

DETECCIÓN DE RUIDO DE DISPERSIÓN EN DATOS SÍSMICOS MULTICOMPONENTE
UTILIZANDO APRENDIZAJE AUTOMÁTICO

HOLGER GIOVANNI QUINTERO SANTANDER

Cod. 2208054

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS BÁSICAS

ESCUELA DE FÍSICA

MAESTRÍA EN GEOFÍSICA

BUCARAMANGA

2022

DETECCIÓN DE RUIDO DE DISPERSIÓN EN DATOS SÍSMICOS MULTICOMPONENTE
UTILIZANDO APRENDIZAJE AUTOMÁTICO

HOLGER GIOVANNI QUINTERO SANTANDER

Trabajo de Grado para optar al título de
Magíster en Geofísica

Director

Ph.D. William Mauricio Agudelo Zambrano
Centro de Innovación y Tecnología - ECOPETROL S.A.

Codirector

Ph.D.(c) Ivan Javier Sánchez Galvis
Universidad Industrial de Santander

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS BÁSICAS

ESCUELA DE FÍSICA

BUCARAMANGA

2022

Elaborado por:	Revisado por:	Revisado por:	Aprobado por:
<p>Holger Giovanni Quintero Santander Estudiante de Maestría</p>	<p>Ph.D. William Mauricio Agudelo Zambrano Director del Trabajo de Investigación</p>	<p>MSc. Ivan Javier Sanchez Galvis Co-Director del Trabajo de Investigación</p>	<p>Comité Asesor de Posgrados Acta No. ____ del ____ de 2020</p>

En memoria de mi padre

AGRADECIMIENTOS

Expreso mis más sinceros agradecimientos a las personas que estuvieron conmigo durante el desarrollo de este trabajo de investigación:

Ph.D. William Mauricio Agudelo Zambrano

Director del trabajo de investigación quien gracias a su amplio conocimiento me fue posible la realización de este proyecto.

MSc. Ivan Javier Sanchez Galvis

Codirector del trabajo y amigo que gracias a su colaboración y paciencia pude lograr cada objetivo en esta etapa de mi vida .

Agradezco a mis compañeros de maestría quienes me colaboraron durante el primer año de mis estudios de posgrado y a mis compañeros de la escuela de física y amigos que brindaron todo su apoyo y amistad en este proceso.

Agradezco a toda mi familia especialmente a mi madre, mis hermanos y mi novia quienes son el motor de mi vida, gracias por todo el apoyo económico y emocional que tuvieron para conmigo durante el desarrollo de mi maestría.

Por último, quiero expresar mi mayor agradecimiento a mi padre, Don Jaime Enrique Quintero Acevedo, quien falleció en la etapa final de mi proyecto de investigación y quien siempre me motivó a cumplir mis sueños. Tu partida ha sido muy difícil de asimilar pero se que me ayudarás a seguir adelante y te llenaré de orgullo donde quiera que estés.

CONTENIDO

	pág.
1. INTRODUCCIÓN	19
1.1. Introducción	19
1.2. Formulación del Problema	21
1.3. Motivación y Justificación	26
1.4. Declaración de la Tesis	28
1.4.1. Objetivo General	28
1.4.2. Objetivos Específicos	28
1.5. Alcance y metodología de trabajo	29
2. MODELADO SÍSMICO	30
2.1. Sobre la Dispersión	30
2.1.1. Visión General	30
2.1.2. Caracterización de la Dispersión	33
2.2. Modelado Numérico en el Dominio 2D P-SV	35
2.2.1. Formulación de Vacío Mejorada	38
2.2.2. Consideración de una Capa Ficticia	39
2.3. Código de Modelado	40
2.4. Modelos Elásticos	41
2.4.1. Modelo terrestre de única capa sobre un semiespacio	42
2.4.2. Modelo terrestre triple capa con topografía irregular.	45
2.5. Modelo TENERIFE	48
2.6. Medio Aleatorio	50
2.7. Generación de Datos Sintéticos Multicomponente	52
2.8. Repositorio	56

3. EXTRACCIÓN DE ATRIBUTOS SÍSMICOS Y AGRUPACIÓN	57
3.1. Cálculo de Atributos Sísmicos	57
3.1.1. Atributos instantáneos	58
3.1.2. Atributos de velocidad	65
3.1.3. Atributos de ventana	66
3.2. Métodos de Selección de Atributos	72
3.2.1. Métodos de envoltura.	73
3.2.2. Métodos de filtrado.	74
3.2.3. Métodos integrados.	74
3.3. Selección de Atributos	75
3.4. Métodos de Agrupación	78
3.4.1. K-means	78
3.4.2. Agrupación Aglomerativa	80
3.4.3. DBSCAN	81
3.5. Elección del Método de Agrupación	83
3.6. Algoritmo K-means en el Proyecto.	87
3.7. Metodología de Agrupación Sobre un Registro Sísmico	88
3.7.1. Lectura de registros	89
3.7.2. Extracción de atributos	90
3.7.3. Agrupación de ondas del registro	91
3.8. Repositorio	93
4. DISPERSIÓN SÍSMICA SINTÉTICA BASADA EN MACHINE LEARNING	94
4.1. Agrupación Simultánea	95
4.2. Atributo de Dispersión Usando Aprendizaje Automático	97
4.3. Demostración del Atributo de Dispersión	100
5. CLASIFICACIÓN	103

5.1. Métodos de clasificación	103
5.1.1. Redes neuronales	104
5.1.2. Árboles de decisión	104
5.1.3. Redes Bayesianas	105
5.1.4. Máquinas de vectores de soporte	106
5.2. Elección del algoritmo de clasificación	106
5.3. Red Neuronal Convolutiva	107
5.3.1. Datos y elección de parámetros	108
5.3.2. Capas convolucionales	110
5.4. Entrenamiento del algoritmo	111
5.5. Repositorio	112
6. APLICACIÓN DEL MÉTODO SOBRE DATOS SÍSMICOS	113
6.1. Detección Sobre Datos Sísmicos Sintéticos	113
6.2. Detección Sobre Datos Sísmicos Reales	116
7. RECOMENDACIONES Y TRABAJO FUTURO	120
7.1. Recomendaciones	120
7.2. Trabajo futuro	121
8. CONCLUSIONES	123
BIBLIOGRAFÍA	125

LISTA DE FIGURAS

	pág.
Figura 1. Registro sísmico multicomponente extraído del programa Tenerife 2009 en el Valle medio del Magdalena.	22
Figura 2. Diagrama de trabajo para la detección del ruido de dispersión sobre datos sísmicos.	23
Figura 3. Ondas dispersas en el área de adquisición de piedemonte.	27
Figura 4. Geometría de dispersión por una heterogeneidad localizada para una onda incidente en la dirección x_3 . Imagen tomada de: [Sato et al., 2012].	30
Figura 5. Geometría de un modelo de dispersión simple con una sola heterogeneidad vista desde planta. Imagen tomada de: [Sato et al., 2012].	31
Figura 6. Diferencial de dispersión para un único objeto dispersor. Imagen adaptada de: [Sato et al., 2012].	33
Figura 7. Medio aleatorio no homogéneo.	34
Figura 8. Esquema de cuadrícula escalonada utilizada para la formulación de vacío. Cada círculo es un nodo de la cuadrícula con sus respectivas posiciones marcadas por los índices. τ_{xx} y τ_{zz} son las componentes normales del tensor de esfuerzos y τ_{xz} el esfuerzo cortante en esa dirección, λ y μ son los coeficientes de Lamé, ρ es la densidad. Los cuadrados y círculos sombreados representan la velocidad de las partículas horizontal y vertical respectivamente. Imagen tomada de: [Zeng et al., 2011].	38
Figura 9. Distribución de cuadrícula en la formulación de vacío mejorada en presencia de topografía superficial.	40
Figura 10. Representación del flujo de trabajo del código de modelado.	41

Figura 11.	Modelo terrestre sintético. Modelo con una capa simple en un semiespacio con dos objetos dispersores circulares $(x,z)=(360 \text{ m}, 15 \text{ m})$ y $(x,z)=(720 \text{ m}, 15 \text{ m})$ respectivamente. Ubicación de la fuente: $(x,z)=(150 \text{ m}, 15 \text{ m})$.	43
Figura 12.	Registro de la propagación del campo de onda total en un tiempo de 0.3 segundos (parte izquierda) y 0.5 segundos (parte derecha). Ubicación de la fuente: $(x,z)=(150 \text{ m}, 15 \text{ m})$. El interfaz entre los dos medios se encuentra marcado por la línea punteada a 200 m de la superficie.	44
Figura 13.	Registro de la propagación del campo de onda total en un tiempo de 0.3 segundos (parte izquierda) y 0.5 segundos (parte derecha). Tomado de: [Abdulaziz et al., 2014]	44
Figura 14.	Registros de disparo obtenidos de la simulación del campo de onda en diferencias finitas de la figura 12. Componente horizontal (V_x) y componente vertical (V_z).	45
Figura 15.	Modelo terrestre sintético con tres capas y topografía irregular gaussiana. Ubicación de la fuente: $(x,y)=(150 \text{ m}, 15 \text{ m})$	46
Figura 16.	Perfil gaussiano utilizado para modelar la topografía del modelo de la figura 15	46
Figura 17.	Propagación del campo de onda durante un segundo en un modelo con topografía irregular. Fuente ubicada a 15 metros de profundidad. (a) Propagación del campo de onda a 0.25 [s]. (b) Propagación del campo de onda a 0.5 [s]. (c) Propagación del campo de onda a 0.75 [s]. (d) Propagación del campo de onda a 1 [s].	47
Figura 18.	Registros de disparo obtenidos para el modelo con topografía irregular.	48
Figura 19.	Ubicación geográfica campo Tenerife (zona de adquisición). Tomado de: [Ospina et al., 2011]	49
Figura 20.	Sección 2D de línea sísmica campo Tenerife.	49

Figura 21.	Variaciones del modelo base para el modelo de velocidad de onda P. (a) Modelo Tenerife. (b) Cambio aleatorio en la magnitud de la velocidad por capas (cambio entre 0-1). (c) Primera capa con tan contrastada. (d) Capas con profundidad uniforme.	53
Figura 22.	Variaciones de los registros verticales para cambios del modelo base. (a) Registro vertical del modelo Tenerife. (b) Registro vertical de modelo con cambios aleatorios en la magnitud de la velocidad por capas (cambio entre 0-1). (c) Registro vertical del modelo con la primera capa con tan contrastada. (d) Registro vertical del modelo con capas con profundidad uniforme.	53
Figura 23.	Medios aleatorios sobre las variaciones del modelo base. (a) Medio aleatorio del ancho del modelo ubicado sobre la capa más superficial. (b) Medio aleatorio de profundidad 20 [m] ubicado en la parte inferior derecha del modelo. (c) Medio aleatorio de profundidad 8 [m] ubicado en la parte superior izquierda el modelo. (d) Medio aleatorio de profundidad 2 [m] ubicado en la parte media-inferior del modelo y de ancho completo.	54
Figura 24.	Registros verticales de los modelos con medios aleatorios de la figura 23.	55
Figura 25.	Clasificación de los atributos sísmicos calculados.	57
Figura 26.	Elipses de polarización para una señal armónica con diferencias de fase de $0, \pi/4, \pi/2, 3\pi/4, \pi$ y $-\pi/4$. Imagen adaptada de: [René et al., 1986] ¹ .	62
Figura 27.	Atributos sísmicos extraídos del modelo Tenerife.	72
Figura 28.	Funcionamiento de los métodos de envoltura.	74
Figura 29.	Funcionamiento de los métodos de filtrado.	74

¹ R. M. René y col. "Multicomponent seismic studies using complex trace analysis". En: *GEOPHYSICS* 51.6 (1986). DOI: 10.1190/1.1442177.

Figura 30.	Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.	75
Figura 31.	Atributos sísmicos para los registros de la figura 30.	76
Figura 32.	Matriz de correlación de atributos.	77
Figura 33.	Funcionamiento del algoritmo k-means en un conjunto de datos sintético. Imagen tomada de la librería mglearn [Müller y Guido, 2016].	79
Figura 34.	Falla del algoritmo k-means al no identificar grupos no esféricos. Imagen tomada de la librería mglearn [Müller y Guido, 2016].	80
Figura 35.	Funcionamiento del agrupamiento aglomerativo (une iterativamente los dos clústers más cercanos deteniéndose cuando encuentra tres grupos). Imagen tomada de la librería mglearn [Müller y Guido, 2016].	81
Figura 36.	Funcionamiento del algoritmo DBSCAN con diferentes configuraciones para los parámetros de muestras mínimas y eps. Imagen tomada de la librería mglearn [Müller y Guido, 2016].	83
Figura 37.	Dos registros sintéticos de onda sísmica con tres diferentes campos de onda cada uno. Imagen tomada de: [Xia et al., 2018]	85
Figura 38.	Resultados de la detección de campos de onda para un registro sísmico. Lado izquierdo: Áreas de un registro sísmico agrupadas por expertos. Lado derecho: Detección del área de registro usando k-means con dos grupos. Imagen tomada de: [Sánchez et al., 2017].	86
Figura 39.	Ejemplo de la agrupación empleando el algoritmo K-means. Imagen tomada de la librería mglearn [Müller y Guido, 2016].	87
Figura 40.	Modelos sintéticos de velocidad y densidad basado en el modelo Tenerife.	89
Figura 41.	Registro sísmico 2D para el modelo de la figura 40.	89
Figura 42.	maska de activación de registros.	90
Figura 43.	Atributos sísmicos para los registros de la figura 41.	90

Figura 44. Agrupación para los tipos de onda presente en los registros de la figura 41.	91
Figura 45. Diagrama de dispersión para cuatro atributos sísmicos del etiquetado de la figura 44.	93
Figura 46. Diferencia entre atributos sísmicos con y sin dispersión para el mismo modelo base.	94
Figura 47. Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.	95
Figura 48. Atributos sísmicos para los registros de la figura 47.	96
Figura 49. Agrupación para los tipos de onda presente en los registros de la figura 47.	97
Figura 50. Separación de las agrupaciones de la figura 49. A) Agrupación para el registro sin dispersión. B) Agrupación para el registro con dispersión.	98
Figura 51. Atributo de dispersión para los registros de la parte derecha en cada gráfica de la figura 47.	99
Figura 52. Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.	100
Figura 53. Separación de las agrupaciones para los registros de la figura 52. A) Agrupación para el registro de la izquierda. B) Agrupación para el registro e la derecha.	101
Figura 54. Atributo de dispersión para los registros de la parte derecha en cada gráfica de la figura 52.	101
Figura 55. Funcionamiento de una red neuronal básica.	104
Figura 56. Funcionamiento de un árbol de decisión.	105

Figura 57. Funcionamiento gráfico de las máquinas de vectores de soporte. Imagen tomada de la librería mglearn [Müller y Guido, 2016].	106
Figura 58. Proceso de clasificación de ondas multicomponente a través de aprendizaje automático.	107
Figura 59. Píxeles de una zona de un registro sísmico.	108
Figura 60. Tríada de datos de entrada.	109
Figura 61. Ejemplo de convolución en una red neuronal convolucional. Imagen adaptada de: [Barrios, 2022].	111
Figura 62. Resultados de la clasificación para un par de datos sísmicos sintéticos.	113
Figura 63. Resultados de la detección del ruido de dispersión en rojo sobre un registro sísmico.	114
Figura 64. Resultados de la clasificación para un par de datos sísmicos sintéticos.	115
Figura 65. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.	116
Figura 66. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.	117
Figura 67. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.	119

LISTA DE TABLAS

	pág.
Tabla 1. Propiedades de los materiales para el modelo y los contrastes de impedancia relativos a la capa I.	43
Tabla 2. Propiedades de los materiales para el modelo de la figura 15.	47
Tabla 3. Métodos de extracción de atributos sísmicos de ventana para datos multicomponente calculados en este trabajo.	71
Tabla 4. Métodos de agrupación con sus parámetros iniciales y la métrica usada.	83
Tabla 5. Propiedades de los métodos de agrupación estudiados.	83

LISTA DE ACRÓNIMOS

CEMOS Grupo de Investigación en Control, Electrónica, Modelado y Simulación

CNN Convolutional Neural Network

DBSCAN Density-based Spatial Clustering of Applications with Noise

DFM Finite Difference Method

FEM Finite Element Method

GPU Graphics Processing Unit

LASSO Least Absolute Shrinkage and Selection Operator

MAE Mean Absolute Error

MSE Mean Square Error

PML Perfectly Matched Layer

PSDF Power Spectral Density Functions

RM Random Media

SEM Spectral Element Method

SVD Singular Value Decomposition

SVM Support Vector Machine

RESUMEN

TÍTULO: DETECCIÓN DE RUIDO DE DISPERSIÓN EN DATOS SÍSMICOS MULTICOMPONENTE UTILIZANDO APRENDIZAJE AUTOMÁTICO *

AUTOR: HOLGER GIOVANNI QUINTERO SANTANDER **

PALABRAS CLAVE: DISPERSIÓN, ESTRATO SOMERO, APRENDIZAJE AUTOMÁTICO, MODELADO NUMÉRICO, ATRIBUTO SÍSMICO.

DESCRIPCIÓN:

La generación de imágenes sísmicas de estructuras complejas en zonas montañosas es uno de los principales desafíos de la exploración sísmica terrestre. Las señales de reflexión deseadas son enmascaradas por las ondas sísmicas dispersadas por heterogeneidades cerca de la superficie logrando que las ondas se superpongan en frecuencia y los filtros convencionales sean poco precisos, por ello, se propone un método de clasificación de datos sintéticos de onda sísmica dispersada, los cuales se procesarán a través de un método de aprendizaje automático. Para esto se modelarán las ondas dispersadas mediante un modelado numérico 2D. Se seleccionarán ciertos atributos sísmicos para resaltar la información geofísica de los datos tales como: frecuencia, velocidad de propagación, polarización, entre otros. Una vez con los atributos seleccionados junto con un algoritmo de clasificación, los datos se dividirán en grupos automáticamente y a partir de esto se inferirá la relación señal-ruido.

* Trabajo de grado, Director: Ph.D. William M. Agudelo, Codirector: Ph.D.(c) Ivan J. Sánchez Galvis.

** Facultad de Ciencias Básicas. Escuela de Física. Maestría en Geofísica.

ABSTRACT

TITLE: BACK-SCATTERING NOISE DETECTION IN MULTICOMPONENT SEISMIC DATA USING MACHINE LEARNING *

AUTHOR: HOLGER GIOVANNI QUINTERO SANTANDER **

KEYWORDS: SCATTERING, NEAR-SURFACE, MACHINE LEARNING, NUMERICAL MODELING, SEISMIC ATTRIBUTE.

DESCRIPTION:

Seismic imaging of complex structures in mountainous areas is one of the main challenges of land seismic exploration. The desired reflection signals are masked by seismic waves scattered by heterogeneities near the surface, causing the waves to overlap in frequency and conventional filters to be inaccurate. For this reason, a method of classification of synthetic data of scattered seismic wave is proposed, which will be processed through a machine learning method. To get this, the scattered waves will be modeled using a 2D numerical modeling. Certain seismic attributes will be selected to highlight the geophysical information of the data such as: frequency, velocity of propagation, polarization, among others. Once with the selected attributes together with a classification algorithm, the data will be divided into groups automatically and from this the signal-to-noise ratio will be inferred.

* M.Sc. Thesis. Advisor: Ph.D. William J. Agudelo, Coadvisor: Ph.D.(c) Ivan J. Sánchez Galvis.

** Faculty of Basic Sciences. Master's Degree in Geophysics.

INTRODUCCIÓN

1.1. Introducción

Durante décadas, el uso de diversos métodos geofísicos en regiones cerca de la superficie ha despertado un amplio interés en sus aplicaciones tales como la exploración de hidrocarburos^{1,2}, la arqueología³, investigación geotécnica⁴ e incluso proyectos de carácter militar y forense^{5,6}. Para estudiar las características sobre esta región se han implementado diversos métodos geofísicos y uno de los más usados es la sísmica. Los registros sísmicos generalmente son utilizadas para plantear modelos de velocidad de onda que se encuentran estrechamente relacionados con propiedades del subsuelo tales como la densidad o el módulo de cizalla, sin embargo, las ondas sísmicas que son dispersadas también pueden ser utilizadas para visualizar y delimitar heterogeneidades en el medio. La relación señal-ruido en un conjunto de datos es un indicador de la calidad de un registro y las ondas dispersadas generalmente están relacionadas con áreas de muy baja relación, esto se debe a que las ondas dispersadas se superponen con las demás ondas en los registros de

¹ Dwain K. Butler. “Near-Surface Geophysics”. En: *Society of Exploration Geophysicists* (2005).

² Richard D. Miller, John H. Bradford y Klaus Holliger. “Advances in Near-Surface Seismology and Ground-Penetrating Radar”. En: *Society of Exploration Geophysicists* (2005).

³ Mario E. Arias y Alexis Mojica. “Prospección Arqueológica en América Central Mediante el Uso de Métodos Geofísicos”. En: *Revista Geológica de América Central* (2011).

⁴ Neil L. Anderson y col. “Geophysical Methods Commonly Employed for Geotechnical Site Characterization”. En: *Transportation Research Board* (2008).

⁵ J. K. Pringle y col. “The Use of Geoscience Methods for Terrestrial Forensic Searches”. En: *Earth-Science Reviews* 114 (2012), págs. 108-123. DOI: 10.1016/j.earscirev.2012.05.006.

⁶ James D. Hansen, Jamie K. Pringlea y Jon Goodwinb. “GPR and Bulk Ground Resistivity Surveys in Graveyards: Locating Unmarked Burials in Contrasting Soil Types”. En: *Forensic Science International* 237 (2014), e14-e29. DOI: 10.1016/j.forsciint.2014.01.009.

disparo, de modo que, estas ondas superpuestas son consideradas como ruido y cuando son filtradas normalmente parte de la señal se elimina, siendo este un problema a considerar en el tratamiento de señales sísmicas.

En este proyecto se plantea un método de clasificación de ondas sísmicas dispersadas para registros multicomponente de manera automática utilizando principalmente la información obtenida de los atributos sísmicos. El funcionamiento del método sobre datos sintéticos abrirá un camino novedoso al tratamiento de datos sísmicos con una elevada relación señal-ruido en regiones cerca de la superficie. Para el desarrollo del trabajo se proponen cuatro etapas:

1. Modelado numérico del medio de propagación y generación de datos sintéticos multicomponente.
2. Extracción de atributos sísmicos en un registros 2C y definición del atributo de ruido de dispersión.
3. Clasificación de zonas en registros sísmicos empleando métodos de aprendizaje automático.
4. Evaluación de resultados (aplicación sobre datos sísmicos sintéticos y datos sísmicos reales).

Se espera que este método pueda ser empleado sobre datos sísmicos reales y junto a herramientas de procesamiento pueda obtener información sobre las heterogeneidades presentes en las zonas de adquisición. La caracterización de las heterogeneidades y la generación de imágenes del subsuelo sobre las zonas de adquisición es muy valiosa, dado que el estrato somero es una región que, aunque compleja, tiene una gran importancia, tiene

un efecto significativo en la calidad de la imagen⁷ a profundidades superiores. Incluso el óptimo desarrollo de este método podría contribuir a un mejoramiento en el diseño de las geometrías de adquisición en regiones complejas⁸.

1.2. Formulación del Problema

La generación de imágenes sísmicas de estructuras complejas en zonas montañosas es uno de los principales desafíos de la exploración sísmica terrestre, sus dificultades radican en la obtención de imágenes de las estructuras más profundas del subsuelo, esto se debe a que la mayor parte de la energía sísmica se queda atrapada en la superficie⁹, de modo que solo una pequeña parte de la energía ilumina las profundidades de interés, por ejemplo; profundidades donde se encuentran recursos tales como hidrocarburos. Además, en áreas de adquisición complejas como el piedemonte, las ondas de reflexión que provienen de estructuras más profundas se dispersan por las heterogeneidades cerca de la superficie, deteriorando la señal junto con la información de interés, por lo tanto, es necesario aprovechar toda la información obtenida y los registros multicomponente son una buena opción para estas zonas complejas.

Aunque los registros multicomponente no son frecuentemente adquiridos en campo, al ofrecer más información que un único registro sísmico, entonces tienen la capacidad de

⁷ E. Díaz y P. Sava. "Understanding the reverse time migration backscattering: noise or signal?" En: *Geophysical Prospecting* 64.3 (2016), págs. 581 -594. DOI: 10.1111/1365-2478.12232.

⁸ C. Stork. "A Model for Land Seismic Surface Scattering Noise has Implications for Acquisition and Processing". En: *European Association of Geoscientists Engineers* (2021). DOI: 10.3997/2214-4609.202112562.

⁹ Christof Stork and David Flentge. "Using 2D ring arrays to remove back-scattered surface noise from land seismic data". En: *SEG Technical Program Expanded* (2015).

permitir extraer una mayor cantidad de atributos sísmicos. La figura 1 muestra un ejemplo de un registro sísmico multicomponente 3C donde cada uno de los modos de onda brindan una información diferente como tiempos de llegada de la onda.

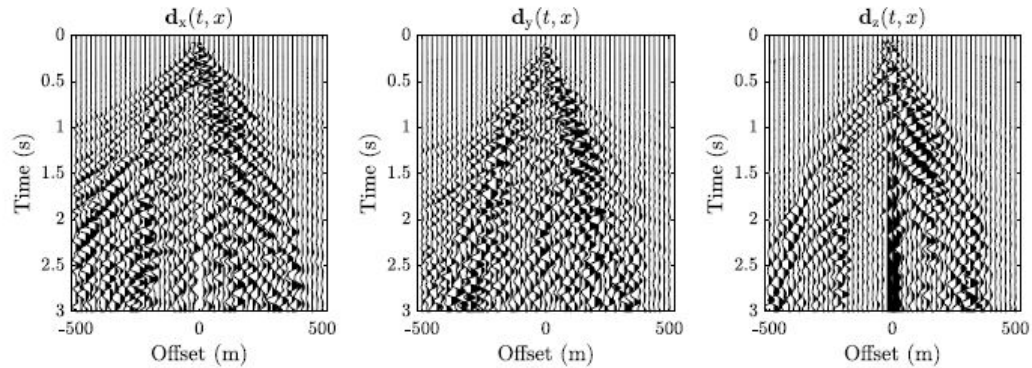


Figura 1. Registro sísmico multicomponente extraído del programa Tenerife 2009 en el Valle medio del Magdalena.

En este trabajo de investigación se detectarán las ondas sísmicas dispersadas para datos sintéticos multicomponente, por lo tanto, la pregunta planteada en esta propuesta de investigación es: **¿Como detectar automáticamente la presencia de ruido de dispersión en datos sísmicos multicomponente y adicionalmente inferir su relación señal-ruido?** Es importante buscar una solución a esta pregunta por que para el éxito de una adquisición y procesamiento de señales sísmicas es fundamental identificar el ruido de dispersión y la relación que existe entre el ruido y la señal. Autores como [Robinson y Al-Husseini,1982]¹⁰ y [Regone,1997]¹¹ analizan el ruido de dispersión producido por topografía y heteroge-

¹⁰ Don K. Robinson y Moujahed I. Al-Husseini. "Technique for reflection prospecting in the Rub' Al-Khali". En: *Society of Exploration Geophysicists* 47.8 (1982), págs. 1135-1234. DOI: 10.1190/1.1441377.

¹¹ Carl J. Regone. "Measurement and identification of 3-D coherent noise generated from irregular surface carbonates". En: *Carbonate Seismology* (1997), pág. 455.

neidades que afectan las ondas reflejadas. [Regone,1998]¹² propone analizar los disparos e identificar rangos de velocidad del ruido para cuantificar la relación señal-ruido empleando un juicio humano experto en el tema.

Aunque el ruido de dispersión es un problema descubierto desde hace varias décadas, actualmente, despierta el interés de diferentes autores y es inevitable no tratar de intervenir el ruido mediante métodos de computación avanzados. Los métodos de aprendizaje automático plantean una nueva forma de observar el ruido de dispersión y cuantificar la relación señal-ruido, no mediante un juicio experto como se había trabajado años atrás, sino mediante un entrenamiento y aprendizaje de la máquina, donde los datos exhiben por ellos mismos sus propios atributos, por lo anterior, es importante destacar el carácter innovador del trabajo propuesto.

La estrategia planteada para atacar el problema del ruido de dispersión es analizar toda la información proporcionada por los registros sísmicos. Para ello, se propone desarrollar un método basado en aprendizaje automático para detectar la presencia de las ondas dispersadas y adicionalmente inferir la relación señal-ruido sobre los datos.

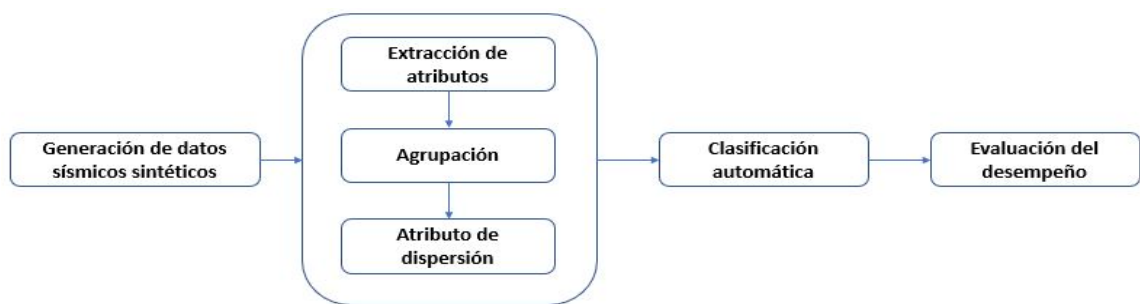


Figura 2. Diagrama de trabajo para la detección del ruido de dispersión sobre datos sísmicos.

¹² Carl J. Regone. "Suppression of coherent noise in 3-D seismology". En: *Carbonate Seismology* 17.11 (1998), págs. 1584-1589. DOI: 10.1190/1.1437900.

El diagrama de la figura 2 muestra el flujo de trabajo que se realizará para la detección del ruido de dispersión.

En un registro de disparo existen diferentes tipos de ondas que pueden ser diferenciadas utilizando atributos sísmicos quienes resaltan la información geofísica de las ondas adquiridas. En consecuencia, **el ruido de dispersión puede ser caracterizado utilizando atributos sísmicos calculados en datos multicomponente. A su vez, los atributos se pueden usar para entrenar algoritmos de aprendizaje automático que detecten la presencia de ruido en registros de disparo.** Es de esperarse que al aplicar este método a un conjunto de datos reales también pueda inferir información sobre el diseño de geometrías de adquisición de datos en lugares altamente complejos y el ruido de dispersión sea minimizado sin necesidad de la aplicación de un filtro electrónico o herramientas de procesamiento.

El método de detección del ruido sísmico de dispersión se considera adecuado ya que se parte del modelado sísmico para generar datos con dispersión y la forma adecuada de estudiar la dispersión en regiones cerca de la superficie es mediante el modelado de las ondas elásticas, empleando heterogeneidades localizadas, topografía irregular, interfaces irregulares o medios aleatorios¹³. El comportamiento de la dispersión depende en gran medida de la estructura cercana a la superficie, por lo que es necesario incluir en lo posible todos los tipos de heterogeneidades para modelarlo con precisión. Además, es importante mencionar que la dispersión es un fenómeno tridimensional, porque de acuerdo con los patrones de radiación la energía se dispersa en todas las direcciones¹⁴, afirmando nueva-

¹³ C. D. Riyanti y G. C. Herman. "Three-dimensional elastic scattering by near-surface heterogeneities". En: *Geophysical Journal International* 160.2 (2005), 609–620. DOI: 10 . 1111 / j . 1365 - 246X . 2005 . 02492 . x.

¹⁴ H. Sato, M. C. Fehler y T. Maeda. *Seismic wave propagation and scattering in the heterogeneous earth*. Springer Science, Business Media, 2012. URL: <https://link.springer.com/book/10.1007/978-3-540-89623-4?noAccess=false>.

mente que los datos multicomponente son los indicados para desarrollar este método. El modelado de los datos la dispersión se realizará en 2D debido al alto costo computacional de los modelados 3D, sin embargo, no se descarta en un futuro trabajar con datos tridimensionales.

Por otro lado, la clasificación de las ondas se realizará empleando métodos de aprendizaje automático, pues el algoritmo entrena con la información brindada exclusivamente por los datos y así se evitan juicios humanos que pueden llegar a ser subjetivos y en ocasiones afectar el procesamiento.

Los métodos de clasificación de aprendizaje automático sobre datos sísmicos han sido empleados para clasificar los tipos de onda presentes en registros sísmicos y las zonas de acción de ondas superficiales¹⁵. Otros trabajos los aplican para delimitar estructuras importantes en la acumulación de hidrocarburos como por ejemplo los domos salinos¹⁶, mientras que trabajos más relacionados con este método se presentan en [Sánchez et al., 2017]¹⁷ y [Stork,2017]¹⁸ donde se clasifican ondas de acuerdo con sus atributos sísmicos y se filtran patrones de dispersión altamente complejos sobre modelos elásticos bastante elaborados.

¹⁵ Keyao Xia, Fred Hilterman y Hao Hua. “Unsupervised machine learning algorithm for detecting and outlining surface waves on seismic shot gathers”. En: *Journal of Applied Geophysics* 157 (2018), págs. 73-86. DOI: 10.1016/j.jappgeo.2018.07.003.

¹⁶ Haibin Di, Muhammad Shafi y Ghassan AlRegib. “Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data”. En: *Geophysical Journal International* 215.3 (2018), 1999–2007. DOI: 10.1093/gji/ggy376.

¹⁷ Ivan Sánchez-Galvis y col. “Seismic attribute selection and clustering to detect and classify surface waves in multicomponent seismic data by using k-means algorithm”. En: *The Leading Edge* 36.3 (2017), págs. 239-248. DOI: 10.1190/tle36030239.1.

¹⁸ Christof Stork. “Removing complex land noise with modern pattern recognition using Machine Learning”. En: *SEG Technical Program Expanded* (2017), pág. 6093.

Este trabajo de investigación propone una clasificación automática del ruido como un atributo sísmico enfocándose en la detección del ruido de dispersión. Esta tarea ha sido lograda únicamente según percepción humana, por lo que su automatización es una idea novedosa y ambiciosa.

1.3. Motivación y Justificación

Los métodos sísmicos son ampliamente usados para problemas de exploración del subsuelo como delimitaciones y mapeos, brindando información importante sobre las propiedades físicas más significativas, estos mapeos pueden aplicarse a secuencias sedimentarias estrechamente relacionadas en la búsqueda de hidrocarburos. En Colombia cerca del 89% de las áreas sedimentarias se encuentran disponibles para la exploración y explotación¹⁹, la industria petrolera ha sido el motor de la economía del país en los últimos años, sin embargo, existen regiones de exploración donde se presenta una compleja topografía y los datos registrados en ellas pueden contener un alto nivel de ruido producido por dispersión. Para cualquier teoría de señales es importante registrar la menor cantidad de ruido posible en los datos, esto contribuye para un mejor tratamiento de ellos y consecuentemente la obtención de mejores resultados, existen diferentes maneras de filtrar o eliminar el ruido en los datos. Este trabajo no propone una técnica de filtrado de ruido común, lo que se busca es clasificar los datos para encontrar correlaciones entre el ruido y los atributos sísmicos de las ondas y de esta manera distinguir los patrones de dispersión y con la información obtenida se puede llegar a perfeccionar geometrías de adquisición para que el país continúe desarrollándose este importante sector.

¹⁹ Asociación Nacional de Hidrocarburos, *Colombia Petrolera*.

En la industria del petróleo y el gas se requiere la generación de imágenes de alta resolución y calidad del subsuelo, lo cual es un problema común principalmente en las zonas más complejas, debido a la gran cantidad de heterogeneidades y topografía irregular en las regiones cerca de la superficie imposibilitando esta tarea y causando que los filtros lineales frecuentemente adaptados al ruido de dispersión sean inefficientes.

La figura 3 muestra recolecciones de tres disparos adquiridas en una zona de piedemonte, en la región plana, el registro exhibe ondas de Rayleigh con un comportamiento homogéneo en un rango de velocidad específico que puede ser fácilmente eliminado por filtros $f-k$; en contraste, los otros dos disparos se recogen en la zona montañosa muestran que la mayor parte de la energía está atrapada y dispersa cerca de la superficie y no es posible observar reflexiones de las ondas de cuerpo.

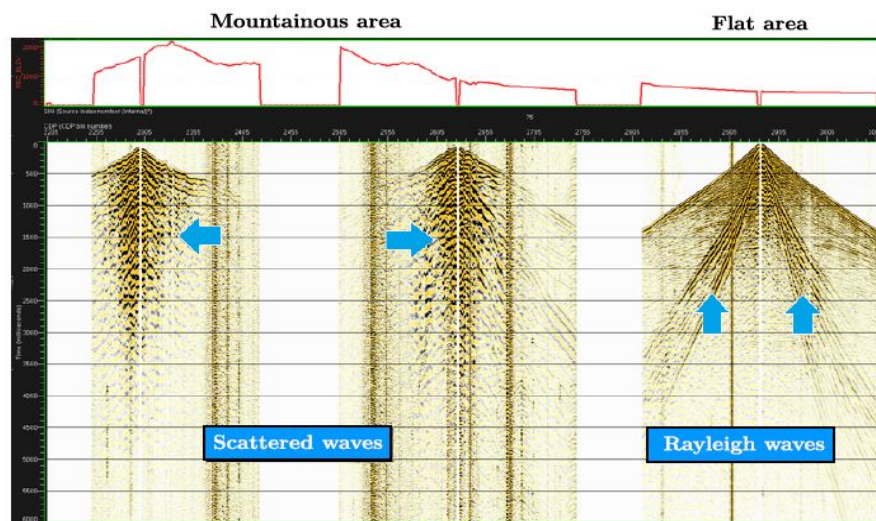


Figura 3. Ondas dispersas en el área de adquisición de piedemonte.

Diseñar nuevos métodos que filtren o permitan distinguir el ruido de los datos es ineludible y las grandes compañías petroleras se han enfocado en impulsar la investigación en este campo, por lo tanto este proyecto de investigación no solo beneficia al sector cientí-

fico sino también al sector de producción energético quien es un pilar económico hoy en día en Colombia.

Por otra parte, una gran ventaja con la que cuentan los proyectos de investigación basados en aprendizaje automático es el acelerado avance en la tecnología, donde se producen masivas y accesibles herramientas informáticas que pueden ser utilizables para mejorar y facilitar el desarrollo del trabajo.

1.4. Declaración de la Tesis

1.4.1. Objetivo General

- Desarrollar un método basado en aprendizaje automático para detectar la presencia de ruido de dispersión y adicionalmente inferir la relación señal-ruido.

1.4.2. Objetivos Específicos

- Modelar el ruido de dispersión utilizando modelado numérico 2D de ondas elásticas en medios con topografía irregular y heterogeneidades cerca de la superficie.
- Seleccionar atributos sísmicos que caractericen el comportamiento del ruido de dispersión en datos multicomponente.
- Implementar un algoritmo de clasificación para detectar automáticamente la presencia de ruido de dispersión en datos multicomponente e inferir la relación señal-ruido utilizando los atributos seleccionados.
- Evaluar el desempeño del algoritmo de clasificación para detectar la presencia de ruido de dispersión en datos sísmicos multicomponente.

1.5. Alcance y metodología de trabajo

Se escogerán diferentes modelos de velocidades y densidad que cuenten con heterogeneidades que produzcan ruido de dispersión. La simulación de la propagación de la onda se efectuará sobre cada modelo para generar los datos sísmicos sintéticos multicomponente que serán usados.

Una vez se cuente con los datos sintéticos de las ondas, se procede a extraer ciertos atributos previamente escogidos. Diferentes métodos de extracción de atributos sísmicos serán empleados, tales como lo son el análisis espectral o descomposición en valores singulares, que escogen características como la amplitud, frecuencia, polarización, energía, planaridad y demás.

La agrupación de las ondas se hace a través de algoritmos de agrupación de datos que son parte de los métodos de aprendizaje automático no supervisado. Los datos se agrupan de acuerdo con una medida o norma que relaciona las características entre sí, creando los denominados clusters. Cuando las regiones de los registros se separan conforme a los diferentes tipos de ondas presentes, entonces se procede a escoger las partes del registro con presencia de ruido de dispersión, se pensó etiquetar las regiones agrupadas en los registros sísmicos, sin embargo, gracias a la estrategia de agrupación simultánea es posible descubrir las regiones con ruido de dispersión para un registro sísmico.

Finalmente, se procede a realizar la clasificación automática de las zonas de los registros sísmicos justificados en la información previamente extraída de los atributos sísmicos, donde se segmentará las imágenes sísmicas de acuerdo a una clasificación binaria basada en el ruido de dispersión, este proceso se llevará a cabo para el conjunto de datos sintéticos y para un registro sísmico de campo para evaluar el desempeño de la clasificación.

2.1. Sobre la Dispersión

2.1.1. Visión General La litosfera de la tierra puede representarse como un medio elástico no homogéneo, por lo tanto, la propagación de las ondas en la tierra debe de tratarse desde el enfoque de la dispersión. La ecuación de onda escalar puede ser escrita como:

$$\left[\Delta - \frac{1}{V(\mathbf{x})^2} \frac{\partial^2}{\partial t^2} \right] u(\mathbf{x}, t) = 0 \quad (1)$$

La propagación de las ondas en el subsuelo puede escribirse en términos del campo de desplazamiento de las partículas y sus esfuerzos en un medio elástico bidimensional, no homogéneo y libre de tracción. Considerando la propagación y la dispersión de la onda elástica incidente sobre una heterogeneidad incrustada y lateralmente homogénea con propiedades físicas contrastantes (ver figura 4), el campo de la onda puede representarse como la superposición del campo de onda de fondo (sin la presencia de la heterogeneidad) y el campo de onda dispersado (producto de la interacción con la heterogeneidad) $u = u_0 + u_1$.

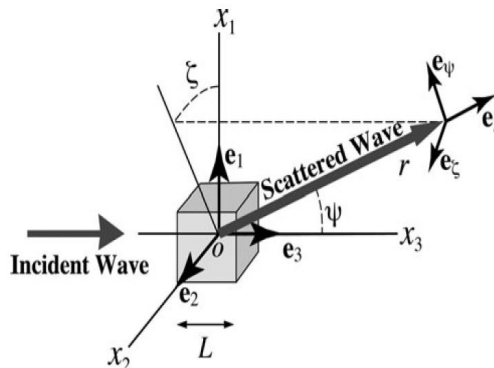


Figura 4. Geometría de dispersión por una heterogeneidad localizada para una onda incidente en la dirección x_3 . Imagen tomada de: [Sato et al., 2012].

Asumiendo que la amplitud de la onda dispersada es mucho menor que la amplitud de la onda incidente $|u_1| \ll |u_0|$ para pequeñas fluctuaciones fraccionadas en la velocidad, se puede escribir la ecuación de onda como:

$$\left[\Delta - \frac{1}{V_0^2} \frac{\partial^2}{\partial t^2} \right] u_1 = -\frac{2}{V_0^2} \xi \frac{\partial^2 u_0}{\partial t^2} \quad (2)$$

Analíticamente la interacción entre la onda incidente y la heterogeneidad se presenta como un término de fuente para la onda dispersada (ver ecuación 2).

Al encontrarse con heterogeneidades en su camino las ondas sísmicas son dispersadas, esto ocurre principalmente en regiones cerca de la superficie donde el medio se distribuye muy heterogéneamente. Existen cuatro tipos de dispersión de las ondas sísmicas: $B \rightarrow B$, $B \rightarrow R$, $R \rightarrow B$ y $R \rightarrow R$, denotando B a las ondas de cuerpo y R a las ondas Rayleigh.

Al considerar un medio no homogéneo, entonces el análisis de la dispersión debe de partir de las ecuaciones no homogéneas de la propagación de las ondas propuestas en [Sato, 1979]²⁰.

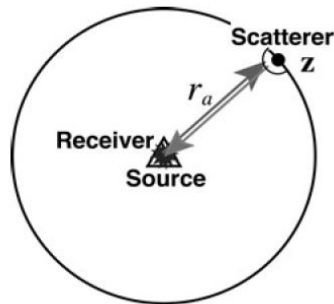


Figura 5. Geometría de un modelo de dispersión simple con una sola heterogeneidad vista desde planta. Imagen tomada de: [Sato et al., 2012].

²⁰ Haruo Sato. "Wave propagation in one dimensional inhomogeneous elastic media". En: *Journal of Physics of the Earth* (1979). DOI: 10.4294/jpe1952.27.455.

Debido a la dispersión, los parámetros elásticos y la densidad pueden ser descompuestos en:

$$\begin{aligned}\rho(\mathbf{x}) &= \rho_0 + \delta\rho(\mathbf{x}) \\ \lambda(\mathbf{x}) &= \lambda_0 + \delta\lambda(\mathbf{x}) \quad \mathbf{x} \in V \\ \mu(\mathbf{x}) &= \mu_0 + \delta\mu(\mathbf{x})\end{aligned}\tag{3}$$

Donde posición es representada por \mathbf{x} , V es la región de volumen de la heterogeneidad y el símbolo δ representa la variación en el parámetro, respectivamente. Como se expresó anteriormente, el campo de desplazamientos se comporta como la suma de un campo inicial \mathbf{u} y un campo dispersado \mathbf{u}_1 .

La dispersión puede ser considerada a través de diferentes enfoques²¹. En este trabajo se considera la dispersión como el efecto sobre el campo de onda producto de diversos objetos dispersores sobre un medio homogéneo. Como se mencionó anteriormente, los objetos dispersores pueden ser considerados como pequeñas fuentes y obedeciendo el principio de Fermat la energía sísmica dispersada llega a los receptores de forma incoherente definiendo el concepto de onda coda. Bajo esta definición, la dispersión es el producto del campo de onda debido a reflexiones, refracciones y difracciones de las ondas sísmicas. Existen diversas formas de clasificar las ondas dispersadas, la clasificación de la dispersión dependiendo del objeto de dispersor es la siguiente:

1. Para dispersores muy pequeños con poca anisotropía se denomina *dispersión cuasi-*

²¹ Seth Stein y Michael Wysession. *An Introduction to Seismology, Earthquakes and Earth Structure*. 1.ª ed. Wiley-Blackwell, 2002.

homogénea.

2. Para dispersores de tamaño considerable en comparación con la longitud de onda se denomina *dispersión Rayleigh*.
3. Para muy rápidas variaciones de velocidad y densidad lateral y vertical con registros sísmicos caóticos se denomina *dispersión de Mie*.

En el caso de estudio se modelaron los tres diferentes tipos de dispersión mencionados para obtener un conjunto de registros muy diversos, pero ninguno de estos tipos de dispersión es distinguido individualmente de los otros al momento de la detección.

2.1.2. Caracterización de la Dispersión Para caracterizar la dispersión se deben usar conceptos como la densidad del flujo de energía que se define como la cantidad de energía que fluye a través de una sección transversal perpendicular a la dirección de propagación por unidad de tiempo. Este proceso se imagina como la densidad del flujo de energía J_0 que interactúa con un objeto dispersor y genera una nueva densidad de flujo de energía J_1 .

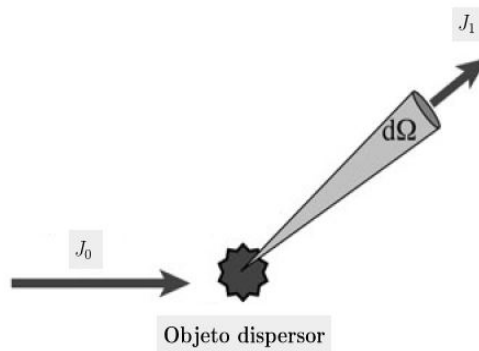


Figura 6. Diferencial de dispersión para un único objeto dispersor. Imagen adaptada de: [Sato et al., 2012].

La cantidad de la energía dispersada por unidad de tiempo viene dada por el elemento del

ángulo sólido, de la siguiente esta forma:

$$d\sigma = \left(\frac{J_1}{J_0} \right) r^2 d\Omega \quad (4)$$

Donde $r^2 d\Omega$ corresponde al elemento de superficie.

Para casos complejos, la dispersión puede considerarse como el efecto producido de la interacción de las ondas sísmicas a medida que se propagan sobre un medio aleatorio. Para modelar un medio aleatorio no homogéneo se suele emplear un medio de fondo homogéneo con velocidad de propagación V que están llenos de puntos dispersores distribuidos con densidad numérica n , como se ilustra esquemáticamente en la figura 7.

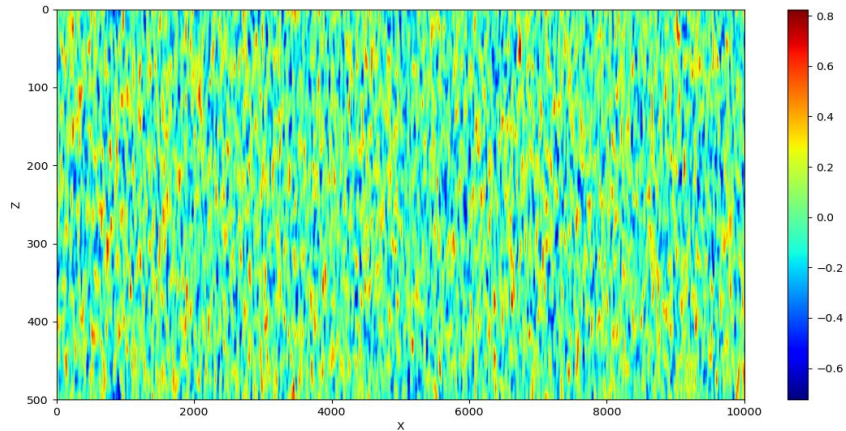


Figura 7. Medio aleatorio no homogéneo.

Precisar un medio aleatorio permite caracterizar la dispersión del medio definiendo su coeficiente de dispersión²²:

$$g = 4\pi n \frac{d\sigma}{d\Omega} \quad (5)$$

²² Keiiti Aki y Bernard Chouet. "Origin of coda waves: Source, attenuation, and scattering effects". En: *Journal of Geophysical Research* 80.23 (1975), págs. 3322-3342.

Para definir el coeficiente de dispersión total se hace un promedio del coeficiente de dispersión para todas las direcciones.

$$g_0 = \frac{1}{4\pi} \oint g d\Omega = n \oint d\sigma = n\sigma_0 \quad (6)$$

Siendo σ_0 la dispersión total de la región transversal. Esto quiere decir que el coeficiente de dispersión total el inverso del camino medio libre del modelo.

2.2. Modelado Numérico en el Dominio 2D P-SV

Las ondas sísmicas son un tipo de onda elástica en la que la energía se propaga a través o sobre una superficie de un medio sin la traslación de material. Para un medio homogéneo e isótropo la ecuación que modela la propagación de ondas elásticas es:

$$(\lambda + 2\mu)\nabla(\nabla \cdot \mathbf{u}) - \mu\nabla \times (\nabla \times \mathbf{u}) = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} + f \quad (7)$$

Donde \mathbf{u} es el vector de desplazamiento, λ y μ junto definen los parámetros de elasticidad del medio (mejor conocidos como parámetros de Lamé), ρ es la densidad y f es un término de fuente.

Las ondas sísmicas se dividen en dos grandes tipos dependiendo de su propagación: Ondas de cuerpo y ondas superficiales, las ondas de cuerpo o internas viajan a través del interior de la tierra y a su vez se clasifican por ondas P y ondas S, de carácter longitudinal y transversal respectivamente. Estas ondas se derivan matemáticamente de la ecuación (1) y se propagan en el medio homogéneo e isótropo con las siguientes velocidades:

$$V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad V_s = \sqrt{\frac{\mu}{\rho}} \quad (8)$$

Cuando las ondas llegan a la superficie se propagan sobre la interfaz de discontinuidad entre dos medios. Las ondas superficiales más estudiadas son las ondas Love y las ondas Rayleigh y ambas se propagan alrededor de un 90% de la velocidad de las ondas S. Existen más modos de propagación de ondas sísmicas, pero son de importancia relativamente menor a las mencionadas anteriormente.

Las ondas Rayleigh son producto de la interferencia entre las ondas de cuerpo, más precisamente las ondas P y la componente vertical de las ondas S a lo largo de una superficie libre. Estas ondas pueden ser simuladas en un dominio 2D resolviendo la ecuación de onda vectorial a través de diferentes métodos (por ejemplo: el método de elementos finitos FEM, el método de diferencias finitas DFM, o el método de elementos espectrales SEM).

El método de diferencias finitas es el escogido para realizar el modelado numérico, más específicamente el método de velocidad-esfuerzo²³. La propagación de onda 2D en la formulación está dada por:

$$\begin{aligned}
 \frac{\partial v_x}{\partial t} &= \frac{1}{\rho} \left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z} \right) + f_{v_x}, \\
 \frac{\partial v_z}{\partial t} &= \frac{1}{\rho} \left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z} \right) + f_{v_z}, \\
 \frac{\partial \sigma_{xx}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_x}{\partial x} + \lambda \frac{\partial v_z}{\partial z} + f_{\sigma_{xx}}, \\
 \frac{\partial \sigma_{zz}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_z}{\partial z} + \lambda \frac{\partial v_x}{\partial x} + f_{\sigma_{zz}}, \\
 \frac{\partial \sigma_{xz}}{\partial t} &= \mu \left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x} \right) + f_{\sigma_{xz}},
 \end{aligned} \tag{9}$$

La ecuación (9) se presenta en coordenadas cartesianas, además σ_{xx} , σ_{zz} y σ_{xz} representan las componentes de los esfuerzos.

²³ Jean Virieux. "P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method". En: *GEOPHYSICS* 51.4 (1986), págs. 889-901. DOI: 10.1190/1.1442147.

La discontinuidad que presenta la superficie terrestre presenta complicaciones a las soluciones numéricas de las ondas elásticas. La formulación de vacío computacional^{24,25} es una buena forma de satisfacer la condición límite de superficie libre de [Aki y Richards, 2002]²⁶. Para un modelo de la tierra libre de tracción, las componentes de los esfuerzos cortante y normal a la superficie son cero, por lo tanto, las ecuaciones se simplifican a:

$$\frac{\partial v_z}{\partial z} = -\frac{\lambda}{\lambda + 2\mu} \frac{\partial v_x}{\partial x}, \quad \frac{\partial v_x}{\partial z} = -\frac{\partial v_z}{\partial x}, \quad (10)$$

Para el modelo del subsuelo computacional, en los bordes izquierdo, derecho e inferior se adjunta una frontera absorbente (C-PML) para solucionar el truncamiento físico de un modelo finito. En el caso del borde superior, el concepto de formulación de vacío es muy atractivo principalmente por su simplicidad en la implementación numérica, sin embargo, existen ciertas dificultades para esta formulación, sobre todo en presencia de una topografía superior irregular, porque la condición límite libre de tracción puede llegar a tener problemas en este caso, es decir, que los esfuerzos normales y de corte pueden no llegar a tender a cero y esto provoca que el campo de onda simulado varíe en un porcentaje del esperado computacionalmente. Para la solución de este problema se han implementado variantes a la formulación de vacío²⁷ (llamada formulación de vacío mejorada).

²⁴ Jirí Zahradník, Peter Moczo y Frantisek Hron. “Testing four elastic finite-difference schemes for behavior at discontinuities”. En: *Bulletin of the Seismological Society of America* 83.1 (1993), 107–129. DOI: 10.1785/BSSA0830010107.

²⁵ Robert W. Graves. “Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences.” En: *Bulletin of the Seismological Society of America* 86.4 (2002), 1091–1106. DOI: 10.1785/BSSA0860041091.

²⁶ Keiiti Aki y Paul G. Richards. *QUANTITATIVE SEISMOLOGY*. 2.^a ed. University Science Books, 2002.

²⁷ Chong Zeng y col. “An improved vacuum formulation for 2D finite-difference modeling of Rayleigh waves including surface topography and internal discontinuities”. En: *GEOPHYSICS* 77.1 (2011). DOI: 10.1190/geo2011-0067.1.

2.2.1. Formulación de Vacío Mejorada La ecuación de onda elástica en un medio isótropo para el sistema de coordenadas cartesiano 2D se presenta en la ecuación (9). El modelo es discretizado empleando el esquema de cuadrícula escalonada de la figura 8 donde i y k son los índices de la cuadrícula para los ejes horizontal y vertical respectivamente.

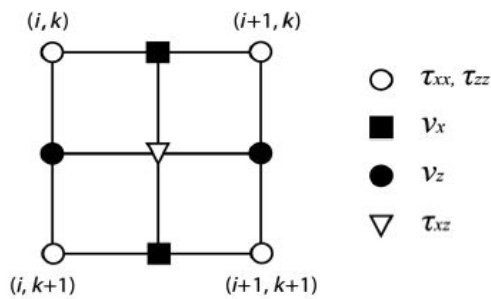


Figura 8. Esquema de cuadrícula escalonada utilizada para la formulación de vacío. Cada círculo es un nodo de la cuadrícula con sus respectivas posiciones marcadas por los índices. τ_{xx} y τ_{zz} son las componentes normales del tensor de esfuerzos y τ_{xz} el esfuerzo cortante en esa dirección, λ y μ son los coeficientes de Lamé, ρ es la densidad. Los cuadrados y círculos sombreados representan la velocidad de las partículas horizontal y vertical respectivamente. Imagen tomada de: [Zeng et al., 2011].

Los componentes de las velocidades de las partículas se encuentran ubicadas en los puntos medios de los nodos de la cuadrícula y las ondas de Rayleigh se pueden simular como parte de la solución a la ecuación de onda elástica 2D.

La posición intermedia de las velocidades de propagación V_x y V_z puede ser arbitrarias dado que las diferencias en sus respectivas simulaciones suele insignificantes²⁸. Usar un

²⁸ Peter Moczo y col. "3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities". En: *Bulletin of the Seismological Society of America* 92.8 (2002), págs. 3042-3066. DOI: 10.1785/0120010167.

promediado de parámetros es fundamental para la estabilidad de la simulación^{29,30}, el promediado de parámetros se presenta en la siguiente ecuación:

$$\begin{aligned}
 B_x &= \begin{cases} 0 & \text{if } \rho_{i,k} = 0 \text{ and } \rho_{i+1,k} = 0; \\ \frac{2}{\rho_{i,k} + \rho_{i+1,k}} & \text{de lo contrario;} \end{cases} \\
 B_z &= \begin{cases} 0 & \text{if } \rho_{i,k} = 0 \text{ and } \rho_{i,k+1} = 0; \\ \frac{2}{\rho_{i,k} + \rho_{i,k+1}} & \text{de lo contrario;} \end{cases} \\
 \mu_{xz} &= \begin{cases} 0 & \text{de lo contrario;} \\ 4 \left(\frac{1}{\mu_{i,k}} + \frac{1}{\mu_{i+1,k}} + \frac{1}{\mu_{i,k+1}} + \frac{1}{\mu_{i+1,k+1}} \right) & \text{if } \mu_{i,k} = \mu_{i+1,k} = \mu_{i,k+1} = \mu_{i+1,k+1} \neq 0; \end{cases}
 \end{aligned} \tag{11}$$

2.2.2. Consideración de una Capa Ficticia El modelo numérico cuenta con una capa superficial donde todos los parámetros elásticos y cantidades físicas serán cero durante el modelado, esta región se conoce como el vacío numérico. Bajo esta capa superior se encuentra una capa considerada como la capa ficticia, donde se encuentra el promedio de los parámetros del modelado y se definen algunos parámetros como cero, por ejemplo, μ para cumplir con las condiciones de superficie libre³¹. Seguido de la capa ficticia se encuentra el modelo original topográfico quien se encuentra desplazado una cuadrícula hacia arriba de su posición original para generar la capa ficticia.

²⁹ Moczo y col., ver n. 28.

³⁰ Rune Mittet. "Free-surface boundary conditions for elastic staggeredgrid modeling schemes". En: *GEOPHYSICS* 67 (2002). DOI: 10.1190/1.1512752.

³¹ Zeng y col., ver n. 27.

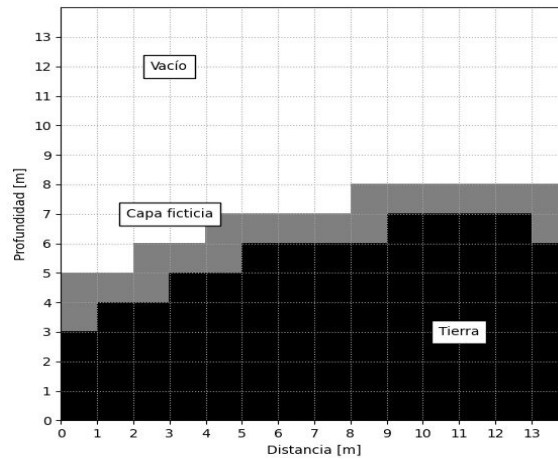


Figura 9. Distribución de cuadrícula en la formulación de vacío mejorada en presencia de topografía superficial.

Esta formulación mejorada no requiere cambios en el programa de implementación por que la configuración de la nueva capa se genera naturalmente en la región desplazada gracias a la técnica de promedio de parámetros.

2.3. Código de Modelado

El grupo de investigación CEMOS adscrito a la Universidad Industrial de Santander cuenta con un código de modelado de propagación de ondas elásticas en diferencias finitas desarrollado en MATLAB que se puede implementado para fines de este trabajo, sin embargo, se desarrollará nuevamente un código de modelado, pero en este caso usando el lenguaje de programación Python con el fin de entregar un código de modelado en fuente abierta.

El código de modelado funciona bajo el diagrama de flujo de la figura 10. El primer paso para obtener los datos sísmicos sintéticos es leer el modelo de velocidad y densidad que se empleará, luego se ajustan los parámetros que se usarán en él, como las característi-

cas de la fuente, la ubicación de los receptores, las distribuciones espaciales y temporales, etc. Se debe definir la zona C-PML para después realizar el promediado de parámetros sobre la cuadrícula escalonada. Finalmente se realizan las actualizaciones de los campos de velocidades y esfuerzos con la inyección de la fuente de acuerdo con el paso temporal escogido. Una vez terminado el bucle temporal se guardan los registros de llegada de las ondas sobre los receptores quienes serán los datos sintéticos del modelo.

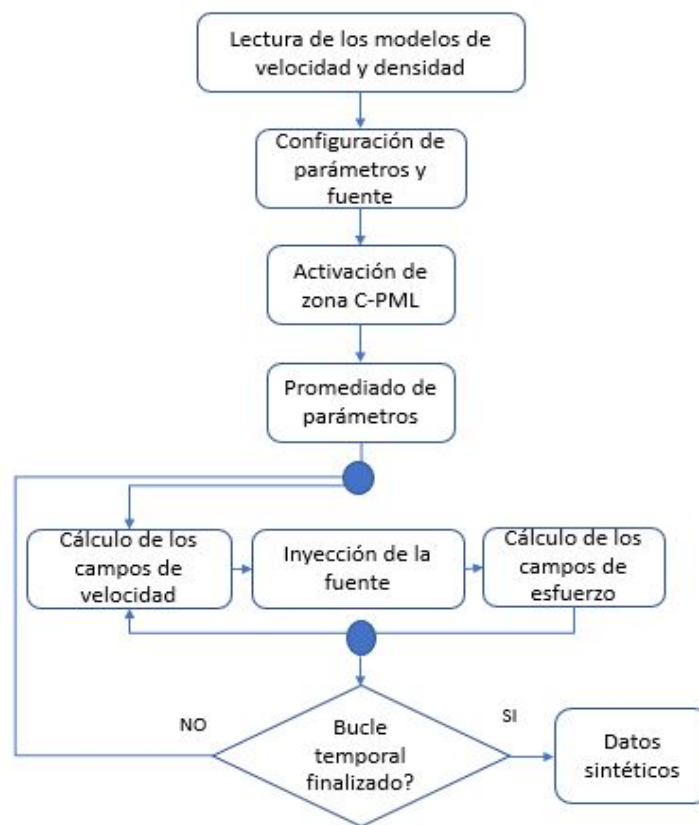


Figura 10. Representación del flujo de trabajo del código de modelado.

2.4. Modelos Elásticos

La mayor parte de la energía sísmica se dispersa en las capas superficiales de la tierra por la presencia de las heterogeneidades cerca de la superficie (por ejemplo, escarpaduras,

lechos de ríos secos o relieves kársticos). Cuando se presentan estas irregularidades, la adquisición de datos sísmicos se ve afectada por las ondas dispersadas³², de aquí la importancia de modelar diferentes tipos de heterogeneidad para detectar el ruido de dispersión y mejorar el procesamiento de datos sísmicos.

Para el modelado numérico se usó un sistema de coordenadas cartesianas 2D, con el eje x positivo apuntando hacia la derecha y el eje z positivo apuntando hacia abajo. Las ecuaciones que describen la propagación de las ondas elásticas bajo ciertas condiciones se presentan en la ecuación (10) [Virieux,1986]³³.

Para comprobar que las simulaciones funcionan de manera correcta, se decidió emular dos tipos de modelos diferentes construidos en el artículo: “*Numerical modeling of elastic-wave scattering by near-surface heterogeneities*”³⁴ y realizar una breve comparación de los resultados.

2.4.1. Modelo terrestre de única capa sobre un semiespacio Para estudiar los efectos de las heterogeneidades sobre el campo de onda se ubicaron dos objetos dispersores circulares sobre la capa poco profunda localizados en $(x,z)=(360 \text{ m}, 15 \text{ m})$ y $(x,z)=(720 \text{ m}, 15 \text{ m})$. Cada uno con 20 m de diámetro y un contraste de impedancia de 0.36. La tabla (1) muestra las propiedades de los materiales para el modelo y sus respectivos contrastes de impedancia.

³² J. Cl. de Bremaecker. “Transmission and reflection of rayleigh waves at corners”. En: *GEOPHYSICS* 23.2 (1956). DOI: 10.1190/1.1438465.

³³ Virieux, ver n. 23.

³⁴ M. Abdulaziz y Nafi Toksöz. “Numerical modeling of elastic-wave scattering by near-surface heterogeneities”. En: *GEOPHYSICS* 79.4 (2014). DOI: 10.1190/geo2013-0208.1.

	V_p [m/s]	V_s [m/s]	Densidad [Kg/m ³]	Contraste de Impedancia
Capa I	1800	1000	1750	-
Capa II	3000	1500	2250	0.36
Dispersor a	3000	1500	2250	0.36
Dispersor b	3000	1500	2250	0.36

Tabla 1. Propiedades de los materiales para el modelo y los contrastes de impedancia relativos a la capa I.

El modelo de la figura 11 cuenta con un dominio de $N_x = 1001$ y $N_z = 501$ puntos sobre la cuadrícula. El espaciamiento entre los puntos de la cuadrícula es uniforme en todo el dominio y un ancho de 1 metro, es decir, el modelo tiene una distancia horizontal de 1000 metros sobre el eje x y una distancia vertical de 500 metros sobre el eje z. El paso temporal utilizado fue de 0.2 ms y la fuente puntual es de tipo Ricker con frecuencia 30 Hz localizada en $(x,z)=(150 \text{ m}, 15 \text{ m})$. Los receptores fueron ubicados en superficie con 5 m de separación entre uno y otro.

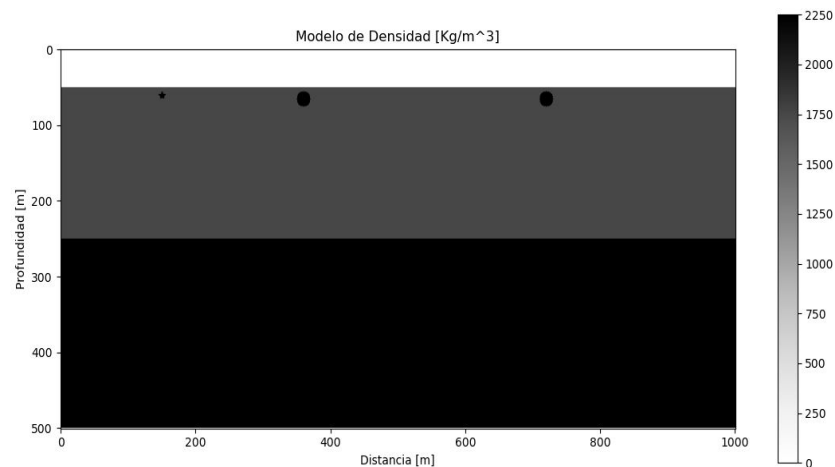


Figura 11. Modelo terrestre sintético. Modelo con una capa simple en un semiespacio con dos objetos dispersores circulares $(x,z)=(360 \text{ m}, 15 \text{ m})$ y $(x,z)=(720 \text{ m}, 15 \text{ m})$ respectivamente. Ubicación de la fuente: $(x,z)=(150 \text{ m}, 15 \text{ m})$.

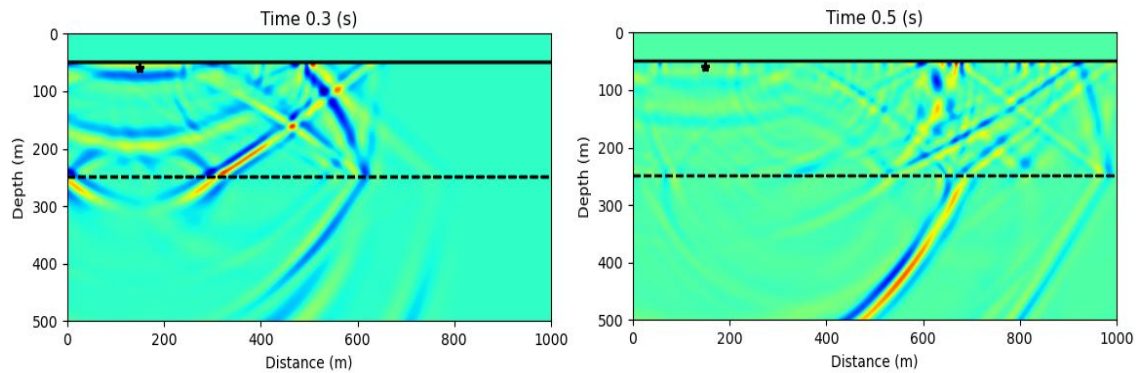


Figura 12. Registro de la propagación del campo de onda total en un tiempo de 0.3 segundos (parte izquierda) y 0.5 segundos (parte derecha). Ubicación de la fuente: $(x,z)=(150\text{ m}, 15\text{ m})$. El interfaz entre los dos medios se encuentra marcado por la línea punteada a 200 m de la superficie.

La propagación del campo de onda se observa en la figura 12.

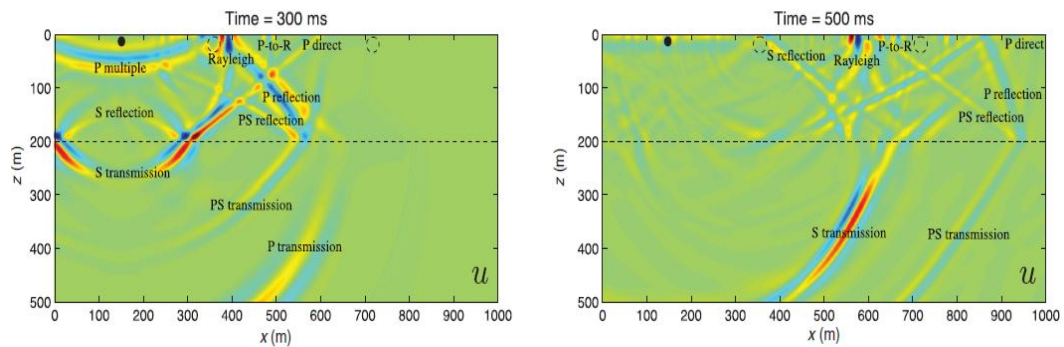


Figura 13. Registro de la propagación del campo de onda total en un tiempo de 0.3 segundos (parte izquierda) y 0.5 segundos (parte derecha). Tomado de: [Abdulaziz et al., 2014]

Comparando estos resultados con los obtenidos para la propagación del campo de onda en el mismo modelo (ver figura 13) obtenidos en el artículo: “*Numerical modeling of elastic-wave scattering by near-surface heterogeneities*”³⁵ se puede concluir que el código

³⁵ Abdulaziz y Toksöz, ver n. 34.

de modelado numérico desarrollado se encuentra bien ejecutado por que los resultados de la simulación son iguales para el mismo instante de tiempo.

La figura 14 muestra los efectos de la dispersión sobre los registros sísmicos sintéticos 2D debido a las heterogeneidades cerca de la superficie para el modelo de la figura 11.

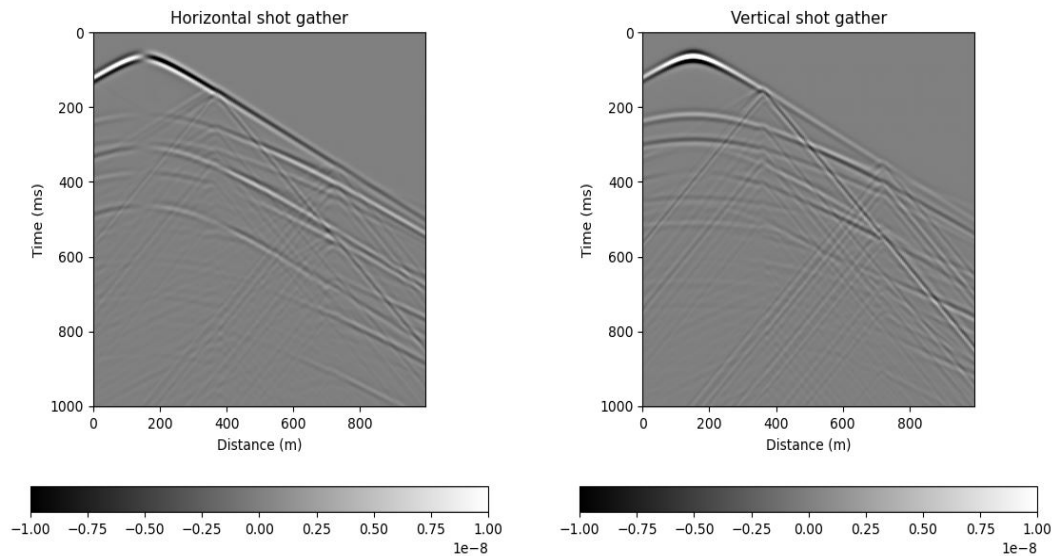


Figura 14. Registros de disparo obtenidos de la simulación del campo de onda en diferencias finitas de la figura 12. Componente horizontal (V_x) y componente vertical (V_z).

2.4.2. Modelo terrestre triple capa con topografía irregular. Para comprobar el funcionamiento del código de modelado bajo una topografía irregular se empleó una topografía de tipo Bedrock en un modelo de tres capas y se observó los efectos producidos. El modelo se presenta en la figura 15 donde topografía irregular fue modelada con un conjunto de números aleatorios no-correlacionados usando una distribución gaussiana con media de cero y una desviación estándar de 20 metros. Los números aleatorios usados en la topografía fueron correlacionados usando un filtro promedio gaussiano de 5 metros de correlación (ver figura 16). Cabe destacar que este modelo fue una reproducción de un

modelo sintético con topografía irregular presentado en [Abdulaziz et al., 2014]³⁶.

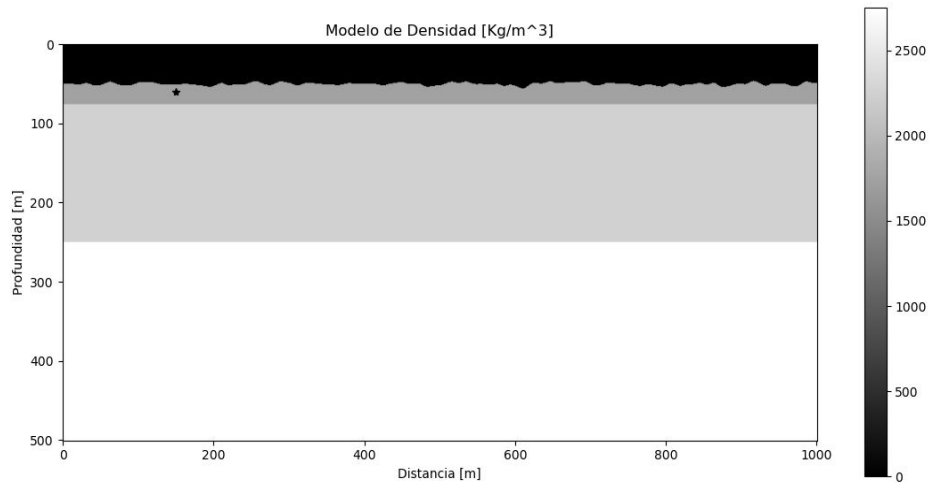


Figura 15. Modelo terrestre sintético con tres capas y topografía irregular gaussiana. Ubicación de la fuente: $(x,y)=(150\text{ m},15\text{ m})$

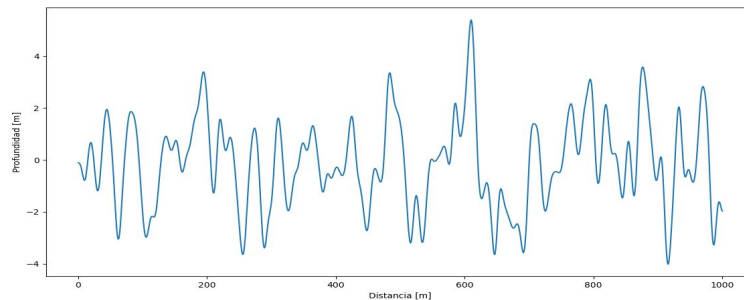


Figura 16. Perfil gaussiano utilizado para modelar la topografía del modelo de la figura 15

La fuente empleada fue una Ricker con frecuencia central de 10 Hz y los receptores ubicados a 10 metros uno del otro en superficie. La tabla 2 presenta la información de las propiedades del modelo con topografía irregular.

³⁶ Abdulaziz y Toksöz, ver n. 34.

	V_p [m/s]	V_s [m/s]	Densidad [Kg/m ³]
Capa I	1800	1000	1750
Capa II	3000	1500	2250
Capa III	5000	2250	2750

Tabla 2. Propiedades de los materiales para el modelo de la figura 15.

La siguiente figura muestra la propagación del campo de onda durante 1 segundo sobre el modelo con topografía irregular.

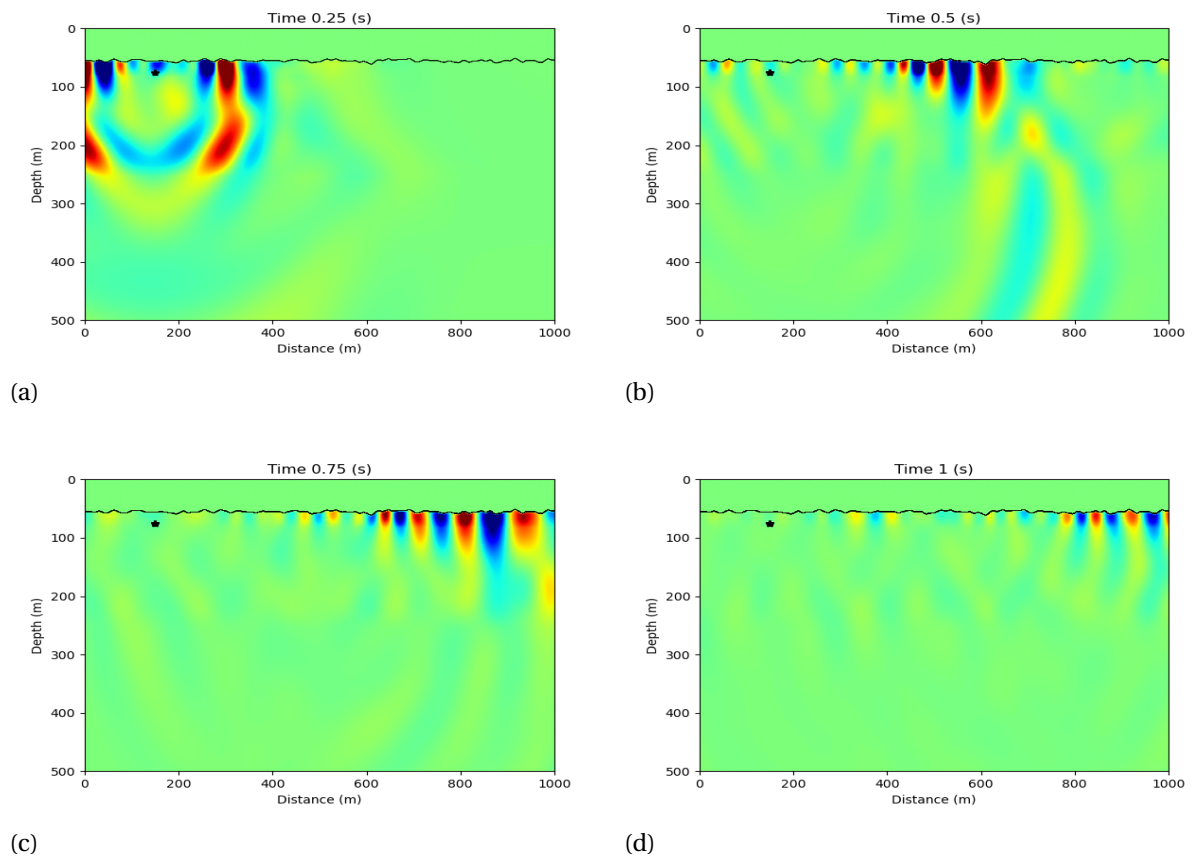


Figura 17. Propagación del campo de onda durante un segundo en un modelo con topografía irregular. Fuente ubicada a 15 metros de profundidad. (a) Propagación del campo de onda a 0.25 [s]. (b) Propagación del campo de onda a 0.5 [s]. (c) Propagación del campo de onda a 0.75 [s]. (d) Propagación del campo de onda a 1 [s].

La figura 18 presenta el registro del disparo.

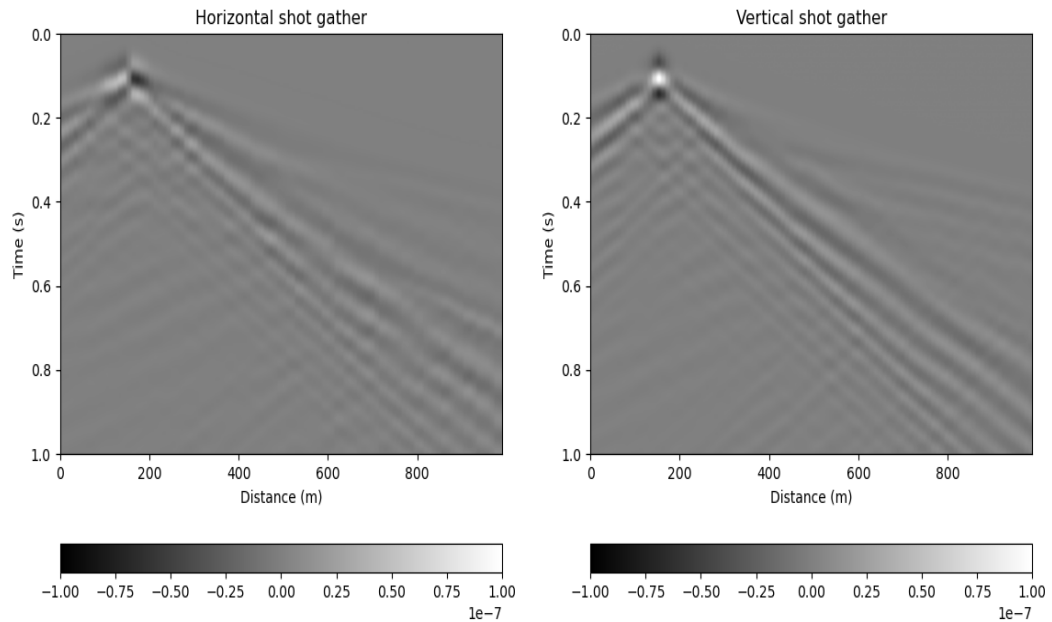


Figura 18. Registros de disparo obtenidos para el modelo con topografía irregular.

2.5. Modelo TENERIFE

El modelo TENERIFE se basa en los datos sísmicos 3D-3C recogidos en el campo Tenerife ubicado en la cuenca del Valle Medio del Magdalena entre la cordillera oriental y central y entre las fallas geológicas de Bucaramanga-Santa Marta y Cambao³⁷.

Los datos sísmicos en el campo Tenerife fueron adquiridos por el Instituto Colombiano del Petróleo en 2003.

³⁷ D. Ospina, W. M. Agudelo y S. Khurama. "Modelo de estrato somero para el campo Tenerife a partir de la integración de datos de refracción sísmica e integración geológica". En: *Universidad Industrial de Santander* (2011).

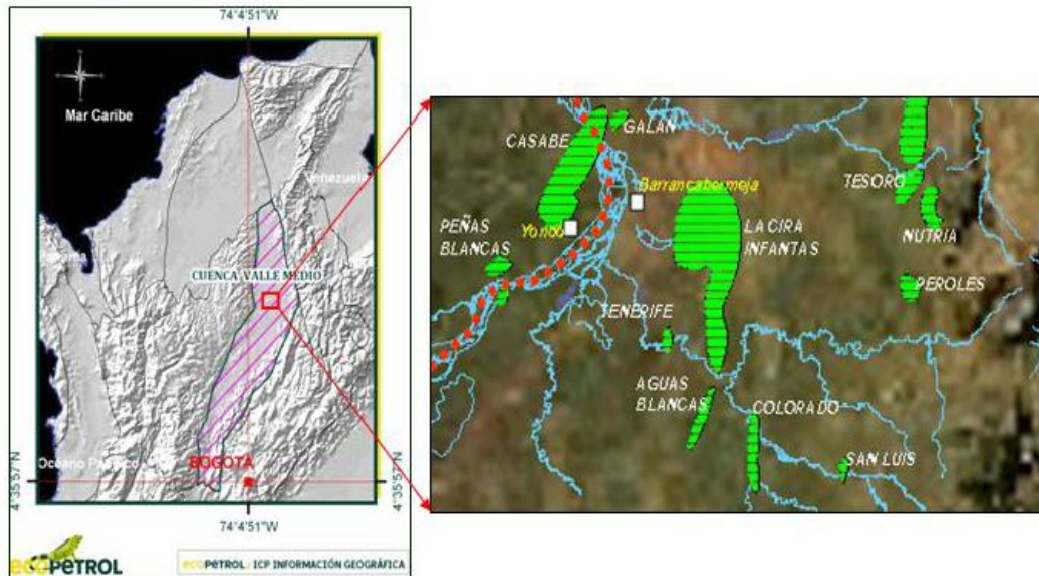


Figura 19. Ubicación geográfica campo Tenerife (zona de adquisición). Tomado de: [Ospina et al., 2011]

Una sección 2D de la inversión de una línea sísmica de registros del campo Tenerife se presenta en la figura 20. Esta sección se denominó *modelo Tenerife* y servirá como punto de partida para la construcción del modelo base o modelo background del conjunto de modelos necesarios para la realización de este trabajo.

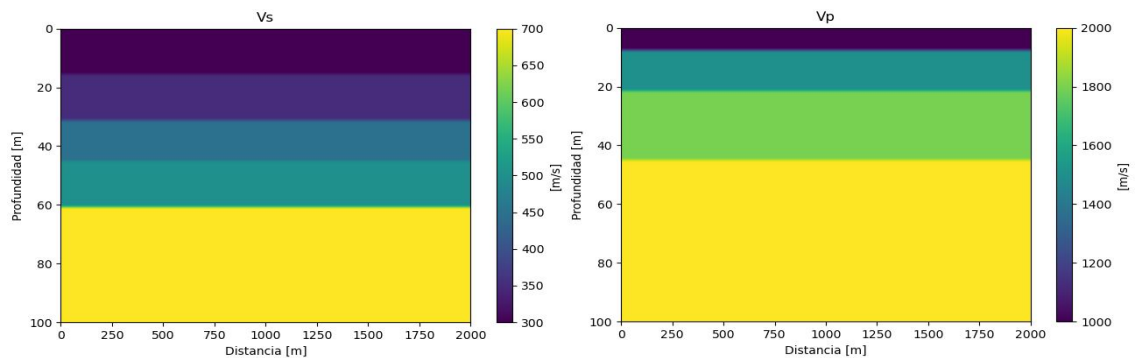


Figura 20. Sección 2D de línea sísmica campo Tenerife.

La sección de la figura 20 muestra que en el campo Tenerife la capa más superficial tie-

ne un alto contraste en la velocidad de propagación respecto a las capas siguientes, esta información es coherente con la geología local de la zona, donde se presentan depósitos aluviales compuesto por arenas finas, limas y materiales poco consolidados.³⁸

2.6. Medio Aleatorio

Las regiones cerca de la superficie tienen la característica de ser muy heterogéneas y esto es fácilmente probado con registros de pozo. Para describir estas zonas computacionalmente es conveniente usar modelos estadísticos de aleatoriedad, como por ejemplo las funciones de densidad espectral de potencia (PSDF por sus siglas en inglés).

Para un medio aleatorio, la velocidad de propagación de las ondas no es constante, es decir, el campo de velocidad en un punto x puede ser visto como la suma entre una velocidad de fondo (velocidad media) y una variación de esta velocidad en el mismo punto.

$$\mathbf{V}(x) = V_0 + \delta\mathbf{V}(x) \quad (12)$$

La variación de la velocidad de la onda puede considerarse como una fluctuación fraccional de la velocidad media de todo el espacio, es decir, $\delta\mathbf{V}(x) = \xi(x)V_0$. La distribución de cada una de las fluctuaciones fraccionales de velocidad en el medio puede modelarse como una función aleatoria con media cero.

³⁸ J. M. Royero y J. Clavijo., *Mapa geológico generalizado del departamento de Santander* (2001).

Para la realización del medio aleatorio en este trabajo, se considerará una aleatoriedad estacionaria o homogénea, lo que sugiere que el producto entre dos fluctuaciones sobre el mismo espacio únicamente va a depender de una sola variable³⁹, en este caso, se puede definir una función de autocorrelación como:

$$\mathbf{R}(x) = \langle \xi(y)\xi(y+x) \rangle \quad (13)$$

La distancia de correlación a es definida como $\mathbf{R}(a)=\varepsilon^2/e$, donde ε^2 es la media cuadrada de la fluctuación fraccional ($\mathbf{R}(0)$).

Ahora bien, la transformada de Fourier de la función de autocorrelación es la función de densidad espectral de potencia:

$$\mathbf{P}(m) = \tilde{\mathbf{R}}(x) = \int \int \int_{-\infty}^{\infty} \mathbf{R}(x) e^{-imx} dx \quad (14)$$

El medio aleatorio presentado en la figura 7 ha pasado por el proceso de construcción matemática. En la generación del medio aleatorio numérico se programaron tres tipos distintos de función de autocorrelación: *Gaussiana*, *exponencial* y *Von Karman*, sin embargo, en el presente trabajo se escogió trabajar con la función Gaussiana como la generadora del medio aleatorio.

³⁹ Sato, Fehler y Maeda, ver n. 14.

2.7. Generación de Datos Sintéticos Multicomponente

Para entrenar un algoritmo de clasificación se necesita un conjunto completo de modelos sísmicos con características variables para detectar el ruido de dispersión automáticamente sobre cualquier registro sísmico. Cabe aclarar que el proyecto cuenta con datos de campo reales adquiridos en el campo Tenerife por el Instituto Colombiano del Petróleo, por esta razón, es tan importante construir el conjunto de datos basado en el *modelo TENERIFE* para aplicar este método sobre los datos reales.

El número total de modelos sintéticos construidos es 100, cada uno con características diferentes a los demás. Los modelos sintéticos fueron creados a partir de 10 variaciones del modelo base, entre las que destacan cambios aleatorios de velocidad y densidad y cambios de profundidad de las capas del modelo. Un ejemplo de las variaciones del modelo base se presenta en la figura 21.

Cada cambio al modelo base representa cambios en los registros luego de ser simulada la propagación. Por ejemplo, se espera que si la velocidad de propagación de la primera capa no es tan baja respecto a las otras no ocurra una guía de onda sobre ella y esto se ve reflejado en el registro (c) de la figura 22.

Cabe aclarar que las variaciones sobre el modelo base no presentan objetos dispersores, ni topografía e interfaces irregulares, ni mucho menos medios aleatorios que produzcan dispersión. Por el contrario, los estratos son horizontales, entonces al detectar automáticamente el ruido de dispersión sobre las 10 variaciones del modelo base se espera que la relación señal-ruido obtenida sobre estos registros sea nula.

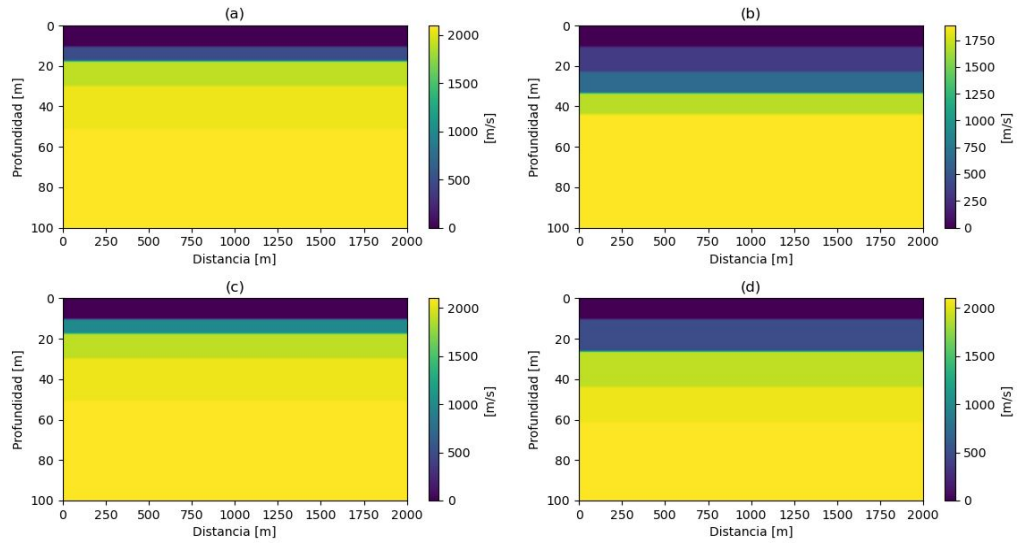


Figura 21. Variaciones del modelo base para el modelo de velocidad de onda P. (a) Modelo Tenerife. (b) Cambio aleatorio en la magnitud de la velocidad por capas (cambio entre 0-1). (c) Primera capa con tan contrastada. (d) Capas con profundidad uniforme.

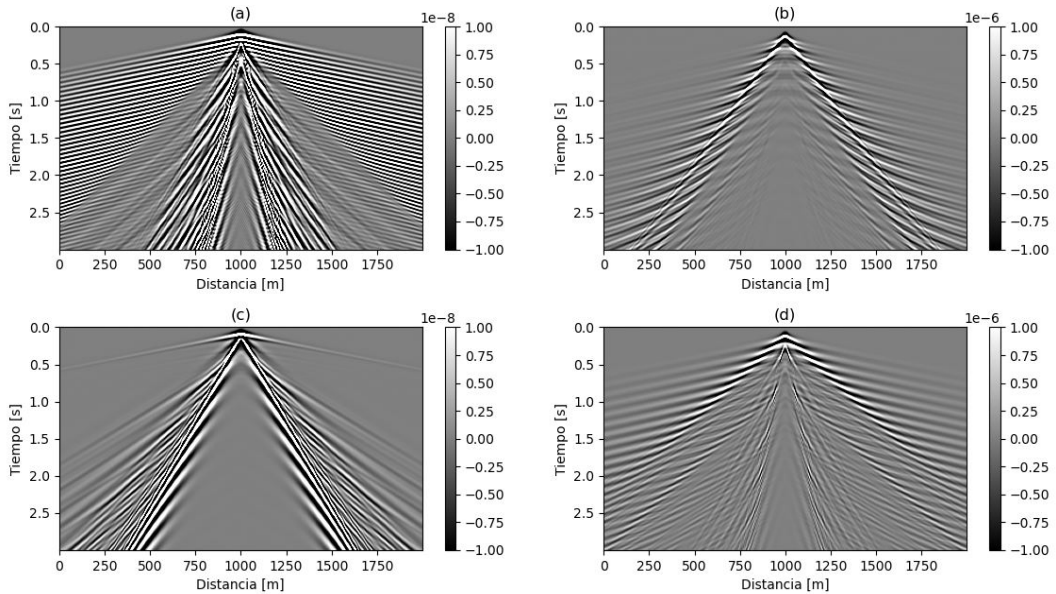


Figura 22. Variaciones de los registros verticales para cambios del modelo base. (a) Registro vertical del modelo Tenerife. (b) Registro vertical de modelo con cambios aleatorios en la magnitud de la velocidad por capas (cambio entre 0-1). (c) Registro vertical del modelo con la primera capa con tan contrastada. (d) Registro vertical del modelo con capas con profundidad uniforme.

Basado en la geología del campo Tenerife presentada en la sección 2.5, se considera un medio aleatorio como la mejor opción para modelar el ruido de dispersión de la zona.

Al ubicar un medio aleatorio sobre los modelos sin dispersión y controlando los parámetros como la fluctuación fraccional, el tamaño de las muestras, la ubicación y el ancho del medio aleatorio sobre el espacio, se construyeron 90 modelos generadores de ruido de dispersión en los registros sintéticos.

La figura 23 muestra cuatro ejemplos de la ubicación de diferentes medios aleatorios sobre las variaciones del modelo base.

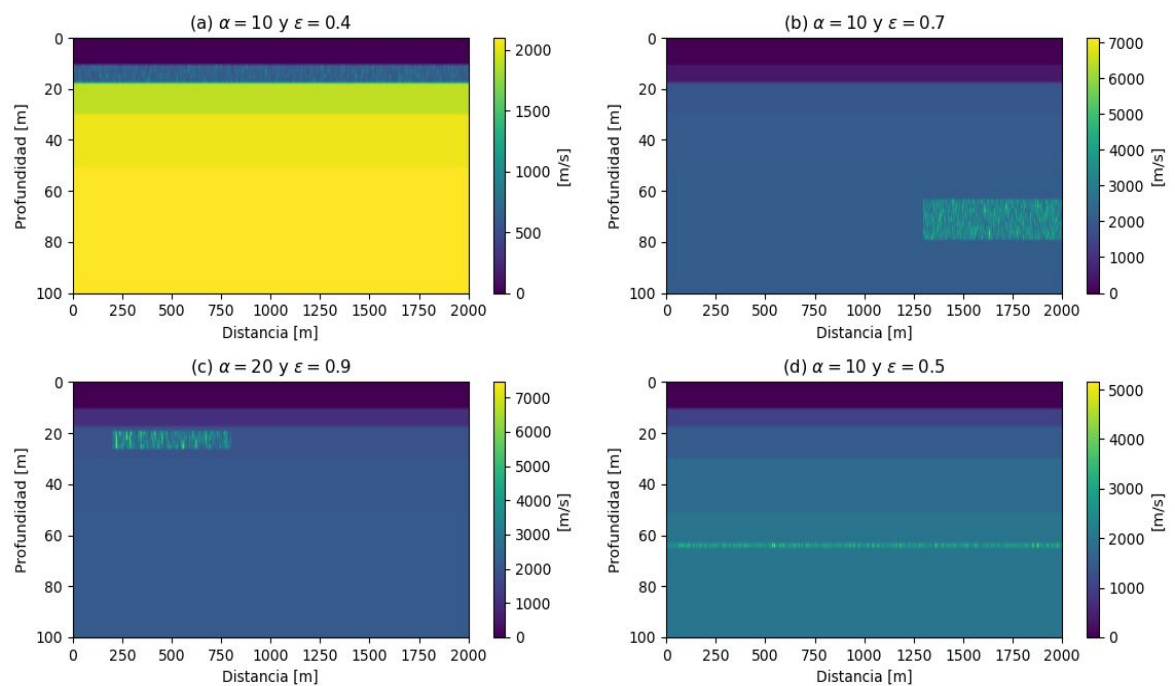


Figura 23. Medios aleatorios sobre las variaciones del modelo base. (a) Medio aleatorio del ancho del modelo ubicado sobre la capa más superficial. (b) Medio aleatorio de profundidad 20 [m] ubicado en la parte inferior derecha del modelo. (c) Medio aleatorio de profundidad 8 [m] ubicado en la parte superior izquierda el modelo. (d) Medio aleatorio de profundidad 2 [m] ubicado en la parte media-inferior del modelo y de ancho completo.

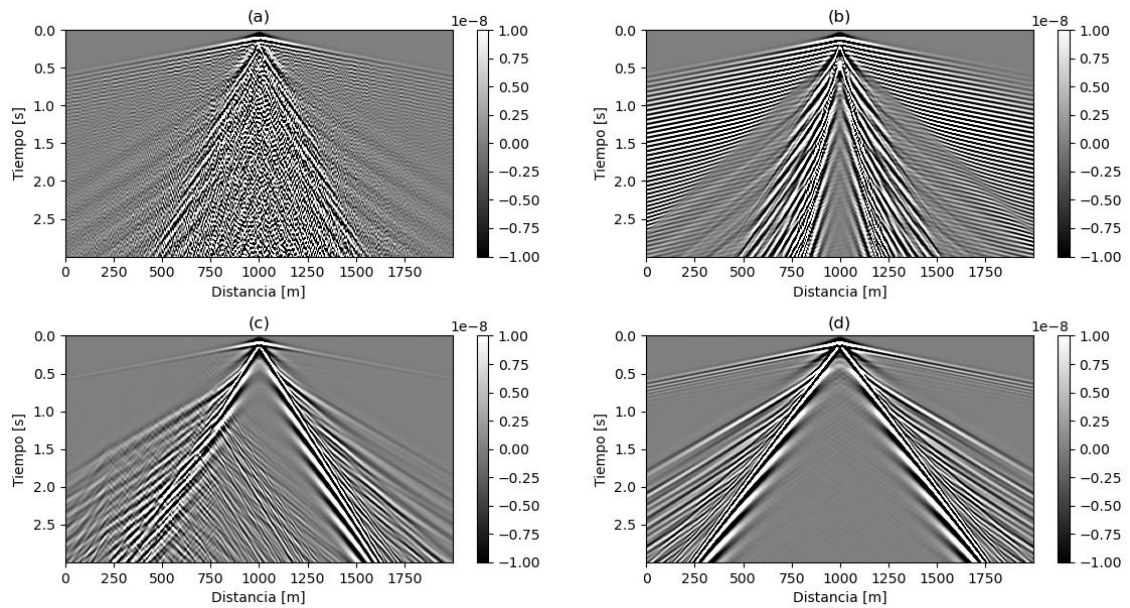


Figura 24. Registros verticales de los modelos con medios aleatorios de la figura 23.

La figura 24 muestra el efecto del medio aleatorio sobre los registros verticales para los modelos de la figura 23.

Los efectos sobre los registros son visibles dependiendo de las características del medio aleatorio. En ocasiones es difícil visualizar a simple vista los efectos de los medios aleatorios sobre los registros sísmicos (figura 23 (d)) ya sea porque el medio aleatorio se encuentra en una posición muy profunda dentro del modelo o porque el medio aleatorio tiene fluctuaciones fraccionales muy bajas. También es posible tener el caso donde los efectos son muy notorios y el ruido de dispersión se enmascara con las reflexiones de las ondas de cuerpo dificultando su procesamiento (figura 23 (d)). Por esta razón, detectar automáticamente el ruido de dispersión en registros sísmicos es muy útil, dado que ayudaría en el procesamiento de señales, se conocería cuales trazas sísmicas presentan el ruido para intentar atenuarlo y con ayuda de campañas sísmicas contribuiría al mejoramiento de las adquisiciones de los datos.

En total fueron 100 modelos sintéticos basados en el *modelo Tenerife*. Todos los modelos cuentan con las mismas características de simulación, es decir, la posición de la fuente (1000 [m], 15[m]), receptores sísmicos ubicados en superficie a 6 [m] uno del otro, ancho: 2 [km] y profundidad: 100 [m], tiempo de muestreo: 2×10^{-3} [s] durante 3 segundos.

2.8. Repositorio

El código de modelado numérico 2D de ondas sísmicas dispersadas elaborado con diferencias finitas se encuentra en el repositorio: <https://github.com/HolgerQuintero/numerical-modeling-of-seismic-wave-scattered->.

3.1. Cálculo de Atributos Sísmicos

Un atributo sísmico es una medida cuantitativa de una característica sísmica de interés⁴⁰. Anteriormente los atributos sísmicos eran utilizados en la construcción de mapas para mejorar interpretación de datos sísmicos. Los atributos sísmicos también han ganado importancia dado que se han convertido en una herramienta clave para la predicción litológica y la caracterización de los yacimientos. En los últimos años, aplicando algoritmos de aprendizaje automático se ha podido establecer relaciones entre los atributos y aspectos físicos y geológicos del subsuelo.

Diversos métodos de extracción se utilizan para calcular los atributos de los datos sísmicos con el fin de categorizar cada señal. El grupo de investigación CEMOS cuenta con un código que calcula atributos sísmicos usando MATLAB y permite extraer un total de 45 diferentes tipos atributos sísmicos sobre datos multicomponente.

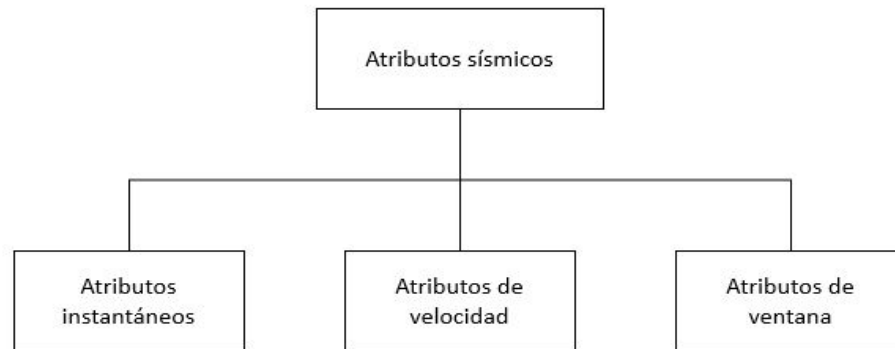


Figura 25. Clasificación de los atributos sísmicos calculados.

⁴⁰ Satinder Chopra y Kurt J. Marfurt. "Seismic attributes - A historical perspective". En: *GEOPHYSICS* 70.5 (2005).

Para este trabajo se extraerán un total de 19 atributos sísmicos para datos multicomponente 2C. A continuación, se presentan los atributos sísmicos calculados quienes se dividen en tres grupos y se presentan en el diagrama de la figura 25.

3.1.1. Atributos instantáneos

Atributos instantáneos Para el cálculo de atributos instantáneos es necesario emplear el análisis de trazas complejas, en este caso, se utilizan trazas multicomponente con una parte real dada por la señal registrada y una parte imaginaria derivada de la transformada de Hilbert de la componente real correspondiente.

$$c(t, x_0) = d(t) + i\hat{d}(t) \quad (15)$$

El análisis de traza compleja es hecho para la traza x_0 , donde $d(t)$ es el valor real de la traza sísmica y $\hat{d}(t)$ es la transformada de Hilbert de la traza real. La transformada de Hilbert es definida como:

$$\hat{d}(t) = \mathcal{H}\{d\}(t) = \frac{1}{i\pi} \int \frac{d(s)}{s-t} ds \quad (16)$$

Los atributos instantáneos se caracterizan por su dependencia temporal, es decir, son calculados traza por traza, por lo que no dependen de su ubicación espacial y la señal utilizada para calcular la traza compleja solo depende del tiempo.

Una vez definida la traza compleja se procede a calcular el valor de su amplitud y su fase instantánea.

$$A(t) = |c(t, x_0)| = [(d(t))^2 + (\hat{d}(t))^2]^{1/2}, \quad (17)$$

y

$$\varphi(t) = \arctan \left[\frac{\hat{d}(t)}{d(t)} \right]. \quad (18)$$

Naturalmente, la traza compleja puede ser escrita en su forma polar, es decir, en términos de su amplitud y fase instantánea, como muestra la ecuación (19)

$$c(t, x_0) = A(t)e^{i\varphi(t)} \quad (19)$$

Para el caso de datos multicomponente, el proceso de la definición de la traza compleja se debe realizar para cada una de las componentes. En el caso de estudio se cuentan con dos componentes denominadas vertical (V) y horizontal (H), por lo que la traza compleja puede escribirse vectorialmente de la siguiente forma:

$$\mathbf{c}(t, x_0) = \begin{pmatrix} c_H(t, x_0) \\ c_V(t, x_0) \end{pmatrix} = \begin{pmatrix} d_H(t) + i\hat{d}_H(t) \\ d_V(t) + i\hat{d}_V(t) \end{pmatrix} \quad (20)$$

Donde c_H y c_V son las trazas complejas para la componente horizontal y vertical respectivamente.

El análisis de traza compleja también nos permite calcular la frecuencia instantánea como la derivada de la fase instantánea respecto al tiempo, como lo muestra la ecuación (21)

$$f(t) = \frac{d\varphi(t)}{dt} \quad (21)$$

Hasta el momento, los atributos sísmicos instantáneos calculados (amplitud, fase y frecuencia) parece que pueden obtenerse del análisis de Fourier de las señales, sin embargo, esto ha sido discutido en⁴¹, mostrando que para señales no estacionarias este análisis puede resultar computacionalmente más costoso respecto al análisis de traza compleja.

Un atributo sísmico que solamente puede ser extraído de trazas multicomponente es la diferencia de fase instantánea, dado que se necesitan al menos dos componentes ortogonales de las señales. En este caso:

$$\Delta\varphi_{V-H}(t) = \varphi_V(t) - \varphi_H(t) \quad (22)$$

⁴¹ J. H. Bodine. "Waveform analysis with seismic attributes". En: *SEG Technical Program Expanded Abstracts* (1984). DOI: 10.1190/1.1894172.

Este atributo fue presentado por primera vez en [René et al., 1986]⁴², donde la diferencia de fase se determinó utilizando su equivalente ecuación (23) para evitar correcciones de fase.

$$\Delta\varphi_{V-H}(t) = \arctan \left[\frac{d_H(t)\hat{d}_V(t) - d_V(t)\hat{d}_H(t)}{d_V(t)d_H(t) - \hat{d}_V(t)\hat{d}_H(t)} \right] \quad (23)$$

Los atributos instantáneos fueron calculados utilizando los métodos de extracción propuestos por [René et al., 1986]⁴³ y [Morozov y Smithson, 1996]⁴⁴. Para la frecuencia y fase instantánea se presenta inestabilidad donde los valores absolutos de los extremos relativos de las trazas real y compleja son pequeños en relación con los valores absolutos de otros extremos relativos cercanos.⁴⁵

Atributos de polarización En [René et al., 1986]⁴⁶ se muestra la relación existente entre la polarización elíptica y la diferencia de fase para una señal armónica. Indicando que los ejes positivos representan los movimientos de las partículas hacia arriba y alejándose de la fuente. Esto indica que para una onda que se aleja de la fuente con un movimiento

⁴² R. M. René y col. "Multicomponent seismic studies using complex trace analysis". En: *GEOPHYSICS* 51.6 (1986). DOI: 10.1190/1.1442177.

⁴³ René y col., ver n. 42.

⁴⁴ Igor B. Morozov y Scott B. Smithson. "Instantaneous polarization attributes and directional filtering". En: *GEOPHYSICS* 61.3 (1996). DOI: 10.1190/1.1444012.

⁴⁵ Bodine, ver n. 41.

⁴⁶ René y col., ver n. 42.

en sentido de las manecillas del reloj, entonces los valores negativos y positivos de la diferencia de fase representan la polarización prógrada y retrógrada, respectivamente.

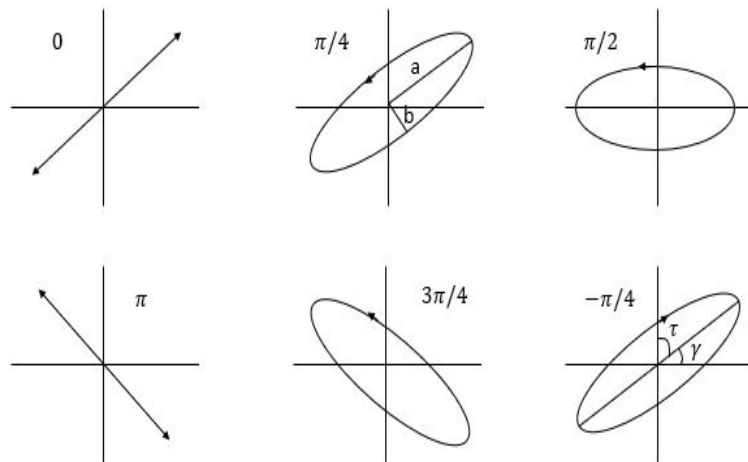


Figura 26. Elipses de polarización para una señal armónica con diferencias de fase de 0 , $\pi/4$, $\pi/2$, $3\pi/4$, π y $-\pi/4$. Imagen adaptada de: [René et al., 1986]⁴⁷.

La diferencia de fase con valores de 0 y π corresponden a una polarización lineal con una inclinación hacia adelante y hacia atrás, respectivamente.

La figura 26 muestra la definición gráfica de los atributos de polarización instantánea: longitud de semieje mayor (a), longitud desemieje menor (b), ángulo de inclinación (τ) y el ángulo de levantamiento (Υ).

Del libro *Principles of Optics*⁴⁸ bajo ciertas modificaciones, son obtenidos los parámetros de Stokes S_0 , S_1 y S_2 empleando las amplitudes y la diferencia de fase de las trazas sísmicas:

⁴⁸ Max Born y Emil Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 2.^a ed. Cambridge University Press, 1999.

$$S_0(t) = A_V^2(t) + A_H^2(t), \quad (24)$$

$$S_1(t) = A_V^2(t) - A_H^2(t) \quad (25)$$

y

$$S_2(t) = 2A_V(t)A_H(t)\cos[\phi(t)]. \quad (26)$$

Usando estos parámetros son obtenidos los cuadrados de las longitudes de los semiejes mayor (a) y menor (b), respectivamente:

$$a^2(t) = \left\{ S_0(t) + [S_1^2(t) + S_2^2(t)]^{1/2} \right\} / 2 \quad (27)$$

y

$$b^2(t) = \left\{ S_0(t) - [S_1^2(t) + S_2^2(t)]^{1/2} \right\} / 2. \quad (28)$$

Donde a(t) y b(t) se denominan como semieje mayor y menor de la traza sísmica, respectivamente. Un atributo calculable con las ecuaciones (27) y (28) es la elipticidad instantánea, definida como la razón:

$$\rho(t) = \frac{a(t)}{b(t)} \quad (29)$$

Para la visualización del atributo es necesario utilizar el recíproco de la elipticidad, porque para el caso de polarización lineal la elipticidad tiene a infinito. De esta forma, $\rho^{-1} = 0$ en presencia de polarización lineal y $\rho^{-1} = 1$ en presencia de una polarización circular.

Para especificar el movimiento de la partícula (progrado o retrógrado) se define el recíproco de la elipticidad con la función signo dada la diferencia de fase (recordando que la elipticidad siempre es positiva porque es la razón entre dos longitudes).

$$\sigma(t) = \text{sign}[\phi(t)]\rho^{-1}(t) \quad (30)$$

El ángulo de inclinación instantánea(τ) se define como:

$$\tau = \{\arctan[S_2(t)/S_1(t)]\} / 2 \quad (31)$$

Una vez definido el ángulo de inclinación, gracias a la figura 26 podemos observar que el ángulo de levantamiento (Υ) viene dado por:

$$\Upsilon(t) = \begin{cases} \pi/2 - \tau(t); & 0 \leq \tau(t) \leq \pi/2 \\ -\pi/2 - \tau(t); & -\pi/2 \leq \tau(t) < 0 \end{cases} \quad (32)$$

Cada uno de los atributos previamente definidos se suavizan utilizando una función de filtro gaussiano móvil que depende de los datos o señales sísmicas.

En total fueron calculados 13 diferentes atributos sísmicos para datos multicomponente 2C.

3.1.2. Atributos de velocidad Las velocidades locales pueden ser estimadas de la pendiente local a través del tensor de estructura. La primera vez que se definió el tensor de estructura fue en [Förstner y Gülch, 1987]⁴⁹. El tensor de estructura para registros sísmicos 2C se expresa como:

$$\mathbf{S} = \begin{bmatrix} \frac{\partial u}{\partial x} \times \frac{\partial u}{\partial x} & \frac{\partial u}{\partial x} \times \frac{\partial}{\partial t} \\ \frac{\partial u}{\partial x} \times \frac{\partial u}{\partial t} & \frac{\partial u}{\partial t} \times \frac{\partial u}{\partial t} \end{bmatrix} \quad (33)$$

Para obtener unos resultados más estables, se realiza un proceso de suavizado sobre los datos con ruido aleatorio. El tensor de estructura se puede descomponer en términos de sus valores y vectores propios, como se observa en la ecuación (34):

$$\mathbf{S} = \sum_{i=1}^2 \lambda_i \mathbf{v}_i \mathbf{v}_i^T \quad (34)$$

⁴⁹ W. Förstner y E. Gülch. "A fast operator for detection and precise location of distinct points, corners and centres of circular features". En: *ISPRS Intercommission workshop. Conference on fast processing of photogrammetric data* (1987), págs. 281-305.

Donde λ_i y \mathbf{v}_i representan los valores y vectores propios respectivamente. Dado que la dirección indicada por el autovector con el autovalor más grande es la variación rápida, la pendiente local es calculada por el vector propio con el valor propio más pequeño que está con la variación más lenta. La velocidad local obtenida gracias al tensor de estructura considera la información tanto temporal como la espacial del registro⁵⁰.

3.1.3. Atributos de ventana Estos atributos sísmicos serán calculados a través del dominio temporal de las trazas sísmicas, el número de ventanas depende del paso temporal ajustado en los registros. Se definen las bandas que son una serie de ventanas superpuestas idénticas por cada una de las componentes del registro. La polarización es estimada para cada una de las ventanas y se asume que, para toda la banda, la polarización es pura durante toda la ventana.

Sea la matriz de datos sobre una ventana temporal $\mathbf{X} = x_{ij}$, con $i=1,2,\dots,N$; y $j=1,2$; siendo i la i -ésima muestra de la componente j . La matriz de covarianza \mathbf{S} es evaluada como:

$$\mathbf{S}_{ij} = \frac{\mathbf{X}\mathbf{X}^T}{N} = \left[\frac{1}{N} \sum_{i=1}^N x_{ij}x_{ik} \right] \quad (35)$$

Donde el superíndice denota traspuesto. En este caso la matriz de covarianza es una matriz 2x2, real y simétrica. Explícitamente:

⁵⁰ Xia, Hilterman y Hua, ver n. 15.

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix} \quad (36)$$

Donde S_{12} representa la covarianza cruzada entre las componentes vertical (1) y horizontal (2). La matriz de covarianza es definida positiva, significa que sus autovalores son todos reales y positivos (aunque en algunos casos pueden ser cero). Resolviendo el problema de valores propios (λ_1, λ_2) y vectores propios ($\mathbf{u}_1, \mathbf{u}_2$) para la matriz de covarianza, se utiliza la siguiente ecuación:

$$(\mathbf{S} - \lambda^2 \mathbf{I}) \mathbf{u} = 0 \quad (37)$$

Donde \mathbf{I} es la matriz identidad 2×2 y $\mathbf{0}$ es el vector columna nulo. El resultado del problema de valores y vectores propios para la matriz de covarianza tiene ciertas particularidades. Los valores propios son reales y los vectores propios son ortogonales, dado que la matriz de covarianza por definición es una matriz hermítica. Los ejes principales del elipsoide de polarización están dados por $\lambda_i \mathbf{u}_i$, donde el valor propio da su longitud y el vector propio la dirección.

Para trabajar con los valores propios es recomendable ordenarlos de mayor a menor, de esta forma es más sencillo predecir el tipo de polarización y los ángulos de incidencia y azimut. En datos sísmicos reales, los valores propios generalmente son diferentes entre ellos y diferentes de cero, de esta forma la polarización exhibida es elíptica. La matriz de covarianza se dice que está bien condicionada al satisfacer las propiedades mencionadas anteriormente, esto se debe gracias a que el ruido sísmico y el ruido de dispersión no se correlacionan entre cada una de las componentes por ventana.

Para el caso 2C, el azimut⁵¹ es estimado de la orientación horizontal del movimiento rectilíneo dado por el vector propio \mathbf{u}_1 del valor propio más grande.

$$P_{azimut} = \arctan \left[\frac{\text{real}\{u_{11}\}}{\text{real}\{u_{12}\}} \right] \quad (38)$$

Siendo u_{11} y u_{12} las componentes del vector propio \mathbf{u}_1 . de igual manera es posible determinar el ángulo incidente⁵² de la orientación vertical del movimiento rectilíneo con la siguiente ecuación:

$$P_{incidente} = \arccos[u_{11}] \quad (39)$$

Con un registro 3C es posible determinar otros atributos de manera similar.

Ahora bien, usando los valores propios se puede definir una medida de la fuerza de la polarización de la señal. Para el caso multicomponente 2C se denomina el grado de la polarización plana de la señal.

⁵¹ Andy Jurkevics. "Polarization analysis of three-component array data". En: *Bulletin of the Seismological Society of America* 78.5 (1988), págs. 1725-1743. DOI: 10.1785/BSSA0780051725.

⁵² John E. Vidale. "Complex polarization analysis of particle motion". En: *Bulletin of the Seismological Society of America* 76.5 (1986), 1393-1405. DOI: 10.1785/BSSA0760051393.

$$P_p = 1 - \frac{\lambda_1}{\lambda_2} \quad (40)$$

Donde P_p es 1 cuando $\lambda_1 \ll \lambda_2$ y es 0 cuando los valores propios son comparables.

Detector de ondas superficiales El método se describe como un filtrado de polarización de las ondas superficiales basado en el método SVD. Los métodos de filtrado de polarización de las ondas superficiales constan principalmente de dos etapas: la detección y el filtrado. Este trabajo está interesado solamente detectar este ruido. Este análisis se aplica a ventanas de tiempo deslizantes superpuestas para un registro 2C.

La matriz de covarianza viene dada por los productos de entre las componentes de cada ventana como se observa en la ecuación (35)

Sea escrita la matriz de covarianza en términos de las amplitudes de las dos componentes ortogonales de las ondas superficiales (a_v y a_h) y la energía del ruido externo⁵³ c^2 , como:

$$\mathbf{C} = \begin{bmatrix} a_v^2 + c^2 & a_v a_h \\ a_v a_h & a_h^2 + c^2 \end{bmatrix} \quad (41)$$

⁵³ Olena Tiapkina y col. "Single-station SVD-based polarization filtering of ground roll: Perfection and investigation of limitations and pitfalls". En: *GEOPHYSICS* 77.2 (2012). DOI: 10.1190/geo2011-0040.1.

Para el caso 3C en [Tiapkina et al., 2012]⁵⁴ luego de resolver el problema de valores propios y descomponiendo la matriz de covarianza en valores singulares se encuentra que el atributo e_{mod} se puede definir como:

$$e_{mod} = \sqrt{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)} = \sqrt{(\sigma_1^2 - \sigma_3^2)(\sigma_2^2 - \sigma_3^2)} \quad (42)$$

Donde σ_i representa los valores singulares y λ_i los valores propios, mostrando la relación existente entre ellos.

Para el caso 2C el tercer valor singular dado por una componente extra es cero, por lo tanto, el atributo e_{mod} en este caso es:

$$e_{mod} = \sqrt{\sigma_1^2 \sigma_2^2} \quad (43)$$

Una forma para mejorar la sensibilidad del detector de onda superficial es empleando el centroide del espectro de potencias, ya que su recíproco reacciona en concordancia con el atributo e_{mod} ⁵⁵. El espectro de potencias es calculado empleando la transformada rápida de Fourier de la traza vertical de los datos 2C. De esta forma:

⁵⁴ Tiapkina y col., ver n. 53.

⁵⁵ Tiapkina y col., ver n. 53.

$$e_{mod(w)}(t) = \frac{e_{mod}(t)}{\bar{w}^m(t)} \quad (44)$$

Donde $\bar{w}^m(t)$ es el centroide del espectro de potencias normalizado.

Método de extracción de atributos	N°	Atributo sísmico
Traza compleja instantánea	1-2	Amplitud instantánea V-H
	3-4	Fase instantánea V-H
	5-6	Frecuencia instantánea V-H
	7	Diferencia de fase instantánea
	8	Semieje mayor
	9	Semieje menor
	10	Elipticidad recíproca instantánea
	11	Elipticidad recíproca instantánea (función signo)
	12	Ángulo de inclinación instantáneo
	13	Ángulo de levantamiento (subida) instantáneo
Velocidad local basado en el tensor de estructura	14-15	Velocidad local V-H
Análisis de covarianza y covarianza compleja	16	Ángulo azimut
	17	Ángulo incidente
	18	Fuerza de polarización
Análisis espectral	19	Detector de onda superficial

Tabla 3. Métodos de extracción de atributos sísmicos de ventana para datos multicomponente calculados en este trabajo.

En total fueron calculados 19 diferentes atributos sísmicos gracias a los métodos de extracción presentados en la tabla 3.

Una vez obtenido los datos sísmicos sintéticos luego de haber simulado la propagación de la onda sobre el modelo, se procede a usar estos datos para extraer sus atributos. La figura 27 muestra la extracción de atributos para los datos del *modelo Tenerife*.

Los atributos sísmicos serán empleados para agrupar los tipos de ondas presentes en los registros. Sin embargo, puede que algunos de los atributos sean linealmente dependientes

otros o algún atributo no contribuya con la agrupación de la dispersión. Por lo tanto, es importante realizar una selección de los atributos que ayuden a agrupar el ruido de dispersión como un grupo único en cada registro.

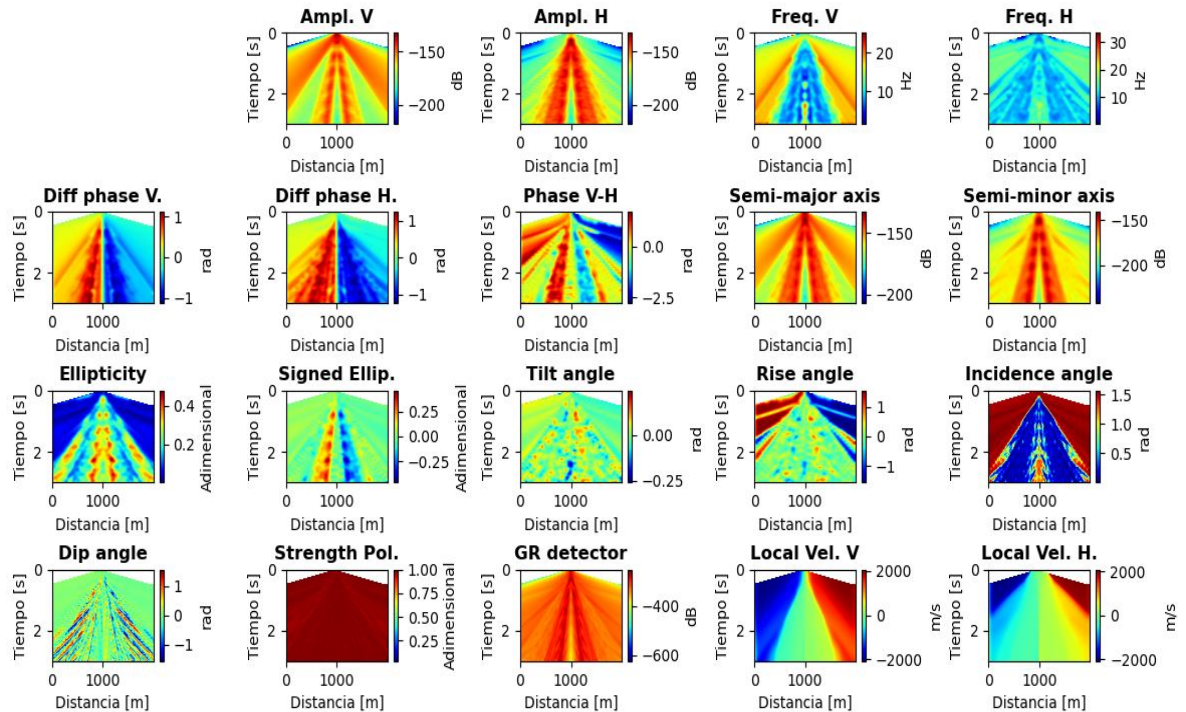


Figura 27. Atributos sísmicos extraídos del modelo Tenerife.

3.2. Métodos de Selección de Atributos

En aprendizaje automático, la selección de atributos es el proceso donde se selecciona un subconjunto de características para su uso, ya sea en modelos de predicción o en la clasificación de datos. Las técnicas de selección de atributos son muy utilizadas principalmente en conjuntos de datos de gran tamaño por tres razones fundamentales:

- El modelo tarda más tiempo en entrenarse (simplificación del modelo).

- Las características adicionales actúan como un ruido para el cual el modelo de Machine Learning puede tener un rendimiento extremadamente bajo.
- Asignación de recursos innecesarios para estas características.

Las ondas pueden agruparse automáticamente empleando técnicas de reconocimiento de patrones permitiendo separar información de acuerdo con el comportamiento de los atributos sísmicos. Se tiene que reconocer que las técnicas de selección de atributos son diferentes de la extracción de atributos, la cual crea características nuevas a partir de funciones de las características originales, mientras que la selección de características devuelve un subconjunto óptimo de las características para entrenar el algoritmo.

Las evaluaciones métricas influyen fuertemente en el rendimiento del algoritmo, y son estas evaluaciones las que distinguen entre las tres categorías principales de selección de características:

3.2.1. Métodos de envoltura. Este método busca un subconjunto de características adecuada para el algoritmo y tiene como objetivo mejorar su rendimiento. El criterio de selección del subconjunto de características adecuado se basa en el método de evaluación del algoritmo de aprendizaje automático que se emplee.

El método realiza diferentes pruebas con diferentes subconjuntos y escoge el subconjunto de características que resultados obtenga.

Los métodos de envoltura generalmente tienen un costo computacional elevado. Algunos ejemplos de estos métodos de selección son: Selección hacia adelante, selección hacia atrás y eliminación recursiva.

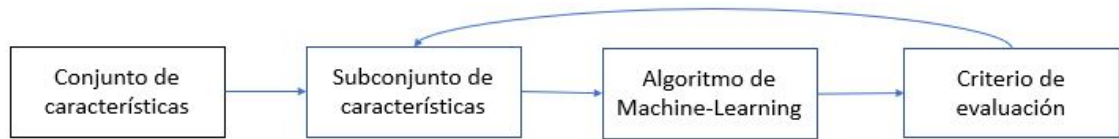


Figura 28. Funcionamiento de los métodos de envoltura.

El gráfico de la figura 28 muestra el funcionamiento de esta clase de métodos.

3.2.2. Métodos de filtrado. Las características se clasifican según los puntajes estadísticos que tienden a determinar la correlación de las características con la variable de resultado, ten en cuenta que la correlación es un término muy contextual y varía de un trabajo a otro. El diagrama de la figura 29 muestra el funcionamiento de este método.

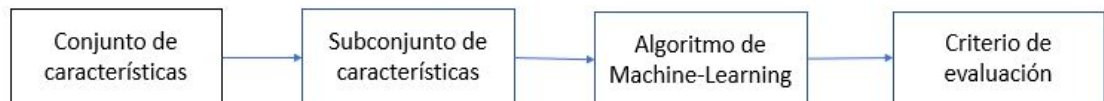


Figura 29. Funcionamiento de los métodos de filtrado.

Este método tiene problemas en detectar dependencias colineales.

3.2.3. Métodos integrados. Son aquellas técnicas que combinan los métodos de filtrado y los métodos de envoltura. generalmente estos métodos ya se encuentran preestablecidos como por ejemplo las regresiones LASSO y RIDGE. A su vez, reducen el sobre ajuste de los algoritmos de aprendizaje automático.

3.3. Selección de Atributos

Ya descritos los posibles métodos de selección que se pueden emplear en este trabajo, ahora se debe elegir uno de ellos para realizar esta tarea. El subconjunto de atributos que sea seleccionado debe ser el que mejor se adapte al problema de agrupación de las ondas dispersadas. Es de esperarse que los atributos sísmicos que no remarcan las zonas de ondas dispersadas no deben ser considerados.

Una primera forma de analizar cuales atributos señalan la dispersión en los registros es haciendo una comparación visual de atributos entre modelos sin dispersión (modelos base) y modelos con medios aleatorios. En este caso se escogieron dos registros del mismo modelo base, pero uno de ellos cuenta con ondas dispersadas dentro de la región de cono de onda superficial (ver figura 30).

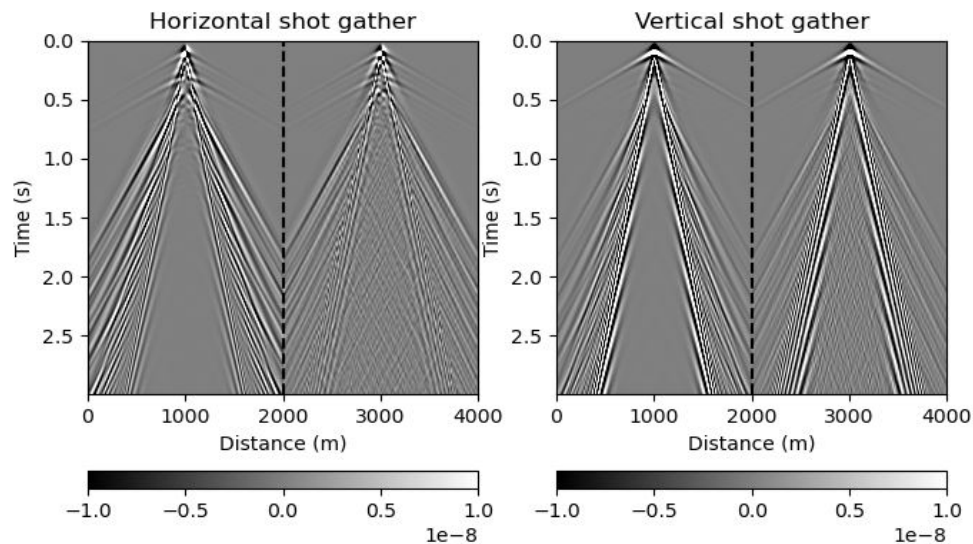


Figura 30. Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.

La idea de mostrar los registros sísmicos uno al lado del otro es para observar el cambio

producto del medio aleatorio, los atributos sísmicos que remarquen la zona de dispersión son aquellos que pueden detectar los efectos del medio aleatorio y apuntan a ser un subconjunto valido para la agrupación de las ondas.

Realizando el respectivo análisis visual de los atributos sísmico de la figura 31 es justo decir que los atributos que se ven afectados por los efectos del medio aleatorio en los registros son: las amplitudes, las frecuencias, las diferencias de fase, cada uno de los semi-ejes, el ángulo de incidencia y el detector de ondas superficiales.

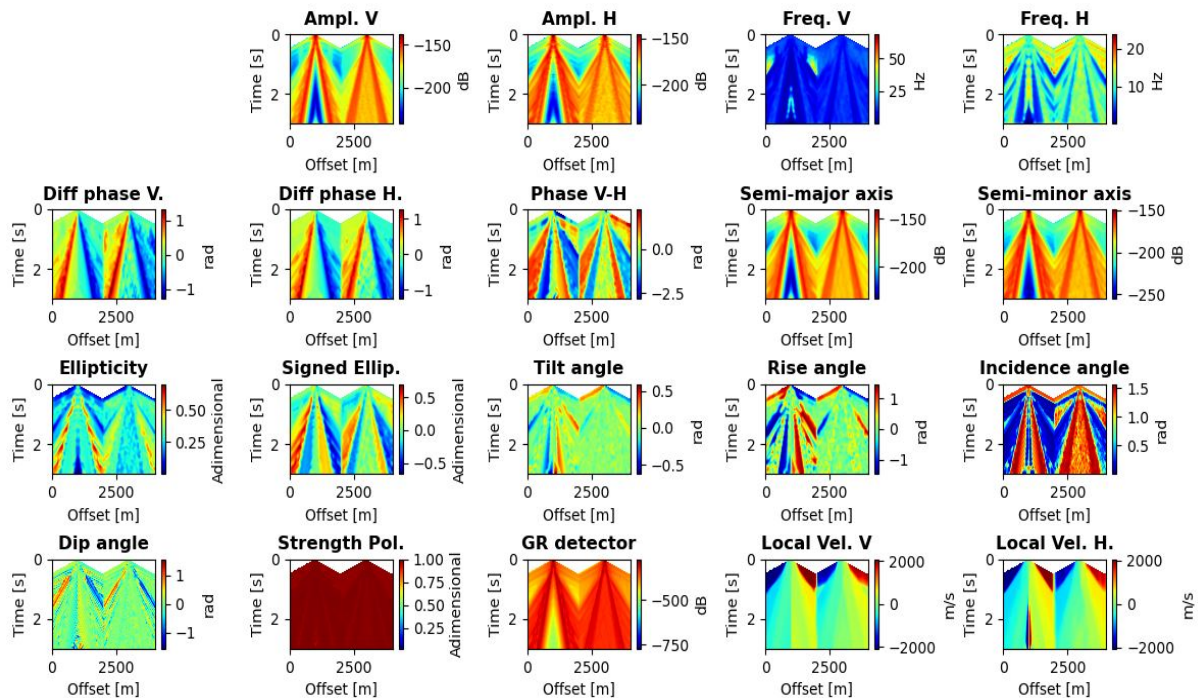


Figura 31. Atributos sísmicos para los registros de la figura 30.

Reconociendo el subconjunto de atributos mencionado anteriormente, es fácil pensar que un método de envoltura más precisamente una selección hacia adelante es adecuado para esta ocasión. El alto costo computacional de este método no es un problema para tratar en este caso porque el número total de atributos es pequeño, sin embargo, no se

considera una buena opción elegir este método de selección porque aunque los atributos detectan los efectos de la dispersión en el registro hay atributos que no se distinguen entre ellos, por lo que se sospecha que pueden estar altamente correlacionados y sean linealmente dependientes unos de los otros. Para evitar la dependencia lineal entre atributos y una posible sobreestimación de los resultados de la agrupación se decidió usar un método de filtrado para realizar un análisis de su correlación.

Con el fin de guardar la independencia entre los atributos, para la selección de atributos [Sánchez et al., 2017]⁵⁶ eligió un método de filtrado de acuerdo con la correlación cruzada entre los atributos. La figura 32 exhibe la correlación cruzada para el conjunto de 19 atributos.

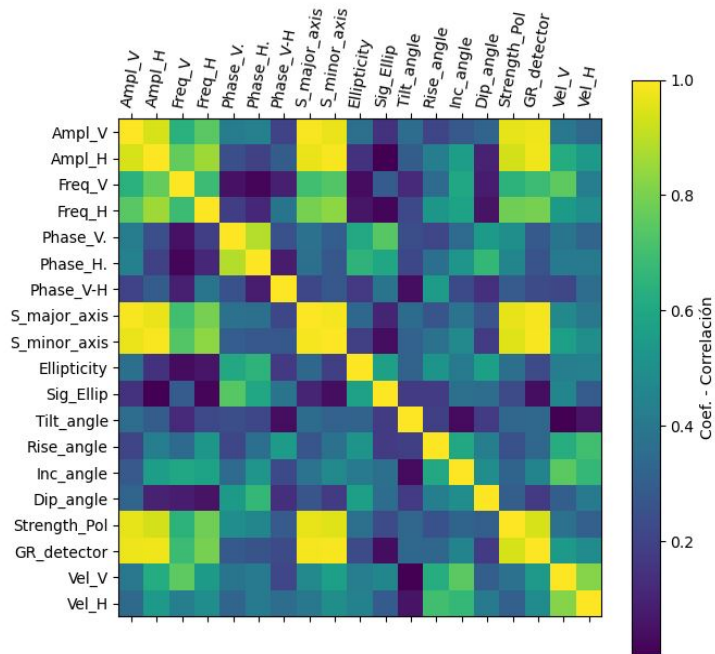


Figura 32. Matriz de correlación de atributos.

⁵⁶ Sánchez-Galvis y col., ver n. 17.

En general el conjunto de características no muestra altas correlaciones entre atributos excepto por el caso de las amplitudes y los semiejes, por lo tanto, se decidió trabajar con un subconjunto de 17 atributos. Existen correlaciones internas entre el mismo atributo, pero para registros diferentes (verticales y horizontales), sin embargo, como los datos son procesados para cada una de las componentes, entonces no es necesario eliminar estos atributos.

Una vez dispuesto el subconjunto de atributos óptimo, continuamos con la realización del algoritmo de agrupación que nos permita distinguir el ruido de dispersión en registros sísmicos.

3.4. Métodos de Agrupación

La información ofrecida por cada uno de los atributos sísmicos debe de agruparse para encontrar los tipos de onda presentes en los registros sísmicos. Para conseguir lo anterior se plantea dividir los datos en conjuntos o regiones de características similares dependiendo de la información ofrecida por los atributos sísmicos. Utilizando algoritmos de agrupación se indica o predice a que grupo pertenece cada punto o dato en particular.

En la metodología se plantea el estudio de diferentes algoritmos de agrupación con el fin de elegir el óptimo a utilizar en este trabajo. A continuación, se presentan tres diferentes algoritmos de agrupación.

3.4.1. K-means El algoritmo k-means es uno de los más simples y más populares algoritmos de agrupación. Se encarga de encontrar los centros de aglomeraciones que sean representativos para ciertas regiones de datos. Para lograr agrupar los conjuntos de datos

el algoritmo utiliza dos pasos: Asignar el centroide de las aglomeraciones de regiones de datos y establecer los puntos asignados a cada centroide. Finalmente, el algoritmo cierra su ciclo cuando las asignación de cada región ya no cambian⁵⁷.

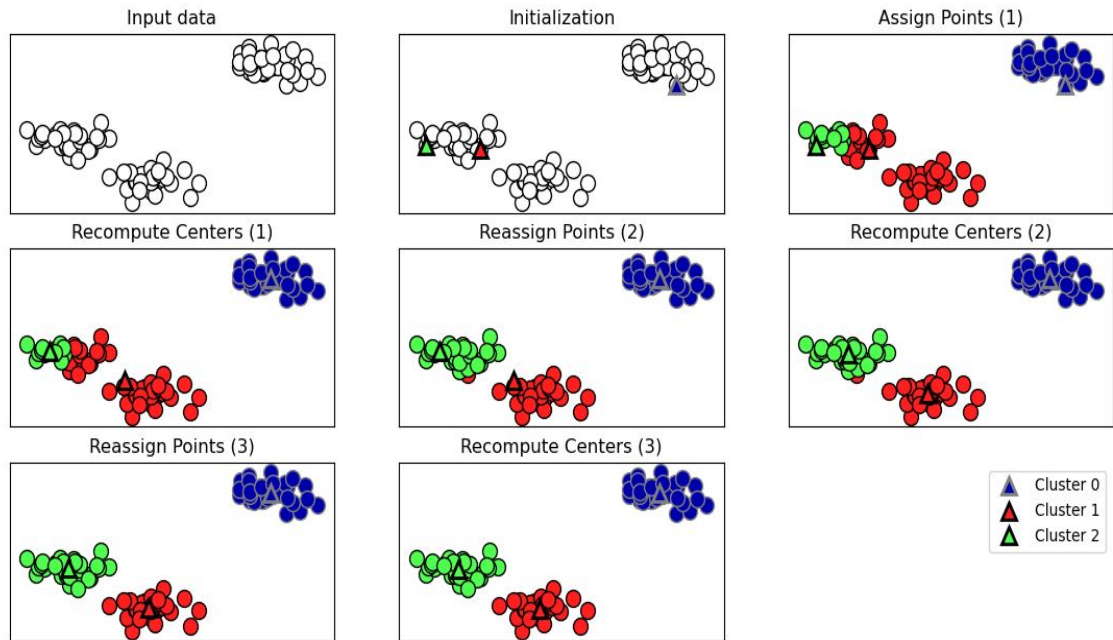


Figura 33. Funcionamiento del algoritmo k-means en un conjunto de datos sintético. Imagen tomada de la librería mglearn [Müller y Guido, 2016].

Una de las fallas de este algoritmo es que solo predice formas de regiones relativamente simples, asumiendo que todos los grupos presentes cuentan con formas similares independientemente si cada grupo cuenta con densidades diferentes o distancias medias (diámetros) diferentes. K-means asume que cualquier dirección es igual de importante que otra, por lo tanto, regiones claramente separadas de datos no son distinguibles aquí.

⁵⁷ Andreas C. Müller y Sarah Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. 1.ª ed. O'Reilly Media, 2016.

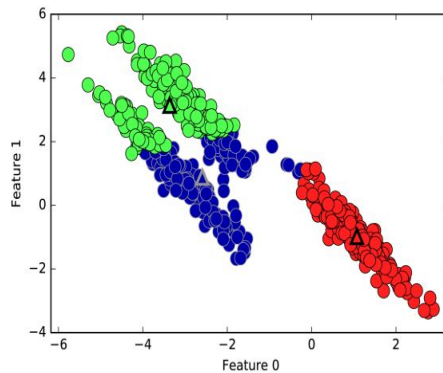


Figura 34. Falla del algoritmo k-means al no identificar grupos no esféricos. Imagen tomada de la librería mglearn [Müller y Guido, 2016].

Una imagen que muestra el comportamiento de las fallas de este algoritmo se presenta en la figura 34.

3.4.2. Agrupación Aglomerativa Los algoritmos aglomerativos inician declarando cada punto como su propio grupo y en cada iteración se fusionan los grupos más similares hasta cumplir el criterio de parada. Este criterio de para es la cantidad de grupos que se quieran obtener.

Generalmente este algoritmo funciona de tres formas diferentes:

- Analizando la varianza en cada iteración dentro de todos grupos haciendo que esta aumente lo menos posible. Esta forma conduce a obtener grupos con tamaños relativamente similares.
- Calculando las distancias promedio entre los grupos y uniendo los grupos con el menor promedio.
- Calculando las distancias máximas entre los miembros de los grupos y uniendo los grupos con las distancias máximas menores.

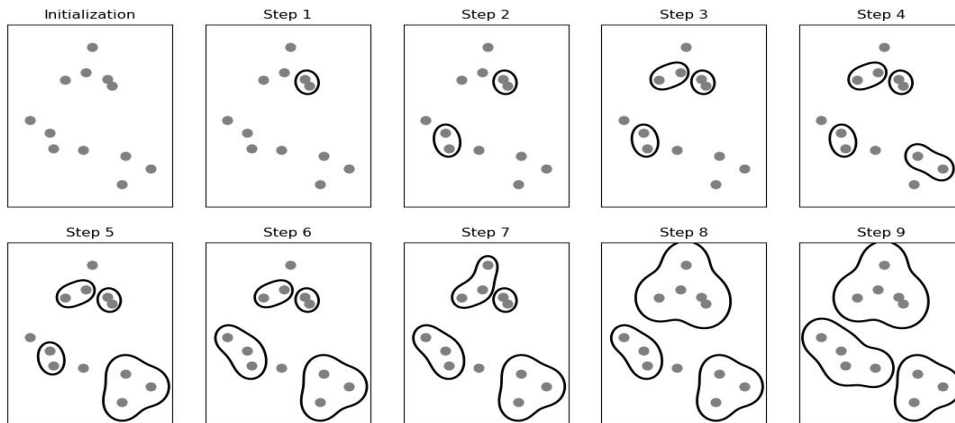


Figura 35. Funcionamiento del agrupamiento aglomerativo (une iterativamente los dos clústers más cercanos deteniéndose cuando encuentra tres grupos). Imagen tomada de la librería *mglearn* [Müller y Guido, 2016].

Un ejemplo del funcionamiento de la agrupación aglomerativa se presenta en la figura 35. Esta figura fue desarrollada en la librería *mglearn*⁵⁸.

Un problema presentado por los algoritmos de agrupación aglomerativa es el conocimiento a priori del número de grupos que se quieren obtener, además, presenta inconvenientes en la separaciones de regiones complejas.

3.4.3. DBSCAN La agrupación espacial de aplicaciones con ruido basado en la densidad (DBSCAN por sus siglas en inglés) es un algoritmo ampliamente utilizado por que no requiere que se establezca un número de grupos a priori, captura las regiones de formas complejas e identifica los puntos que no son parte de ninguna región.

⁵⁸ Andreas C. Müller and Sarah Guido. *mglearn library*. 2016. URL: https://github.com/amueller/introduction_to_ml_with_python.

DBSCAN necesita de dos parámetros iniciales para funcionar: el número de muestras mínimas por región y la distancia *eps*. El algoritmo funciona eligiendo un punto arbitrario para comenzar. Luego encuentra todos los puntos con distancia *eps* o menos desde ese punto. Si hay menos puntos que la muestra mínima a una distancia *eps* del punto de inicio, este punto se etiqueta como ruido, lo que significa que no pertenece a ningún grupo. Si hay más puntos que muestras mínimas dentro de una distancia de *eps*, el punto se etiqueta como muestra central y se le asigna una nueva etiqueta de grupo. Luego, se visitan todos los vecinos (dentro de *eps*) del punto. Si aún no se les ha asignado un clúster, se les asigna la nueva etiqueta de clúster que se acaba de crear. Si son muestras de núcleo, sus vecinos son visitados a su vez, y así sucesivamente. El grupo crece hasta que no hay más muestras de núcleo dentro de una distancia *eps* del grupo. Luego se selecciona otro punto que aún no ha sido visitado y se repite el mismo procedimiento.⁵⁹

Finalmente, DBSCAN divide el conjunto de datos en tres tipos diferentes: puntos centrales, puntos a distancia y ruido. Las diferentes asignaciones producto del agrupamiento gracias a DBSCAN pueden observarse en la figura 36.

La agrupación empleando DBSCAN es mucho más lenta en comparación con los algoritmos de agrupación previamente mencionados, sin embargo, también puede tratar conjuntos de datos relativamente grandes y la identificación de grupos de puntos que se encuentran en regiones pobladas en comparación a otras donde muchos puntos de datos están muy juntos es su principal ventaja.

⁵⁹ Müller y Guido, ver n. 57.

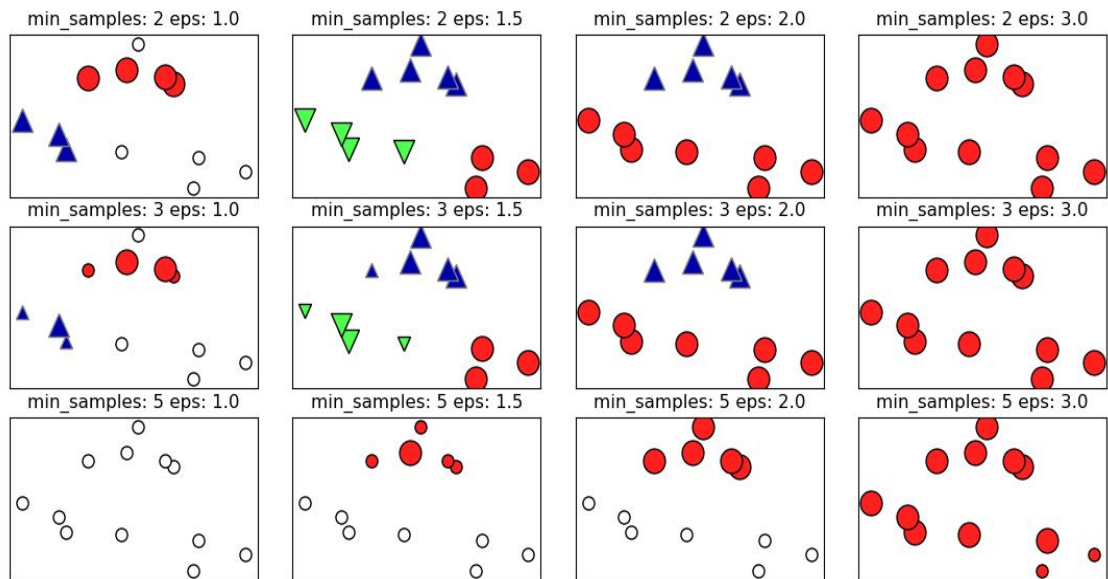


Figura 36. Funcionamiento del algoritmo DBSCAN con diferentes configuraciones para los parámetros de muestras mínimas y eps. Imagen tomada de la librería mglearn [Müller y Guido, 2016].

3.5. Elección del Método de Agrupación

Para sintetizar la información expuesta anteriormente para los tres algoritmos de agrupación se desarrollarán dos tablas donde se exponen cada uno de los métodos que servirán de ayuda para escoger el método de agrupación que mejor se adapte al problema.

Nombre del método	Parámetros	Métrica usada
K-means	Número de grupos	Distancia entre puntos
Agrupación Aglomerativa	Número de grupos o distancia	Distancia entre puntos
DBSCAN	Tamaño de vecindario	Distancia entre puntos (más cercanos)

Tabla 4. Métodos de agrupación con sus parámetros iniciales y la métrica usada.

Nombre del método	Escala de muestras			Escala de grupos			Regiones Irregulares	Regiones Semejantes	Regiones Densas	Detección de Outliers	Aprendizaje	
	Bajo	Medio	Alto	Bajo	Medio	Alto					Inductivo	Transductivo
K-means			✓			✓		✓			✓	
Agrupación Aglomerativa		✓				✓		✓				✓
DBSCAN			✓		✓		✓		✓	✓	✓	

Tabla 5. Propiedades de los métodos de agrupación estudiados.

La tabla 5 expone las propiedades de los algoritmos de agrupación estudiados. Cabe destacar que los espacios marcados indican el buen desempeño de cada algoritmo según la tarea y los espacios vacíos no indican la imposibilidad del algoritmo (solo indican que no cuenta con el mejor desempeño en esa tarea).

Para el caso de estudio de la agrupación de un registro sísmico 2D, el número de muestras que se tiene es muy elevado con un número de grupos medianamente bajo. Para poder entrenar un algoritmo de clasificación debe proporcionarse manualmente un conjunto de etiquetas al proceso de agrupación. De acuerdo con la comprensión geofísica de la propagación de ondas sísmicas en un modelo sintético se proponen cinco clases diferentes de tipos de onda:

1. Ondas P: Incluyendo refracciones y reflexiones.
2. Ondas P-S: Incluyendo refracciones y reflexiones.
3. Ondas superficiales: Incluyendo ondas Rayleigh y modos superiores.
4. Ruido de dispersión: Incluyendo ondas superficiales dispersadas.
5. Posibles reverberaciones.

Dadas las escalas que se están manejando en este trabajo para el número de muestras y el número de grupos, es recomendable que la agrupación aglomerativa sea descartada por que, aunque maneja un número de grupos alto, el número de muestras que soporta para un buen desempeño es mediano. De esta forma, se está tentado a elegir entre el algoritmo K-means y el DBSCAN para la agrupación.

Para decidir que algoritmo se empleará se compararan los campos de onda presentes en un registro sísmico y estos quienes dirán como pueden ser las regiones del registro. Dos ejemplos de la agrupación en registros sísmicos se presentan en [Xia et al., 2018]⁶⁰ y [Sánchez et al., 2017]⁶¹.

La figura 37 muestra tres campos de onda diferentes en un registro sísmico sintético donde se aprecia que las regiones son realmente irregulares, sin embargo, las regiones muestran una semejanza en sus tamaños, propiedad que también es observada gracias a la figura 38 del lado izquierdo.

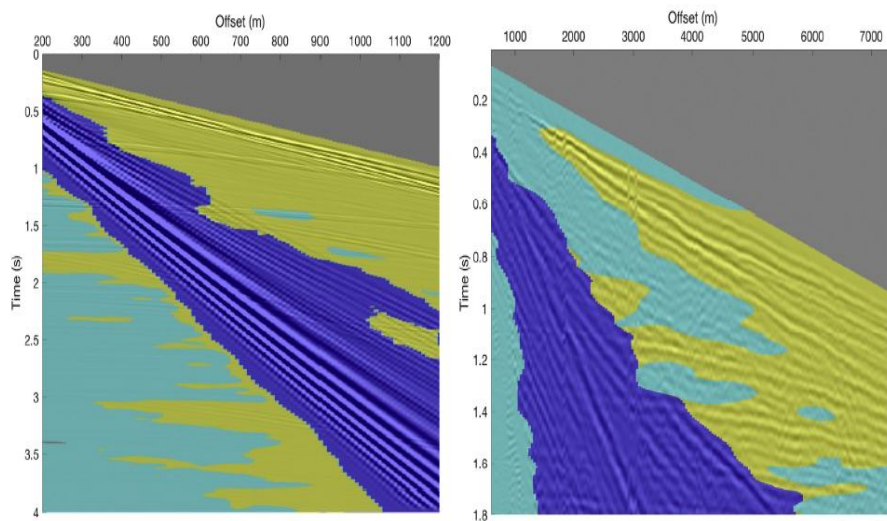


Figura 37. Dos registros sintéticos de onda sísmica con tres diferentes campos de onda cada uno. Imagen tomada de: [Xia et al., 2018]

La figura 38 del lado derecho muestra la detección de los campos de onda del registro, aunque comparando con la imagen del lado izquierdo esta detección puede parecer in-

⁶⁰ Xia, Hilterman y Hua, ver n. 15.

⁶¹ Sánchez-Galvis y col., ver n. 17.

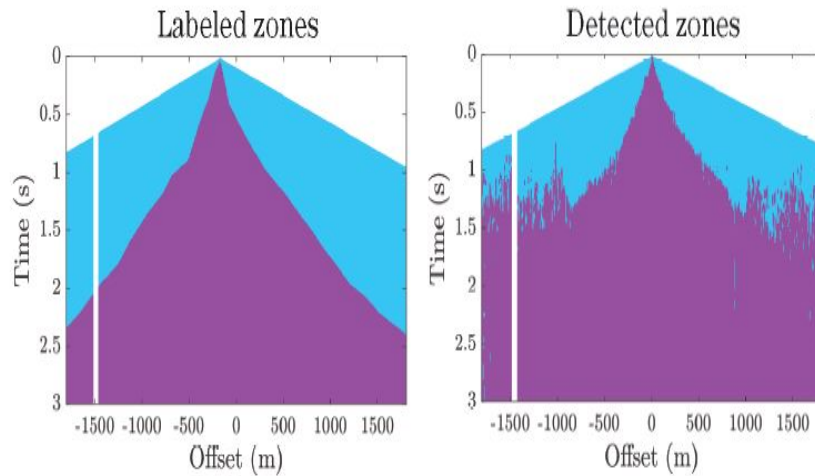


Figura 38. Resultados de la detección de campos de onda para un registro sísmico. Lado izquierdo: Áreas de un registro sísmico agrupadas por expertos. Lado derecho: Detección del área de registro usando k-means con dos grupos. Imagen tomada de: [Sánchez et al., 2017].

correcta, cabe destacar que en un registro sísmico no solo se poseen dos campos de onda diferentes, indicando que el algoritmo k-means se ve obligado a etiquetar las ondas en solo dos grupos y eso puede ocasionar problemas en la agrupación (al igual que el bajo número de grupos que se forzó a agrupar).

La comparación de las regiones puede que no sea muy determinante al momento de elegir que algoritmo de agrupación que se empleará, entonces este algoritmo se definirá teniendo en cuenta que, si bien en un registro sísmico se cuenta con ruido de dispersión, este tipo de ruido es coherente, entonces la detección de ruido de dispersión que permite el algoritmo DBSCAN no sería la mejor, además su tiempo de ejecución y alto costo computacional hace que k-means sea la mejor opción para trabajar.

3.6. Algoritmo K-means en el Proyecto.

Como se mencionó anteriormente, el algoritmo *K-means* es el más empleado para la agrupación de un conjunto de datos n en cierto número de grupos. Una ventaja de usar este método es que el usuario no debe tener conocimiento a priori de la clasificación de los datos.

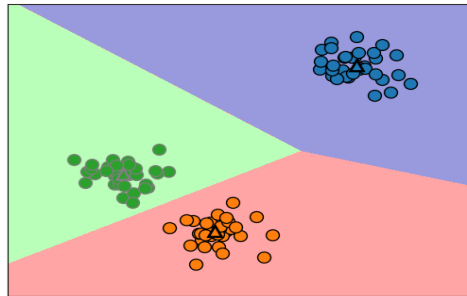


Figura 39. Ejemplo de la agrupación empleando el algoritmo K-means. Imagen tomada de la librería *mglearn* [Müller y Guido, 2016].

El algoritmo inicia con la información del número de grupos k que se van a agrupar. Cuando el usuario no puede suministrar el número de grupos, entonces el algoritmo hace un barrido entre un rango de valores de k y compara los resultados obtenidos para los diferentes grupos. Este método se conoce como el *Codo de Jambú* o *Elbow method* en inglés y fue usado por primera vez por [Thorndike, 1953]⁶² sobre estudios psicosociales basados en la estadística.

Una vez sea escogido el número de grupos óptimo a trabajar entonces es seleccionado una semilla aleatoriamente entre el conjunto de datos para cada grupo (es decir, se esco-

⁶² R. L. Thorndike. *Who Belongs in the Family?* 18.^a ed. Psychometrika, 1953.

gen k semillas), esta semilla se conocerá como el centroide. Cada dato será agrupado de acuerdo con su centroide más cercano haciendo que para esta primera iteración todos los datos correspondan con un centroide.

Durante cada iteración y para cada k grupo será calculada una función objetivo quien recordará los valores medios de la distancia entre los puntos de datos de cada región a su centroide. Para la siguiente iteración se hará el cambio de ubicación del centroide y se cambiará entonces la distribución de datos por cada región y nuevamente se calculará la función objetivo por grupo. Las iteraciones se harán siempre y cuando la diferencia entre las funciones objetivo sea mayor que la tolerancia (esto quiere decir que el algoritmo funcionará siempre y cuando la función objetivo tenga cambios significantes). La función objetivo se presenta en la ecuación (45)

$$J = \sum_{j=1}^k \sum_{i=1}^n \left\| \mathbf{x}_i^{(j)} - \mathbf{C}_j \right\| \quad (45)$$

Donde J es la función objetivo, $\mathbf{x}_i^{(j)}$ es el vector de atributos correspondientes a la observación i del grupo j y \mathbf{C}_j es la ubicación del centroide del grupo j .

3.7. Metodología de Agrupación Sobre un Registro Sísmico

En esta sección se presenta la metodología para realizar la agrupación para registro sísmico. Los modelos sintéticos de velocidad y densidad de los cuales se obtuvieron los datos con los que se trabajará la metodología de agrupación se presentan en la figura 40.

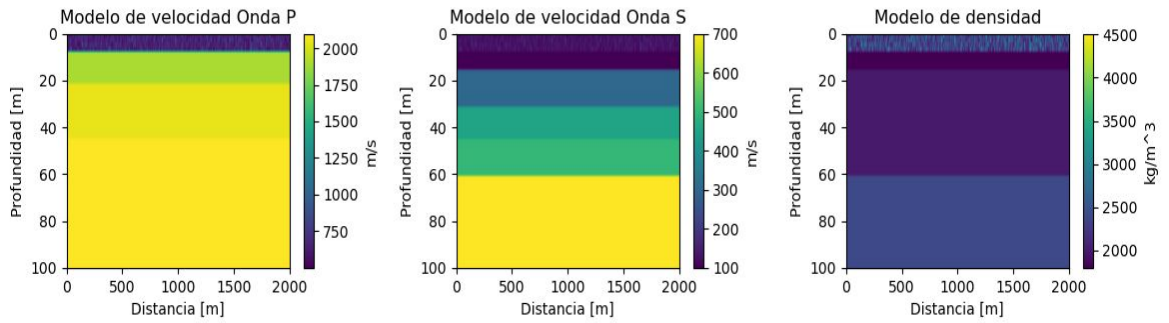


Figura 40. Modelos sintéticos de velocidad y densidad basado en el modelo Tenerife.

3.7.1. Lectura de registros El primer paso es leer los datos de los registros sísmicos con los que se vaya a trabajar. En el caso de estudio los registros sísmicos se presentan en la figura 41.

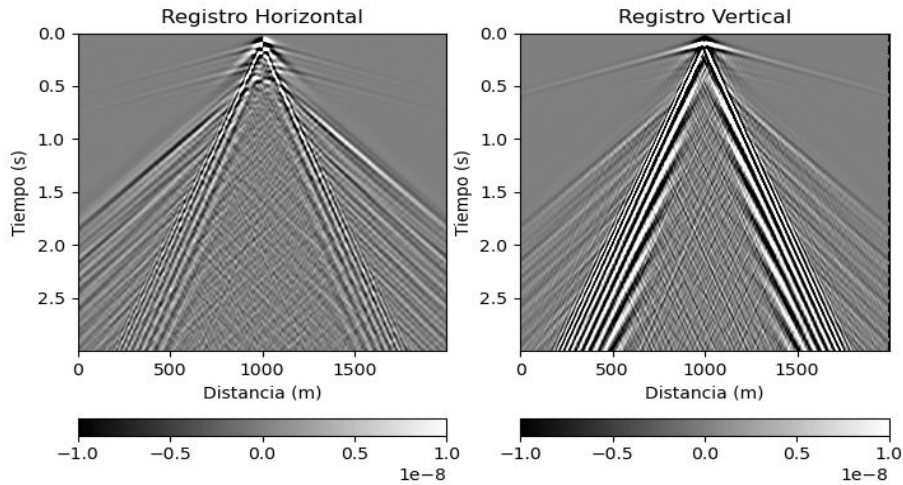


Figura 41. Registro sísmico 2D para el modelo de la figura 40.

Antes de pasar a la extracción de atributos se debe de asegurar de trabajar sobre los datos del registro, por esta razón se crea una máscara que se encienda únicamente donde se activen las trazas sísmicas. La máscara se presenta en la figura 42.

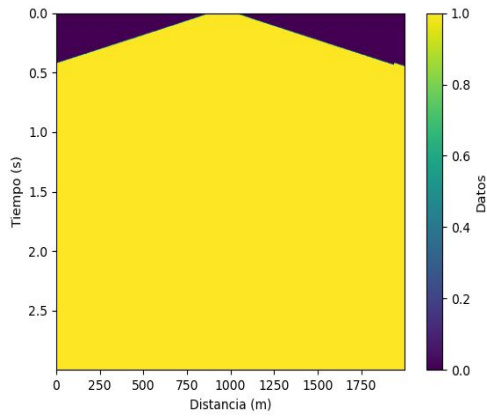


Figura 42. mascara de activación de registros.

3.7.2. Extracción de atributos A partir de los registros sísmicos se extraen sus atributos. La figura 43 presenta los 19 atributos sísmicos extraídos de los registros de la figura 41.

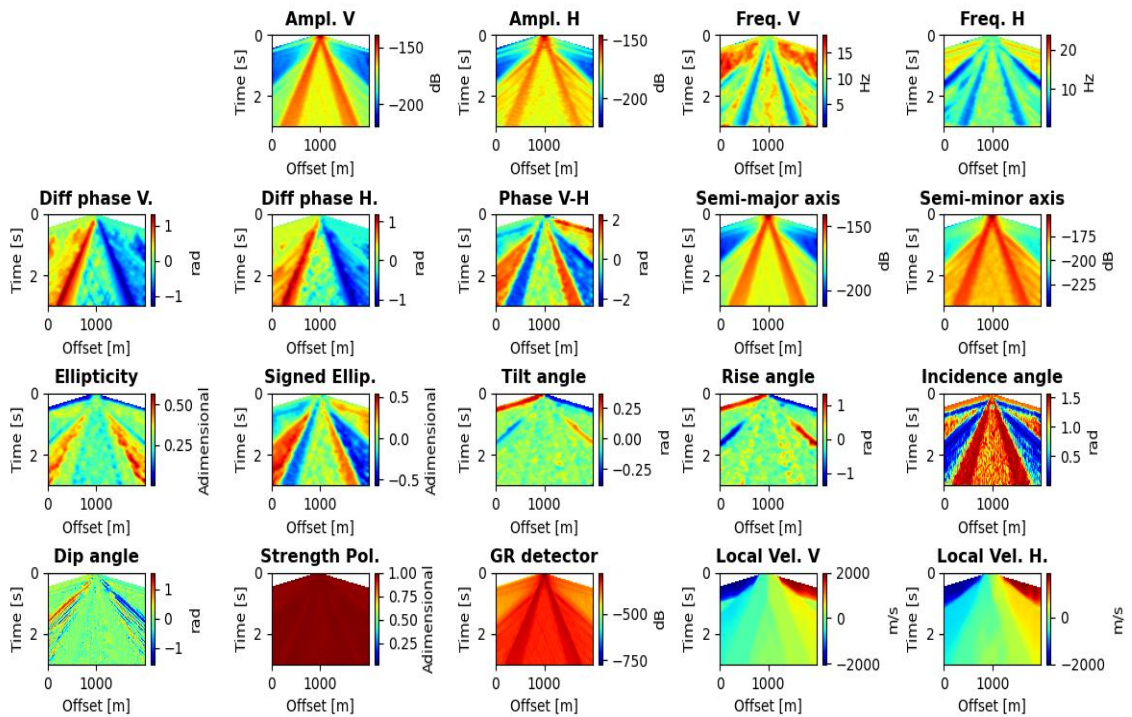


Figura 43. Atributos sísmicos para los registros de la figura 41.

3.7.3. Agrupación de ondas del registro Ahora se procede a realizar la agrupación de las ondas del registro. En este caso fueron detectados seis tipos de ondas diferentes presentadas en la figura 44.

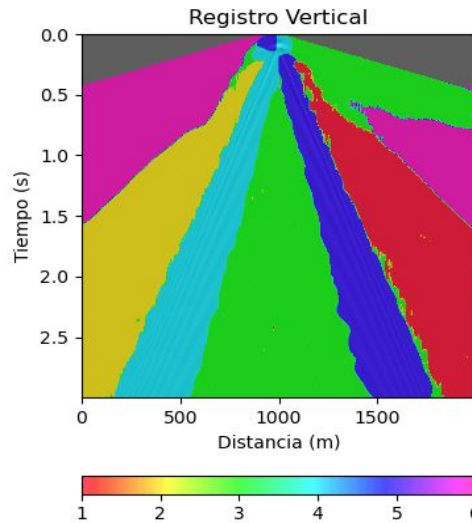


Figura 44. Agrupación para los tipos de onda presente en los registros de la figura 41.

La agrupación enmarca una buena delimitación de las zonas del registro, sin embargo, las etiquetas proporcionadas no muestran que tipo de onda contiene cada grupo y la región de ondas dispersadas no es única en los datos (región de color verde), es decir que la agrupación no registra las ondas dispersadas como un solo grupo con características diferentes a las demás ondas. Esta situación era de esperarse, ya que las ondas dispersadas son producto de la interacción de las ondas sísmicas con los objetos dispersores, por esta razón, las ondas dispersadas deben contener la misma información sísmica de las ondas primarias predecesoras imposibilitando la tarea de la agrupación.

La agrupación de la figura 44 es exitosa, sin embargo, genera la impresión que el algoritmo K-means no cuenta con las herramientas suficientes para enmarcar las zonas de dispersión como un grupo diferente porque muchos de los atributos extraídos (principalmente

los atributos mono-traza) no presentan una diferencia sobresaliente del caso sin medio aleatorio.

Se consideran dos maneras de lograr detectar el ruido de dispersión como un grupo unico:

1. **Aumento o disminución del peso de los atributos:** La extracción de más atributos con relaciones entre trazas, por ejemplo, correlaciones o filtros convolucionales, o eliminación de más atributos mono-traza (ya se han eliminado dos de ellos).
2. **Agrupación simultanea:** Examinar las diferencias entre las agrupaciones de los registros con dispersión y el registro de su modelo base respectivo.

El diagrama de dispersión de los datos para cuatro diferentes atributos sísmicos se presenta en la figura 45 junto con las respectivas funciones de distribución de las etiquetas para una muestra de datos por cada uno de los atributos.

Antes de crear la red neuronal que detecte automáticamente las zonas de ondas dispersadas sobre un registro sísmico es necesario definir de manera correcta las regiones del registro con ruido de dispersión usando el algoritmo de agrupación, pero como el algoritmo de agrupación no cuenta con las herramientas suficientes para detectar estas regiones sobre los registros, entonces esta tarea es imposible de lograr.

Una manera de lograr que el algoritmo K-means agrupe la dispersión es escogiendo un subconjunto aún más pequeño de atributos sísmicos que solo resalte la información de la dispersión, sin embargo, el algoritmo estaría forzado y las regiones etiquetadas estarían delimitadas de la forma incorrecta. También se podría extraer muchos más atributos sísmicos, pero esto aumentaría el costo computacional del proceso de agrupación y no

asegura que el ruido de dispersión sea detectado.

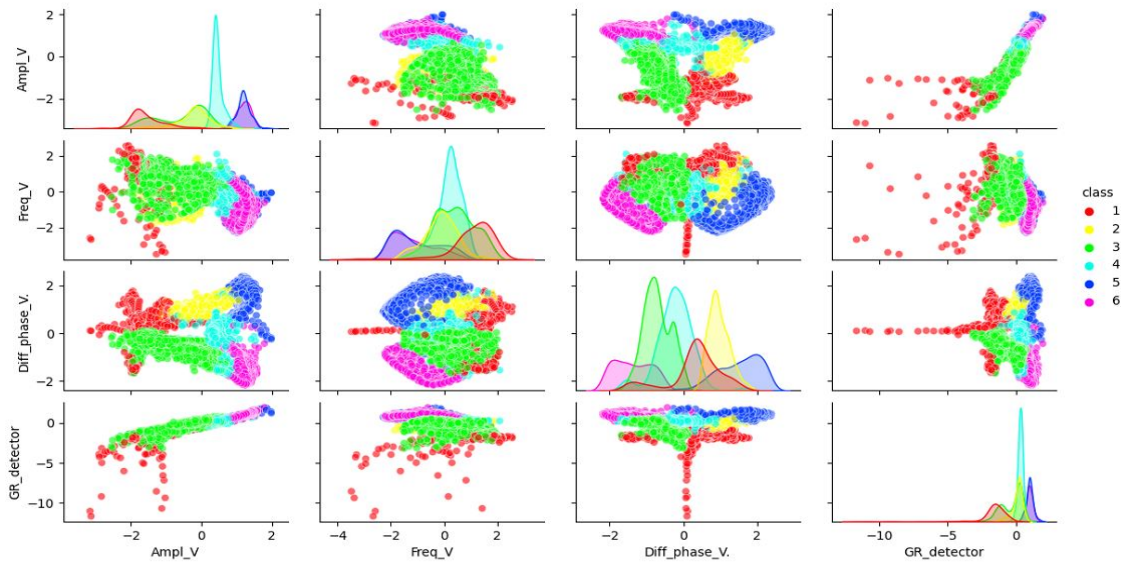


Figura 45. Diagrama de dispersión para cuatro atributos sísmicos del etiquetado de la figura 44.

Para solucionar el problema de agrupación se debe diferenciar los resultados de la agrupación para los registros con y dispersión. En el capítulo 4 se encuentra la solución al problema de la agrupación y también la definición de la dispersión numéricamente.

3.8. Repositorio

El código de extracción de atributos sísmicos junto con al algoritmo de agrupación K-means empleado en este proyecto se encuentra en el repositorio: [.https://github.com/HolgerQuintero/seismic_attributes](https://github.com/HolgerQuintero/seismic_attributes).

DISPERSIÓN SÍSMICA SINTÉTICA BASADA EN MACHINE LEARNING

Es fácil pensar que la resta entre el registro sísmico con dispersión y el registro para un modelo base resaltaré la información sísmica del ruido de dispersión, sin embargo, esta afirmación no es correcta. Si bien, la resta entre los registros sísmicos con y sin medio aleatorio tendrán la información de las ondas sísmicas dispersadas, estos registros también obtendrán las ondas que presentan corrimientos de fase debido a reflexiones junto con las ondas que presentan cambios de amplitud y estos tipos de ondas no son considerados como dispersión (Ver figura 46).

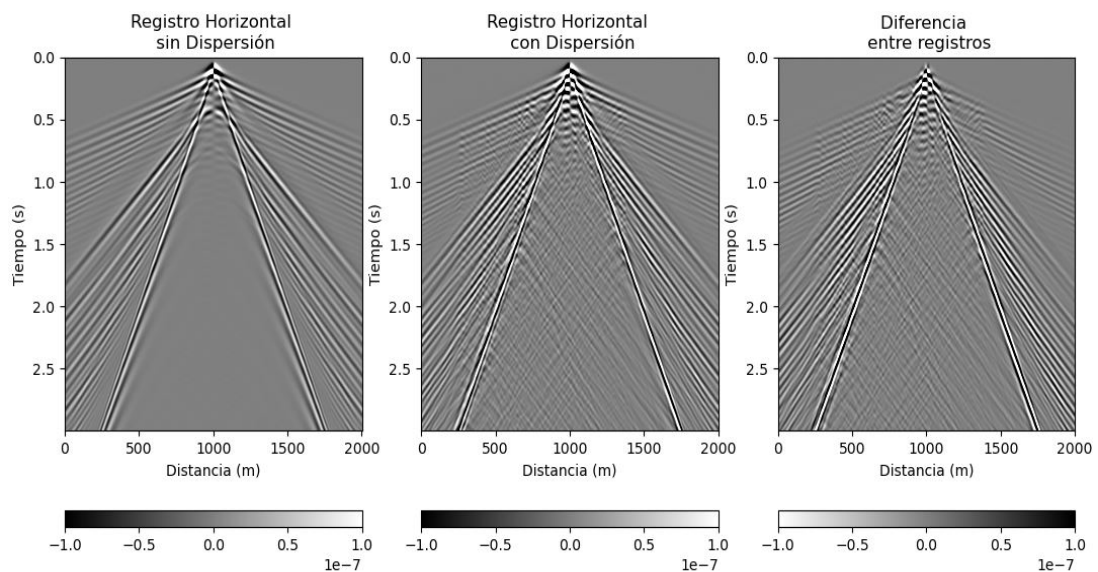


Figura 46. Diferencia entre atributos sísmicos con y sin dispersión para el mismo modelo base.

En este capítulo se plantea una solución para determinar las regiones de las ondas dispersadas de los registros sísmicos usando el algoritmo K-means. Esta solución se basa en el funcionamiento intrínseco del algoritmo de agrupación, aunque el algoritmo K-means no agrupa la dispersión como un grupo único, si ocurre que para diferentes pruebas de agrupación se obtiene la misma cantidad de grupos óptimos y la diferencia de forma de las

regiones es casi imperceptible. Considerando lo anterior y aprovechando que se cuenta con la información sísmica del modelo sin dispersión, se procede a realizar la agrupación simultánea de un registro sísmico para un medio con medio aleatorio junto con su respectivo modelo base (sin medio aleatorio). Este procedimiento se realizará sobre dos registros sísmicos unidos como los de la figura 30.

4.1. Agrupación Simultánea

El modelo usado cuenta con un medio aleatorio ubicado en la mitad de la profundidad del modelo y recostado hacia la izquierda. La figura 47 muestra el modelo empleado.

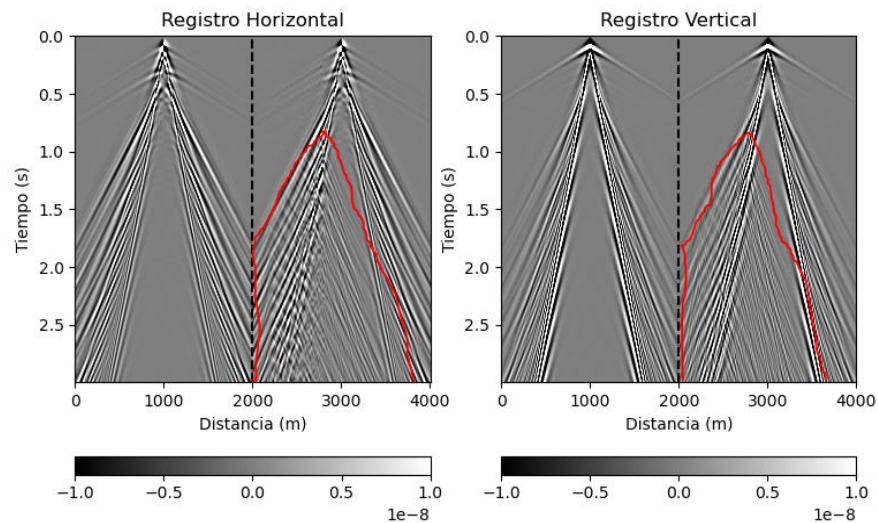


Figura 47. Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.

Notoriamente cada registro del medio aleatorio exhibe la zona de dispersión en su parte izquierda del registro, la región encerrada en rojo no es del todo dispersión, también se presentan desplazamientos en los tiempos de llegada de las ondas por las variaciones de

velocidad que se encuentra en el medio. Se espera que la zona de dispersión se encuentre próximo a la región encerrada.

Ahora se procede a extraer los atributos sísmicos para este registro (ver figura 48).

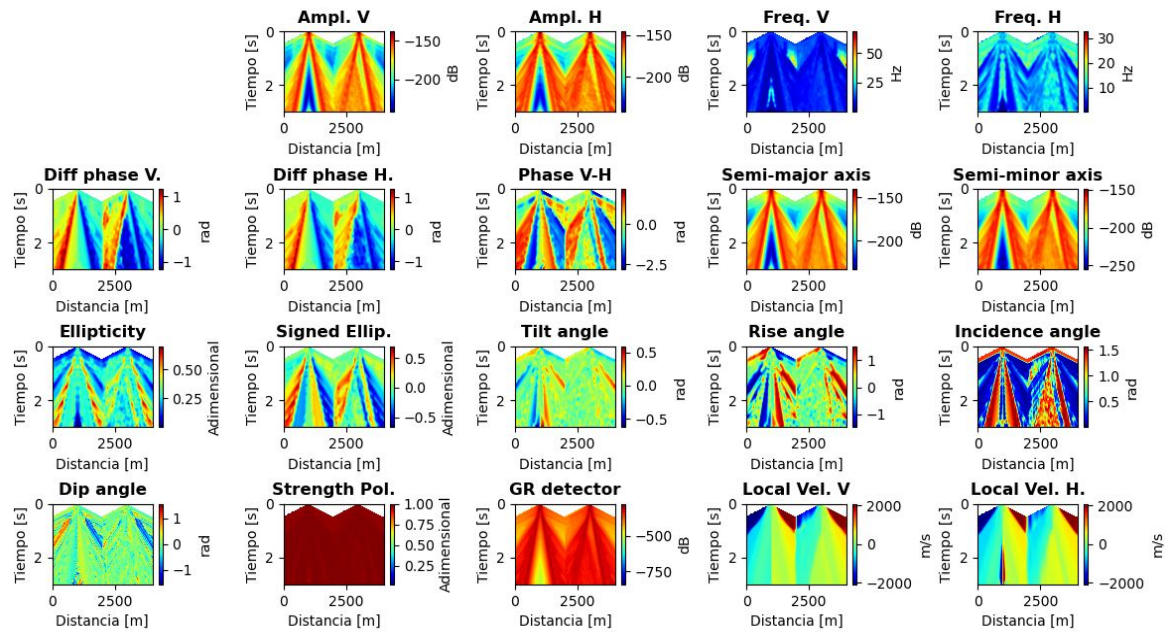


Figura 48. Atributos sísmicos para los registros de la figura 47.

La agrupación de las ondas en los registros sísmicos se muestra en la figura 49. Si bien las agrupaciones para ambos registros son muy diferentes, las etiquetas de cada zona son la misma, esta es la principal razón de trabajar los registros de forma simultánea.

Nuevamente se comprobó que algoritmo K-means no es capaz de encontrar por sí solo una zona exclusiva en los registros para el ruido de dispersión. A continuación, se procede a determinar el denominado atributo de dispersión.

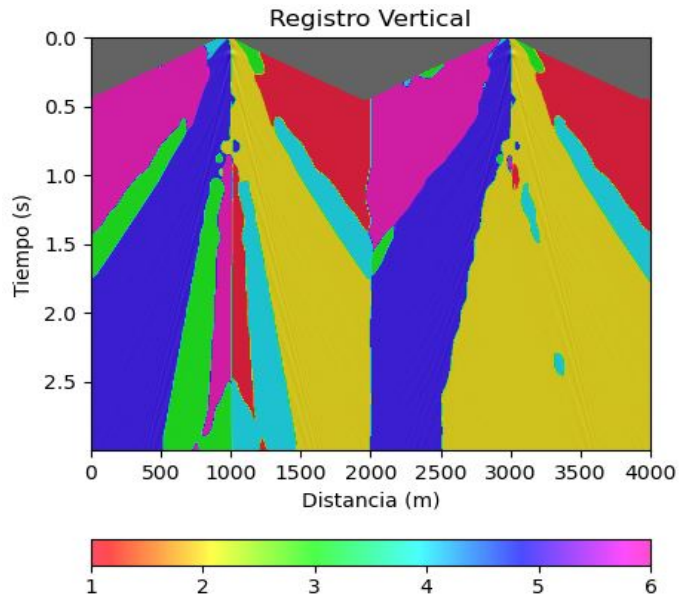


Figura 49. Agrupación para los tipos de onda presente en los registros de la figura 47.

4.2. Atributo de Dispersión Usando Aprendizaje Automático

Aprovechando que la agrupación simultánea de la sección 4.1 proporciona un etiquetado conjunto entre las regiones de los registros con y sin dispersión, se procede a separar en dos matrices (**A** y **B**) ambas agrupaciones, donde la matriz **A** es la agrupación del registro sin dispersión y **B** es la agrupación del registro con dispersión.

Una vez extraídas las matrices **A** y **B** de la agrupación simultánea, se procede a definir la matriz **C** como la diferencia binaria entre las etiquetas de los grupos de ambas matrices de la siguiente manera:

$$C = \|B - A\| \text{ donde } \|B - A\| = \begin{cases} 1, & \text{si } (B - A) \neq 0 \\ 0, & \text{si } (B - A) = 0 \end{cases} \quad (46)$$

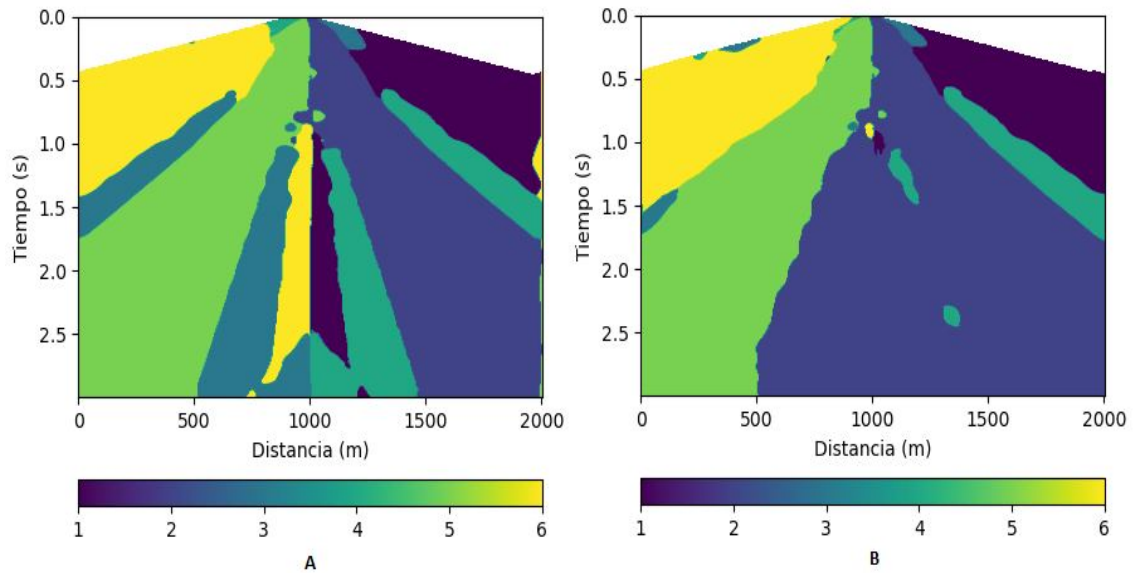


Figura 50. Separación de las agrupaciones de la figura 49. A) Agrupación para el registro sin dispersión. B) Agrupación para el registro con dispersión.

La matriz **C** de la ecuación (46) indica en que parte del registro sísmico hay diferencia entre las agrupaciones. La diferencia se consideró de manera binaria porque solo es importante saber dónde cambiaron las regiones agrupadas para el medio en presencia del medio aleatorio. Esta definición numérica es un indicativo de los cambios en los registros sísmicos producto de la dispersión por que la única discrepancia entre los registros de la figura 47 es la dispersión producto del medio aleatorio.

Cabe aclarar que esta definición solo puede ser aplicada a datos sísmicos sintéticos porque es reconocido de primera mano el modelo base o modelo background, no obstante, esta definición es útil para fines de este trabajo porque se emplean modelos sintéticos basados en la previa inversión de los datos sísmicos obtenidos en el campo TENERIFE.

La figura 51 muestra la región del registro sísmico que obtuvo cambios en la agrupación de las ondas luego de usar el algoritmo K-means.

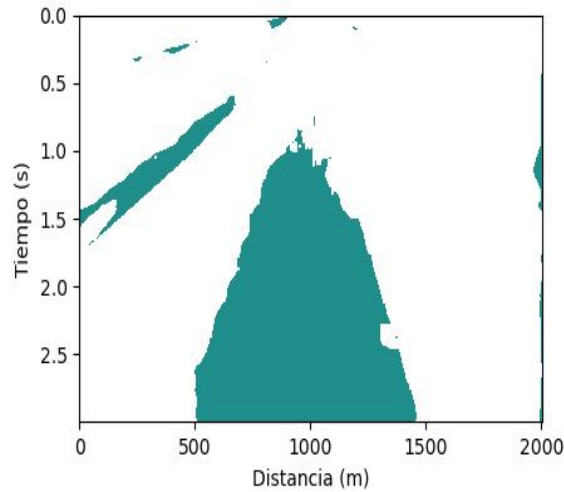


Figura 51. Atributo de dispersión para los registros de la parte derecha en cada gráfica de la figura 47.

Esta imagen sísmica se denominó como el *atributo de dispersión*. Notoriamente se observa que el cambio en la agrupación se encuentra recostado a la izquierda, así como la región enmarcada de la figura 47.

El atributo de dispersión también permite calcular la relación señal-ruido del conjunto de datos sísmicos. Para este caso la relación señal-ruido es: 25.64 %

No es posible definir el atributo de dispersión como la dispersión total en el registro sísmico por que el atributo también presenta desfases debido a cambios de velocidad de las ondas y problemas de frontera numéricos, sin embargo, las contribuciones de estas particularidades sobre el atributo de dispersión son mínimas en comparación con la contribución de los efectos de la dispersión, además el atributo no considera la estática debido a reflexiones en los registros.

4.3. Demostración del Atributo de Dispersión

Para comprobar que el atributo de dispersión resalta únicamente las zonas de ondas dispersadas en las agrupaciones simultáneas (las zonas diferentes entre los registros de los modelos base y los modelos con medios aleatorios), se consideraron dos registros exactamente iguales como los de la figura 52.

En este caso, al no tener diferencia alguna en la agrupación simultánea, el atributo de dispersión debería ser nulo.

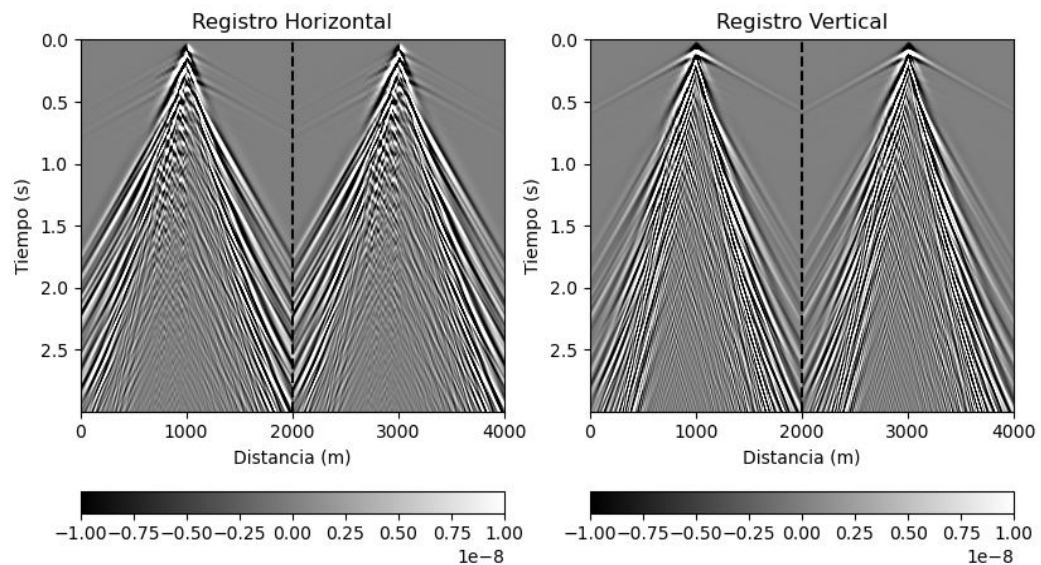


Figura 52. Registros sísmicos de un modelo sintético. Parte derecha de cada registro: Registro sísmico para un modelo con un medio aleatorio incrustado. Parte izquierda de cada registro: Registro sísmico para un modelo sin medio aleatorio.

Las agrupaciones luego de ser separadas se presentan en la figura 53 mientras que el atributo de dispersión para este caso se presenta en la figura 54. El atributo de dispersión en este caso solo reconoció unos cuantos problemas de frontera numérica no muy significativos. Para este caso la relación señal-ruido es: 0.44 %

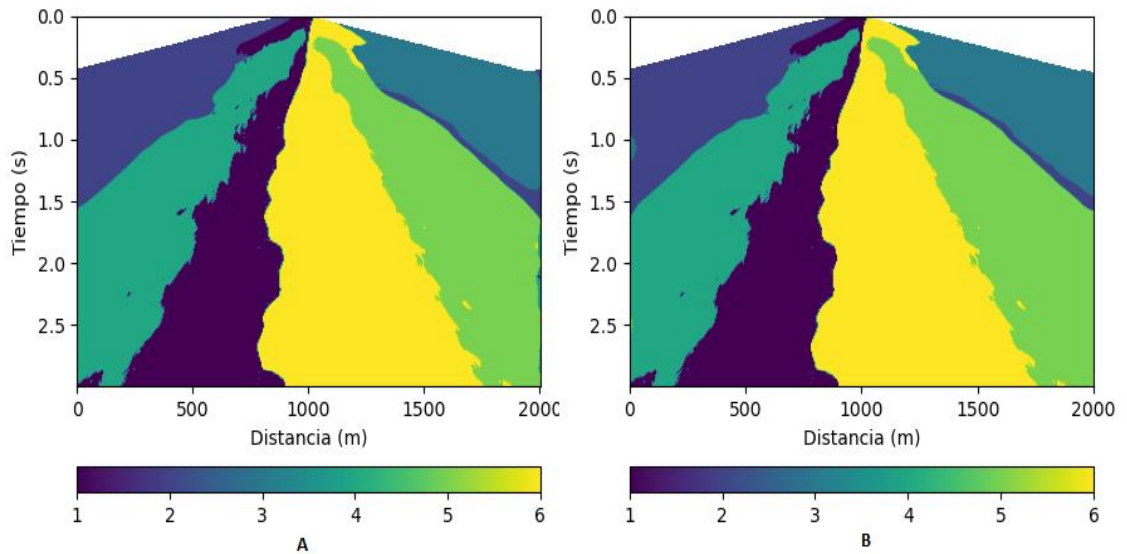


Figura 53. Separación de las agrupaciones para los registros de la figura 52. A) Agrupación para el registro de la izquierda. B) Agrupación para el registro e la derecha.

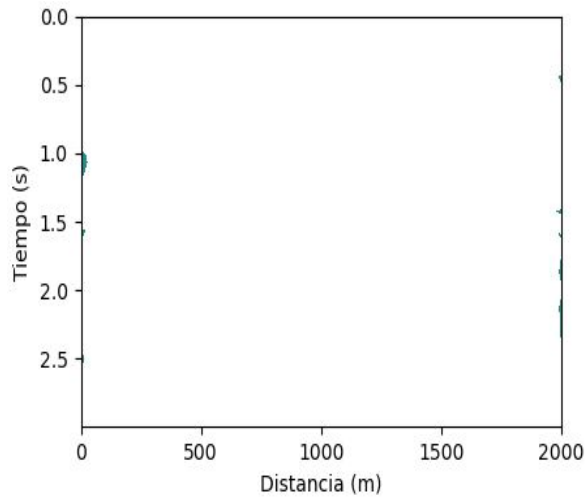


Figura 54. Atributo de dispersión para los registros de la parte derecha en cada gráfica de la figura 52.

Este resultado es muy relevante porque el atributo de dispersión mide el cambio en las regiones de los registros sísmicos sintéticos, entonces sí solo se presentan cambios debido a la dispersión de las ondas sísmicas producto del medio aleatorio, entonces el atributo reconocerá las zonas del ruido de dispersión en los registros.

El atributo de dispersión es detectado automáticamente usando la agrupación simultánea y esto evita que una persona etiquete manualmente los grupos del registro.

Se procederá a almacenar el conjunto de datos sísmicos sintéticos junto con su respectivo atributo de dispersión y así se entrenará la red neuronal para detectar automáticamente las regiones con ruido de dispersión en los registros sísmicos.

En este capítulo se busca desarrollar un sistema que detecte las regiones de los registros sísmicos con presencia de dispersión, este sistema debe aprender a tomar decisiones propias y debe tener un alto rendimiento computacional todo en función de los datos sísmicos sintéticos que se introduzcan.

Para esto, se expondrán algunos de los métodos de clasificación usando aprendizaje automático.

5.1. Métodos de clasificación

El aprendizaje automático cuenta con dos tipos de algoritmos de aprendizaje: supervisado y no supervisado. El aprendizaje no supervisado se empleó en el capítulo 3 con el algoritmo K-means, donde el entrenamiento se basó netamente en las características de los datos. Ahora se usará el aprendizaje de tipo supervisado con un algoritmo de clasificación, donde se enseñará a la maquina a tomar las decisiones.

La clasificación de datos cuenta con varios algoritmos o métodos entre los que resaltan:

- Redes neuronales.
- Árboles de decisión.
- Redes Bayesianas.
- Máquinas de vectores de soporte.

Es importante estudiar los diferentes algoritmos de clasificación, a continuación, se presenta un breve resumen de cada uno.

5.1.1. Redes neuronales Consiste en un conjunto de unidades, llamadas neuronas artificiales, conectadas entre sí para transmitirse señales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones) produciendo unos valores de salida. Las operaciones matemáticas son ordenadas de acuerdo con la estructura que se desea seguir. La figura 55 presenta el funcionamiento de una red neuronal básica.

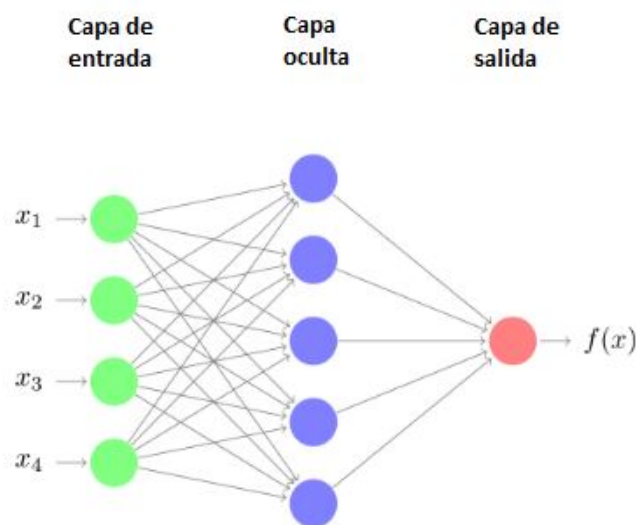


Figura 55. Funcionamiento de una red neuronal básica.

5.1.2. Árboles de decisión Conocidos como árboles de decisión o de clasificación son algoritmos empleados para clasificar utilizando particiones sucesivas. Son apropiados cuando hay un número elevado de datos, siendo una de sus ventajas su carácter descriptivo que permite entender e interpretar fácilmente las decisiones tomadas por el modelo, revelando formas complejas en la estructura de datos que no se pueden detectar con los métodos

convencionales de regresión.

Los árboles de decisión parten de una base de datos que crea diagramas de construcciones lógicas que nos ayudan a resolver problemas. Es un proceso de división secuencial, iterativo y descendente que, partiendo de una variable dependiente, forma grupos homogéneos definidos específicamente mediante combinaciones de variables independientes en las que se incluyen la totalidad de los casos recogidos en la muestra.

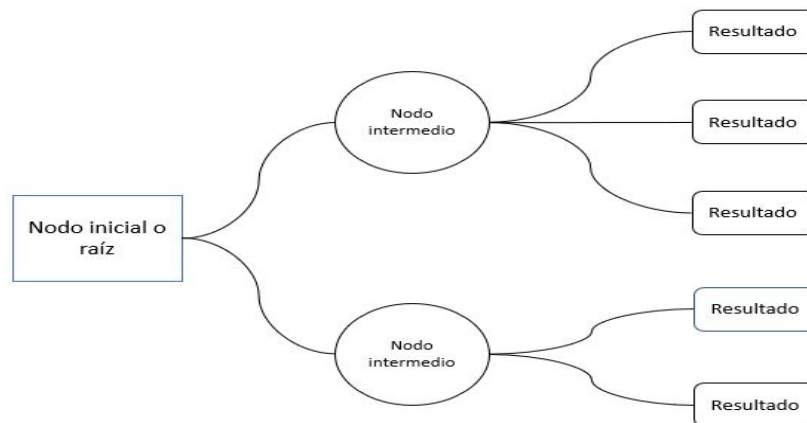


Figura 56. Funcionamiento de un árbol de decisión.

5.1.3. Redes Bayesianas Es uno de los métodos de clasificación más utilizados por su simplicidad y rapidez. Se trata de una técnica que construye modelos que predicen la probabilidad de posibles resultados, basados en el teorema de Bayes o teorema de probabilidad condicional.

Una red bayesiana se considera como un gráfico no cíclico donde cada variable tiene asociada una función de probabilidad. Su estructura provee información respecto a la dependencia de las variables.

5.1.4. Máquinas de vectores de soporte Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM (por sus siglas en inglés Support Vector Machines) para construir un modelo que prediga la clase de una nueva muestra.

La SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano (o hiperplanos) de separación definido como el vector entre los 2 puntos, de las clases, más cercanos al que se llama vector soporte. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o a otra clase.

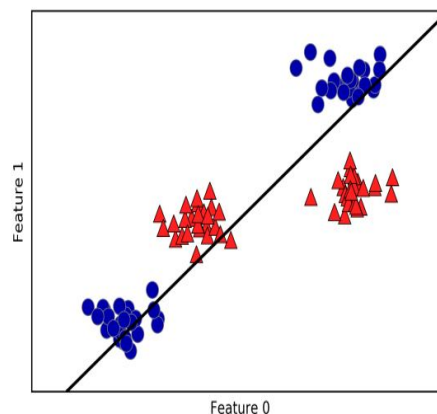


Figura 57. Funcionamiento gráfico de las máquinas de vectores de soporte. Imagen tomada de la librería mglearn [Müller y Guido, 2016].

5.2. Elección del algoritmo de clasificación

Si se consideran los registros sísmicos como imágenes, donde el resultado de la clasificación es una segmentación binaria considerando como única solución las zonas en los

registros con presencia de dispersión, entonces una red neuronal es considerada la mejor opción para realizar este trabajo, principalmente por que el resultado de la segmentación de acuerdo con la forma es considerado el trabajo más usual realizado por redes neuronales⁶³.

El proceso de la segmentación de la imagen sísmica con redes neuronales se muestra en la figura 58.

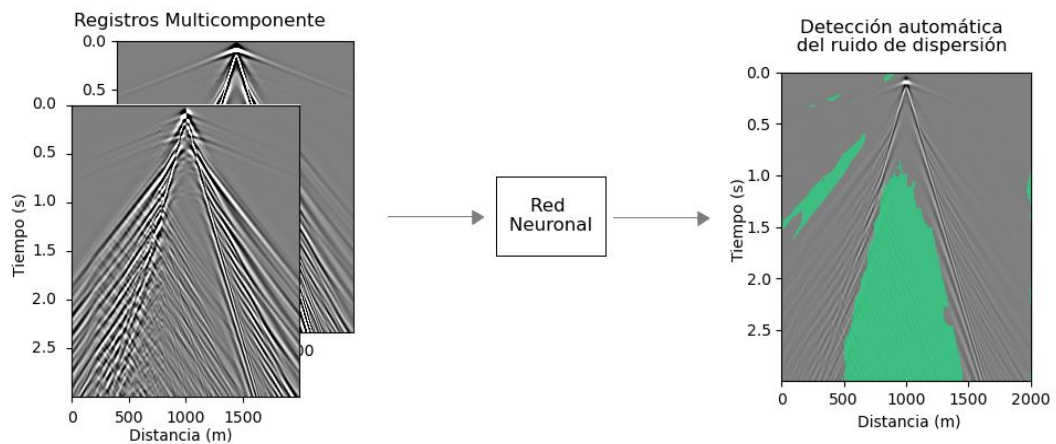


Figura 58. Proceso de clasificación de ondas multicomponente a través de aprendizaje automático.

5.3. Red Neuronal Convolutacional

Dada la complejidad de los modelos que se usan para este proyecto se empleará una red neuronal convolutacional para realizar la segmentación de la imagen sísmica porque este tipo de redes son aplicables a matrices bidimensionales. El *framework* **PyTorch** es usado

⁶³ J. Ramirez y M. Chacón. “Redes neuronales artificiales para el procesamiento de imágenes, una revisión de la última década”. En: *REVISTA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y COMPUTACIÓN* 9.1 (2011), págs. 7-16.

como el motor de la red.

La red neuronal convolucional consiste en múltiples capas con filtros basados en convoluciones de una o más dimensiones. Como cualquier otra red neuronal primero se deben extraer las características, este trabajo se lleva a cabo en las diferentes capas convolucionales, luego cada muestra es reducida para finalmente realizar la clasificación convencional.

A medida que los datos son procesados se presentan los cambios de dimensionalidad de cada muestra. Las reducciones de dimensión facilitan el trabajo de las capas de neuronas más lejanas por que se vuelven mucho menos sensibles a las perturbaciones en los datos de entrada.

5.3.1. Datos y elección de parámetros Las redes neuronales convolucionales reconocen objetos dentro de las imágenes, convirtiendo cada imagen en una matriz de píxeles, su entrenamiento hace que sea capaz de reconocer los píxeles de las matrices e infiera similitudes en todo el dominio. Las imágenes de los registros sísmicos serán descompuestas en píxeles como lo muestra la figura 59.

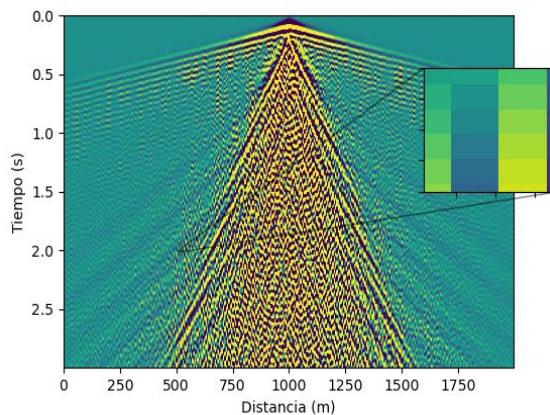


Figura 59. Píxeles de una zona de un registro sísmico.

Cada píxel es equivalente a una neurona en la red, dado que cada matriz de registro sísmico cuenta con 1500 filas y 334 columnas, entonces cada imagen proporcionará a la red un total de 501000 neuronas. Los registros sísmicos son dados en escala de grises, por lo que no es necesario obtener cada registro en los canales RGB (red, green, blue). Por otro lado, al contar con registros multicomponente 2C, entonces se tiene que utilizar dos canales para ambas componentes, lo que quiere decir que el total de neuronas por muestra será 1002000.

La red neuronal se alimentará de 200 registros sísmicos divididos entre componentes verticales y horizontales. Es conveniente realizar una normalización de los datos de entrada para obtener un mejor resultado de la clasificación. A su vez, a la red se incorporará los atributos de dispersión determinados para los 100 modelos sintéticos, este atributo es de carácter binario, por lo que no es necesaria su normalización. Los datos de entrada son empaquetados en tres matrices como lo muestra la figura 60.

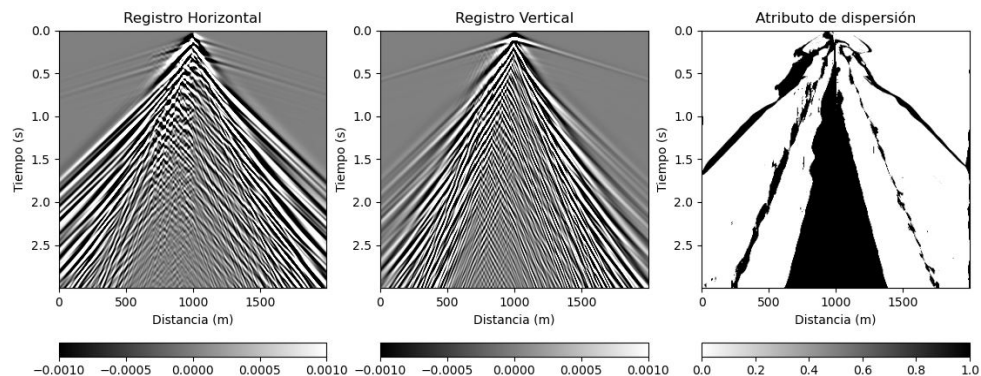


Figura 60. Tríada de datos de entrada.

Los parámetros de la red son configurados a conveniencia, cabe aclarar que estos parámetros pueden ser sustituidos tanto como sea necesario para mejorar el rendimiento de la red. A continuación, se presentan los parámetros para la configuración inicial de la red neuronal.

- Se escogió un total de 50 *épocas* quien va a representar el número de veces en las que la red se entrenará o en otras palabras el número de veces de entrenamiento.
- La *razón de aprendizaje* fue escogida constante para cada una de las épocas, su valor fue 1×10^{-3} para que el modelo de entrenamiento no salte regiones y sea capaz de aprender por sí mismo.
- Para calcular el error del modelo en cada uno de los entrenamientos se empleó el método del gradiente descendiente. Aquí se divide cada entrenamiento en lotes gracias al parámetro *Batch*=6. Cuando se cumple el número total de lotes entonces se cumple una época.
- El tamaño de la imagen (*IMG*) es un parámetro que funciona como razón de escala de cada imagen sísmica, esto quiere decir que cada imagen ingresada (tanto de entrenamiento como de prueba) es automáticamente escalada en un valor de 320×320 píxeles.
- El tipo de codificación usado fue *timm-efficientnet-b0*, usualmente usada en ambientes GPU.

5.3.2. Capas convolucionales Para realizar las convoluciones se toman subconjunto de píxeles adyacentes formando una matriz que será operada por el *kernel*. No existe un único kernel, por lo que cada sección de las imágenes será operada con su respectivo kernel, aquí también radica la importancia de re-escalar cada imagen sísmica para que el conjunto de operaciones matriciales sea exacto y no sobre ni falte ninguna sección de la imagen.

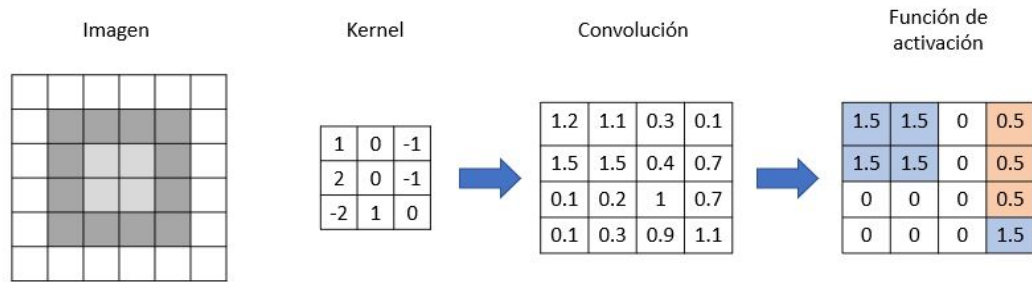


Figura 61. Ejemplo de convolución en una red neuronal convolucional. Imagen adaptada de: [Barrios, 2022].

5.4. Entrenamiento del algoritmo

Del conjunto total de 100 modelos sintéticos se escogió aleatoriamente el 85 % de los modelos para entrenar y los modelos restantes para realizar pruebas. Como se mencionó anteriormente, para crear el modelo de entrenamiento se usó el framework Pytorch. Aquí se empleó los modelos tipo *Unet* empleado usualmente para la detección de máscaras binarias segmentadas de una imagen.

El entrenamiento del modelo se realizó usando la función de optimización ADAM y el entorno de ejecución GPU para los 85 modelos de entrenamiento, una vez entrenado el modelo se procede a verificar su funcionamiento con la función de validación para los 15 modelos sintéticos restantes.

Esta fase se realiza para cada una de las 50 épocas y el rendimiento del entrenamiento es calculado gracias al error medio cuadrado (MSE) que va disminuyendo a medida que se reduzcan las épocas, además el mejor modelo encontrado de acuerdo con la función de validación utilizando el error medio absoluto (MAE) es guardado.

5.5. Repositorio

El código de detección del ruido de dispersión sobre registros sísmicos 2C empleado en este proyecto se encuentra en el repositorio: `https://github.com/HolgerQuintero/seismic-image-segmentation`.

La evaluación del desempeño del método de detección se realizará sobre datos sísmicos sintéticos y reales. Los datos sintéticos usados son generados basados en los datos reales obtenidos en el programa Tenerife el año 2009 en la región del Valle Medio del Magdalena.

Un ejemplo de uno de los 85 datos sísmicos para el entrenamiento de la red neuronal convolucional son los presentados en la figura 60.

6.1. Detección Sobre Datos Sísmicos Sintéticos

Como se ha mencionado anteriormente, el algoritmo de clasificación se entrena con 85 de los 100 registros sísmicos modelados. Una vez es entrenado el modelo bajo la configuración expresada en el capítulo 5 solo basta con ingresar los datos de entrada para los 15 modelos restantes y obtener los resultados de la clasificación.

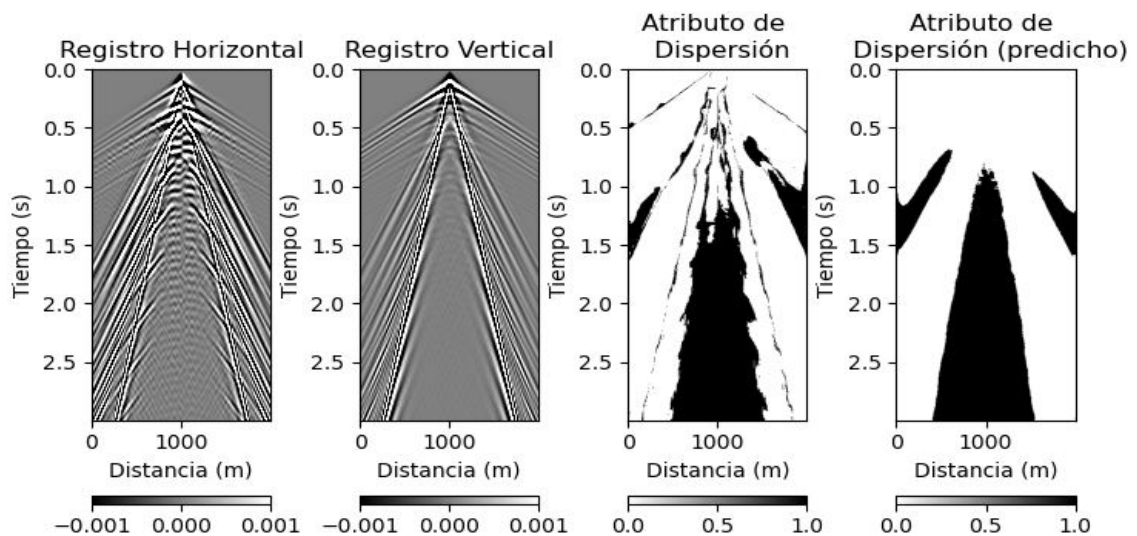


Figura 62. Resultados de la clasificación para un par de datos sísmicos sintéticos.

El modelo obtuvo un valor de error sobre los datos de entrenamiento de 0.51 y un valor de error sobre los datos de prueba de 0.54.

Los resultados presentados en la figura 62 para uno de los modelos sintéticos luego de la clasificación visualmente son favorables, las regiones sombreadas que fueron detectadas o predichas corresponden en su mayoría a las mismas regiones presentes en el atributo de dispersión.

La relación señal ruido determinada para el modelo de la figura 62 fue 28.9%.

Al sobreponer las regiones con ruido de dispersión sobre el registro sísmico se aprecia mejor los resultados de este método (Ver figura 63).

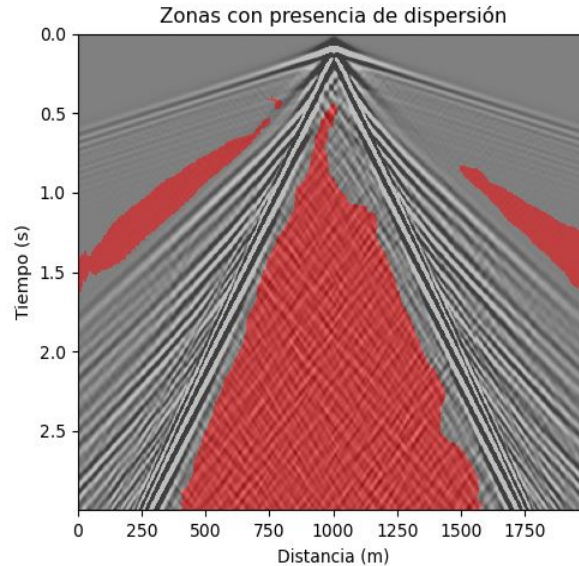


Figura 63. Resultados de la detección del ruido de dispersión en rojo sobre un registro sísmico.

Se observa que la detección del ruido de dispersión sobre registros sintéticos es correcta y la red neuronal es capaz de detectar las regiones con presencia de ruido de dispersión con

facilidad para registros sísmicos sintéticos.

Analizando los posibles resultados sobre otros tipos de modelos como por ejemplo los modelos sin ruido de dispersión, es de esperarse que la solución de la detección del ruido de dispersión para un registro sísmico sin medio aleatorio incrustado debe ser cero o muy cercano a cero.

Los resultados de la detección del ruido de dispersión para un registro sin el medio aleatorio se presentan en la figura 64.

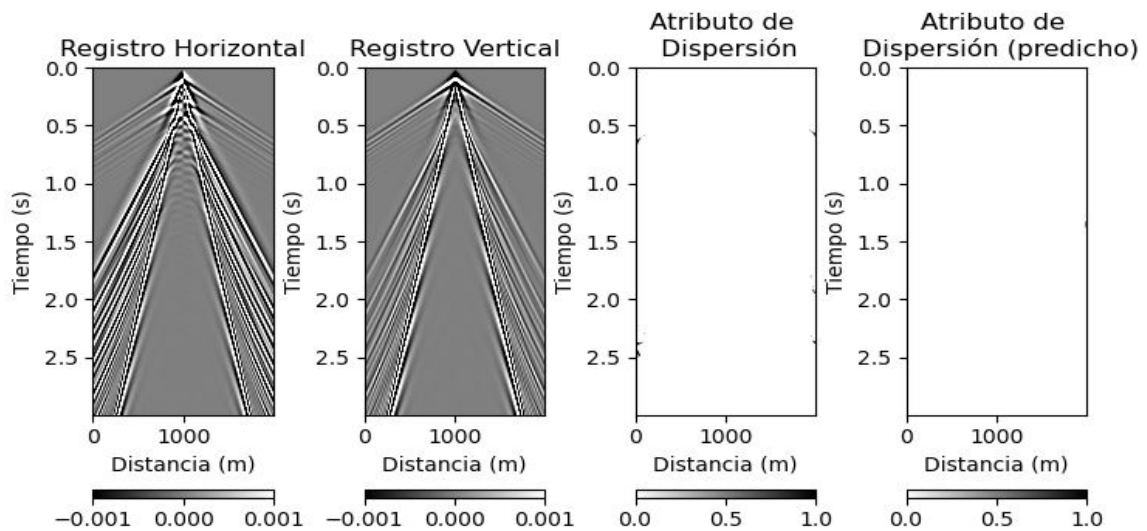


Figura 64. Resultados de la clasificación para un par de datos sísmicos sintéticos.

En este caso se obtuvo una relación señal-ruido de 0.04% y los resultados mostraron que la detección sobre un registro sísmico sin medio aleatorio también se ve afectada o emula los problemas de frontera que presenta el atributo de dispersión, sin embargo, los resultados en este caso también son exitosos.

Finalmente se debe proceder a aplicar este método y realizar un análisis de resultados so-

bre el conjunto de datos reales. Se considera ambicioso obtener unos buenos resultados sobre datos sísmicos reales con tan bajo número de modelos, sin embargo, en la sección 6.2 se presenta este desarrollo.

6.2. Detección Sobre Datos Sísmicos Reales

El desarrollo de este método de detección de ruido de dispersión sobre datos sísmicos reales fue empleado bajo el mismo modelo de clasificación de la sección 6.1 con un total de 85 modelos o datos de entrenamiento usados.

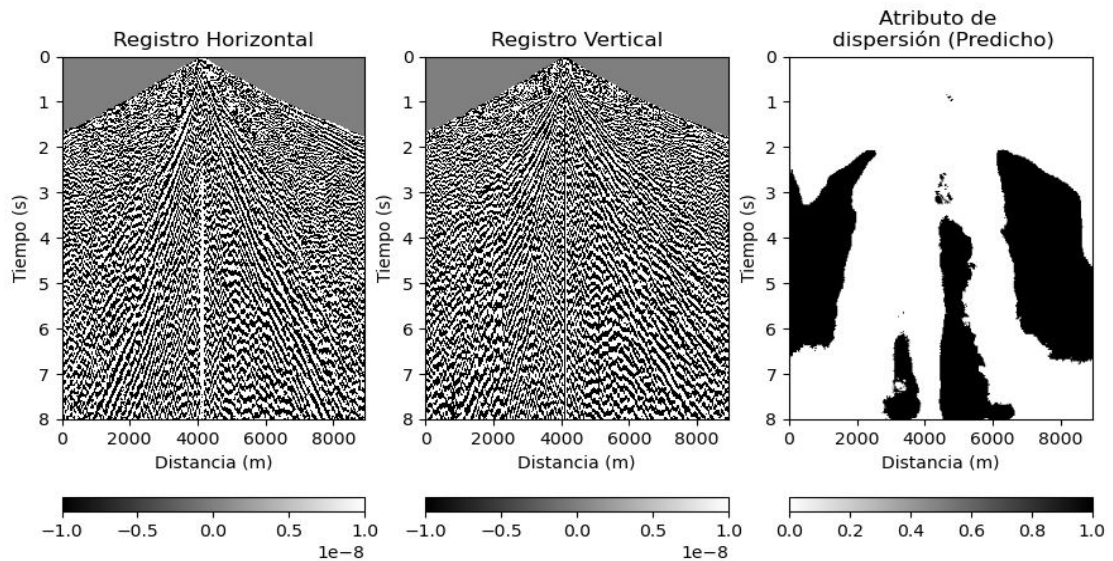


Figura 65. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.

La figura 65 muestra los resultados de la detección de ruido de dispersión. Se logra apreciar que fueron detectadas cuatro regiones con ruido de dispersión en los datos, las dos regiones centrales se atribuyen al ruido presente dentro de las zonas del cono de ondas superficiales, mientras que las regiones fuera del cono son debido a la dispersión registra-

da por los receptores debido a las ondas reflejadas y luego reflejadas.

La relación señal ruido para este conjunto de datos con el modelo para 85 datos de entrenamiento fue del 33.03% , esta cantidad de ruido es considerada baja de acuerdo con lo que se logra apreciar en los registros.

Para ver el comportamiento del método de detección se realizó un nuevo modelo de aprendizaje pero en este caso con 100 datos de entrenamiento. En este caso el modelo obtuvo un valor de error sobre los datos de entrenamiento de 0.41 y un valor de error sobre los datos de prueba de 0.36.

La figura 66 muestra la respectiva clasificación con 100 datos de entrenamiento. La relación señal ruido obtenida para este caso fue 82.4%, un resultado mucho mayor que para el caso de 85 datos de entrenamiento.

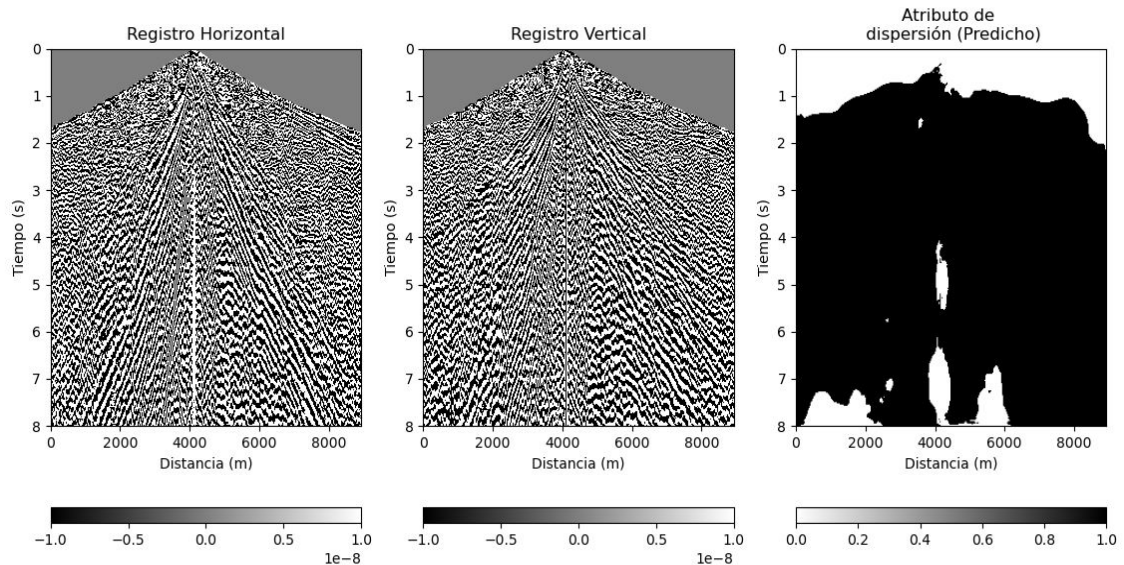


Figura 66. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.

Las regiones detectadas como ruido de dispersión tienden a compactarse en una sola región que abarca la mayoría del registro. También se aprecian zonas o huellas que no son afectadas por la dispersión en regiones de las ondas con mayor tiempo de llegada a los receptores.

Otra información importante que se rescata el uso de más datos de entrenamiento es el clasificador no es capaz de delimitar el ruido de dispersión eficazmente principalmente en las regiones superiores del registro sísmico.

Es de esperarse que a medida que son usados muchos más datos de entrenamiento, la mayoría del registro sísmico se reconocerá como ruido de dispersión, si bien esta afirmación es perjudicial para el objetivo principal de este trabajo, no está del todo equivocado porque en las regiones como el campo Tenerife la mayoría de los datos adquiridos se ven afectados por la dispersión.

Como última prueba se realizará un nuevo modelo de clasificación con un total de 150 datos de entrenamiento (quienes realmente son 300 por las dos componentes de los registros).

En este caso el modelo obtuvo un valor de error sobre los datos de entrenamiento de 0.43 y un valor de error sobre los datos de prueba de 0.23.

La relación señal ruido obtenida fue de 81.8%. Nuevamente se observa que las zonas con ruido de dispersión abarcan casi que el total de datos en el registro. Aquí realizar cambios en el modelo de clasificación del método es necesario para que todas las señales no sean reconocidas como ruido de dispersión (ver figura 67).

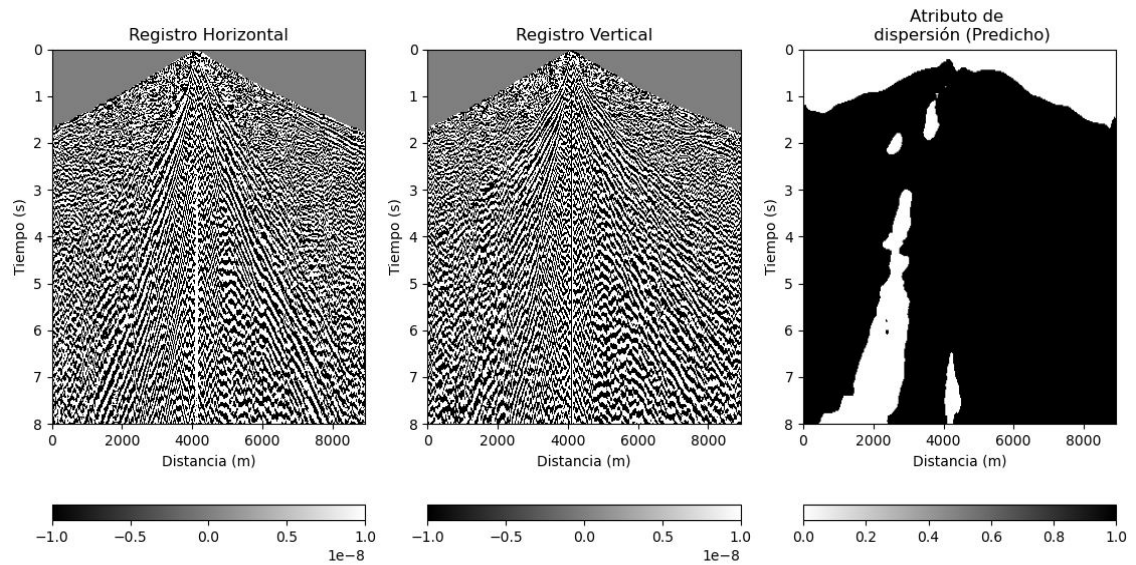


Figura 67. Resultados de la clasificación para un par de datos sísmicos reales adquiridos en el campo Tenerife.

Una forma de mejorar los resultados de la clasificación es realizar un algoritmo que permita crear modelos sísmicos sintéticos que permitan variar las propiedades del modelo base o background y las propiedades del medio aleatorio, en total se estiman $4^5 = 1024$ variaciones de las propiedades del medio o grados de libertad, siendo este número una aproximación para obtener un conjunto más general del número total de modelos óptimos para tener una solución más real.

Para este proyecto se consideró un aproximado del 15% de modelos necesarios estimados para realizar la clasificación.

7.1. Recomendaciones

La detección del ruido de dispersión sobre datos sísmicos sintéticos fue exitosa, sin embargo, la detección del ruido de dispersión para los datos reales puede mejorar. Para este fin es recomendable generar más modelos sísmicos sintéticos basados específicamente en la geología del campo Tenerife, pero evitando caer en un sobreajuste del modelo de clasificación.

Si el conjunto de datos de entrenamiento engloba cualquier tipo de modelo, es decir, si se cuenta con una base de datos de modelos sísmicos que sea muy amplia y general, considerando diferentes tipos de dispersión y diferentes tipos de modelo base o modelo background, entonces este método puede ser aplicable a conjuntos de datos sísmicos de campo de toda índole.

Asimismo, la extracción de atributos sísmicos ha mostrado algunos errores principalmente en los atributos de ventana o atributos multitraza, donde es necesario realizar un manejo de outliers para controlar la estabilidad de estos. La principal razón de la inconsistencia de los atributos de ventana es la estrategia de agrupación simultánea, porque al realizar la elección de las ventanas sobre los registros sísmicos pueden mezclarse trazas sísmicas entre los registros con y sin dispersión, dificultando la extracción del atributo. La agrupación simultánea fue el método desarrollado para detectar la dispersión sísmica numérica, entonces el único camino para evitar la inestabilidad en estos atributos es controlando manualmente los límites de las ventanas a analizar y detenerla justo antes de que las trazas sísmicas de diferentes registros sean combinadas en una misma ventana.

Finalmente, para obtener mejores resultados en la definición del atributo de dispersión

numérico, se podría considerar un conjunto mayor de atributos extraídos que resalten la información sísmica entre trazas, es decir, otorgando más peso a los atributos multitraza. Se intuye que lo anterior contribuirá a detectar un grupo único que resalte la información de la dispersión sin necesidad de realizar una agrupación simultánea.

7.2. Trabajo futuro

Para optimizar el desarrollo del método de detección automática del ruido de dispersión es necesario el acompañamiento de un software de modelado que genere de manera rápida y automática múltiples modelos con dispersión para que de esta forma el método desarrollado se enfoque solamente en la detección del ruido de dispersión en cualquier tipo de registro sísmico sin invertir tiempo y recursos en la parte del modelado. En efecto, existen múltiples softwares de modelado y generación de datos geofísicos, sin embargo, sería sobresaliente la construcción de un software que solo sea empleado para problemas de dispersión sísmica, simulación y generación de datos que minimice los tiempos de construcción del conjunto de datos sísmicos a entrenar.

Asimismo, el método de detección automática de la dispersión sobre registros sísmicos puede llegar convertirse en un atractivo software que ayude al procesamiento de datos sísmicos, cabe aclarar que primero deben ser resueltos todos los problemas con los que aún cuenta el método.

La clasificación binaria permite simplificar los problemas que se presentan al momento de desarrollar la red neuronal convolucional; esta clasificación es útil para definir la cantidad de ruido de dispersión que se estima en los datos sísmicos, sin embargo, el ruido de dispersión también puede ser clasificado dependiendo del tipo de dispersión que presenten los datos. Si logramos clasificar los diferentes tipos de dispersión internamente en este

método y más generalmente si es posible clasificar cualquier tipo de onda en los registros, esto permitiría una mejor interpretación y análisis de cada registro haciendo al método más beneficioso.

Otro cambio que puede hacer el método más atractivo es diseñarlo para cualquier tipo de dato sísmico que se posea, es decir, que el método llegue a ser adaptable a datos sísmicos 1C, 2C y 3C, expandiendo la capacidad de ocupación del método de detección.

El método desarrollado permite detectar la dispersión presente en los registros sísmicos dentro del cono de ondas superficiales, a su vez, es detectado otros tipos de dispersión como por ejemplo las ondas refractadas que se reflejan por los interfaces de las capas más profundas del modelo numérico y al regresar a superficie sufren de dispersión nuevamente, estas ondas se presentan en regiones fuera del cono de onda superficiales (micro-scattering)⁶⁴, esto quiere decir que el método también permite clasificar los diferentes tipos de dispersión presentes en un registro sísmico (por ejemplo dispersión asociada a fuentes o asociada a receptores).

Los resultados de la segmentación binaria de los registros sísmicos fueron exitosos para los datos sintéticos, donde con solo un subconjunto de 85 modelos de entrenamiento se logró obtener una imagen casi igual del atributo de dispersión, sin embargo, esto no ocurrió de la misma manera para los datos sísmicos reales, infiriendo que la cantidad de modelos que se deben emplear para este tipo de datos debe ser mayor debido a su complejidad.

Generalmente los registros sísmicos de baja calidad están asociados a zonas con presencia de ondas dispersadas, estas ondas se ocultan con las reflexiones de las ondas de cuerpo y los sistemas de filtros convencionales no son capaces de detectarlas y filtrarlas. El método desarrollado en este trabajo permitió detectar las trazas sísmicas y las zonas en los datos sísmicos con presencia de dispersión, una vez son detectadas las ondas dispersadas, es posible clasificar las ondas dependiendo de su mecanismo de dispersión, esto además contribuye a mejorar los métodos de filtrado de ruido de dispersión con el fin de obtener

⁶⁴ Stork, "A Model for Land Seismic Surface Scattering Noise has Implications for Acquisition and Processing", ver n. 8.

mejores resultados luego del procesamiento de datos.

Durante décadas, la detección del ruido de dispersión se ha realizado a través de un juicio experto, donde un jurado calificador define las regiones y trazas de un registro sísmico con ruido de dispersión basado en su experiencia, este método es considerado ineficiente y anticuado teniendo en cuenta la existencia de diferentes métodos de automatización de procesos. Actualmente se han planteado métodos de detección para clasificar y agrupar los tipos de ondas sísmicas en shot-gathers^{65,66}, sin embargo, hasta el momento no se ha podido detectar la dispersión en los datos sísmicos de forma automática. Este trabajo contribuye a facilitar la detección de la dispersión en los datos sísmicos pues se hace de manera rápida y automática evitando juicios que pueden llegar a ser subjetivos.

La energía del ruido de dispersión que deteriora las señales es reincidente, por lo tanto, esta energía puede ser predicha⁶⁷, por lo que las variaciones en las señales registradas pueden ser producto de las ubicaciones tanto de receptores como de la fuente, esto quiere decir que cierta cantidad de dispersión registrada en los datos sísmicos puede estar relacionada con el tipo de geometría empleada en las campañas de adquisición, por tanto, el detectar rápidamente el ruido de dispersión en registros sísmicos es significativamente benéfico por que puede ayudar a evitar que los instrumentos sean ubicados donde se registrará la mayor cantidad de ruido en la adquisición en zonas complejas y esto a su vez contribuye potencialmente en el tratamiento de la dispersión al momento de procesar los datos.

⁶⁵ Xia, Hilterman y Hua, ver n. 15.

⁶⁶ Sánchez-Galvis y col., ver n. 17.

⁶⁷ Stork, "A Model for Land Seismic Surface Scattering Noise has Implications for Acquisition and Processing", ver n. 8.

Bibliografía

- Asociación Nacional de Hidrocarburos, *Colombia Petrolera*. (vid. pág. 26).
- J. M. Royero y J. Clavijo., *Mapa geológico generalizado del departamento de Santander* (2001). (vid. pág. 50).
- Abdulaziz, M. y Nafi Toksöz. “Numerical modeling of elastic-wave scattering by near-surface heterogeneities”. En: *GEOPHYSICS* 79.4 (2014). DOI: 10.1190/geo2013-0208.1 (vid. págs. 42, 44, 46).
- Aki, Keiiti y Bernard Chouet. “Origin of coda waves: Source, attenuation, and scattering effects”. En: *Journal of Geophysical Research* 80.23 (1975), págs. 3322-3342 (vid. pág. 34).
- Aki, Keiiti y Paul G. Richards. *QUANTITATIVE SEISMOLOGY*. 2.^a ed. University Science Books, 2002 (vid. pág. 37).
- Anderson, Neil L. y col. “Geophysical Methods Commonly Employed for Geotechnical Site Characterization”. En: *Transportation Research Board* (2008) (vid. pág. 19).
- Andreas C. Müller and Sarah Guido. *mglearn library*. 2016. URL: https://github.com/amueller/introduction_to_ml_with_python (vid. pág. 81).
- Arias, Mario E. y Alexis Mojica. “Prospección Arqueológica en América Central Mediante el Uso de Métodos Geofísicos”. En: *Revista Geológica de América Central* (2011) (vid. pág. 19).
- Bodine, J. H. “Waveform analysis with seismic attributes”. En: *SEG Technical Program Expanded Abstracts* (1984). DOI: 10.1190/1.1894172 (vid. págs. 60, 61).
- Born, Max y Emil Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 2.^a ed. Cambridge University Press, 1999 (vid. pág. 62).
- Bremaecker, J. Cl. de. “Transmission and reflection of rayleigh waves at corners”. En: *GEOPHYSICS* 23.2 (1956). DOI: 10.1190/1.1438465 (vid. pág. 42).
- Butler, Dwain K. “Near-Surface Geophysics”. En: *Society of Exploration Geophysicists* (2005) (vid. pág. 19).

- Chopra, Satinder y Kurt J. Marfurt. “Seismic attributes - A historical perspective”. En: *GEOPHYSICAL* 70.5 (2005) (vid. pág. 57).
- David Flentge, Christof Stork ans. “Using 2D ring arrays to remove back-scattered surface noise from land seismic data”. En: *SEG Technical Program Expanded* (2015) (vid. pág. 21).
- Di, Haibin, Muhammad Shafi y Ghassan AlRegib. “Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data”. En: *Geophysical Journal International* 215.3 (2018), 1999–2007. DOI: 10.1093/gji/ggy376 (vid. pág. 25).
- Díaz, E. y P. Sava. “Understanding the reverse time migration backscattering: noise or signal?” En: *Geophysical Prospecting* 64.3 (2016), págs. 581 -594. DOI: 10.1111/1365-2478.12232 (vid. pág. 21).
- Förstner, W. y E. Gülch. “A fast operator for detection and precise location of distinct points, corners and centres of circular features”. En: *ISPRS Intercommission workshop. Conference on fast processing of photogrammetric data* (1987), págs. 281-305 (vid. pág. 65).
- Graves, Robert W. “Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences.” En: *Bulletin of the Seismological Society of America* 86.4 (2002), 1091–1106. DOI: 10.1785/BSSA0860041091 (vid. pág. 37).
- Hansen, James D., Jamie K. Pringlea y Jon Goodwinb. “GPR and Bulk Ground Resistivity Surveys in Graveyards: Locating Unmarked Burials in Contrasting Soil Types”. En: *Forensic Science International* 237 (2014), e14-e29. DOI: 10.1016/j.forsciint.2014.01.009 (vid. pág. 19).
- Jurkevics, Andy. “Polarization analysis of three-component array data”. En: *Bulletin of the Seismological Society of America* 78.5 (1988), págs. 1725-1743. DOI: 10.1785/BSSA0780051725 (vid. pág. 68).
- Miller, Richard D., John H. Bradford y Klaus Holliger. “Advances in Near-Surface Seismology and Ground-Penetrating Radar”. En: *Society of Exploration Geophysicists* (2005) (vid. pág. 19).

- Mittet, Rune. “Free-surface boundary conditions for elastic staggered-grid modeling schemes”. En: *GEOPHYSICS* 67 (2002). DOI: 10.1190/1.1512752 (vid. pág. 39).
- Moczo, Peter y col. “3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities”. En: *Bulletin of the Seismological Society of America* 92.8 (2002), págs. 3042-3066. DOI: 10.1785/0120010167 (vid. págs. 38, 39).
- Morozov, Igor B. y Scott B. Smithson. “Instantaneous polarization attributes and directional filtering”. En: *GEOPHYSICS* 61.3 (1996). DOI: 10.1190/1.1444012 (vid. pág. 61).
- Müller, Andreas C. y Sarah Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. 1.^a ed. O’Reilly Media, 2016 (vid. págs. 79, 82).
- Ospina, D., W. M. Agudelo y S. Khurama. “Modelo de estrato somero para el campo Tenerife a partir de la integración de datos de refracción sísmica e integración geológica”. En: *Universidad Industrial de Santander* (2011) (vid. pág. 48).
- Pringle, J. K. y col. “The Use of Geoscience Methods for Terrestrial Forensic Searches”. En: *Earth-Science Reviews* 114 (2012), págs. 108-123. DOI: 10.1016/j.earscirev.2012.05.006 (vid. pág. 19).
- Ramirez, J. y M. Chacón. “Redes neuronales artificiales para el procesamiento de imágenes, una revisión de la última década”. En: *REVISTA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y COMPUTACIÓN* 9.1 (2011), págs. 7-16 (vid. pág. 107).
- Regone, Carl J. “Measurement and identification of 3-D coherent noise generated from irregular surface carbonates”. En: *Carbonate Seismology* (1997), pág. 455 (vid. pág. 22).
- “Suppression of coherent noise in 3-D seismology”. En: *Carbonate Seismology* 17.11 (1998), págs. 1584-1589. DOI: 10.1190/1.1437900 (vid. pág. 23).
- René, R. M. y col. “Multicomponent seismic studies using complex trace analysis”. En: *GEOPHYSICS* 51.6 (1986). DOI: 10.1190/1.1442177 (vid. pág. 61).

- Riyanti, C. D. y G. C. Herman. “Three-dimensional elastic scattering by near-surface heterogeneities”. En: *Geophysical Journal International* 160.2 (2005), 609–620. DOI: 10.1111/j.1365-246X.2005.02492.x (vid. pág. 24).
- Robinson, Don K. y Moujahed I. Al-Husseini. “Technique for reflection prospecting in the Rub’ Al-Khali”. En: *Society of Exploration Geophysicists* 47.8 (1982), págs. 1135-1234. DOI: 10.1190/1.1441377 (vid. pág. 22).
- Sato, H., M. C. Fehler y T. Maeda. *Seismic wave propagation and scattering in the heterogeneous earth*. Springer Science, Business Media, 2012. URL: <https://link.springer.com/book/10.1007/978-3-540-89623-4?noAccess=false> (vid. págs. 24, 51).
- Sato, Haruo. “Wave propagation in one dimensional inhomogeneous elastic media”. En: *Journal of Physics of the Earth* (1979). DOI: 10.4294/jpe1952.27.455 (vid. pág. 31).
- Stein, Seth y Michael Wysession. *An Introduction to Seismology, Earthquakes and Earth Structure*. 1.^a ed. Wiley-Blackwell, 2002 (vid. pág. 32).
- Stork, C. “A Model for Land Seismic Surface Scattering Noise has Implications for Acquisition and Processing”. En: *European Association of Geoscientists Engineers* (2021). DOI: 10.3997/2214-4609.202112562 (vid. págs. 21, 123, 124).
- Stork, Christof. “Removing complex land noise with modern pattern recognition using Machine Learning”. En: *SEG Technical Program Expanded* (2017), pág. 6093 (vid. pág. 25).
- Sánchez-Galvis, Ivan y col. “Seismic attribute selection and clustering to detect and classify surface waves in multicomponent seismic data by using k-means algorithm”. En: *The Leading Edge* 36.3 (2017), págs. 239-248. DOI: 10.1190/tle36030239.1 (vid. págs. 25, 77, 85, 124).
- Thorndike, R. L. *Who Belongs in the Family?* 18.^a ed. Psychometrika, 1953 (vid. pág. 87).
- Tiapkina, Olena y col. “Single-station SVD-based polarization filtering of ground roll: Perfection and investigation of limitations and pitfalls”. En: *GEOPHYSICS* 77.2 (2012). DOI: 10.1190/geo2011-0040.1 (vid. págs. 69, 70).

- Vidale, John E. “Complex polarization analysis of particle motion”. En: *Bulletin of the Seismological Society of America* 76.5 (1986), 1393–1405. DOI: 10.1785/BSSA0760051393 (vid. pág. 68).
- Virieux, Jean. “P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method”. En: *GEOPHYSICS* 51.4 (1986), págs. 889-901. DOI: 10.1190/1.1442147 (vid. págs. 36, 42).
- Xia, Keyao, Fred Hilterman y Hao Hua. “Unsupervised machine learning algorithm for detecting and outlining surface waves on seismic shot gathers”. En: *Journal of Applied Geophysics* 157 (2018), págs. 73-86. DOI: 10.1016/j.jappgeo.2018.07.003 (vid. págs. 25, 66, 85, 124).
- Zahradník, Jirí, Peter Moczo y Frantisek Hron. “Testing four elastic finite-difference schemes for behavior at discontinuities”. En: *Bulletin of the Seismological Society of America* 83.1 (1993), 107–129. DOI: 10.1785/BSSA0830010107 (vid. pág. 37).
- Zeng, Chong y col. “An improved vacuum formulation for 2D finite-difference modeling of Rayleigh waves including surface topography and internal discontinuities”. En: *GEOPHYSICS* 77.1 (2011). DOI: 10.1190/geo2011-0067.1 (vid. págs. 37, 39).