos. El estudio involucró la evaluación de un conjunto de 10000 señales sintéticas, distribuidas uniformemente, con 2500 señales correspondientes a cada tipo de evento microsísmico.

La estrategia de clasificación se basó en la implementación de tres diferentes algoritmos de aprendizaje supervisado. En primer lugar, se empleó un clasificador basado en el algoritmo de árbol de decisión (DT, del inglés *Decision Tree*), el cual demostró una capacidad de

clasificación con un 88.00% de exactitud, esto evidencia su eficacia en el proceso. El parámetro de tasa de error equilibrada (BER, del inglés *Balanced Error Rate*), fue calculado y arrojó un valor de 0.08, esto indica una aceptable tasa de clasificación errónea. Los resultados se observan en la Figura 53 y Tabla 12.

Además, se aplicó el algoritmo de clasificación *k-Nearest Neighbors* (*k-NN*), que logró resultados sumamente prometedores en la tarea de clasificación de las señales sintéticas. Con una exactitud sobresaliente del 99.90% y un BER de 0, este enfoque demostró su capacidad para diferenciar eficazmente entre los diferentes tipos de eventos microsísmicos. Los resultados del clasificador con *k-NN* se visualizan en la Figura 54 y Tabla 13.

Finalmente, se incorporó un tercer clasificador basado en la Máquina de Vectores de Soporte (SVM, del inglés *Support Vector Machine*), otro algoritmo de aprendizaje supervisado ampliamente utilizado. Este enfoque alcanzó una exactitud de clasificación del 99.90% y un BER de 0, lo que resalta su efectividad en la tarea de clasificación de eventos microsísmicos. Los resultados del clasificador con SVM se pueden observar en la Figura 55 y Tabla 14.

Después de haber implementado los tres clasificadores mencionados, se procedió a integrarlos en un sistema de votación con el propósito de determinar cuál de ellos sobresale como el enfoque más eficaz. Tras llevar a cabo diversas pruebas, el grupo de investigación llegó a una conclusión: el clasificador SVM demostró ser el más sobresaliente. Los resultados del sistema de votación se pueden observar en la Figura 56 y Tabla 15.

Las Tablas a continuación presentan las diversas métricas de rendimiento: Exactitud (A, del inglés *Accuracy*), Precisión (P, del inglés *Precision*), Sensibilidad (R, del inglés *Recall*), Especificidad (S, del inglés *Specificity*) y BER. Estas métricas son utilizadas para la evaluación de la eficacia y el comportamiento del clasificador, y se definen como:

$$A (\%) = \frac{N_C}{N_T} \times 100, \quad ( 22 )$$

$$P (\%) = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100, \quad ( 23 )$$

$$R (\%) = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100, \quad ( 24 )$$

$$S (\%) = \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100, \quad ( 25 )$$

Dónde  $N_C$  corresponde al número de eventos correctamente clasificados,  $N_T$  representa al total de eventos utilizados para entrenar el clasificador,  $N_{TP}$  corresponde a la cantidad de verdaderos positivos,  $N_{FN}$  a la cantidad de falsos negativos,  $N_{TN}$  corresponde a la cantidad de verdaderos negativos y  $N_{FP}$  corresponde a la cantidad de falsos positivos.

### **Figura 53**

*Resultados de la clasificación de las 10000 señales sintéticas al utilizar DT*

Output Class	LP	TC	TR	VT	
LP	1556 15.6%	51 0.5%	0 0.0%	0 0.0%	96.8% 3.2%
TC	924 9.2%	2449 24.5%	202 2.0%	0 0.0%	68.5% 31.5%
TR	17 0.2%	0 0.0%	2298 23.0%	0 0.0%	99.3% 0.7%
VT	3 0.0%	0 0.0%	0 0.0%	2500 25.0%	99.9% 0.1%
	62.2% 37.8%	98.0% 2.0%	91.9% 8.1%	100% 0.0%	88.0% 12.0%
	LP	TC	TR	VT	

*Nota.* La Figura muestra la matriz de confusión que refleja el resultado de la clasificación de las 10000 señales sintéticas, esta evidencia una exactitud del 88.00%.

La matriz de confusión presentada en la Figura 53 exhibe los resultados del modelo de clasificación de los eventos: LP, TC, TR y VT. En la fila correspondiente al evento TC, se observa que ha clasificado correctamente 2449 señales; sin embargo, en su mayoría se ha confundido con 924 señales del evento LP, en una cantidad menor con 202 señales del evento TR y no ha presentado ninguna confusión con el evento VT. En consecuencia, se puede inferir que el modelo principalmente se confunde con la categoría LP.

### Tabla 12

*Métricas de evaluación del rendimiento del clasificador DT*

Eventos	A (%)	P (%)	S (%)	R (%)	BER
LP	90.00	97.00	99.00	62.00	0,19
TC	88.00	69.00	85.00	98.00	0,09
TR	98.00	99.00	100.00	92.00	0,04
VT	100.00	100.00	100.00	100.00	0.00
General	88.00	91.00	96.00	88.00	0,08

*Nota.* La Tabla muestra las métricas de rendimiento; Exactitud, Precisión, Especificidad, Sensibilidad y BER que permiten evaluar el desempeño y comportamiento del clasificador DT en las diferentes clases.

#### Figura 54

*Resultados de la clasificación de las 10000 señales sintéticas al utilizar k-NN*

Output Class	LP	TC	TR	VT	
LP	2499 25.0%	4 0.0%	1 0.0%	3 0.0%	99.7% 0.3%
TC	0 0.0%	2494 24.9%	0 0.0%	0 0.0%	100% 0.0%
TR	0 0.0%	2 0.0%	2499 25.0%	0 0.0%	99.9% 0.1%
VT	1 0.0%	0 0.0%	0 0.0%	2497 25.0%	100.0% 0.0%
	100.0% 0.0%	99.8% 0.2%	100.0% 0.0%	99.9% 0.1%	99.9% 0.1%
	LP	TC	TR	VT	
	Target Class				

*Nota.* La Figura muestra la matriz de confusión que refleja el resultado de la clasificación de las 10000 señales sintéticas, esta evidencia una exactitud del 99.9%.

**Tabla 13**

*Métricas de evaluación del rendimiento del clasificador k-NN*

Eventos	A (%)	P (%)	S (%)	R (%)	BER
LP	100.00	100.00	100.00	100.00	0.00
TC	100.00	100.00	100.00	100.00	0.00
TR	100.00	100.00	100.00	100.00	0.00
VT	100.00	100.00	100.00	100.00	0.00
General	100.00	100.00	100.00	100.00	0.00

*Nota.* La Tabla muestra las métricas de rendimiento; Exactitud, Precisión, Especificidad, Sensibilidad y BER que permiten evaluar el desempeño y comportamiento del clasificador *k*-NN en las diferentes clases.

**Figura 55**

*Resultados de la clasificación de las 10000 señales sintéticas al utilizar SVM*

LP	2500 25.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
TC	0 0.0%	2487 24.9%	0 0.0%	0 0.0%	100% 0.0%
TR	0 0.0%	0 0.0%	2500 25.0%	0 0.0%	100% 0.0%
VT	0 0.0%	13 0.1%	0 0.0%	2500 25.0%	99.5% 0.5%
	100% 0.0%	99.5% 0.5%	100% 0.0%	100% 0.0%	99.9% 0.1%
	LP	TC	TR	VT	
	Target Class				

*Nota.* La Figura muestra la matriz de confusión que refleja el resultado de la mejor clasificación de las 10000 señales sintéticas al utilizar SVM, el cual evidencia una exactitud del 99.90%.

**Tabla 14**

*Métricas de evaluación del rendimiento del clasificador SVM*

Eventos	A (%)	P (%)	S (%)	R (%)	BER
LP	100.00	100.00	100.00	100.00	0.00
TC	100.00	100.00	100.00	99.00	0.00
TR	100.00	100.00	100.00	100.00	0.00
VT	100.00	99.00	100.00	100.00	0.00
General	100.00	100.00	100.00	100.00	0.00

*Nota.* La Tabla muestra las métricas de rendimiento; Exactitud, Precisión, Especificidad, Sensibilidad y BER que permiten evaluar el desempeño y comportamiento del clasificador SVM en las diferentes clases.

**Figura 56**

*Resultados del sistema de votación de los tres diferentes clasificadores*

Confusion Matrix					
Output Class	LP	TC	TR	VT	
LP	2497 25.0%	4 0.0%	1 0.0%	0 0.0%	99.8% 0.2%
TC	0 0.0%	2494 24.9%	0 0.0%	0 0.0%	100% 0.0%
TR	0 0.0%	2 0.0%	2499 25.0%	0 0.0%	99.9% 0.1%
VT	3 0.0%	0 0.0%	0 0.0%	2500 25.0%	99.9% 0.1%
	99.9% 0.1%	99.8% 0.2%	100.0% 0.0%	100% 0.0%	99.9% 0.1%
	LP	TC	TR	VT	
	Target Class				



*Nota.* La Figura exhibe la matriz de confusión del sistema de votación, esto evidencia una exactitud destacada del 99.90%. Esto refuerza la conclusión de que el clasificador SVM sobresale como la opción más efectiva.

### **Tabla 15**

*Métricas de evaluación del sistema de votación*

<b>Eventos</b>	<b>A (%)</b>	<b>P (%)</b>	<b>S (%)</b>	<b>R (%)</b>	<b>BER</b>
LP	100.00	100.00	100.00	100.00	0.00
TC	100.00	100.00	100.00	100.00	0.00
TR	100.00	100.00	100.00	100.00	0.00
VT	100.00	100.00	100.00	100.00	0.00
General	100.00	100.00	100.00	100.00	0.00

*Nota.* La Tabla muestra las métricas de rendimiento; Exactitud, Precisión, Especificidad, Sensibilidad y BER del sistema de votación y destaca valores de 100%.

## Capítulo V

### Conclusiones y Trabajos Futuros

#### Conclusiones

Mediante la utilización del método de remuestreo Bootstrap en conjunción con la Transformada de Fourier, detallado en el capítulo 4, se ha logrado la generación de señales sintéticas que representan eventos microsísmicos del volcán Llaima, estas señales sintéticas mantienen sus características distintivas tanto en el dominio temporal como en el de frecuencia, en concordancia con las señales originales.

Se optó por seleccionar fases reales que se encontraban dentro de un margen cercano a una fase ideal. Esta fase ideal, al mantener una naturaleza lineal, exhibe una menor presencia de ruido, lo que resulta en una preservación más efectiva de la información de la envolvente. Como consecuencia, se logra capturar de manera más precisa la forma distintiva de la señal en función del tiempo al momento de generar las señales sintéticas.

El desfase que emerge en la señal sintética en el plano temporal es resultado del proceso de transformación de la IFFT. Este fenómeno deriva de la variabilidad inherente a la selección aleatoria de fases ideales. Para rectificar esta situación, se lleva a cabo un proceso de corrección mediante el cálculo de la media de las fases ubicadas en un rango específico y se revierte el desplazamiento originado. La señal recupera su forma de manera satisfactoria debido a que el cálculo de la media toma en consideración los desfases individuales de cada señal.

El análisis de los resultados obtenidos de la encuesta aplicada a los expertos del IGEPN revela hallazgos satisfactorios. En concreto, los resultados indican que fue posible engañar a los expertos en un 94.45% de los casos con las señales sintéticas LP, en un 55.53% de los casos con las señales TC y en un 71.44% de los casos con las señales sintéticas VT. Cabe

destacar que las señales TC no presentaron ningún porcentaje de engaño (0.00%). Esta ausencia de engaño se debe a que los expertos catalogaron estas señales como "otras señales". Esta confusión podría atribuirse posiblemente a que este tipo de evento no está relacionado con la actividad volcánica y exhibe diversas frecuencias predominantes en todo el espectro, lo que facilita su confusión con otros tipos de eventos. Los eventos TC tienden a confundirse principalmente con eventos TR, registrando un 66.65% de confusión con eventos TR reales y un 25% con eventos TR sintéticos. Además, se observa una minoritaria confusión del 8.23% con eventos LP sintéticos.

Las evaluaciones sistemáticas realizadas revelaron que el clasificador SVM no solo mantenía su alta tasa de exactitud del 99.90%, sino que también presentaba un BER cercano a 0, esto destaca su habilidad para minimizar los errores de clasificación de manera excepcional. En el proceso de votación, este enfoque se destacó significativamente por encima de los otros dos clasificadores, lo que afirma su superioridad en la tarea de clasificación de eventos microsísmicos de las señales sintéticas

### **Trabajos Futuros**

Nuestro grupo de investigación está interesado en realizar una evaluación de la precisión de la técnica Bootstrap en la generación de las señales TC con el objetivo de mejorar su respuesta en el dominio temporal y así procurar engañar a los expertos del IGEPN haciéndolas parecer señales genuinas.

Se desea explorar alternativas de remuestreo como la especificación de histograma, el método de Monte Carlo y el método de remuestreo de jackknife para la creación de señales sintéticas. Estas técnicas generan nuevas muestras a partir de datos iniciales, y estima sus propiedades estadísticas al brindar un enfoque distinto en la generación de señales.

El grupo de investigación está interesado en utilizar enfoques innovadores de estimación espectral al crear señales sintéticas, centrándose en la correlación.

Se propone en nuestro grupo de investigación utilizar la PSD y su transformada inversa para generar señales sintéticas que capturen las propiedades de las señales reales en términos de su distribución de energía.

## Referencias

- Aguinaga Navarreta, D. S. (2018). *ANÁLISIS Y CARACTERIZACIÓN DE MICROSISMOS PRODUCIDOS EN EL VOLCÁN COTOPAXI MEDIANTE EL USO DE TÉCNICAS ESPECTRALES*.
- Ali, M., Dewan, A., Sahu, A. K., & Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* 2023, Vol. 12, Page 91, 12(5), 91. <https://doi.org/10.3390/COMPUTERS12050091>
- Belío Miranda, J. (2020). *Métodos Bootstrap y sus aplicaciones* [Universidad de Zaragoza]. <https://zaguán.unizar.es/record/98153/files/TAZ-TFG-2020-1954.pdf>
- Cabrera, A., & Rosero, E. (2019). *CARACTERIZACIÓN DE SEÑALES VOLCÁNICAS PARA LA CLASIFICACIÓN SUPERVISADA DE EVENTOS SÍSMICOS*.
- Cabrera Riobamba, A. P., & Rosero Narvaez, E. G. (2019). *CARACTERIZACIÓN DE SEÑALES VOLCÁNICAS PARA LA CLASIFICACIÓN SUPERVISADA DE EVENTOS SÍSMICOS* [Universidad de Nariño]. [https://www.diegopeluffo.com/Theses/2019\\_DegreeThesis\\_Volcano\\_classification.pdf](https://www.diegopeluffo.com/Theses/2019_DegreeThesis_Volcano_classification.pdf)
- Cadena, O. (2021). *Modelos de fuente de sismicidad LP para la actividad del volcán Galeras (Colombia)*. Universidad Nacional de Colombia.
- Camacho, K. (2022). *Generación de señales sintéticas de eventos sismo-volcánicos usando el método de Reconstrucción de señales a través de su Fase*.
- Canário, J. (2021). Conjunto de datos del volcán Llaima: comparación en profundidad de arquitecturas de redes neuronales artificiales profundas en la clasificación de eventos sísmicos. *Data in Brief*. [https://www.data-in-brief.com/article/S2352-3409\(20\)30521-7/fulltext](https://www.data-in-brief.com/article/S2352-3409(20)30521-7/fulltext)

- Canário, J. P., Mello, R., Curilem, M., Huenupan, F., & Rios, R. (2020). In-depth comparison of deep artificial neural network architectures on seismic events classification. *Journal of Volcanology and Geothermal Research*, 401.
- Carmona, A. (2017). *Volcán Llaima | Pontificia Universidad Católica de Chile*.  
<https://villarrica.uc.cl/volc%C3%A1n-llaima>
- Cortés Moreno, G. (2015). *Reconocimiento de señales sismo-volcánicas mediante canales específicos basados en modelos ocultos de Markov*.  
<https://digibug.ugr.es/handle/10481/42050>
- Cruz, R. (2022). *Transformada Inversa de Fourier - Convolución*.
- Curilem, M., Canário, J., & Franco, L. (2018). *Using CNN to classify spectrograms of seismic events from Llaima volcano (Chile)*. International Joint Conference on Neural Networks (IJCNN).
- Curilem, M., Canário, J. P., Franco, L., & Rios, R. A. (2018). *Using CNN to classify spectrograms of seismic events from Llaima volcano (Chile)*.
- Duoandikoetxea, J. (2003). LECCIONES SOBRE LAS SERIES Y TRANSFORMADAS DE FOURIER. UNAN.
- Franco, L. (2019). Comportamiento eruptivo del volcán Llaima (2007-2010) e incidencia del terremoto del Maule MW 8.8 en la actividad volcánica y tectónica local. *Universidad de Concepción - Chile*.  
<http://repositorio.udec.cl/jspui/bitstream/11594/1139/1/Tesis%20comportamiento%20eruptivo%20del%20volcan%20Llaima.pdf>

- Gil, C. (2018). *MÉTODOS DE REMUESTREO Y VALIDACIÓN DE MODELOS: VALIDACIÓN CRUZADA Y BOOTSTRAP*. [https://rstudio-pubs-static.s3.amazonaws.com/390382\\_949f84c9e0ef47cf8feb25a8060acf3e.html](https://rstudio-pubs-static.s3.amazonaws.com/390382_949f84c9e0ef47cf8feb25a8060acf3e.html)
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.  
<https://doi.org/10.1109/TKDE.2008.239>
- Hesterberg, T. C. (2015). What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum. *American Statistician*, 69(4), 371–386.  
<https://doi.org/10.1080/00031305.2015.1089789>
- Kennett, B. L. N. (2001). *The Seismic Wavefield: Volume 1, Introduction and Theoretical Development*. Cambridge University Press.  
<https://books.google.co.vi/books?id=88x8jgEACAAJ&printsec=copyright#v=onepage&q=Hz&f=false>
- Ledesma, R. (2008). *Introducción al Bootstrap*. [En línea]. Available:  
<https://www.tqmp.org/RegularArticles/vol04-2/p051/p051.pdf>.
- Mendoza, F. (2014). *Cordillera de Los Andes, una oportunidad para la integración y desarrollo de América del Sur*. [www.fao.org/publications](http://www.fao.org/publications)
- Minakami, T. (1974). *Seismology of Volcanes in Japan: Vol. vol.6*. Devp. in Solid Earth Geophys.
- Navarrete, P. (2017). Volcán Llaima: Antecedentes, Amenaza volcánica y evaluación de la gestión del riesgo. *Universidad de Chile, Santiago - Chile*.
- Pérez, B. (2022). *Generación de señales sintéticas de eventos sismo-volcánicos del volcán Cotopaxi mediante el método de Bootstrap*.

- Rice University. (1999). *8.1 Un intervalo de confianza para una desviación típica de la población, con un tamaño de muestra conocido o grande - Introducción a la estadística empresarial | OpenStax*. <https://openstax.org/books/introducci%C3%B3n-estad%C3%ADstica-empresarial/pages/8-1-un-intervalo-de-confianza-para-una-desviacion-tipica-de-la-poblacion-con-un-tamano-de-muestra-conocido-o-grande>
- Rocco S., C. M., & Zio, E. (2005). Bootstrap-based techniques for computing confidence intervals in Monte Carlo system reliability evaluation. *Proceedings - Annual Reliability and Maintainability Symposium*, 303–307. <https://doi.org/10.1109/RAMS.2005.1408379>
- Schuster, E., & Millar, G. (2020, agosto 18). *Volcán Llaima - Andeshandbook*. <https://www.andeshandbook.org/montanismo/cerro/146/Llaima/descripcion>
- Stoica, P., & Moses, R. (2005). *SPECTRAL ANALYSIS OF SIGNALS*.
- UNA. (2023). *Acerca de sismos asociados con actividad volcánica*. Universidad Nacional Costa Rica. <http://www.ovsicori.una.ac.cr/index.php/faqs/vulcanologia/acerca-de-sismos-asociados-con-actividad-volcanica>
- Zapata, J. (2022). *GENERACIÓN DE SEÑALES VOLCÁNICAS ARTIFICIALES DE TIPO LP (LONG-PERIOD) Y VT (VOLCANO-TECTONIC) A PARTIR DE UNA BASE DE DATOS DEL VOLCÁN COTOPAXI USANDO LA TÉCNICA DE BOOTSTRAPPING*. Escuela Politécnica Nacional.



## Apéndices