

DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES CARRERA DE TELECOMUNICACIONES

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

AUTOR: Henry Fabricio Caisaluisa Hurtado

DIRECTOR DEL PROYECTO: Ing. Lara Cueva, Román Alcides, PhD.

07 de septiembre de 2023

VERSIÓN: 1.1

Agenda

1.- Introducción

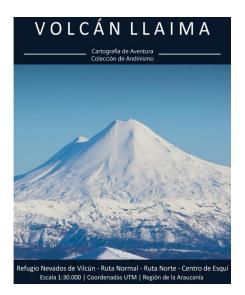
2.- Desarrollo

3.- Pruebas y Resultados

4.- Conclusiones y Trabajos futuros



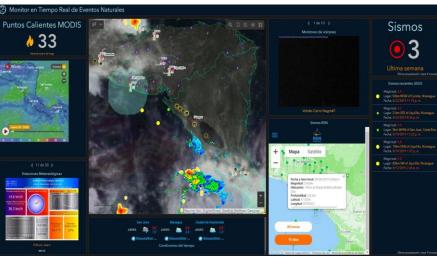
1. Introducción Motivación e Importancia



Situado en una región altamente propensa a actividad sísmica

Resulta crucial establecer un Sistema de Monitorización efectiva







Motivación primordial

- Preservación de vidas
 - Anticipar riesgos





1. Introducción Antecedentes

Generación de **señales sintéticas** de eventos microsísmicos usando el método de reconstrucción de señales a través de su fase.



APLICACIÓN DE GENERACIÓN Y RECONOCIMIENTO DE SEÑALES MICRO-SÍSMICAS

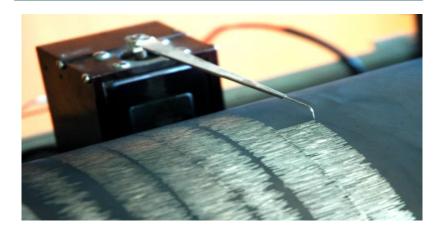


Autor: Brenda Elizabeth Pérez Moreno

Director: Dr. Román Lara

Colaboradores: Ing. Julio Larco

Utilización de técnicas de **Bootstrap** para establecer intervalos de confianza en cada tipo de evento.



Generación de **señales sintéticas** de eventos sismo volcánicos del volcán Cotopaxi mediante el método de Bootstrap.



1. Introducción

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 Volcanológico 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 Boostraping.
- Pruebas y evaluación de desempeño.



2. Desarrollo Materiales





Procesador Intel® Core™ i7 8th Generación 2.00 GHz

Memoria RAM 12 GB DDR4

Gráficos Integrados Intel UHD Graphics 620

Almacenamiento Disco SSD M.2, 500 GB

Disco HDD, 500 GB

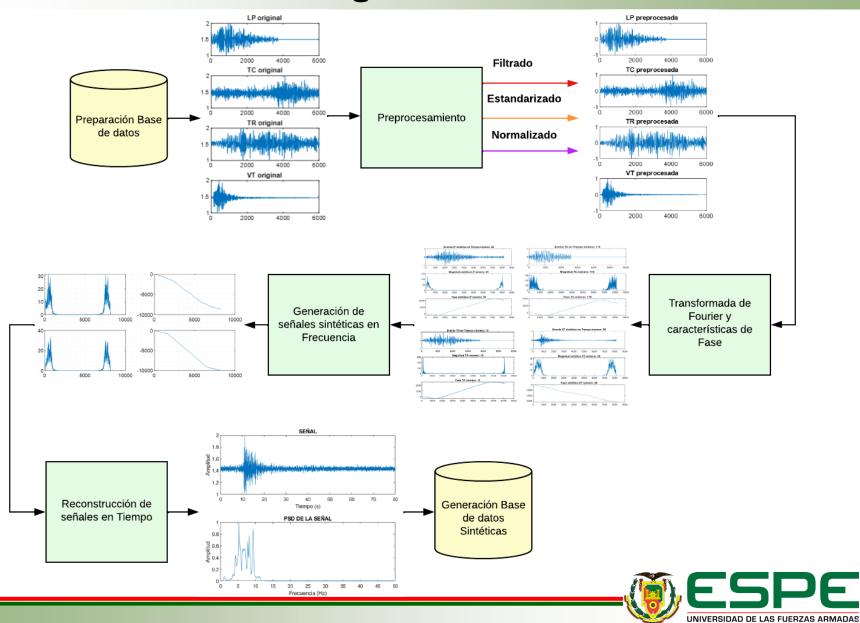
Sistema Operativo Windows 10 Home



Versión: R2020a



2. Desarrollo - Metodología



INNOVACIÓN PARA LA EXCELENCIA

Preparación de datos

Base de datos:

La base de datos provista por Observatorio Vulcanológico de los Andes Sur (OVDAS).

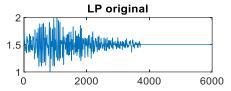
Presenta las siguientes características:

- Número de señales: M = 3592
- Señales 60 [s], filtradas a 100 Hz
- Tipo de eventos: 4

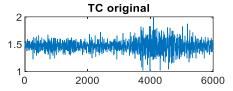
Dónde:

Tipo de Evento	Número de señales
Largo Periodo (LP)	1310
Tectónicos (TC)	1488
Tremor (TR)	490
Volcano Tectónicos (VT)	304

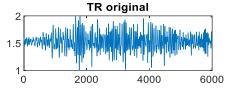
• **LP:** Asociados a la resonancia de flujos volcánicos que se desplazan a través de fracturas y grietas.



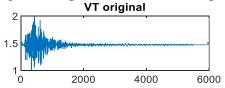
• TC: No se encuentra relacionado con la actividad volcánica, se produce por la dinámica de las fallas geológicas.



 TR: Origen a procesos como la liberación de gases, resonancia de cavidades y cambios en la presión y temperatura en sistemas hidrotermales.

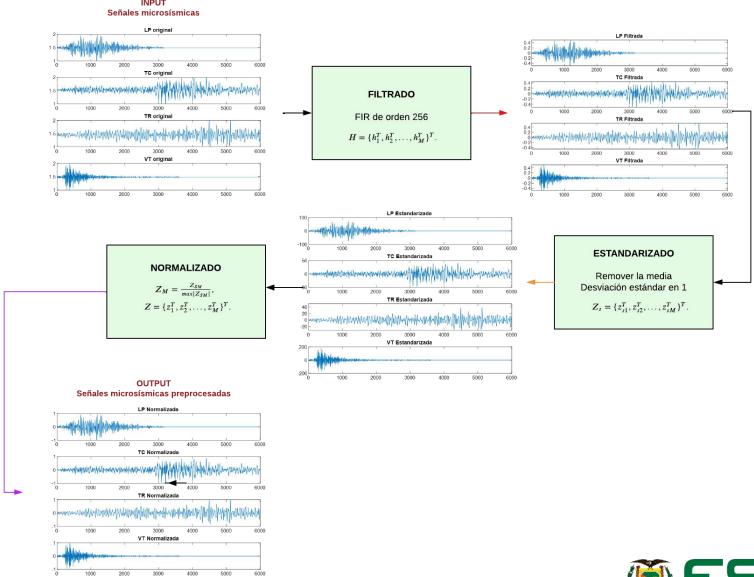


VT: Resultado de las fracturas de las paredes volcánicas debido a la presión generada por el ascenso del magma.



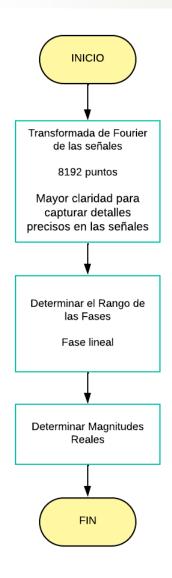


Preprocesamiento

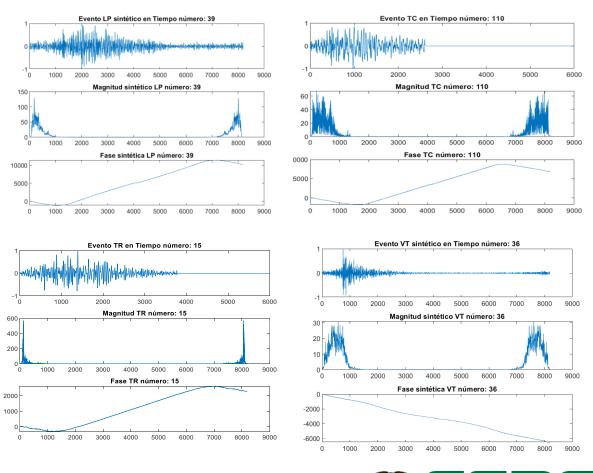




Transformada de Fourier y su Inversa



Descomposición en Magnitud y Fase obtenida al realizar la Transformada de Fourier.

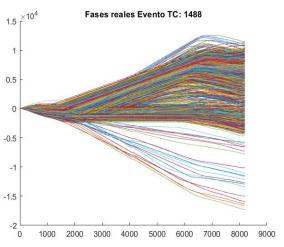


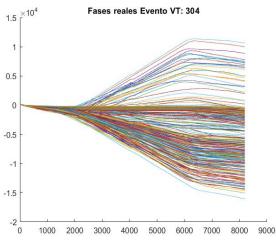


Determinar el rango de las fases

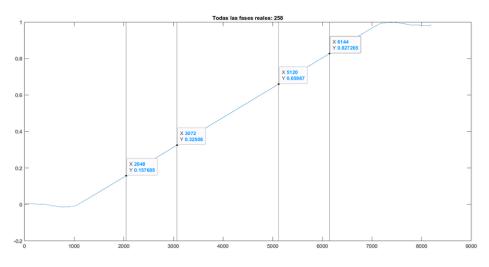
• Fase contiene la forma de la envolvente.

Concentración de todas las fases para cada tipo de evento





Modelado de la "Fase Lineal"

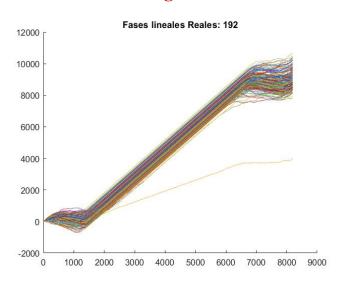


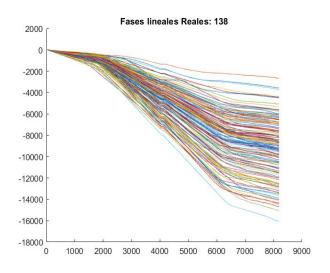
Posición	Posición Amplitud LP Am		Amplitud TR	Amplitud VT
2048 -0.13		0.13	0.16	-0.14
3072	3072 -0.25		0.32 0.33	
5120	-0.75	0.68	0.66	-0.68
6144	-0.88	0.86	0.83	-0.86



Determinar el rango de las fases

Rango de fases alrededor del 20% de intervalo de confianza





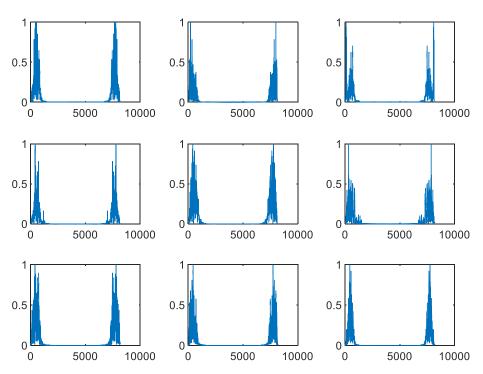
Cantidad de "Fases ideales"

Tipo de evento	Cantidad de Fases ideales
LP	157
тс	192
TR	106
VT	138



Normalización

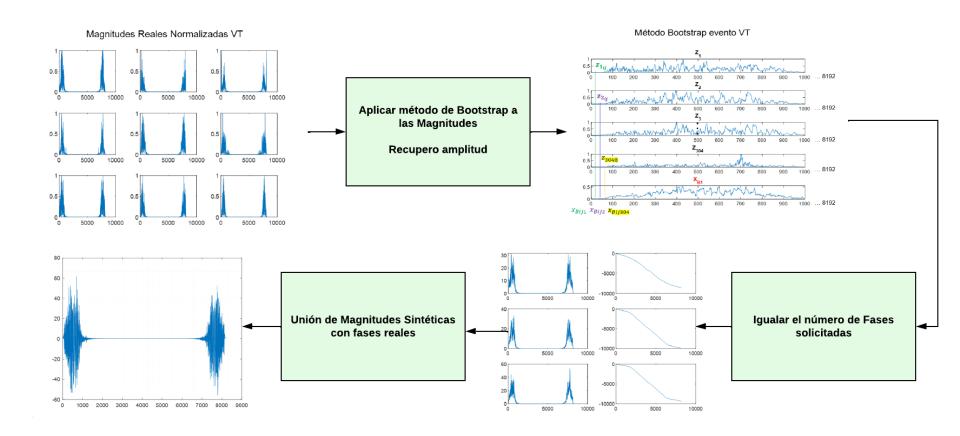
Magnitudes Reales Normalizadas VT



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.

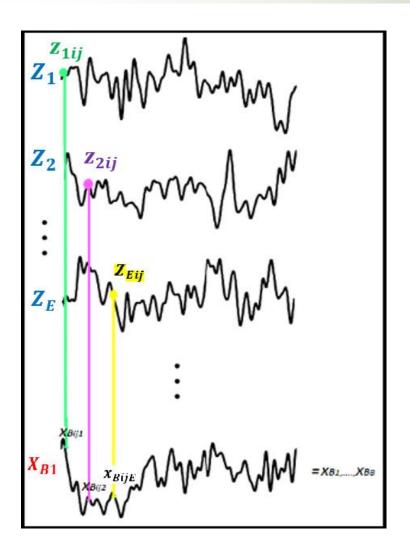


Generación de señales sintéticas





Bootstrap



Se parte de una magnitud perteneciente a un evento Z_1 y se toma un punto z_{1ij} para conformar el punto x_{Bij1} de la señal X_{B1} correspondiente a la magnitud sintética, se toma otro punto z_{Eij} que pertenece al evento Z_E , dónde E, es la cantidad de señales originales de cada tipo de evento microsísmico.

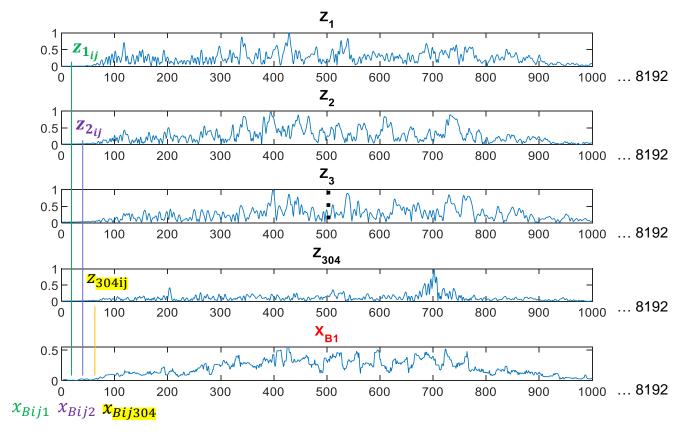
El cual forma otro punto x_{BijE} de señal sintética X_{B1} , este proceso se repite de forma aleatoria desde las magnitudes Z_1 hasta Z_E , el proceso toma puntos en orden hasta completar la primera réplica o magnitud sintética.

$$X_B = \left\{ x_{B1}^T, x_{B2}^T, \dots x_{BB}^T \right\}^T.$$



Bootstrap

Método Bootstrap evento VT



$$X_{B1} = \{x_{Bij1}, x_{Bij2}, \dots x_{BijN}\}.$$

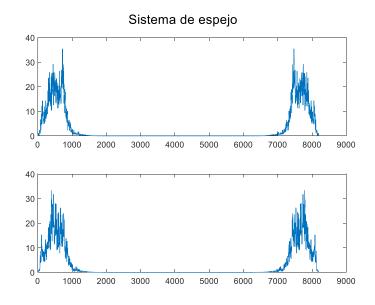


Bootstrap

• Parámetros mínimos de Bootstrap

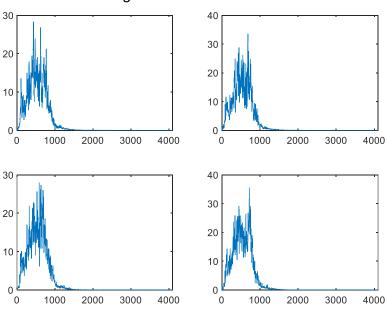
Variable		Valor
Cantidad de señales sintéticas	(B)	10000
Tamaño de las muestras sintéticas	(N)	8192

• Sistema de espejo a los 4096 puntos



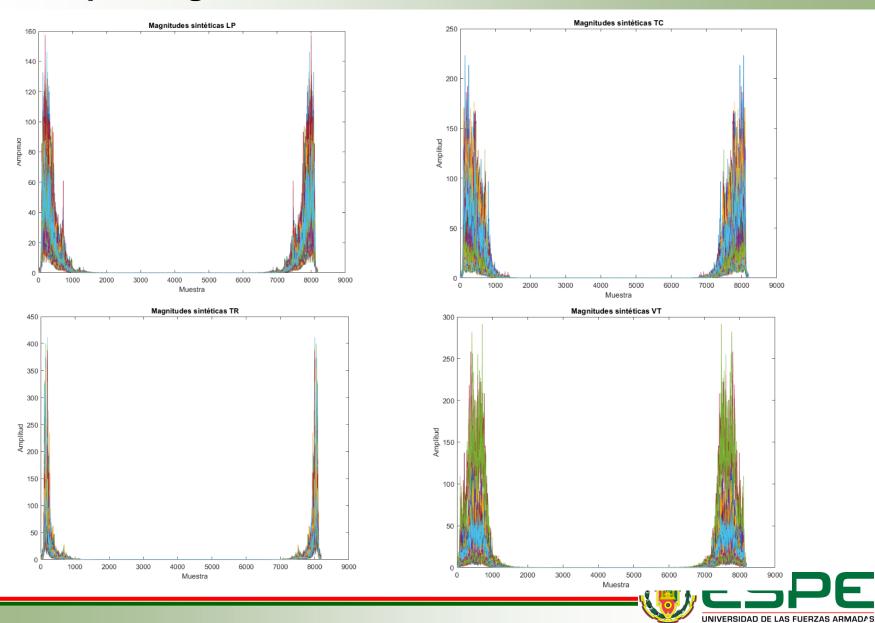
- Se toma los primeros 4096 puntos
- Amplifico la Magnitud sintética

Magnitudes Sintéticas VT





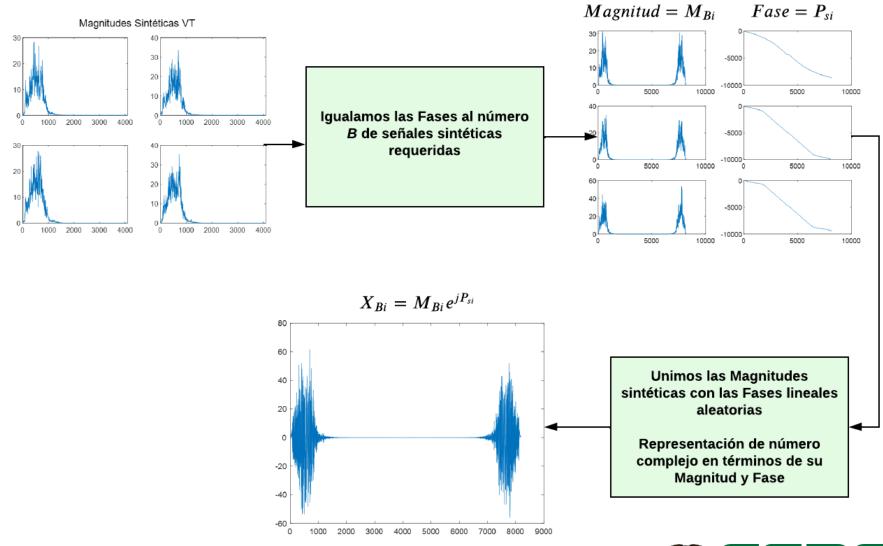
Bootstrap – Magnitudes Sintéticas



9000

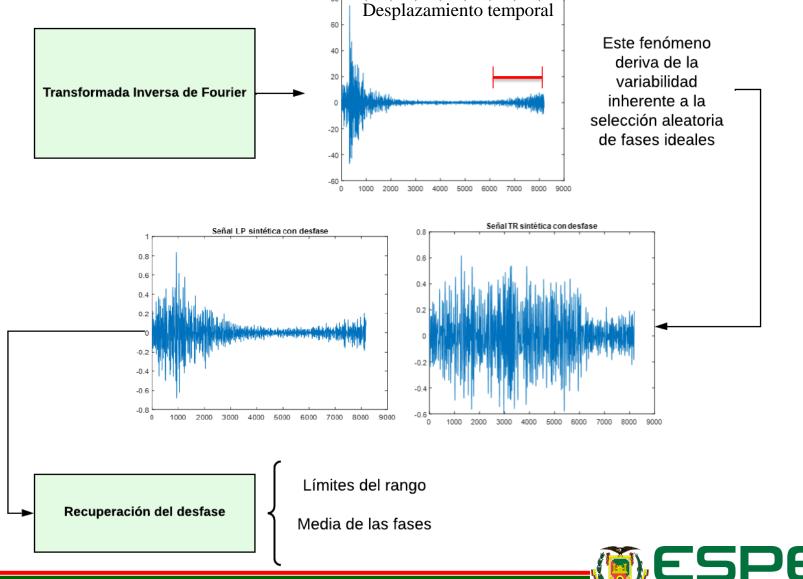
INNOVACIÓN PARA LA EXCELENCIA

Unión de magnitudes sintéticas con fases lineales



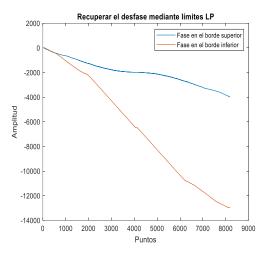


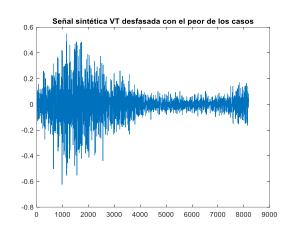
Reconstrucción de la señal sintética



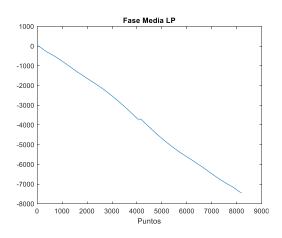
3. Pruebas y resultados – Recuperación desfase

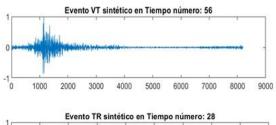
• Límites del rango (Peor de los casos)

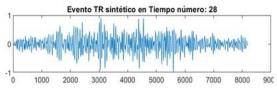


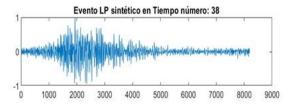


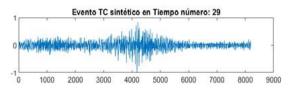
Media de las fases





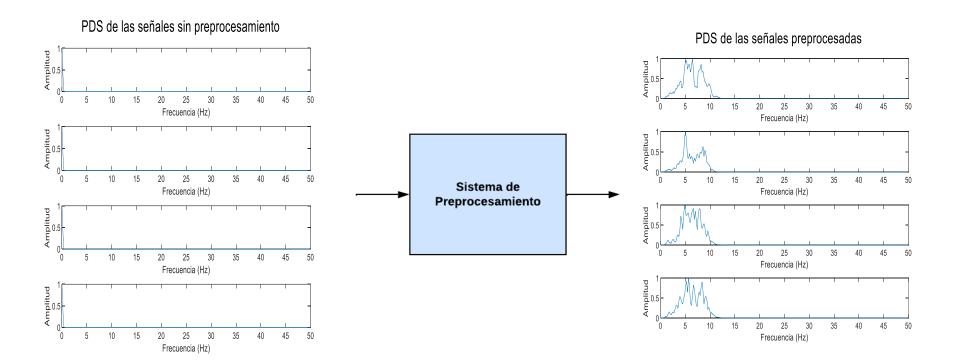






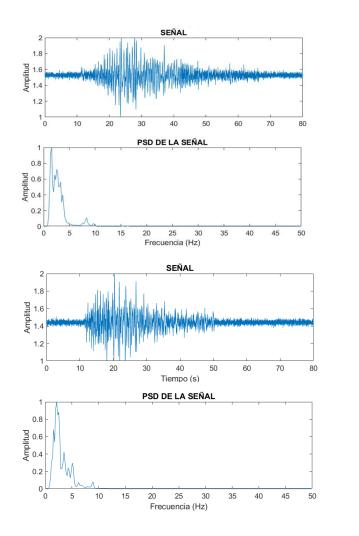


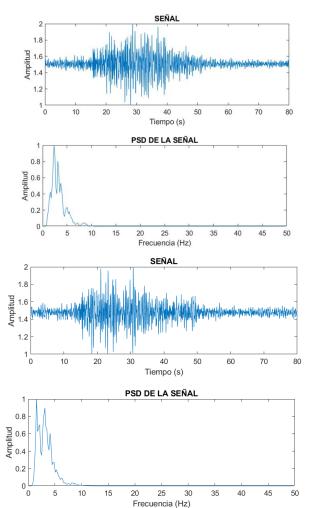
Preprocesamiento





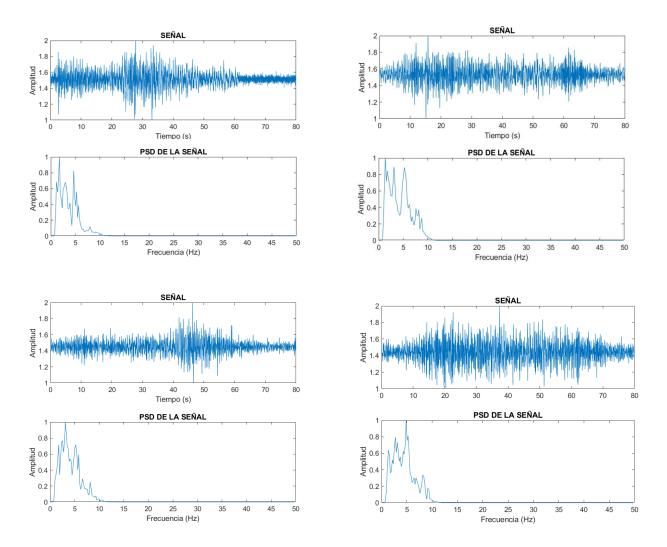
Señales sintéticas LP





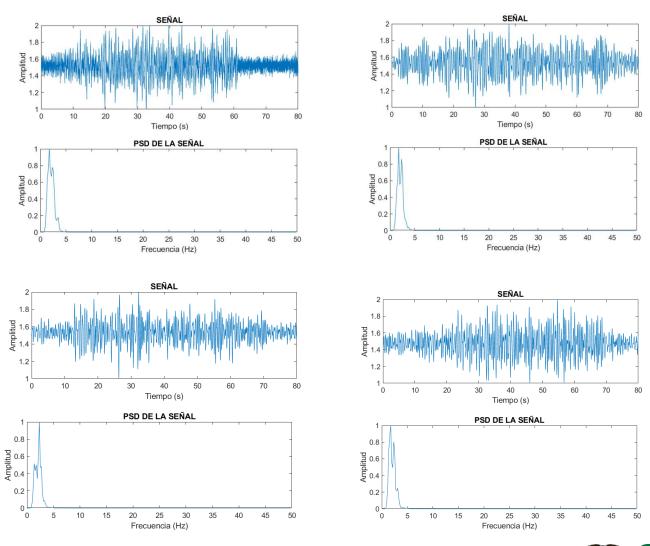


Señales sintéticas TC



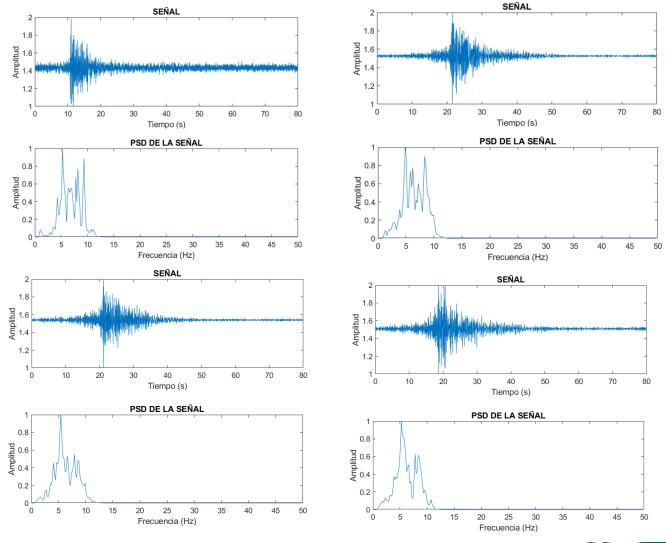


Señales sintéticas TR





Señales sintéticas VT



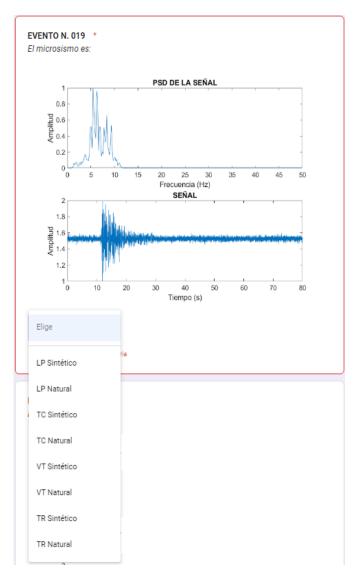




1. EXPERTOS IGEPN: El 19 de junio de 2023, se implementó un cuestionario, que contenía un conjunto de 40 señales. Estas señales incluían tanto señales sintéticas como señales reales.

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.





Tipo de evento	Cantidad sintéticos	Cantidad reales		
LP	6	4		
тс	4	6		
TR	6	4		
VT	7	3		



PSD DE LA SEÑAL PSD DE LA SEÑAL PSD DE LA SEÑAL 0.8 Amplitud 0.4 Amplitud 0.4 0.2 Señales 25 25 Frecuencia (Hz) Frecuencia (Hz) originales Frecuencia (Hz) SEÑAL SEÑAL SEÑAL Tiempo (s) Tiempo (s) Tiempo (s) PSD DE LA SEÑAL PSD DE LA SEÑAL PSD DE LA SEÑAL Amplitud 0.4 Señales sintéticas 25 20 25 Frecuencia (Hz) Frecuencia (Hz) Frecuencia (Hz) SEÑAL SEÑAL SEÑAL 1.2

Tiempo (s)



			Respu	esta	
Numero	Tipo de evento	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	3330	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	

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		
Promedio pon	derado	60.87%		

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



Número de	Tipo Evento	
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%)

Tipo de señal	Evento que causa confusión					
Tipo de Seliai	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		

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 Clasificador

Métricas de evaluación

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

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

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

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

$$BER = 1 - \frac{R+S}{200}$$

- Exactitud (A, del inglés Accuracy)
- Precisión (P, del inglés Precision)
- Sensibilidad (R, del inglés Recall)
- Especificidad (S, del inglés Specificity)
- Tasa de Error de Bit (BER, del inglés Bit Error Rate)

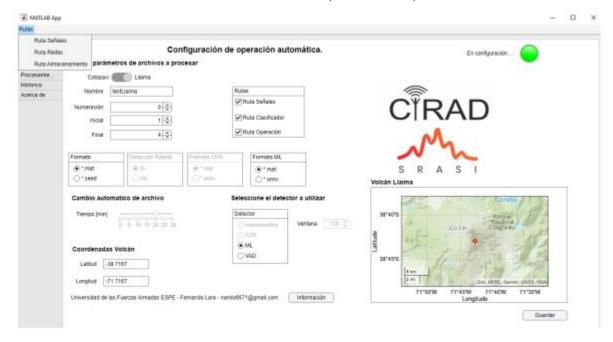
Dónde:

- 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
- N_{FP} corresponde a la cantidad de falsos positivos



Clasificador

OE3: Creación de algoritmos inteligentes basados en la teoría de Machine Learning tradicional para la clasificación de los eventos sísmicos en el volcán Llaima (multiclase).



Clasificador basado en tres algoritmos de Machine Learning

- Árboles de Decisión (DT)
- k-Nearest Neighbors (k-NN)
- Máquinas de Vectores de Soporte (SVM)
- Algoritmo de votación



Clasificador

Proceso:

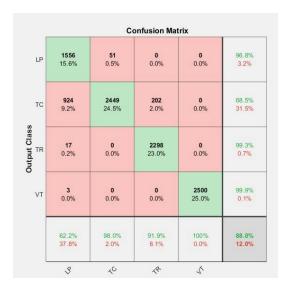
- Antes de ingresar al clasificador, las señales fueron filtradas.
- Se extrajeron 84 características entre:
 - ✓ 13 tiempo
 - ✓ 21 frecuencia
 - ✓ 50 escala
- El clasificador emplea 3 modelos de Machine Learning y algoritmo de votación.
- Se utilizó DT con tan solo 6 características de las 84, debido que DT tiene su propio selector de características.
- *k*-NN *y* SVM emplea las 29 características, las cuales fueron adquiridas mediante el método de envoltura hacia adelante.
- Después estos modelos fueron ingresados al algoritmo de votación y se determina una A(%) = 99,90 y BER = 0,01



Clasificador

Utilizando una base de datos de 10000 señales sintéticas se obtuvo los siguientes resultados:

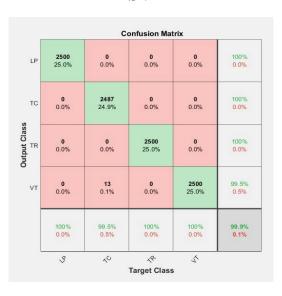
DT



k-NN



SVM



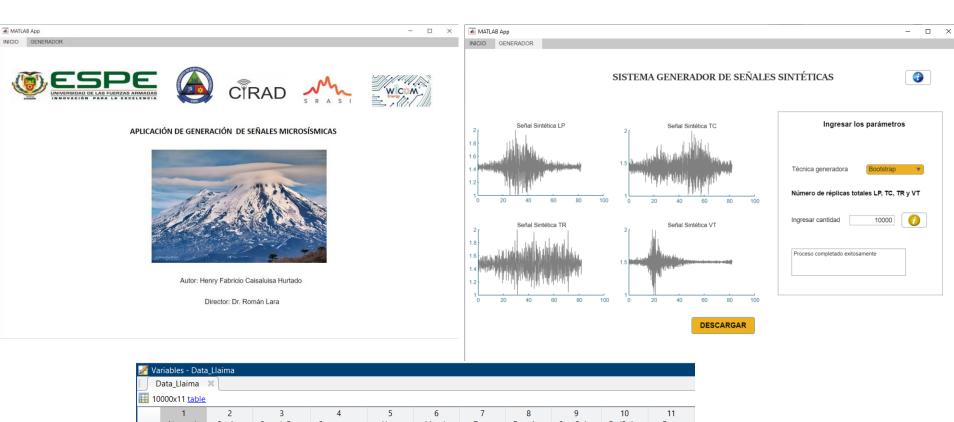
Sistema de votación

Eventos	ventos A (%)		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

Sistema de votación: A(%) = 99.90



Aplicativo Computacional



4	ranables - Dai	_									
	Data_Llaima	×									
	10000x11 <u>table</u>	2									
	1 2 3 4 5 6 7 8 9 10									11	
	Network	Station	SampleRate	Component	Year	Month	Туре	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



- Mediante la utilización del método de remuestreo Bootstrap en conjunto con la Transformada de Fourier, 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 TR 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.



4. 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.
- Existe gran interés en 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.



4. Trabajos Futuros

- 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.







DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

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

AUTOR: Henry Fabricio Caisaluisa Hurtado

DIRECTOR DEL PROYECTO: Ing. Román Alcides Lara Cueva, PhD.

07 de septiembre de 2023

