

ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS ADQUIRIDOS
MEDIANTE GEOMETRÍAS COMPRESIVAS USANDO UN ENFOQUE DE
APRENDIZAJE PROFUNDO INTERNO Y EXTERNO

SEBASTIAN RIVERA LEÓN

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2023

ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS ADQUIRIDOS
MEDIANTE GEOMETRÍAS COMPRESIVAS USANDO UN ENFOQUE DE
APRENDIZAJE PROFUNDO INTERNO Y EXTERNO

SEBASTIAN RIVERA LEÓN

Trabajo de Grado para optar al título de
Ingeniero de Sistemas

Director:

Laura Viviana Galvis Carreño

Ph. D. Electrical and Computer Engineering

Codirectores:

Tatiana Carolina Gelvez Barrera

Ph.D. in Engineering

Henry Arguello Fuentes

Ph.D. Electrical and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2023

DEDICATORIA

*Este trabajo esta dedicado para mi familia,
para mis amigos y compañeros a lo largo de mi carrera.
para todos aquellos que me han apoyado a lo largo de mi vida.
Mi titulo profesional es el resultado del trabajo en conjunto
con todas las personas importantes y especiales para mi.*

AGRADECIMIENTOS

Este trabajo de grado fue realizado en el marco del proyecto con código 110287780575 - “Nuevas tecnologías computacionales para el diseño de geometrías de adquisición sísmica 3D con muestreo compresivo para la reducción de costos económicos e impactos ambientales en la exploración de hidrocarburos en cuencas terrestres colombianas”, financiado con recursos provenientes del Convenio 785-2019, suscrito entre la Agencia Nacional de Hidrocarburos, el MINISTERIO y el Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas.

A Tatiana Gelvez por ser mi mentora durante la mayoría del tiempo en mi carrera y formarme tanto como profesional como persona.

A Laura Galvis por ser mi apoyo en el desarrollo de este trabajo.

A los miembros del grupo HDSP por brindarme su apoyo y conocimientos para superarme de manera académica.

A a la Universidad Industrial de Santander por ser mi alma mater.

A Juan Sebastian, Geison y Anderson por su amistad y compañía durante todos estos años.

CONTENIDO

	pág.
INTRODUCCIÓN	13
Planteamiento del problema	15
1 OBJETIVOS	18
2 MARCO DE REFERENCIA	19
2.1 FUNDAMENTOS TEÓRICOS	19
2.1.1 Datos sísmicos.	19
2.1.2 Adquisiciones sísmicas.	20
2.1.2.1 Geometrías de adquisición 2D.	20
2.1.2.1.1 Split-spread.	20
2.1.2.1.2 End-on spread.	21
2.1.2.2 Geometrías de adquisición 3D.	22
2.1.2.2.1 Geometría ortogonal.	22
2.1.3 Aprendizaje profundo.	23
2.1.3.1 Aprendizaje profundo externo.	24
2.1.3.2 Aprendizaje profundo interno.	26
2.1.4 Promax.	27
2.2 ANTECEDENTES DEL TEMA	28
3 METODOLOGÍA PROPUESTA	31
3.1 FORMULACIÓN MATEMÁTICA DEL MODELO DE ADQUISICIÓN	31
3.1.1 Modelo en notación vectorial.	31
3.2 SELECCIÓN DE CONJUNTOS DE DATOS	32

3.3	ALGORITMO DE APRENDIZAJE PROFUNDO DE RECONSTRUCCIÓN DE DISPAROS SÍSMICOS FALTANTES	35
3.3.1	APRENDIZAJE EXTERNO	36
3.3.2	APRENDIZAJE INTERNO	38
3.3.3	APRENDIZAJE EXTERNO E INTERNO	40
3.3.4	INFERENCIA	41
4	CONFIGURACIÓN EXPERIMENTAL	43
4.1	MÉTRICAS	43
4.2	ALGORITMOS DE COMPARACIÓN	44
4.2.1	Algoritmos de reconstrucción de disparos sísmicos.	44
4.2.2	Algoritmos de reconstrucción de datos.	45
4.2.2.1	Eliminación de ruido.	45
4.2.2.2	Escasez.	46
4.2.2.3	Suavidad.	47
4.3	EXPERIMENTOS	47
4.3.1	Configuración experimental.	47
4.3.1.1	Configuración estudio de ablación.	48
4.3.1.2	Configuración con software PROMAX.	48
5	SIMULACIONES Y RESULTADOS	49
5.1	VALIDACIÓN DE HIPERPARÁMETROS	49
5.2	RESULTADOS DE SIMULACIONES	52
5.3	ESTUDIO DE ABLACIÓN	56
5.4	EVALUACIÓN DE RECONSTRUCCIÓN A TRAVÉS DE POST-PROCESAMIENTO MEDIANTE SOFTWARE PROMAX	57
6	CONCLUSIONES	64

7 TRABAJO FUTURO

67

BIBLIOGRAFÍA

68

LISTA DE FIGURAS

	pág.
Figura 1 Visualización de los diferentes tipos de datos sísmicos dependiendo de su dimensionalidad.	21
Figura 2 Geometrías de adquisición 2D.	22
Figura 3 Geometría de adquisición ortogonal junto con dato adquirido en el dominio cross-spread.	23
Figura 4 Esquema de aprendizaje profundo.	25
Figura 5 Esquema de aprendizaje profundo externo para resolver el problema inverso de superresolución.	26
Figura 6 Esquema de aprendizaje profundo interno para resolver el problema inverso de superresolución.	27
Figura 7 Software SeisSpace ProMAX.	28
Figura 8 Modelo de adquisición en el dominio <i>cross-spread</i> . a) <i>Cross-spread</i> con submuestreo de fuentes sísmicas. b) Matriz de adquisición o muestreo Φ . c) Cubo de datos adquirido.	32
Figura 9 Conjuntos de datos de prueba en el dominio <i>cross-spread</i> . a) Un cubo de datos de SEAM phase II . b) Stratton .	34
Figura 10 Conjunto de datos Alaska .	34
Figura 11 Esquema de aprendizaje externo e interno para la reconstrucción de disparos sísmicos.	35
Figura 12 Esquema aprendizaje externo.	37
Figura 13 Esquema de aprendizaje interno.	40
Figura 14 Visualización de red neuronal empleada.	41

Figura 15	Proceso de inferencia para la reconstrucción de disparos en un cubo sísmico.	42
Figura 16	Mapas de calor de las métricas de PSNR y SSIM evaluadas sobre el conjunto de datos SEAM phase II para diferentes combinaciones de ρ , τ y η .	51
Figura 17	Resultados visuales para el disparo 30 del conjunto de datos SEAM phase II .	54
Figura 18	Resultados visuales para el disparo 16 del conjunto de datos Stratton .	55
Figura 19	Resultados visuales para el disparo 5 del conjunto de datos Alaska .	59
Figura 20	Flujos de trabajo empleados en PROMAX para generar la imagen apilada del conjunto de datos Alaska .	60
Figura 21	Imágenes de apilamiento bruto obtenidas con el software PROMAX del a) cubo submuestreado, b) el cubo con disparos reocnstruidos y c) el cubo original completo.	62
Figura 22	Imágenes de apilamiento bruto obtenidas con el software PROMAX del cubo con disparos reocnstruidos con a) el método propuesto, b) Aprendizaje externo c) Aprendizaje Interno.	63

LISTA DE CUADROS

	pág.
Cuadro 1 Resultados obtenidos de la recuperación de disparos sísmicos obtenidos sobre los conjuntos de datos de SEAM phase II y Stratton removiendo el 25 % de sus disparos. Los valores mostrados son el promedio de las métricas obtenidas por cada disparo reconstruido en el cubo de datos. El PSNR y SNR se mide en decibelios [dB] y el MSE se encuentra en un escala de 10^{-3} .	53
Cuadro 2 Comparación de los resultados obtenidos con el método propuesto y los demás trabajos en la recuperación de disparos sísmicos sobre los conjuntos de datos SEAM phase II y Stratton . El PSNR se mide en decibelios [dB].	57
Cuadro 3 Comparación de los resultados obtenidos con el método propuesto sobre los demás trabajos en la recuperación de disparos sísmicos sobre los conjuntos de datos de Alaska . El PSNR se mide en decibelios [dB] y el MSE esta en una escala de 10^{-3} .	58

RESUMEN

TÍTULO: ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS ADQUIRIDOS MEDIANTE GEOMETRÍAS COMPRESIVAS USANDO UN ENFOQUE DE APRENDIZAJE PROFUNDO INTERNO Y EXTERNO *

AUTOR: SEBASTIAN RIVERA LEÓN **

PALABRAS CLAVE: reconstrucción de datos sísmicos, geometría de adquisición, aprendizaje profundo, aprendizaje interno y externo, ProMax.

DESCRIPCIÓN:

La obtención de imágenes sísmicas requiere una configuración densa de receptores y fuentes para obtener datos interpretables de alta calidad para la exploración de hidrocarburos. Sin embargo, las condiciones económicas y medioambientales suelen limitar el número de fuentes empleadas en la adquisición. Los métodos actuales de adquisición de datos sísmicos se centran en reconstruir o interpolar computacionalmente las fuentes que faltan, reduciendo los costos. Sin embargo, la mayoría de los trabajos se basan en algoritmos de optimización, donde la calidad depende de la propia naturaleza del dato; o en el aprendizaje profundo, donde la disponibilidad limitada de datos de entrenamiento también limita el proceso de entrenamiento. Por lo tanto, en este trabajo se propone la integración de aprendizaje profundo interno y externo en las siguientes dos etapas: un paso de aprendizaje de una arquitectura de red neuronal en la tarea de recuperación de fuentes sísmicas empleando aprendizaje profundo externo; y un paso de reentrenamiento de la misma arquitectura de red neuronal bajo el esquema de aprendizaje profundo interno. Los resultados experimentales muestran que el método propuesto supera a los trabajos del estado del arte al evaluar diferentes métricas de error y similitud en los conjuntos de datos sísmicos estudiados.

* Trabajo de grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Laura Viviana Galvis Carreño. Codirectores: Tatiana Carolina Gelvez Barrera y Henry Arguello Fuentes.

ABSTRACT

TITLE: SEISMIC DATA RECONSTRUCTION ALGORITHM ACQUIRED BY COMPRESSIVE GEOMETRIES VIA INTERNAL AND EXTERNAL DEEP LEARNING APPROACH *

AUTHOR: SEBASTIAN RIVERA LEÓN **

KEYWORDS: seismic data reconstruction, acquisition geometry, deep learning, external learning, internal learning, ProMax.

DESCRIPTION:

Seismic imaging requires a dense configuration of receivers and sources to obtain high-quality interpretable data for hydrocarbon exploration. However, economic and environmental constraints often limit the number of sources used in the acquisition. Current seismic data acquisition methods focus on computationally reconstructing or interpolating missing sources, reducing costs. However, most of the work is based on optimization algorithms, where the quality depends on the nature of the data itself, or on deep learning, where the limited amount of training data availability also limits the training process. Therefore, this work proposes the integration of two frameworks, external deep learning, and internal deep learning, to reconstruct missing sources. The proposed scheme consists of two stages: a learning step of a neural network architecture in the seismic source reconstruction task employing external deep learning; and a re-training step of the same neural network architecture under the internal deep learning scheme. Experimental results show that the proposed method outperforms state-of-the-art methods when evaluating several error and similarity metrics in seismic datasets.

* Bachelor Thesis

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Laura Viviana Galvis Carreño. Co-advisors: Tatiana Carolina Gelvez Barrera y Henry Arguello Fuentes

INTRODUCCIÓN

La adquisición de datos sísmicos permite la estimación de propiedades del subsuelo, así como descubrir reservas de gas y petróleo. El proceso de adquisición emplea fuentes de ondas acústicas que viajan a través del suelo y que son reflejadas al cruzar las diferentes capas del subsuelo. Luego, las ondas reflejadas son capturadas por los receptores situados en la superficie del suelo.

Existen dos tipos de adquisición sísmica: 2D o en dos dimensiones y 3D o tridimensional. La principal diferencia entre estos tipos es la distancia entre las líneas sísmicas o densidad de malla que es mayor en la sísmica 3D, por lo que la sísmica 2D aporta información solo en un plano vertical y en su lugar la sísmica 3D permite mayor detalle para verificar la presencia o no de hidrocarburos. La geometría ortogonal, es la más empleada en sísmica 3D para la exploración de suelos colombianos, posiciona líneas densas de receptores y líneas densas de fuentes que se intersecan entre sí de manera ortogonal. La intersección entre una línea de fuentes y una línea de receptores se denomina cross-spread, cuyos datos adquiridos son organizados en un cubo de datos con dimensiones de tiempo, receptores y fuentes. Específicamente, las muestras registradas a través del tiempo por los receptores, producidas por una única fuente, forman una matriz denominada disparo. Las muestras de cada disparo se concatenan en profundidad para organizar el cubo de datos sísmicos.

Las fuentes sísmicas generan ondas acústicas son por medio explosiones de dinamita o por un camión vibroseis. Sin embargo, las limitaciones económicas, dificultad de acceso al terreno, o restricciones ambientales, impiden la implementación de líneas densas de fuentes, afectando la calidad del cubo de datos sísmicos. Debido a esto, el cubo adquirido presenta pocos disparos o grandes distancias de muestreo

entre disparos. Por lo tanto, para el área investigativa es de interés desarrollar técnicas computacionales que permitan estimar los disparos faltantes de un cubo de datos incompleto resolviendo un problema inverso.

La recuperación de disparos ha sido abordada por enfoques variacionales incluyendo propiedades físicas para regularizar el problema inverso. Por ejemplo, muestreo compresivo ^{1,2}, equilibrio consenso (en inglés consensus equilibrium) ³ y la reciprocidad de la función de Green ⁴. No obstante, estos métodos fallan al intentar reconstruir disparos con estructuras bastante complejas como las que normalmente presentan los datos reales. Por otro lado, el aprendizaje profundo es una herramienta computacional que permite reconstruir datos mediante el entrenamiento de una red neuronal a partir de datos externos previamente adquiridos ⁵. Sin embargo, la cantidad reducida de datos sísmicos de entrenamiento disponibles ha limitado la aplicabilidad del aprendizaje profundo externo en la recuperación de disparos ⁶. Por

-
- ¹ Kareth León López et al. "Regular multi-shot subsampling and reconstruction on 3d orthogonal symmetric seismic grids via compressive sensing". En: *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE. 2019, págs. 1-5.
 - ² Ofelia P Villarreal et al. "Compressive sensing seismic acquisition by using regular sampling in an orthogonal grid". En: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, págs. 1-5.
 - ³ Paul Goyes-Peñafiel et al. "A Consensus Equilibrium Approach for 3-D Land Seismic Shots Recovery". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), págs. 1-5.
 - ⁴ Benfeng Wang et al. "Intelligent missing shots' reconstruction using the spatial reciprocity of Green's function based on deep learning". En: *IEEE Transactions on Geoscience and Remote Sensing* 58.3 (2019), págs. 1587-1597.
 - ⁵ Shuhang Tang et al. "Reconstruction of sparsely sampled seismic data via residual U-Net". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2020), págs. 1-5.
 - ⁶ Yuanming Li et al. "Seismic data augmentation based on conditional generative adversarial networks". En: *Sensors* 20.23 (2020), pág. 6850.

su parte, el esquema de aprendizaje profundo interno supera tal limitación ya que emplea únicamente el dato observado para ajustar los parámetros del modelo de aprendizaje y resolver el problema ⁷. Sin embargo, el enfoque de aprendizaje profundo interno no ha sido implementado para la recuperación de disparos sísmicos, y aún presenta falencias en la recuperación coherente de estructuras complejas como las que se presentan en los datos sísmicos.

Por consiguiente, este trabajo de grado propone el desarrollo de un algoritmo de reconstrucción de datos sísmicos faltantes que combine los enfoques de aprendizaje profundo externo e interno.

Planteamiento del problema

La adquisición de datos sísmicos captura información de la estructura del subsuelo, útil en la exploración de recursos como hidrocarburos y otros minerales ⁸. La industria petrolera representa el 40 % de las exportaciones colombianas, por lo que la exploración sísmica en cuencas sedimentarias es un área de interés en el país ⁹. La geometría de adquisición más empleada en Colombia es la geometría ortogonal, en la cual, se intersecan ortogonalmente un conjunto de líneas paralelas de

⁷ Dmitry Ulyanov, Andrea Vedaldi y Victor Lempitsky. “Deep image prior”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 9446-9454.

⁸ Ofelia Villarreal et al. “Seismic source reconstruction in an orthogonal geometry based on local and non-local information in the time slice domain”. En: *Journal of Applied Geophysics* 170 (2019), pág. 103846.

⁹ Edgar Eduardo Yáñez Angarita et al. “Rapid screening and probabilistic estimation of the potential for CO₂-EOR and associated geological CO₂ storage in Colombian petroleum basins”. En: *Petroleum Geoscience* 28.1 (2022), petgeo2020-110.

fuentes y receptores ¹⁰. En la adquisición sísmica 2D y 3D ortogonal el dato sísmico consiste en un volumen de datos de dimensiones tiempo, receptores y fuentes, el cual permite analizar los eventos de reflexión del subsuelo ¹¹. La calidad de los eventos observados depende de la cantidad de receptores y fuentes utilizados en la geometría. No obstante, tal cantidad está limitada por factores ambientales y económicos ^{2, 8}. Particularmente, Colombia presenta una topografía irregular, por lo que requiere la detonación de explosivos como fuentes de energía, generando un impacto negativo al medio ambiente, e incurriendo en altos costos de equipos especializados, detonadores y explosivos ¹². Recientemente, la comunidad científica e industria se han enfocado en el diseño de geometrías que reduzcan la cantidad de receptores o fuentes, manteniendo la calidad del dato sísmico. De particular interés, la teoría de adquisición compresiva se ha aplicado en datos sísmicos tal que es posible eliminar fuentes o receptores de forma estratégica y recuperar la información faltante mediante algoritmos computacionales ¹. Sin embargo, la recuperación de datos sísmicos a partir de medidas comprimidas es un problema mal condicionado con soluciones infinitas ³. Por lo que el desarrollo de algoritmos de reconstrucción es un área de investigación que requiere la incorporación de diferentes estrategias para obtener soluciones fiables para la industria.

El aprendizaje profundo es una herramienta computacional que permite reconstruir

¹⁰ John Mora y Mauricio Rocha. "Implementación del dominio del CROSS SPREAD en procesamiento sísmico 3D en cuencas colombianas". En: *11th Simposio Bolivariano-Exploracion Petrolera en las Cuencas Subandinas*. European Association of Geoscientists & Engineers. 2012, cp-330.

¹¹ CPA Wapenaar. "3-D migration of cross-spread data: Resolution and amplitude aspects". En: *Geophysics* 62.4 (1997), págs. 1220-1225.

¹² A Mikhailiuk y A Faul. "Deep learning applied to seismic data interpolation". En: *80th EAGE Conference and Exhibition 2018*. Vol. 2018. 1. European Association of Geoscientists & Engineers. 2018, págs. 1-5.

datos sísmicos mediante el entrenamiento de una red neuronal a partir de datos previamente adquiridos, aprendiendo a estimar los receptores y fuentes faltantes ¹³. No obstante, la cantidad de datos de entrenamiento disponibles es muy reducida debido a la dificultad de adquisición de datos sísmicos ¹. Como consecuencia, han surgido trabajos del estado del arte que usan un esquema de trabajo denominado aprendizaje profundo interno donde se entrenan redes neuronales para recuperar datos sísmicos haciendo uso únicamente del dato sub-muestreado adquirido ¹³. Aunque los algoritmos mencionados no dependen de una base de datos para entrenar, estos tienen falencias en cuanto a la estimación de recuperación consistente de la estructura de los datos sísmicos ¹³.

Actualmente existen algoritmos de aprendizaje profundo que integran los esquemas tanto de aprendizaje profundo externo como el interno, consiguiendo un modelo que alcanza un nivel aceptable de generalización para realizar una tarea con cualquier dato, y que a la vez se ajusta a la distribución a la que pertenece el dato sub-muestreado. Por lo tanto, la precisión obtenida es más alta en el desarrollo de diferentes tareas tales como repintado de imágenes, transferencia de estilo, recuperación de imágenes hiperespectrales, etc ^{14,15,16}.

¹³ Yanghua Wang. "Seismic trace interpolation in the fxy domain". En: *Geophysics* 67.4 (2002), págs. 1232-1239.

¹⁴ Tao Zhang et al. "Hyperspectral image reconstruction using deep external and internal learning". En: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, págs. 8559-8568.

¹⁵ Haibo Chen et al. "Artistic style transfer with internal-external learning and contrastive learning". En: *Advances in Neural Information Processing Systems* 34 (2021), págs. 26561-26573.

¹⁶ Tengfei Wang, Hao Ouyang y Qifeng Chen. "Image inpainting with external-internal learning and monochromic bottleneck". En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, págs. 5120-5129.

1. OBJETIVOS

Objetivo general

- Desarrollar un algoritmo de aprendizaje profundo con enfoque de aprendizaje interno y externo para reconstruir datos sísmicos a partir de medidas adquiridas con geometrías compresivas.

Objetivos específicos

1. Formular matemáticamente el modelo de adquisición de datos sísmicos usando geometrías compresivas.
2. Seleccionar un conjunto de bases de datos sísmicos sintéticos para realizar simulaciones de reconstrucción de datos adquiridos con geometrías compresivas.
3. Diseñar e implementar un esquema de aprendizaje profundo que permita la reconstrucción de datos sísmicos incorporando un enfoque de aprendizaje interno y externo.
4. Evaluar la efectividad del algoritmo implementado en la reconstrucción de datos sísmicos adquiridos con geometrías compresivas respecto a la calidad de reconstrucción de métodos del estado del arte.
5. Evaluar la reconstrucción de datos sísmicos utilizando el esquema de aprendizaje profundo en imágenes obtenidas a través del post-procesamiento mediante el software ProMax.

2. MARCO DE REFERENCIA

2.1. FUNDAMENTOS TEÓRICOS

En el desarrollo de este trabajo de grado se abordará el estudio de datos sísmicos y se trabajará con conceptos del área como geometrías de adquisición, específicamente se enfocará en la geometría 2D y de adquisición ortogonal 3D, siendo estas las más empleadas en el territorio colombiano. A su vez, se detallará sobre el aprendizaje profundo, así como sus diferentes metodologías de implementación como lo son el aprendizaje externo y el aprendizaje interno.

2.1.1. Datos sísmicos. Los **datos** sísmicos son el resultado de la captación de las ondas acústicas, generadas por una fuente sísmica (detonación de dinamita o camión vibroseis) que son adquiridas por los receptores (geófonos) que se encuentran situados en el terreno. Estas ondas acústicas transmitidas por las fuentes son percibidas por los receptores debido a que estas se reflejan en el suelo cada vez que viajan hacia una nueva capa del subsuelo o chocan con algún nuevo medio presente en el terreno. Los datos sísmicos contienen información acerca de las propiedades del suelo y permiten encontrar reservas tanto de petróleo o de gas.

Los receptores captan las ondas reflejadas a través del suelo y miden las amplitudes y velocidades con las que se reflejan estas ondas acústicas permitiendo determinar las estructuras geológicas del subsuelo, este registro se conoce como **traza sísmica**.

La concatenación de cada una de las trazas sísmicas captadas por cada receptor genera un dato sísmico de dos dimensiones conocido como **disparo sísmico**, el

cual presenta la información obtenida por los receptores correspondiente a las ondas generadas por una sola fuente sísmica.

Por último, el dato sísmico con mayor cantidad de información del subsuelo es el dato sísmico tridimensional conocido también como **cubo sísmico**, estos consisten en el ordenamiento consecutivo de disparos sísmicos. Los cubos sísmicos manifiestan cada uno de los registros obtenidos por un arreglo de receptores (disparo sísmico) por cada una de las fuentes usadas durante el estudio sísmico.

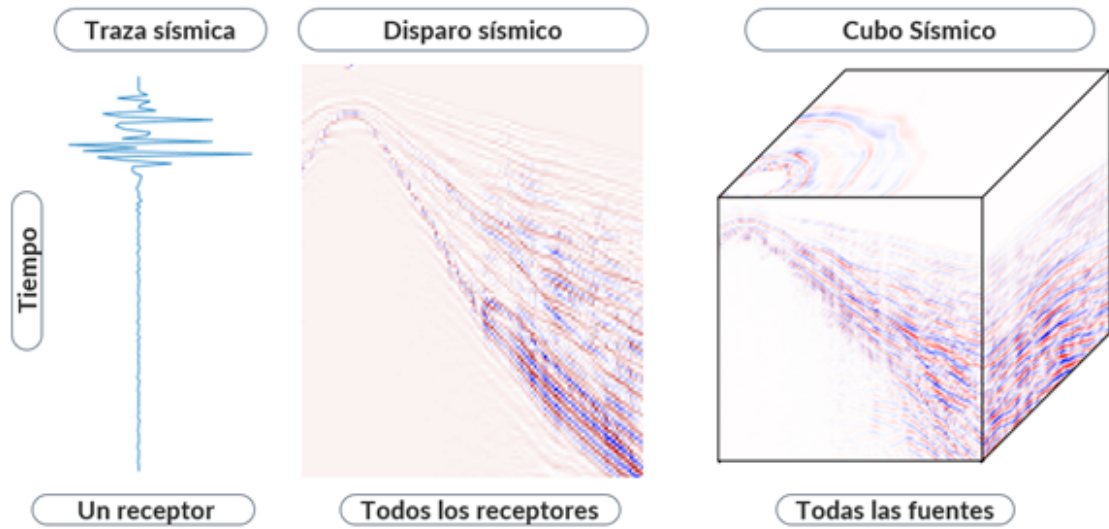
En la **Figura 1** se pueden observar los distintos tipos de datos sísmicos que se pueden obtener en el proceso de exploración sísmica. En esta se presenta una medición de un receptor, la cual es una traza sísmica, de igual manera, se presenta una concatenación de todas las trazas sísmicas que conforman un disparo. Por último, está el apilamiento consecutivo de disparos, el cual resulta en un cubo de datos sísmicos.

2.1.2. Adquisiciones sísmicas.

2.1.2.1. Geometrías de adquisición 2D. Estas geometrías se basan en el posicionamiento de todos los receptores sobre una misma línea y las fuentes generadoras de las ondas acústicas también son posicionadas sobre esta misma línea resultando en una imagen 2D del suelo.

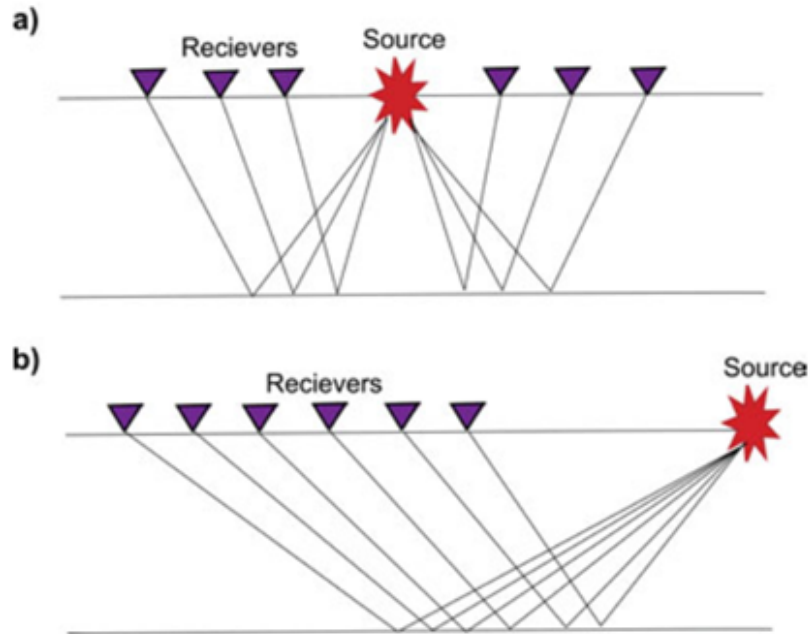
2.1.2.1.1. Split-spread. En la geometría de adquisición *split-spread* la fuente se posiciona justo en la mitad de la línea de receptores y así los receptores se posicionan de manera equiespaciada de manera que se obtenga la resolución de la imagen del suelo deseada.

Figura 1. Visualización de los diferentes tipos de datos sísmicos dependiendo de su dimensionalidad.



2.1.2.1.2. End-on spread. En la geometría *end-on spread*, la fuente de las ondas acústicas está situada a uno de los extremos de la línea de receptores. Por ejemplo, la fuente puede estar al inicio o al final de la línea de receptores. Esta geometría es una de las más empleadas debido a que la onda viaja en la misma dirección para todos los receptores.

Figura 2. Geometrías de adquisición 2D.



En la **Figura 2** a) se ilustra la geometría de adquisición split-spread y b) detalla la geometría de adquisición end-on spread.

2.1.2.2. Geometrías de adquisición 3D. En el caso de adquisiciones sísmicas 3D, se tienen diferentes geometrías de adquisición de datos sísmicos de acuerdo con el modo en el que se posicionan los arreglos de receptores y fuentes en el terreno para ejecutar el proceso de adquisición. Existen diferentes tipos de geometrías de adquisición tales como la geometría ortogonal, paralela, en zigzag, etc. Sin embargo, debido a que la geometría de adquisición más empleada en Colombia es la geometría ortogonal, se profundizará acerca de esta.

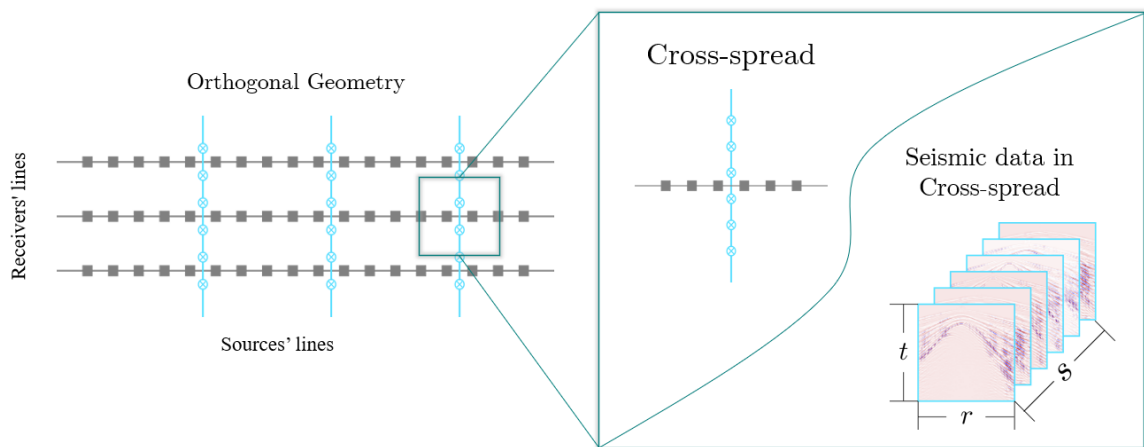
2.1.2.2.1. Geometría ortogonal. La **geometría ortogonal** consiste en un arreglo de líneas de receptores ubicadas paralelamente entre sí que se cruzan con líneas de fuentes de manera ortogonal, las líneas de fuentes también son paralelas

entre sí.

Los datos adquiridos bajo este tipo de geometría se encuentran en el dominio *cross-spread* donde se obtiene un cubo de datos por cada una de las intersecciones que se produzcan entre una línea de receptores y una línea de fuentes.

En la **Figura 3** se visualiza el posicionamiento de las fuentes y los receptores en la geometría de adquisición ortogonal junto con la unidad mínima de adquisición de esta geometría, la cual se da por la intersección entre una línea de fuentes y una línea de receptores. El dato muestreado sobre esta intersección es un cubo de datos en el dominio *cross-spread*.

Figura 3. Geometría de adquisición ortogonal junto con dato adquirido en el dominio *cross-spread*.



2.1.3. Aprendizaje profundo. El aprendizaje profundo (DL, por sus siglas en inglés Deep Learning) es un conjunto de algoritmos de aprendizaje automático donde se hace uso de un conjunto de datos de entrenamiento para “enseñar” al algoritmo de aprendizaje profundo a resolver una tarea específica.

Estos algoritmos de aprendizaje profundo (denominados también redes neuronales) consisten en neuronas, las cuales realizan una operación específica, conectadas con otras neuronas de manera consecutiva hasta que se produce la salida deseada por el algoritmo. Cada neurona de una red neuronal presenta parámetros optimizables de modo que se ajusten para resolver la tarea especificada y a la naturaleza del conjunto de datos.

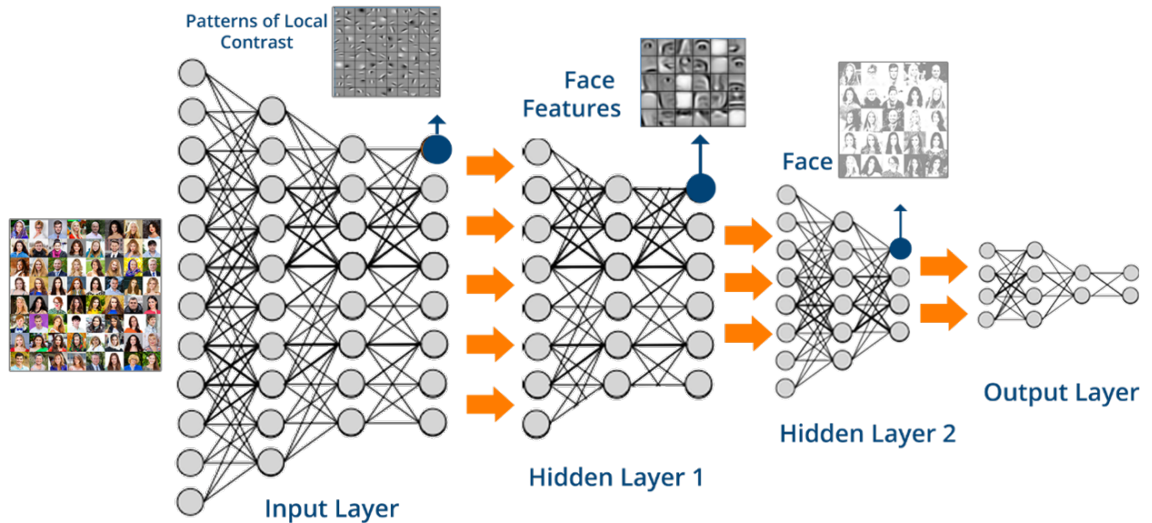
El conjunto de datos de entrenamiento consiste en una gran cantidad de parejas de información que contienen la entrada y etiqueta, donde la entrada es aquello que el algoritmo de aprendizaje profundo recibe; mientras que la etiqueta es lo que se espera que el modelo genere como salida.

Para optimizar los parámetros de la red neuronal se emplea una función de error entre la salida de la red neuronal y la etiqueta real, esta función es minimizada durante el proceso de entrenamiento; la red recibe todo el conjunto de datos de entrenamiento y, por medio de un algoritmo de retropropagación se actualiza el valor de los parámetros de cada una de las neuronas de la red de manera iterativa. De esta manera, se busca que la red aprenda a generar salidas similares a las etiquetas reales al minimizar la función de error definida.

En la **Figura 4** se visualiza el proceso de aprendizaje tradicional (externo) sobre una red neuronal que recibe una gran cantidad de datos de entrada, en este caso imágenes, para ajustar sus parámetros de manera iterativa para aumentar la precisión al momento de ejecutar la tarea específica del esquema.

2.1.3.1. Aprendizaje profundo externo. Esta metodología basada en aprendizaje profundo aborda un esquema de trabajo supervisado, el cual consiste en el

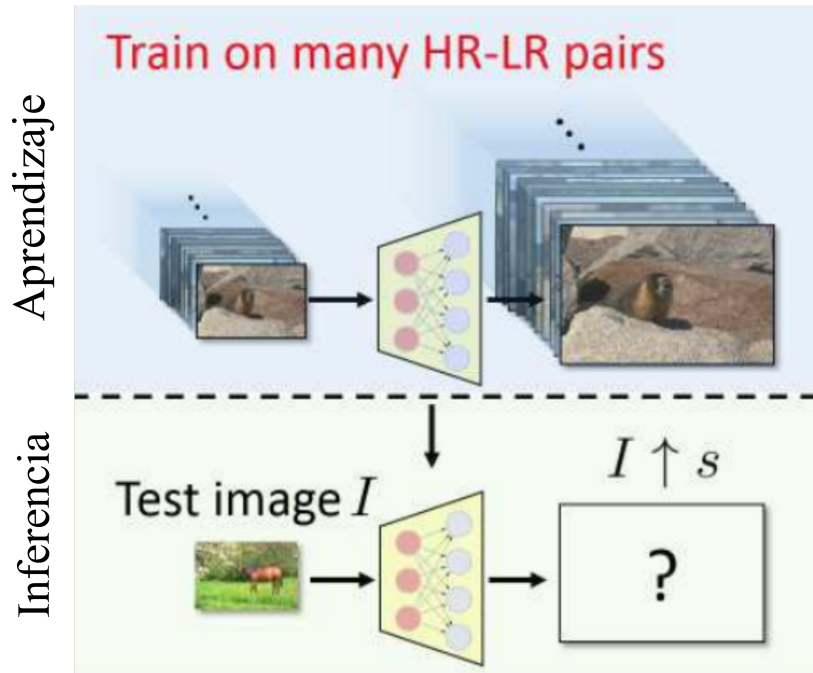
Figura 4. Esquema de aprendizaje profundo.



aprendizaje de una red neuronal a resolver una tarea específica haciendo uso de un conjunto de datos etiquetados, es decir, cada dato de entrada está emparejado con la salida esperada de la red neuronal. El aprendizaje profundo externo se enfoca en que la red neuronal aprenda una representación general de los datos, extrayendo sus características más importantes, esto se logra gracias a la gran cantidad de datos que procesa la red neuronal durante su entrenamiento.

La **Figura 5** ilustra el proceso de aprendizaje profundo externo sobre una red neuronal para hacer la tarea de superresolución; para esto la red neuronal recibe como entrada una gran cantidad de imágenes en baja resolución y ajusta sus parámetros comparando las imágenes de salida con las imágenes de alta resolución reales. En la etapa de prueba o inferencia, se usa como entrada a la red neuronal una imagen que no haya sido usada en la etapa de aprendizaje y se espera que la salida sea una versión de alta resolución de la imagen en cuestión.

Figura 5. Esquema de aprendizaje profundo externo para resolver el problema inverso de superresolución.



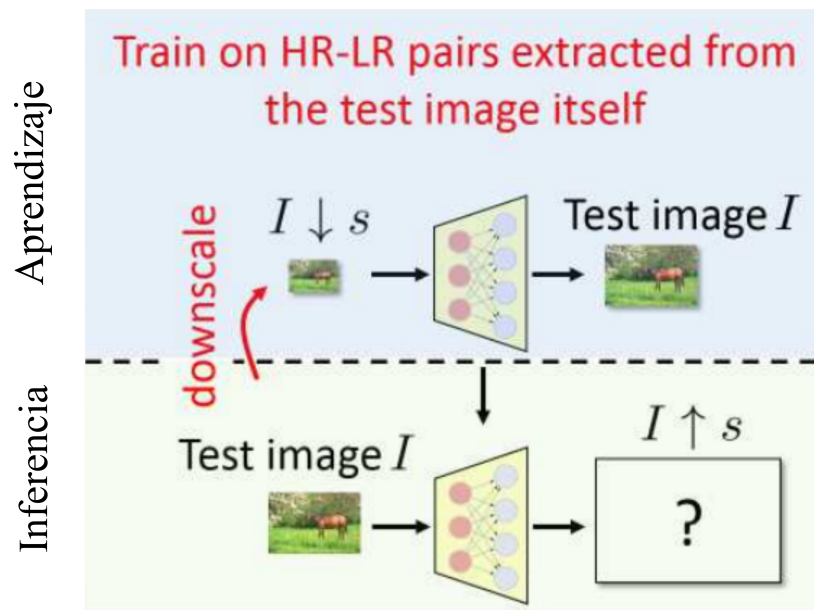
2.1.3.2. Aprendizaje profundo interno. El aprendizaje profundo interno consiste en una metodología de entrenamiento no supervisado, esto quiere decir que los datos de entrenamiento no están etiquetados. Así mismo, este esquema de entrenamiento, aparte de no poseer datos etiquetados de entrenamiento, solo cuenta con un único dato para el aprendizaje. Por lo tanto, en el aprendizaje profundo interno se busca que la red neuronal aprenda la estructura y la distribución interna de los datos con el fin de estimar la etiqueta desconocida del dato único disponible.

La implementación de este tipo de aprendizaje se ha hecho de diferentes maneras, la primera de ellas inicializa los parámetros de las redes neuronales de manera aleatoria, usa el dato de entrada como la etiqueta de entrenamiento y ruido gaussiano como dato de entrada ⁷. También existen implementaciones de aprendizaje interno donde se busca aprender la distribución de color de una imagen con el fin

de restaurar el color en zonas reconstruidas por un algoritmo de reconstrucción de imágenes ¹⁶.

En la **Figura 6**, siguiendo el ejemplo de aprendizaje externo presentado previamente, se muestra el proceso de aprendizaje profundo sobre una red neuronal para hacer la tarea de superresolución. Para la etapa de entrenamiento, se toma el único dato disponible y se reducen sus dimensiones, de esta manera se genera una muestra de aprendizaje artificial. Esta imagen con dimensiones reducidas se pasa como entrada a la red y se compara la salida con la imagen original. En la etapa de inferencia, se toma como entrada la imagen original y se espera que la salida de la red neuronal sea una versión de alta resolución de la imagen de entrada.

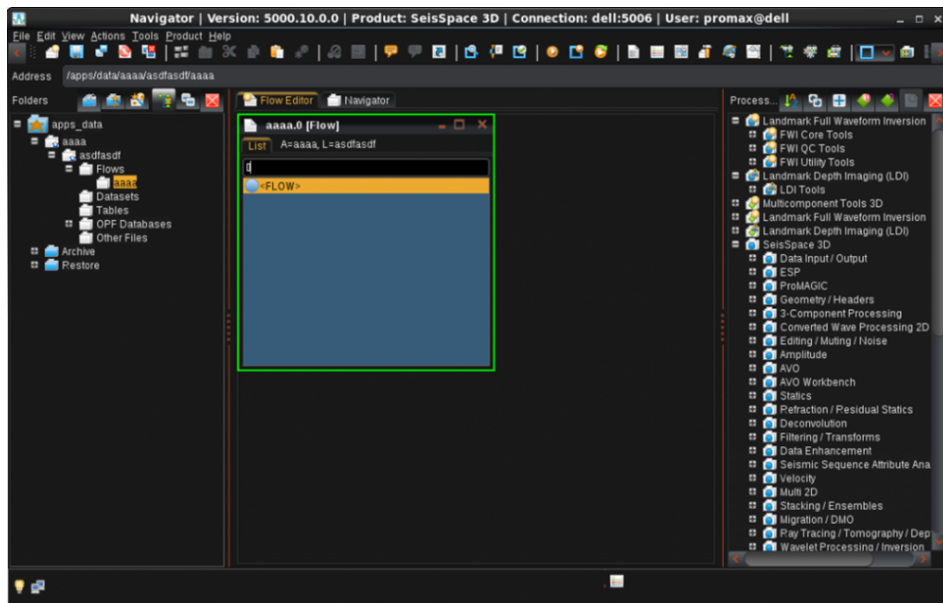
Figura 6. Esquema de aprendizaje profundo interno para resolver el problema inverso de superresolución.



2.1.4. Promax. SeisSpace ProMAX es un software de procesamiento, el cual permite eliminar ruido presente en datos sísmicos adquiridos durante las explora-

ciones de suelos. ProMAX permite mantener múltiples flujos de trabajo de manera paralela; también se puede realizar un mejor análisis de las amplitudes presentes en los datos sísmicos. ProMAX habilita la visualización y aplicación de diferentes técnicas de procesamiento a datos sísmicos 3D, por lo que es utilizado dentro del flujo de post-procesamiento de datos sísmicos para crear lo que se conoce como imagen sísmica. A partir de esta imagen sísmica se pueden efectuar comparaciones de datos sísmicos con datos faltantes y datos recuperados, de tal forma que se puede realizar una evaluación adicional a las reconstrucciones obtenidas a través del método propuesto. En la **Figura 7** se puede visualizar la interfaz del software de post-procesamiento ProMAX junto con las funciones que este dispone además de otros complementos tales como un manejador de archivos.

Figura 7. Software SeisSpace ProMAX.



2.2. ANTECEDENTES DEL TEMA

En el caso de recuperación de trazas sísmicas sobre datos sísmicos bidimensionales adquiridos bajo geometrías compresivas, es decir, cuando se reduce la cantidad

de receptores durante la adquisición, se han abordado una gran cantidad de métodos de interpolación para estimar las trazas faltantes. Algunos métodos han utilizado transformaciones de dominio para realizar la estimación de trazas desconocidas. Por ejemplo, Spitz ¹⁷ propuso una interpolación en el dominio de la frecuencia, permitiendo que las trazas faltantes sean estimadas resolviendo un sistema de ecuaciones lineales; Gülünay ¹⁸ transformó los datos al dominio de Fourier para posteriormente interpolarlos teniendo en cuenta los componentes de más baja frecuencia. Otros ejemplos de interpolación en un dominio transformado de los datos se presentan en ^{19, 13}.

Para realizar interpolación de trazas sobre datos sísmicos bidimensionales también se han empleado métodos relacionados con el aprendizaje profundo y redes neuronales, ejemplo de esto, es el trabajo de Kaur ²⁰, que empleó un tipo de red neuronal denominada Generative Adversarial Network (GAN) para estimar las trazas faltantes en datos sísmicos 2D; Mandelli ²¹ empleó una arquitectura de red neuronal del estado del arte denominada U-net para estimar las trazas faltantes y además hacer una eliminación de ruido de los datos. Dentro de los trabajos relacionados con estima-

¹⁷ Simon Spitz. "Seismic trace interpolation in the FX domain". En: *Geophysics* 56.6 (1991), págs. 785-794.

¹⁸ Necati Gülünay. "Seismic trace interpolation in the Fourier transform domain". En: *Geophysics* 68.1 (2003), págs. 355-369.

¹⁹ Mostafa Naghizadeh y Mauricio D Sacchi. "f-x adaptive seismic-trace interpolation". En: *Geophysics* 74.1 (2009), págs. V9-V16.

²⁰ Harpreet Kaur, Nam Pham y Sergey Fomel. "Seismic data interpolation using CycleGAN". En: *SEG technical program expanded abstracts 2019*. Society of Exploration Geophysicists, 2019, págs. 2202-2206.

²¹ Sara Mandelli et al. "Interpolation and denoising of seismic data using convolutional neural networks". En: *arXiv preprint arXiv:1901.07927* (2019).

ción de trazas sobre datos sísmicos bidimensionales, Wang ²² empleó un esquema de aprendizaje interno en el cual a un disparo con trazas faltantes se le eliminan trazas nuevamente, de modo que una red neuronal aprenda a estimar trazas usando el nuevo dato generado artificialmente.

La recuperación de disparos faltantes ha sido estudiada sobre datos sísmicos 2D y 3D, debido a que se ha estudiado su adquisición a través de geometrías compresivas. En particular, O. P. Villarreal ² propuso muestrear de manera regular los disparos y usar la teoría del muestreo compresivo para recuperar los disparos faltantes en el cubo de datos sísmicos; K. L. López ¹ empleó un esquema de muestreo regular combinándose con la teoría de muestreo compresivo, pero diseñándose una matriz de submuestreo ortogonal y simétrica. Además, en P. Goyes-Peñañiel ³, se propuso el uso de la teoría de equilibrio consenso, siendo el esquema de trabajo con mayor flexibilidad al permitir la incorporación de distintos regularizadores al momento de estimar los disparos faltantes.

Por otro lado, también se han propuesto trabajos basados en aprendizaje profundo para estimar los disparos sísmicos faltantes en cubos de datos sísmicos. Por ejemplo, B. Wang ⁴ reconstruye los disparos faltantes con redes neuronales transversalmente a través del cubo de datos basándose en la reciprocidad de la función de Green. Finalmente, S. Tang ⁵ emplea una arquitectura U-net ²³ para estimar las fuentes faltantes en datos sísmicos escasamente muestreados.

²² Qin Wang et al. "Seismic data interpolation using deep internal learning". En: *Exploration Geophysics* 51.6 (2020), págs. 683-697.

²³ Olaf Ronneberger, Philipp Fischer y Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". En: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, págs. 234-241.

3. METODOLOGÍA PROPUESTA

3.1. FORMULACIÓN MATEMÁTICA DEL MODELO DE ADQUISICIÓN

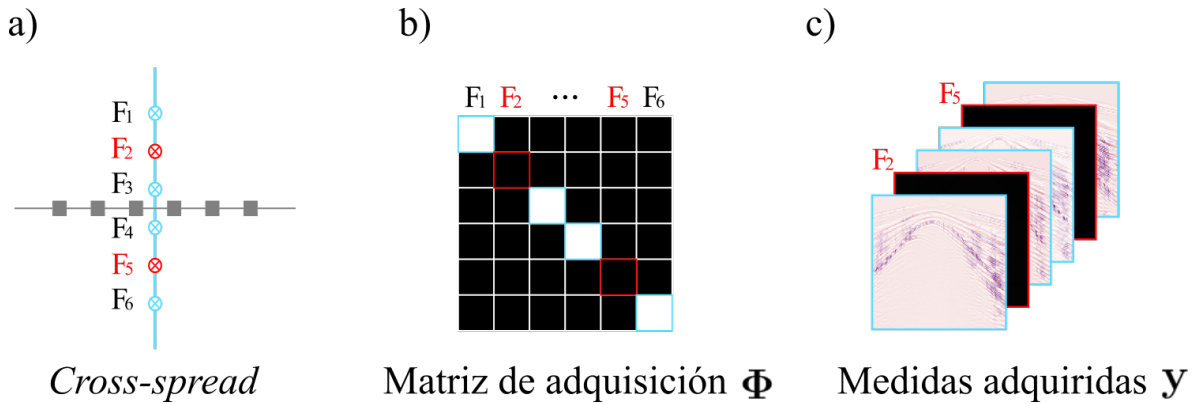
La primera etapa de este proyecto consiste en el modelamiento matemático del proceso de adquisición de datos sísmicos empleando geometrías compresivas removiendo cierta cantidad de fuentes sísmicas, cabe resaltar que tanto la información adquirida por medio de una geometría 2D como la adquirida con una geometría 3D ortogonal en el dominio *cross-spread*, puede ser almacenada en un dato sísmico 3D o cubo sísmico. Un cubo sísmico se denota de manera vectorial como $\mathbf{x} \in \mathbb{R}^{t \cdot r \cdot s}$ donde t representa las muestras de tiempo, r indexa la cantidad de receptores y s la cantidad de fuentes utilizadas en el proceso de adquisición.

3.1.1. Modelo en notación vectorial. En la literatura, el modelamiento de una geometría de adquisición compresiva se ha formulado matemáticamente por medio de un sistema lineal de la forma:

$$\mathbf{y} = \Phi \mathbf{x}, \quad (1)$$

donde $\mathbf{y} \in \mathbb{R}^{t \cdot r \cdot s}$ denota el cubo de datos sísmicos vectorizado, este vector se produce al remover siguiendo alguna estrategia de muestreo un porcentaje de disparos, $\Phi \in \{0, 1\}^{t \cdot r \cdot s \times t \cdot r \cdot s}$ denota la matriz diagonal binaria de muestreo donde los elementos cuyo valor es uno representa los datos medidos y cero los que no fueron adquiridos o presentaron algún problema en la adquisición. En el desarrollo de este proyecto la variable objetivo a recuperar es $\mathbf{X} \in \mathbb{R}^{t \times r \times s}$ ya que ésta representa los datos reales completos del subsuelo.

Figura 8. Modelo de adquisición en el dominio *cross-spread*. a) *Cross-spread* con submuestreo de fuentes sísmicas. b) Matriz de adquisición o muestreo Φ . c) Cubo de datos adquirido.



En la Figura 8 se visualiza la relación entre el esquema de muestreo y la matriz de submuestreo empleada en el modelado matemático. En la Figura 8c) se muestra el cubo sísmico resultante de eliminar durante el muestreo los disparos F_2 y F_5 , los cuales se muestran como matrices de ceros dentro del cubo o medidas adquiridas.

3.2. SELECCIÓN DE CONJUNTOS DE DATOS

Las bases de datos seleccionadas corresponden a datos sísmicos tridimensionales adquiridos bajo geometrías 2D y geometrías 3D ortogonales en el dominio *cross-spread*. Respectivamente se empleó un conjunto de datos adquirido bajo geometrías 2D y dos conjuntos de datos adquiridos bajo geometrías 3D ortogonales. Estos conjuntos de datos fueron los empleados para evaluar el desempeño del algoritmo propuesto en diferentes aspectos.

1. El conjunto de datos **SEAM phase II** ²⁴ corresponde a 72 cubos sísmicos

²⁴ *SEAM Phase II Land Seismic Datasets*. <https://seg.org/News-Resources/Research-Data/>

sintéticos-realistas de dimensiones $1034 \times 100 \times 61$ modelados por un programa de la SEG, ejecutado durante el desarrollo del proyecto "*SEAM Phase II–Land Seismic Challenges*". Durante las pruebas 71 cubos sísmicos fueron empleados para entrenamiento en todos los trabajos que emplean aprendizaje externo y el cubo restante se utilizó para las pruebas que simulan la adquisición submostrada en la Sección 5. Todos los cubos sísmicos fueron redimensionados en ancho y alto a 128×128 , lo que indica que se utilizaron 128 muestras en el tiempo y 128 receptores por cada disparo.

2. **Stratton** es un conjunto de datos sísmicos 3D liberados de manera pública en 2014 por motivos investigativos, adquiridos durante el estudio "*Stratton 3D survey*" realizado en el estado de Texas, Estados Unidos. El cubo de datos en el dominio *cross-spread* tiene dimensiones de $1001 \times 80 \times 18$, sin embargo por facilidades computacionales durante los experimentos, éste fue redimensionado a $128 \times 128 \times 18$, este conjunto de datos fue empleado únicamente para pruebas en la Sección 5.
3. El conjunto de datos **Alaska** es un cubo sísmico 2D, específicamente adquirido con una geometría *split-spread*, compuesto originalmente por 55 fuentes, 96 receptores y 3000 muestras de tiempo. Sin embargo para las simulaciones realizadas en este proyecto, se redujo a 16 fuentes y 500 muestras de tiempo, manteniendo los 96 receptores originales, este conjunto de datos solo fue empleado para las pruebas en la Subsección 5.4.

Land-Seismic-Data. (Accessed on 11/30/2022).

Figura 9. Conjuntos de datos de prueba en el dominio *cross-spread*. a) Un cubo de datos de **SEAM phase II**. b) **Stratton**.

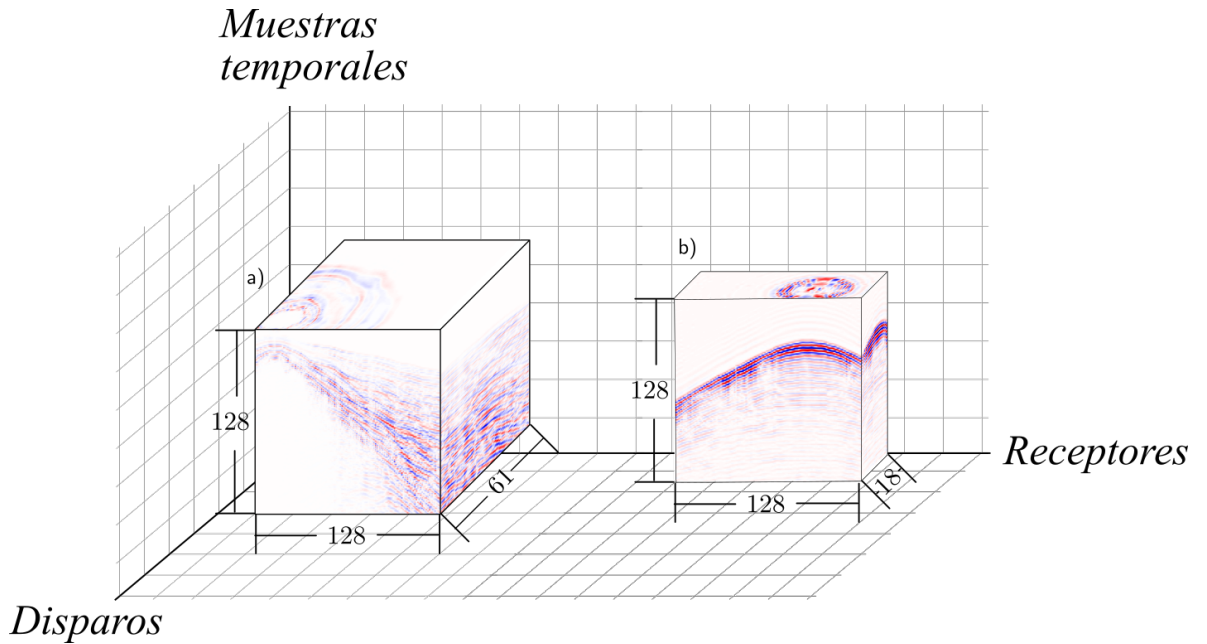
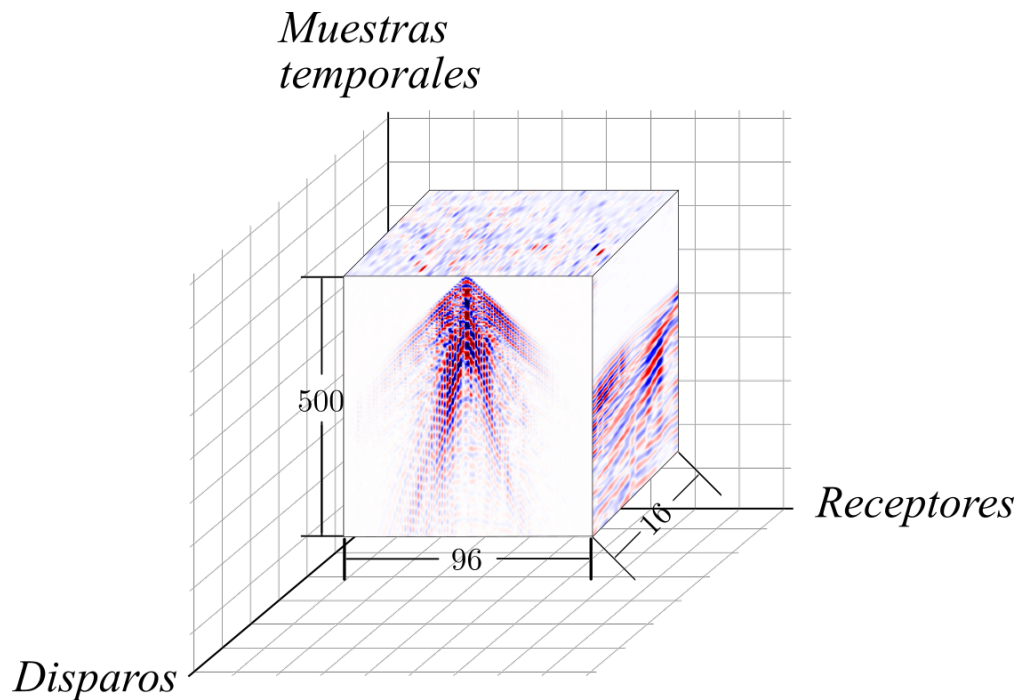


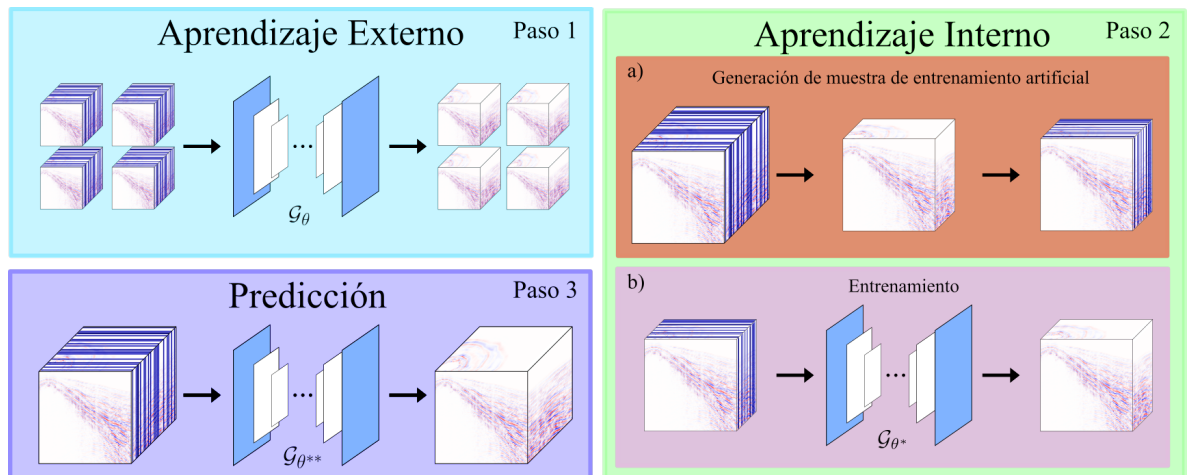
Figura 10. Conjunto de datos **Alaska**.



3.3. ALGORITMO DE APRENDIZAJE PROFUNDO DE RECONSTRUCCIÓN DE DISPAROS SÍSMICOS FALTANTES

El algoritmo de reconstrucción basado en la integración de aprendizaje externo e interno consta de dos etapas principales: (i), aprendizaje externo y (ii) aprendizaje interno. En la Figura 11 se detalla el esquema general para la integración del aprendizaje externo e interno para la recuperación de disparos sísmicos. El primer paso, corresponde al aprendizaje externo en el cual se emplea un conjunto de datos de cubos sísmicos para entrenar una red neuronal cuya tarea es la de reconstruir disparos sísmicos. Posteriormente, en el paso 2 de aprendizaje interno, se genera una sola pareja de entrenamiento a partir del dato incompleto que se busca reconstruir y se realiza nuevamente un reentrenamiento de la red neuronal, ajustando sus parámetros únicamente a esta nueva pareja de datos de entrenamiento. Por último, para el proceso de inferencia, se pasa como entrada el dato submuestreado a la red neuronal y ésta buscará reconstruir los disparos faltantes.

Figura 11. Esquema de aprendizaje externo e interno para la reconstrucción de disparos sísmicos.



3.3.1. APRENDIZAJE EXTERNO En este trabajo, la etapa de aprendizaje externo consiste en emplear un conjunto de datos sísmicos para realizar un pre-entrenamiento de una red neuronal en la tarea de reconstrucción de disparos sísmicos. Específicamente cada uno de los cubos del conjunto de cubos sísmicos disponibles se submuestra siguiendo la ecuación (5) de manera que sean removidos una fracción $\alpha \in [0, 1]$ de todos los disparos en cada cubo. Previamente, los valores contenidos dentro del conjunto de datos son normalizados.

Posteriormente, se emplea una arquitectura de red neuronal, la cual durante la fase de entrenamiento busca resolver el siguiente problema de optimización para hallar los mejores parámetros θ^* que resuelven el problema de reconstrucción de disparos sísmicos:

$$\theta^* \in \arg \min_{\theta} \mathcal{L}(\mathbf{X}, \mathcal{G}_{\theta}(\mathbf{Y})), \quad (2)$$

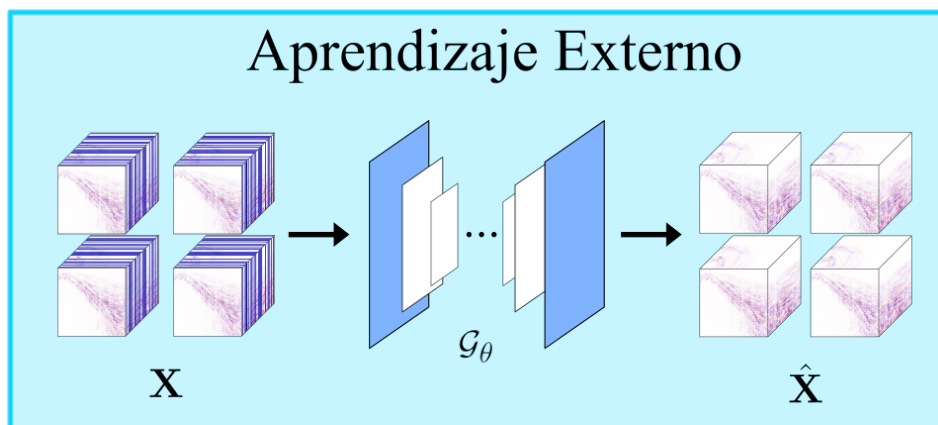
donde \mathbf{X} corresponde al conjunto de datos con las medidas completas y $\mathcal{G}_{\theta}(\mathbf{Y})$ a las reconstrucciones realizadas por la red neuronal de los cubos sísmicos submuestreados. \mathcal{L} es la función de pérdida denominada como error cuadrático medio y se formula matemáticamente de la siguiente manera:

$$\mathcal{L}(\mathbf{X}, \mathcal{G}_{\theta}(\mathbf{Y})) = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathcal{G}_{\theta}(\mathbf{y}_i))^2, \quad (3)$$

donde n corresponde al número de cubos disponibles para entrenamiento. De esta manera el problema de optimización en (2) busca minimizar la diferencia entre los datos sísmicos reales del conjunto de datos de entrenamiento y las reconstrucciones realizadas por la red neuronal. \mathbf{x}_i y $\mathcal{G}_{\theta}(\mathbf{y}_i)$ corresponden a los i -ésimos datos reales del conjunto de datos y la reconstrucción generada por la red neuronal respectivamente.

El Algoritmo 1 detalla el proceso de aprendizaje externo el cual se puede observar en la Figura 12. En la línea 4 se realiza la reconstrucción de los disparos desconocidos de un cubo sísmico, posteriormente en las líneas 5-6 por medio del optimizador Adam ²⁵ \mathcal{A}_{dam} se busca minimizar la función de costo con base al gradiente de la misma respecto a los parámetros de la red neuronal.

Figura 12. Esquema aprendizaje externo.



Algoritmo 1 Enfoque de aprendizaje externo.

- 1: **Entrada:** Conjunto de entrenamiento $\{\mathbf{X}, \mathbf{Y}\}$ y red neuronal \mathcal{G}_θ .
 - 2: Inicialización aleatoria de parámetros θ
 - 3: **para** época = 1: \mathcal{E} **hacer** ▷ \mathcal{E} épocas
 - 4: $\hat{\mathbf{X}} \leftarrow \mathcal{G}_\theta(\mathbf{Y})$ ▷ Reconstrucción
 - 5: $\mathcal{L}_\theta = \mathcal{L}(\mathbf{X}, \hat{\mathbf{X}})$ ▷ Función de costo
 - 6: $\theta \leftarrow \mathcal{A}_{dam}(\theta, \nabla_\theta \mathcal{L}_\theta)$ ▷ Optimización sobre θ
 - 7: **fin para**
 - 8: **Salida:** Parámetros óptimos de la red neuronal θ^* .
-

²⁵ Diederik P Kingma y Jimmy Ba. "Adam: A method for stochastic optimization". En: *arXiv preprint arXiv:1412.6980* (2014).

3.3.2. APRENDIZAJE INTERNO En esta etapa, se propuso un esquema de aprendizaje interno en el que se generan muestras de entrenamiento artificiales a partir de los datos que se desean reconstruir. Este esquema consiste en tomar un cubo sísmico submuestreado con disparos desconocidos $\mathbf{y} \in \mathbb{R}^{t \cdot r \cdot s}$ y concatenar de manera consecutiva solo los disparos conocidos, obteniendo así un cubo artificial completo $\mathbf{z} \in \mathbb{R}^{t \cdot r \cdot (1-\alpha)s}$ por medio de la ecuación:

$$\mathbf{z} = \Psi \mathbf{y} \quad (4)$$

donde $\Psi \in \mathbb{R}^{t \cdot r \cdot s \times t \cdot r \cdot (1-\alpha)s}$. Luego, siguiendo el esquema de submuestreo en (5), se elimina una fracción $\alpha \in [0, 1]$ de los disparos totales, obteniendo un cubo sísmico con una fracción de disparos removidos $\mathbf{z}' \in \mathbb{R}^{t \cdot r \cdot (1-\alpha)s}$ por medio de la ecuación:

$$\mathbf{z}' = \mathbf{H} \mathbf{z} \quad (5)$$

donde $\mathbf{H} \in \{0, 1\}^{t \cdot r \cdot (1-\alpha)s \times t \cdot r \cdot (1-\alpha)s}$. En la Figura 11 en el paso 2a se presenta este proceso, denominado generación de muestra de entrenamiento artificial. Con las nuevas muestras artificiales de entrenamiento, se ajustan los parámetros θ de la red neuronal con el fin de resolver el siguiente problema de optimización:

$$\theta^* \in \arg \min_{\theta} \mathcal{L}(\mathbf{z}, \mathcal{G}_{\theta}(\mathbf{z}')), \quad (6)$$

donde \mathcal{L} corresponde al error cuadrático medio entre el dato sísmico densamente muestreado artificial \mathbf{z} y la reconstrucción realizada por la red neuronal $\mathcal{G}_{\theta}(\mathbf{z}')$. Matemáticamente, esta función de error se denota de la siguiente manera:

$$\mathcal{L}(\mathbf{z}, \mathcal{G}_{\theta}(\mathbf{z}')) = \|\mathbf{z} - \mathcal{G}(\mathbf{z}')\|_2^2. \quad (7)$$

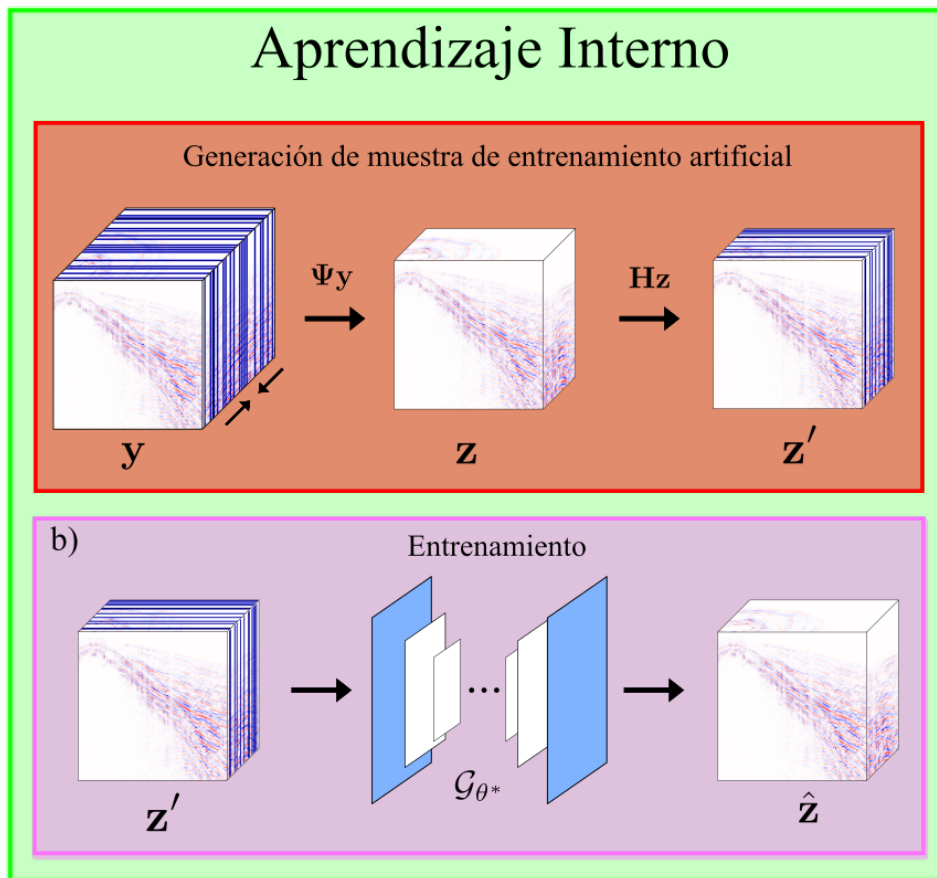
El Algoritmo 2 presenta el proceso completo del aprendizaje interno, el cual se

puede visualizar en la Figura 13. En las líneas 2-3 se generan los datos de entrenamiento artificiales para la red neuronal. En la línea 2, se realiza una multiplicación entre el cubo incompleto vectorizado $\mathbf{y} \in \mathbb{R}^{t \cdot r \cdot s}$ y la matriz de submuestreo $\Psi \in \mathbb{R}^{t \cdot r \cdot (1-\alpha)s \times t \cdot r \cdot s}$, la cual selecciona únicamente los disparos conocidos del cubo de datos. Posteriormente, en la línea 3, se sigue el esquema de submuestro en la ecuación 5 removiendo una fracción $\alpha \in [0, 1]$ de los disparos totales en el cubo de datos. Finalmente, en las líneas 5-7, se llevan a cabo todos los pasos necesarios para actualizar los parámetros de la red neuronal mediante el cálculo de la función de pérdida y su gradiente.

Algoritmo 2 Enfoque de aprendizaje interno.

- 1: **Entrada:** Cubo submuestreado \mathbf{y} , matriz de submuestreo \mathbf{H} , matriz de reducción de tamaño Ψ y red neuronal preentrada \mathcal{G}_θ^*
 - 2: $\mathbf{z} \leftarrow \Psi \mathbf{y}$ ▷ Concatenación de disparos conocidos
 - 3: $\mathbf{z}' \leftarrow \mathbf{H} \mathbf{z}$ ▷ Eliminación de disparos
 - 4: **para** época = 1: \mathcal{E} **hacer** ▷ \mathcal{E} épocas
 - 5: $\hat{\mathbf{z}} \leftarrow \mathcal{G}_\theta^*(\mathbf{z}')$ ▷ Reconstrucción
 - 6: $\mathcal{L}_{\theta^*} = \mathcal{L}(\mathbf{z}, \hat{\mathbf{z}})$ ▷ Función de costo
 - 7: $\theta^* \leftarrow \mathcal{A}_{dam}(\theta^*, \nabla_{\theta^*} \mathcal{L}_{\theta^*})$ ▷ Optimización sobre θ^*
 - 8: **fin para**
 - 9: **Salida:** Parámetros óptimos de la red neuronal θ^{**} .
-

Figura 13. Esquema de aprendizaje interno.



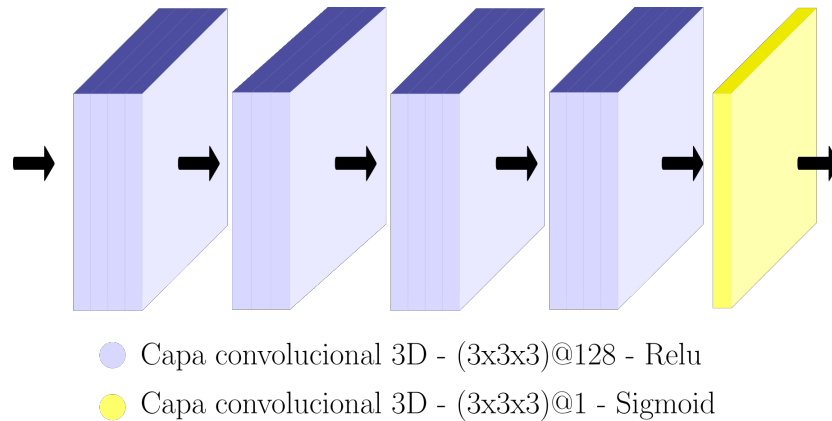
3.3.3. APRENDIZAJE EXTERNO E INTERNO En esta sección se muestra como se integran los dos aprendizajes en un solo esquema de reconstrucción de disparos para cubos sísmicos. En el Algoritmo 3 se puede encontrar esta integración de esquemas de manera más detallada.

Algoritmo 3 Reconstrucción de disparos con enfoque de aprendizaje externo e interno.

- 1: **Entrada:** Conjunto de entrenamiento $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^n$ con n cubos sísmicos, cubo submuestreado \mathbf{y} , matriz de submuestreo \mathbf{H} y red neuronal \mathcal{G}_θ
 - 2: $\mathcal{G}_{\theta^*} \leftarrow \text{AprendizajeExterno}(\mathbf{Y}, \mathbf{X}, \mathcal{G}_\theta)$ ▷ Ajuste de parámetros con aprendizaje externo
 - 3: $\mathcal{G}_{\theta^*} \leftarrow \text{AprendizajeInterno}(\mathbf{y}, \mathbf{H}, \mathcal{G}_{\theta^*})$ ▷ Ajuste de parámetros con aprendizaje interno
 - 4: $\hat{\mathbf{x}} \leftarrow \mathcal{G}_{\theta^*}(\mathbf{y})$
 - 5: **Salida:** Cubo con disparos reconstruidos $\hat{\mathbf{x}}$.
-

En esta etapa se empleó una red neuronal convolucional 3D con los hiperparámetros definidos en la Sección 4.3.1 y la Sección 5.1. Además se empleó como función de pérdida el error cuadrático medio para ambas etapas del método propuesto. En la Figura 14 se muestra una representación visual de la red neuronal junto con algunos de sus hiperparámetros.

Figura 14. Visualización de red neuronal empleada.



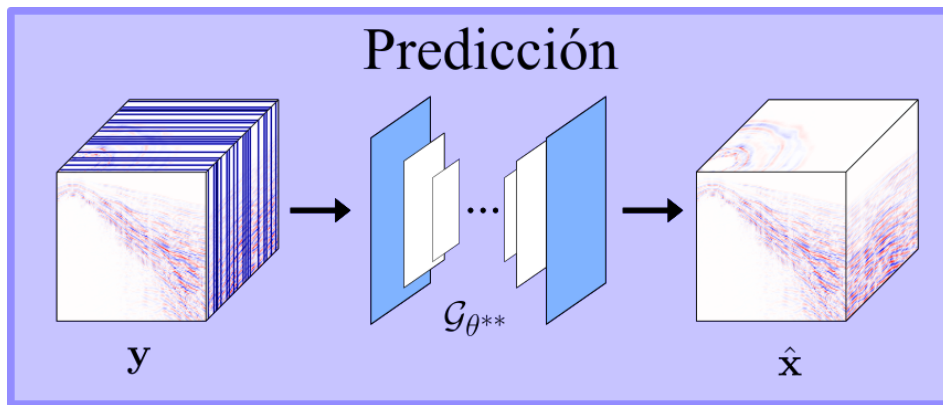
3.3.4. INFERENCIA Finalmente, en la fase de inferencia se utiliza la red neuronal pre-entrenada en las dos etapas anteriores, \mathcal{G}_{θ^*} para estimar los disparos faltantes

del cubo sísmico y y de acuerdo con la ecuación que se presenta a continuación:

$$\mathcal{G}_{\theta^{**}}(y) = \hat{x} \quad (8)$$

donde $\hat{x} \in \mathbb{R}^{t \cdot r \cdot s}$ es una aproximación del cubo sísmico real con los disparos faltantes reconstruidos mediante la utilización de la red neuronal pre-entrenada $\mathcal{G}_{\theta^{**}}$. Esta aproximación se utiliza para completar los datos faltantes en el cubo sísmico original y poder obtener una representación más precisa y completa de las características del subsuelo. En la Figura 15 se presenta el esquema para reconstruir los disparos faltantes del cubo de datos sísmicos.

Figura 15. Proceso de inferencia para la reconstrucción de disparos en un cubo sísmico.



4. CONFIGURACIÓN EXPERIMENTAL

En esta sección se presentan los resultados obtenidos a partir del método propuesto para la reconstrucción de disparos sísmicos empleando una esquema de aprendizaje profundo externo e interno y su comparación con esquemas de reconstrucción del estado del arte. Además se muestran las métricas de evaluación.

4.1. MÉTRICAS

La calidad de los disparos reconstruidos por el esquema propuesto se cuantificaron siguiendo las métricas a continuación:

- Medida del índice de similitud estructural (SSIM):

$$\text{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{2(\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + C_1) + (2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{x}} + \mu_{\hat{\mathbf{x}}} + C_1)(\sigma_{\mathbf{x}}\sigma_{\hat{\mathbf{x}}} + C_1)}. \quad (9)$$

Esta métrica se maneja en el rango de $[0, 1]$ donde los valores más altos indican una mayor similitud entre \mathbf{x} y $\hat{\mathbf{x}}$. Donde $(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}})$ y $(\mu_{\hat{\mathbf{x}}}, \sigma_{\hat{\mathbf{x}}})$ representan la media y la varianza de \mathbf{x} y $\hat{\mathbf{x}}$ respectivamente, además $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$ es la covarianza entre ellos. Para evitar una indeterminación en la división están las constantes $C_1 = (k_1L)^2$ y $C_2 = (k_2L)^2$ con $k_1 = 0,01$ y $k_2 = 0,03$.

- Error cuadrático medio (MSE):

$$\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2. \quad (10)$$

Esta métrica tiene valores en el rango de $[0, \infty)$, donde a menor valor, mejor es la estimación debido a que esta mide el promedio de los errores elevados al cuadrado entre dos señales.

- Relación de señal a ruido máxima (PSNR):

$$\text{PSNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}(\mathbf{x}, \hat{\mathbf{x}})} \right), \quad (11)$$

donde $\text{MAX}(\cdot)$ denota el rango dinámico de la imagen. Esta métrica tiene valores en el rango de $[0, \infty)$, donde a mayor valor, mejor es la estimación dado a que esta define la relación de potencia máxima entre una señal y el ruido que la afecta. Su unidad métrica son los decibelios [dB].

- Relación señal a ruido (SNR):

$$\text{SNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \cdot \log_{10} \left(\frac{\|\mathbf{x}\|_2^2}{\text{MSE}(\mathbf{x}, \hat{\mathbf{x}})} \right). \quad (12)$$

Esta métrica tiene valores en el rango de $[0, \infty)$, donde a mayor valor, mejor es la estimación esto porque mide la relación entre la señal original y el ruido. Esta medida al igual que el PSNR aumenta a medida que el MSE disminuye. Su unidad métrica son los decibelios [dB].

4.2. ALGORITMOS DE COMPARACIÓN

En la literatura la cantidad de trabajos enfocados en la reconstrucción de disparos sísmicos son limitados, al momento de realizar este trabajo, se seleccionaron 3 trabajos hallados enfocados en esta tarea para realizar la comparación. Además se seleccionaron otros 6 trabajos enfocados en la tarea de reconstrucción de datos basados en diferentes principios. Adicionalmente, se presentan los resultados de comparación de las reconstrucciones obtenidas al emplear las dos etapas del método propuesto, aprendizaje externo e interno de manera separada.

4.2.1. Algoritmos de reconstrucción de disparos sísmicos.

- P. Goyes-Peñafiel ³ (CE) propuso un enfoque denominado "*Consensus Equilibrium*", el cual reconstruye disparos sísmicos encontrando un equilibrio entre la escasez de los datos y la eliminación de ruido.
- Wang ²² (ResNet) empleó un enfoque de reconstrucción de disparos mediante algoritmos de inteligencia artificial utilizando aprendizaje interno aprovechando la alta correlación entre las estructuras internas del dato sísmico para reconstruir los disparos faltantes.
- El trabajo "*Multichannel Singular Spectrum Analysis*" ²⁶ (MSSA) emplea un algoritmo basado en el rango bajo en el dominio de Hankel ofreciendo una estimación de los disparos faltantes en un cubo sísmico.
- Externo corresponde a la implementación únicamente de la etapa de aprendizaje externo para reconstruir disparos sísmicos.
- Interno hace referencia a emplear únicamente la segunda etapa del método propuesto para reconstruir los disparos de un cubo de datos.

4.2.2. Algoritmos de reconstrucción de datos.

4.2.2.1. Eliminación de ruido. Este principio se basa en que durante el proceso de optimización de un algoritmo de reconstrucción es posible añadir un nuevo término regularizador para eliminar el ruido de la reconstrucción. Este término regularizador, puede ser un algoritmo del estado del arte como el Ajuste de bloques y filtrado 3D (o BM3D por las siglas en inglés de *Block-matching and 3D filtering*)

²⁶ Yangkang Chen et al. "An open-source Matlab code package for improved rank-reduction 3D seismic data denoising and reconstruction". En: *Computers & Geosciences* 95 (2016), págs. 59-66.

²⁷. Sin embargo, para este trabajo se empleó un algoritmo Plug and Play (PnP) ²⁸ con un filtro recursivo (RF) donde la salida depende de los resultados de previas iteraciones .

4.2.2.2. Escasez. Al transformar una señal a otro dominio o base de representación específica la mayoría de su información se concentra en algunos pocos coeficientes mientras los demás tienden a ser cero, esto se conoce como escasez ²⁹. Los algoritmos que normalmente emplean este principio formulan su problema de optimización bajo dos términos, el primero que corresponde a un término de fidelidad de los datos y el segundo que induce la escasez de la reconstrucción sobre una transformación establecida. En este trabajo se emplearon los resultados ofrecidos por el método de multiplicadores en dirección alternada (ADMM) ²⁸, el algoritmo iterativo rápido de reducción y umbralización (FISTA) ³⁰, el algoritmo de reducción y umbralización iterativa en dos pasos (TWIST) ³¹ y el algoritmo de proyección alternante generalizada para minimización ponderada-2,1 (GAP) ³² para realizar una

²⁷ Kostadin Dabov et al. "Image denoising with block-matching and 3D filtering". En: *Image processing: algorithms and systems, neural networks, and machine learning*. Vol. 6064. SPIE. 2006, págs. 354-365.

²⁸ Stephen Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122.

²⁹ Elaine Crespo Marques et al. "A review of sparse recovery algorithms". En: *IEEE access* 7 (2018), págs. 1300-1322.

³⁰ Amir Beck y Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". En: *SIAM journal on imaging sciences* 2.1 (2009), págs. 183-202.

³¹ José M Bioucas-Dias y Mário AT Figueiredo. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration". En: *IEEE Transactions on Image processing* 16.12 (2007), págs. 2992-3004.

³² Xuejun Liao, Hui Li y Lawrence Carin. "Generalized alternating projection for weighted-2,1 minimization with applications to model-based compressive sensing". En: *SIAM Journal on Imaging Sciences* 7.2 (2014), págs. 797-823.

comparación con el método propuesto.

4.2.2.3. Suavidad. Dada una estructura de datos como una matriz, este principio se constituye cuando las diferencias entre elementos consecutivos de la estructura de datos son muy mínimas. Al igual que con el principio de escasez este es inducido por medio de un término regularizador conocido como *Total Variation*³³ (TV).

4.3. EXPERIMENTOS

4.3.1. Configuración experimental. En esta subsección se presentan los parámetros usados para desarrollar las simulaciones de reconstrucción de disparos empleando el método propuesto. Los experimentos fueron llevados a cabo en un equipo de cómputo con una GPU NVIDIA Quadro RTX 8000 con 48GB de VRAM y CPU Intel(R) Xeon(R) W-3223 CPU @3.50GHz. El lenguaje de programación empleado fue *Python 3.8* junto con la librería de *Tensorflow 2.8*³⁴. Respecto a las bases de datos de prueba, se removieron el 25% de los disparos presentes en cada cubo de datos de manera aleatoria, es decir: $\alpha = 0,25$. En cuanto a la red neuronal se empleó una red neuronal convolucional 3D con filtros de tamaño $3 \times 3 \times 3$ usando un optimizador Adam durante $\mathcal{E} = 100$, los parámetros del entrenamiento fueron los mismos tanto en la etapa de aprendizaje externo como en la de aprendizaje interno. Los trabajos de la literatura elegidos para realizar la comparación fueron configurados según Ortiz³⁵.

³³ Stanley Osher et al. "An iterative regularization method for total variation-based image restoration". En: *Multiscale Modeling & Simulation* 4.2 (2005), págs. 460-489.

³⁴ Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

³⁵ Iván Ortiz et al. *Algoritmo de reconstrucción de datos sísmicos faltantes adquiridos mediante geometrías compresivas en el dominio cross-spread empleando optimización convexa con apro-*

4.3.1.1. Configuración estudio de ablación. Se realizó un estudio de ablación donde se varió la tasa de submuestra $\alpha \in \{0,15, 0,25, 0,35, 0,50\}$, que corresponde a remover el 15 %, 25 %, 35 % y 50 % de los disparos en los cubos de datos sísmicos de manera uniforme, con el fin de validar la robustez del método propuesto sobre la cantidad de información faltante comparándose con los demás trabajos de reconstrucción.

4.3.1.2. Configuración con software PROMAX. Debido a que el uso de PROMAX se hace para datos 2D, la configuración realizada se utilizó para el conjunto de datos Alaska. Para esto, se eliminó el 25 % de los disparos de manera aleatoria ($\alpha = 0,25$) con el objetivo de reconstruir los disparos faltantes mediante el método propuesto. Una vez realizada la reconstrucción de los disparos faltantes, estos se reemplazan en el archivo original de la adquisición y este queda con 55 disparos. Este proceso de reemplazo se realiza para mantener la coherencia de los metadatos que PROMAX utiliza para manejar el conjunto de datos en todas las herramientas de post-procesamiento disponibles. Posteriormente, se cargaron el cubo original, el cubo estimado y el cubo sin los shots reconstruidos en el PROMAX, con el fin de realizar una comparación de la imagen sísmica producida por el PROMAX.

5. SIMULACIONES Y RESULTADOS

En esta sección se presentan las simulaciones realizadas junto con los resultados obtenidos por el método propuesto para la reconstrucción de disparos sísmicos faltantes. Aquí se evalúa de manera cuantitativa y cualitativa la metodología propuesta, así como el algoritmo desarrollado en comparación con trabajos del estado del arte que resuelven la misma tarea. De igual manera, se presenta un estudio sobre los mejores parámetros para la red neuronal y, a su vez, se presentan los resultados del estudio de ablación sobre la robustez del algoritmo frente al porcentaje de disparos eliminados en un cubo sísmico. Por último, se hace uso del software ProMax para generar imágenes sísmicas y realizar una comparativa de la imagen resultante al realizar reconstrucción con el método propuesto o generar la imagen con la información (fuentes) faltante.

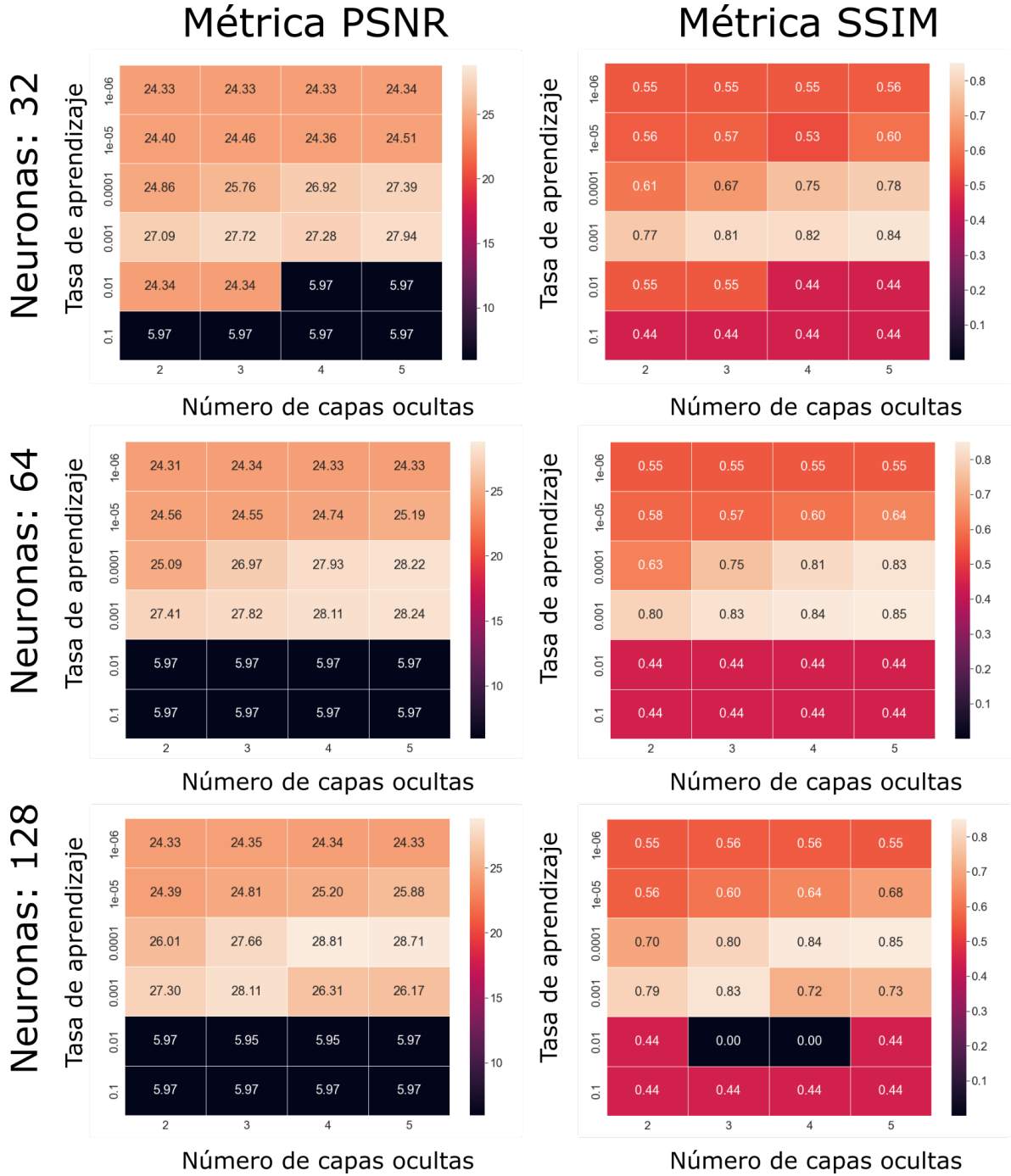
5.1. VALIDACIÓN DE HIPERPARÁMETROS

En la etapa de búsqueda de los mejores hiperparámetros para configurar la red neuronal, se hizo una investigación sobre el desempeño de la red y de la influencia de la tasa de aprendizaje de la red (τ), la cantidad de capas ocultas de la red (ρ) y la cantidad de neuronas en cada capa (η) en los resultados obtenidos variando cada uno de estos parámetros. Por lo tanto, se realizaron pruebas con valores de $\tau \in \{1e - 6, 1e - 5, 1e - 4, 1e - 3, 1e - 2, 1e - 1\}$, $\rho \in \{2, 3, 4, 5\}$ y $\eta \in \{32, 64, 128\}$, así realizando una variación de todas las posibles combinaciones de estas variables para determinar la mejor configuración de hiperparámetros para la red neuronal.

En la Figura 16 se observan los resultados obtenidos de la validación cruzada para el método propuesto en forma de mapas de calor. El comportamiento es similar usando 32, 64, y 128 neuronas tanto en PSNR como en SSIM. La tasa de aprendizaje

se comporta similarmente en todos los casos para valores de $\tau = 1e-3$ y $\tau = 1e-4$. Respecto al número de capas ocultas, los mejores resultados en PSNR y SSIM se obtuvieron al utilizar 4 y 5 capas. De estos resultados obtenidos se determinan los valores óptimos de estos hiperparámetros a utilizar: $\tau = 1e-4$, $\rho = 4$ y $\eta = 128$.

Figura 16. Mapas de calor de las métricas de PSNR y SSIM evaluadas sobre el conjunto de datos **SEAM phase II** para diferentes combinaciones de ρ , τ y η .



5.2. RESULTADOS DE SIMULACIONES

La evaluación de la metodología y algoritmos se realizó a través de simulaciones utilizando el método propuesto y métodos del estado del arte. Los resultados se evaluaron utilizando las diferentes métricas reportadas previamente. Igualmente, con el fin de evaluar los resultados de reconstrucción, se realizó la comparación de la imagen sísmica final obtenida al aplicar el método propuesto de reconstrucción. Es por esto que las simulaciones, pruebas y evaluaciones se dividieron en dos grupos, el primero evalúa las reconstrucciones obtenidas por todos los métodos utilizando las bases de datos en el dominio *cross-spread*, es decir, adquiridas bajo geometrías 3D ortogonales. El segundo grupo evalúa tanto la reconstrucción como la imagen sísmica apilada obtenida con el método propuesto para la base de datos Alaska 2D.

En el Cuadro 1 se presentan los resultados obtenidos utilizando las métricas presentadas previamente, tanto para las reconstrucciones obtenidas por el método propuesto como con los demás métodos de comparación. El método propuesto logra ganancias sobre los demás métodos de comparación tanto para el conjunto de datos **SEAM phase II** como para **Stratton** logrando superar en al menos 1,76 [dB] y 0,81 [dB] en la métrica PSNR respectivamente. En la Figura 17 y en la Figura 18 se muestran los resultados visuales del método propuesto y algunos métodos de comparación para el conjunto de dato **SEAM phase II** y **Stratton** respectivamente. En las reconstrucciones obtenidas con el método propuesto se logra visualizar una continuidad más fuerte en sus eventos, al igual que una estructura del dato más consistente con la del dato real. La transformación de cada disparo reconstruido en el dominio de la frecuencia revela una similitud estructural superior en las reconstrucciones obtenidas por parte del método propuesto, lo que indica que la información importante se concentra en las mismas regiones manteniendo una coherencia y semejanza al dato real. También, se destaca el desempeño del método del aprendizaje

externo, sin embargo puede deberse a principalmente a que los cubos de entrenamiento pertenecen a la misma distribución de datos debido a que fueron obtenidos de la misma base de datos.

Cuadro 1. Resultados obtenidos de la recuperación de disparos sísmicos obtenidos sobre los conjuntos de datos de **SEAM phase II** y **Stratton** removiendo el 25 % de sus disparos. Los valores mostrados son el promedio de las métricas obtenidas por cada disparo reconstruido en el cubo de datos. El PSNR y SNR se mide en decibelios [dB] y el MSE se encuentra en un escala de 10^{-3} .

Conjunto de datos	Métrica	Propuesto	Interno	Externo	CE	ResNet	MSSA	ADMM	FISTA	TWIST	GAP	TV	RF
SEAM phase II	MSE ↓	0.96	2.56	<u>1.96</u>	2.29	21.7	2.2	2.87	3.58	2.63	2.74	3.42	3.20
	PSNR ↑	28.81	26.01	<u>27.05</u>	26.4	16.66	26.57	25.42	24.26	25.7	25.61	24.66	24.94
	SNR ↑	23.81	20.06	<u>21.10</u>	20.45	10.71	20.62	19.47	18.51	19.75	19.66	18.71	19.99
	SSIM ↑	0.84	0.71	<u>0.75</u>	0.74	0.37	0.65	0.53	0.45	0.59	0.55	0.57	0.62
Stratton	MSE ↓	2.56	4.00	3.51	<u>3.29</u>	4.1	3.52	4.24	4.6	3.66	4.07	3.66	3.56
	PSNR ↑	25.63	23.97	24.53	<u>24.82</u>	23.87	24.54	23.72	23.38	24.37	23.91	24.37	24.49
	SNR ↑	19.75	18.03	18.66	<u>18.94</u>	14.99	18.67	17.85	17.5	18.49	18.03	18.5	18.61
	SSIM ↑	0.70	0.56	0.62	<u>0.64</u>	0.61	0.56	0.43	0.43	0.51	0.45	0.59	0.61

Figura 17. Resultados visuales para el disparo 30 del conjunto de datos **SEAM phase II**.

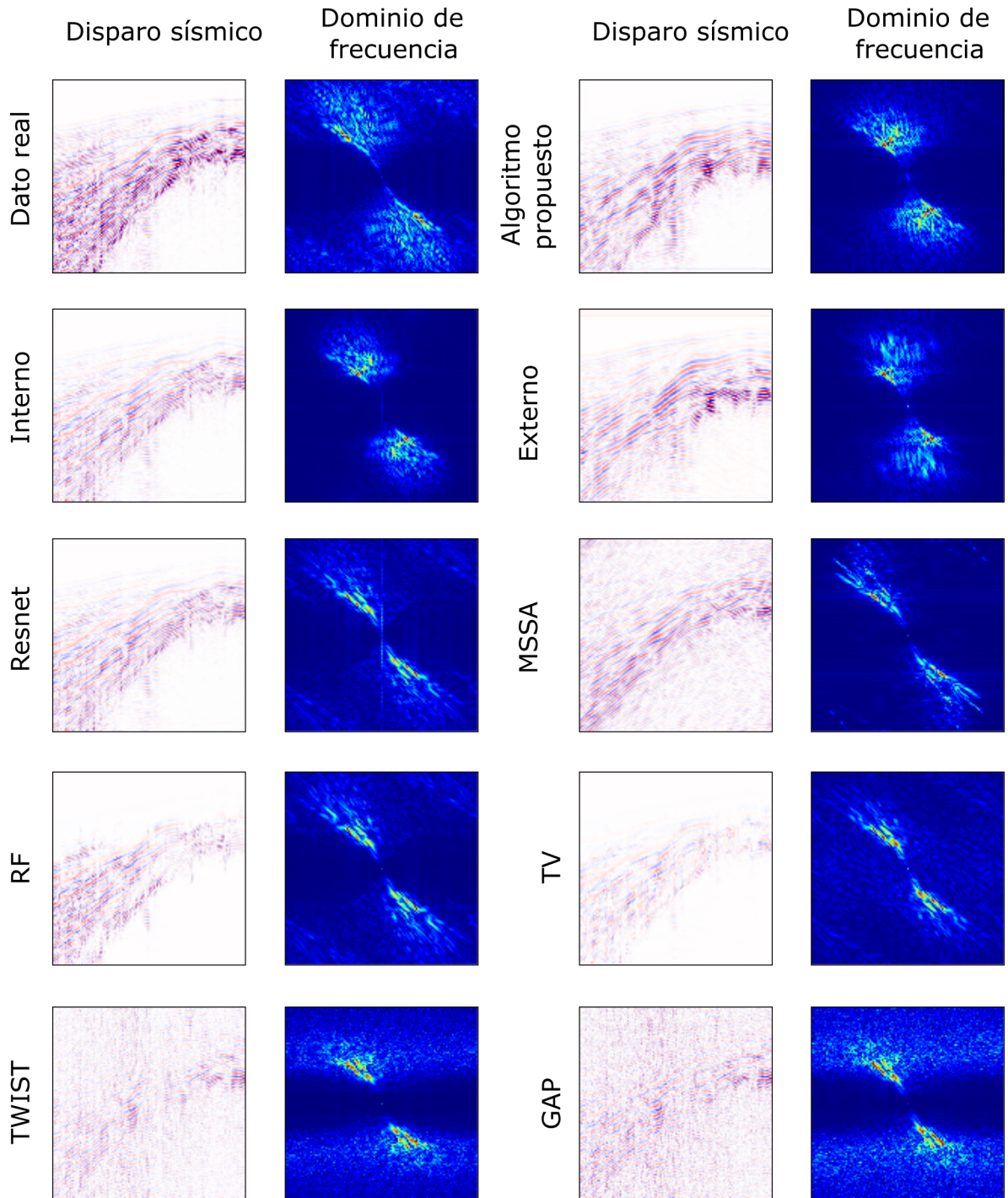
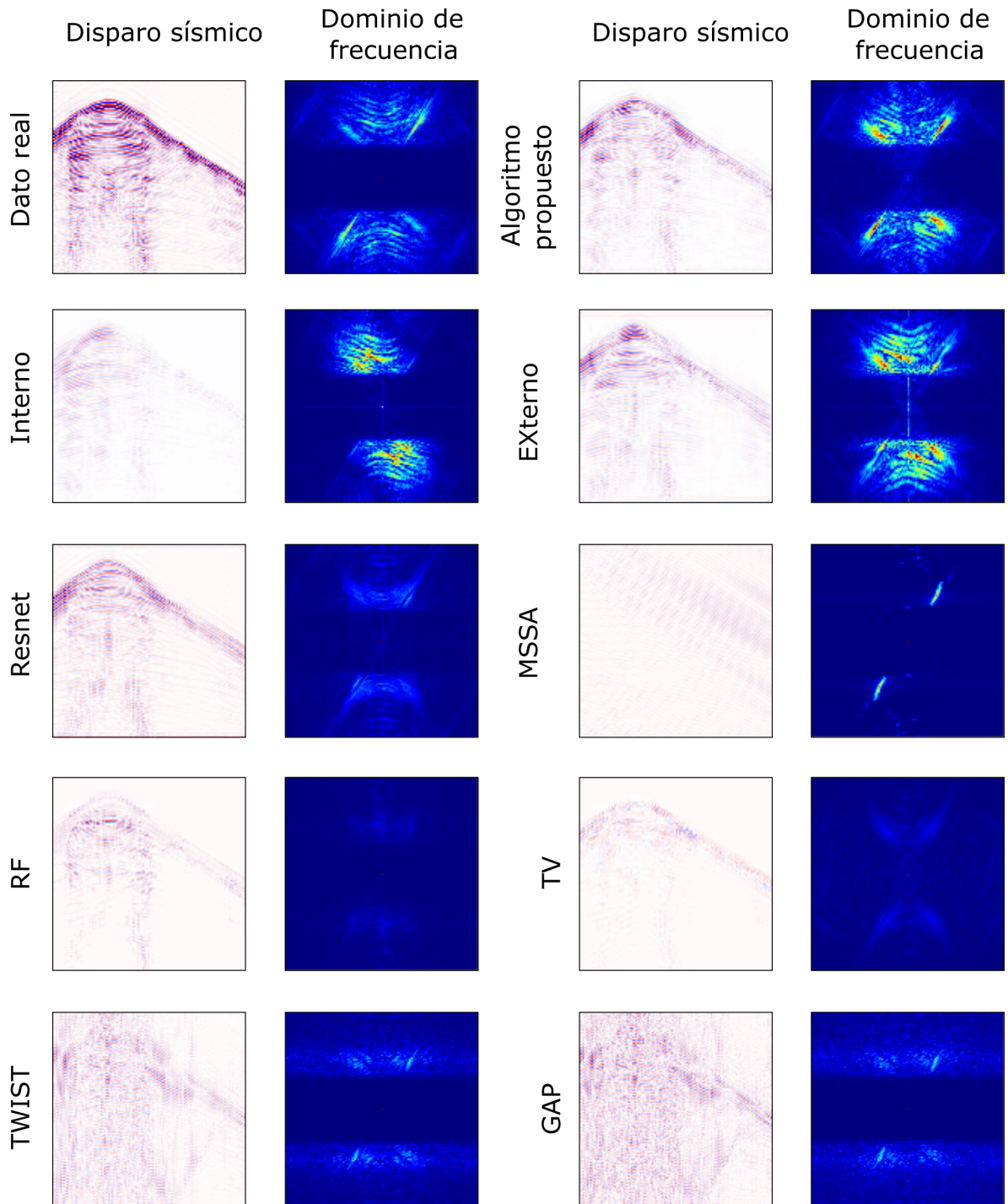


Figura 18. Resultados visuales para el disparo 16 del conjunto de datos **Stratton**.



5.3. ESTUDIO DE ABLACIÓN

Debido a que la calidad de reconstrucción de disparos faltantes depende de la cantidad de información cercana presente, se realizó un estudio de ablación que evalúa el comportamiento de las reconstrucciones al variar el porcentaje de disparos removidos. En el Cuadro 2 se muestran las métricas de PSNR y SSIM obtenidas al evaluar los resultados del algoritmo propuesto junto con los demás trabajos de comparación en función de la cantidad de disparos sísmicos removidos. Se puede observar como el método propuesto consigue una ganancia sobre los demás métodos en la mayoría de los casos para los conjuntos de datos adquiridos bajo geomtetrías ortogonales, **SEAM phase II** y **Stratton**, consiguiendo un mejor rendimiento cuando la cantidad de disparos removidos es menor. Sin embargo las métricas obtenidas son bastante similares para porcentajes de submuestreo iguales o por debajo del 35 % observándose un decaimiento más significativo para un submuestreo del 50 %. Comparando los resultados obtenidos en el Cuadro 1 con los obtenidos en el Cuadro 2 cuando son removidos el 25 % de los disparos, se aprecia que los resultados son similares, lo que permite concluir que el algoritmo no es muy sensible al tipo de submuestreo empleado para eliminar los disparos sísmicos, sin embargo y como es de esperarse un menor número de disparos eliminados produce mejores resultados, tanto en PSNR como en SSIM.

Cuadro 2. Comparación de los resultados obtenidos con el método propuesto y los demás trabajos en la recuperación de disparos sísmicos sobre los conjuntos de datos **SEAM phase II** y **Stratton**. El PSNR se mide en decibelios [dB].

Conjunto de datos	Disparos eliminados	Métrica	Propuesto	Interno	Externo	CE	ResNet	MSSA	ADMM	FISTA	TWIST	GAP	TV	RF
Seam phase II	15 % (9 disparos)	PSNR↑ SSIM↑	30.87 0.88	27.09 0.70	<u>28.52</u> <u>0.80</u>	19.67 0.71	25.36 0.54	27.84 0.69	26.12 0.47	25.73 0.444	26.86 0.57	26.39 0.50	26.46 0.59	26.8 0.61
	25 % (15 disparos)	PSNR↑ SSIM↑	30.67 0.88	26.81 0.69	<u>28.40</u> <u>0.80</u>	27.83 0.74	25.35 0.60	27.19 0.67	25.68 0.45	25.17 0.41	26.46 0.57	25.94 0.49	26.02 0.60	26.36 0.62
	35 % (21 disparos)	PSNR↑ SSIM↑	29.63 0.85	26.69 0.69	<u>27.76</u> <u>0.78</u>	26.90 0.7	25.13 0.62	26.75 0.66	25.43 0.44	24.64 0.38	26.2 0.56	25.68 0.47	25.91 0.62	26.10 0.63
	50 % (30 disparos)	PSNR↑ SSIM↑	24.79 0.66	<u>26.59</u> <u>0.68</u>	26.68 0.72	25.86 0.64	24.88 0.62	25.52 0.55	24.72 0.40	6.08 0.17	25.65 0.54	24.96 0.43	25.87 0.63	25.21 0.61
Stratton	15 % (2 disparos)	PSNR↑ SSIM↑	28.27 0.79	25.55 0.62	25.47 0.64	18.77 <u>0.69</u>	24.81 0.67	25.55 0.60	24.53 0.44	24.3 0.42	25.29 0.54	24.73 0.47	25.57 0.63	<u>25.81</u> 0.67
	25 % (4 disparos)	PSNR↑ SSIM↑	27.43 0.77	24.72 0.62	24.69 0.63	<u>24.99</u> <u>0.66</u>	23.89 0.60	24.54 0.56	23.62 0.42	22.91 0.35	24.31 0.51	23.85 0.45	24.37 0.60	24.57 0.62
	35 % (6 disparos)	PSNR↑ SSIM↑	27.01 0.75	24.94 0.63	24.85 0.63	<u>25.74</u> <u>0.67</u>	24.43 0.64	24.55 0.54	23.51 0.40	23.18 0.35	24.29 0.49	23.72 0.42	24.66 0.61	24.94 0.63
	50 % (9 disparos)	PSNR↑ SSIM↑	25.63 0.70	<u>24.93</u> <u>0.63</u>	24.46 0.61	24.60 0.61	23.27 0.56	24.75 0.58	23.17 0.36	5.84 0.01	24.11 0.47	23.4 0.39	24.55 0.60	24.60 0.61

5.4. EVALUACIÓN DE RECONSTRUCCIÓN A TRAVÉS DE POST-PROCESAMIENTO MEDIANTE SOFTWARE PROMAX

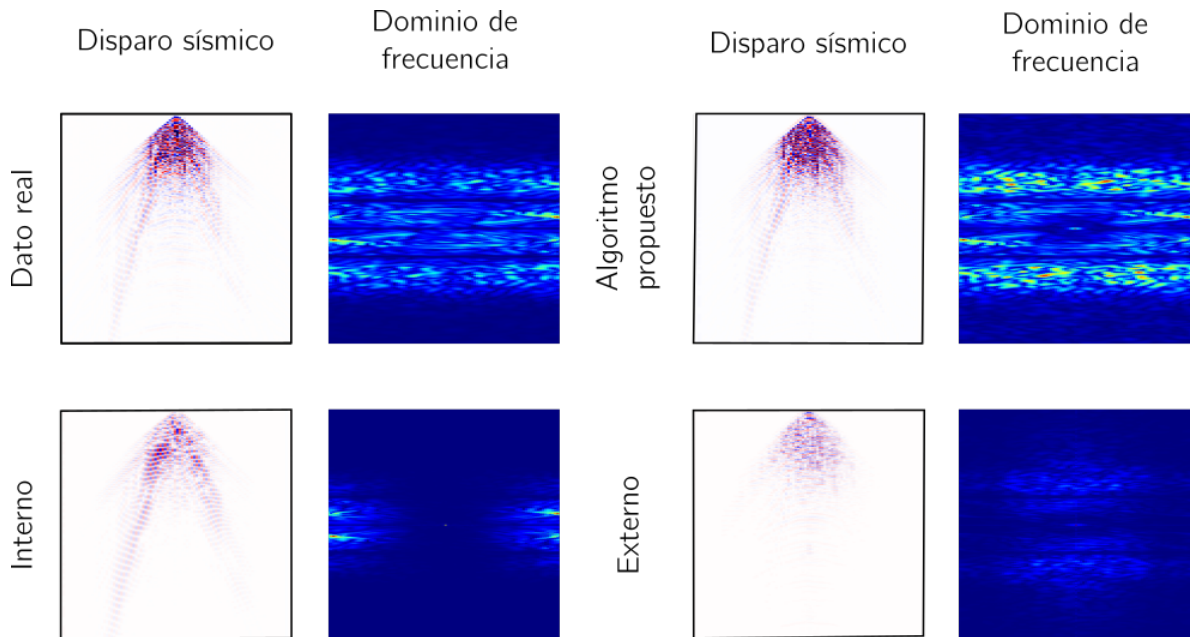
Con el objetivo de validar la usabilidad del método propuesto, se llevará a cabo una reconstrucción de imágenes sísmicas a partir de los datos obtenidos. Estas imágenes serán interpretadas por expertos para evaluar las características del suelo y detectar la presencia de reservorios de gas y petróleo. Se realizó una primera reconstrucción utilizando el conjunto de datos Alaska, según se detalla en la Sección 4.3.1.2.

Cuadro 3. Comparación de los resultados obtenidos con el método propuesto sobre los demás trabajos en la recuperación de disparos sísmicos sobre los conjuntos de datos de **Alaska**. El PSNR se mide en decibelios [dB] y el MSE esta en una escala de 10^{-3} .

Conjunto de datos	Métrica	Propuesto	Interno	Externo
Alaska	PSNR ↑	24.20	<u>24.11</u>	24.01
	SNR ↑	18.24	<u>18.16</u>	18.06
	SSIM ↑	0.83	<u>0.82</u>	0.83
	MSE ↓	3.80	<u>3.87</u>	3.96

En el Cuadro 3 se presentan los resultados obtenidos en la recuperación de disparos sísmicos utilizando el método propuesto en comparación con las dos etapas del mismo empleadas por separado. Estas reconstrucciones se realizaron con el fin de validar la usabilidad del método propuesto en procesamientos sísmicos posteriores. Los resultados se presentan para el conjunto de datos **Alaska** y se evalúan utilizando las métricas PSNR, SNR, SSIM y MSE. Los valores en la tabla indican que el método propuesto obtuvo mejores resultados en las métricas PSNR, SNR y MSE en comparación con los otros dos métodos, con una diferencia especialmente notable en el PSNR y el MSE. En general, se puede concluir que el método propuesto sigue siendo efectivo en la recuperación de disparos sísmicos en comparación con los otros métodos evaluados en el conjunto de datos **Alaska**.

Figura 19. Resultados visuales para el disparo 5 del conjunto de datos **Alaska**.



En la Figura 19 se observan las reconstrucciones realizadas por el método propuesto junto con los otros métodos de comparación al igual que el disparo real. Sin embargo, como se puede apreciar el método propuesto es el que consigue una similitud mucho mayor al igual que su transformación al dominio de la frecuencia donde se observa que la información más importante del dato se acumula en las mismas zonas que la del disparo real.

Después de generar la reconstrucción de los disparos sísmicos faltantes en el cubo de datos, estos se reemplazaron en el cubo completo de datos original para ingresarlo en el software PROMAX. Posteriormente, se generó un flujo de post-procesamiento con el objetivo de utilizar estos datos para obtener un dato apilado (imagen sísmica) del cual ya se puede extraer información acerca de las características del suelo. La Figura 20 ilustra los cuatro procesos necesarios para transformar el cubo de datos de Alaska en una imagen sísmica apilada.

Figura 20. Flujos de trabajo empleados en PROMAX para generar la imagen apilada del conjunto de datos **Alaska**.



La Figura 21 muestra las imágenes de apilamiento bruto generadas con el cubo de datos submuestreado, con los disparos reconstruidos y con el cubo original completo. La imagen generada con el cubo submuestreado presenta una evidente pérdida de información, lo que dificulta el análisis de los intervalos faltantes y provoca discontinuidades en los eventos sísmicos. En cambio, la imagen obtenida con los disparos reconstruidos permite analizar los intervalos no muestreados y garantiza una continuidad en los eventos sísmicos. Además, las estimaciones realizadas por el método propuesto muestran una alta similitud con los eventos reales muestreados en campo, como se aprecia en las regiones ampliadas de la imagen. Se puede observar que en las zonas superficiales no hay más intervalos con información desconocida, y en el subsuelo se logra una continuidad de los eventos sísmicos incluso más fuerte que la obtenida en los datos originales.

En la Figura 22 se observan las imágenes apiladas obtenidas de los cubos de datos reconstruidos con el esquema propuesto y otros métodos de comparación. En las áreas ampliadas, se puede ver cómo el método propuesto logra resultados más coherentes, con un equilibrio en las intensidades de los eventos sísmicos en la imagen. Además, los resultados obtenidos por el método de aprendizaje interno son destacables debido a que presentan un desequilibrio menos notable. Sin embargo, en la parte inferior de la imagen, los métodos comparativos presentan un artefacto, mientras que el método propuesto lo muestra de manera más tenue, conservando la coherencia en la información de la imagen.

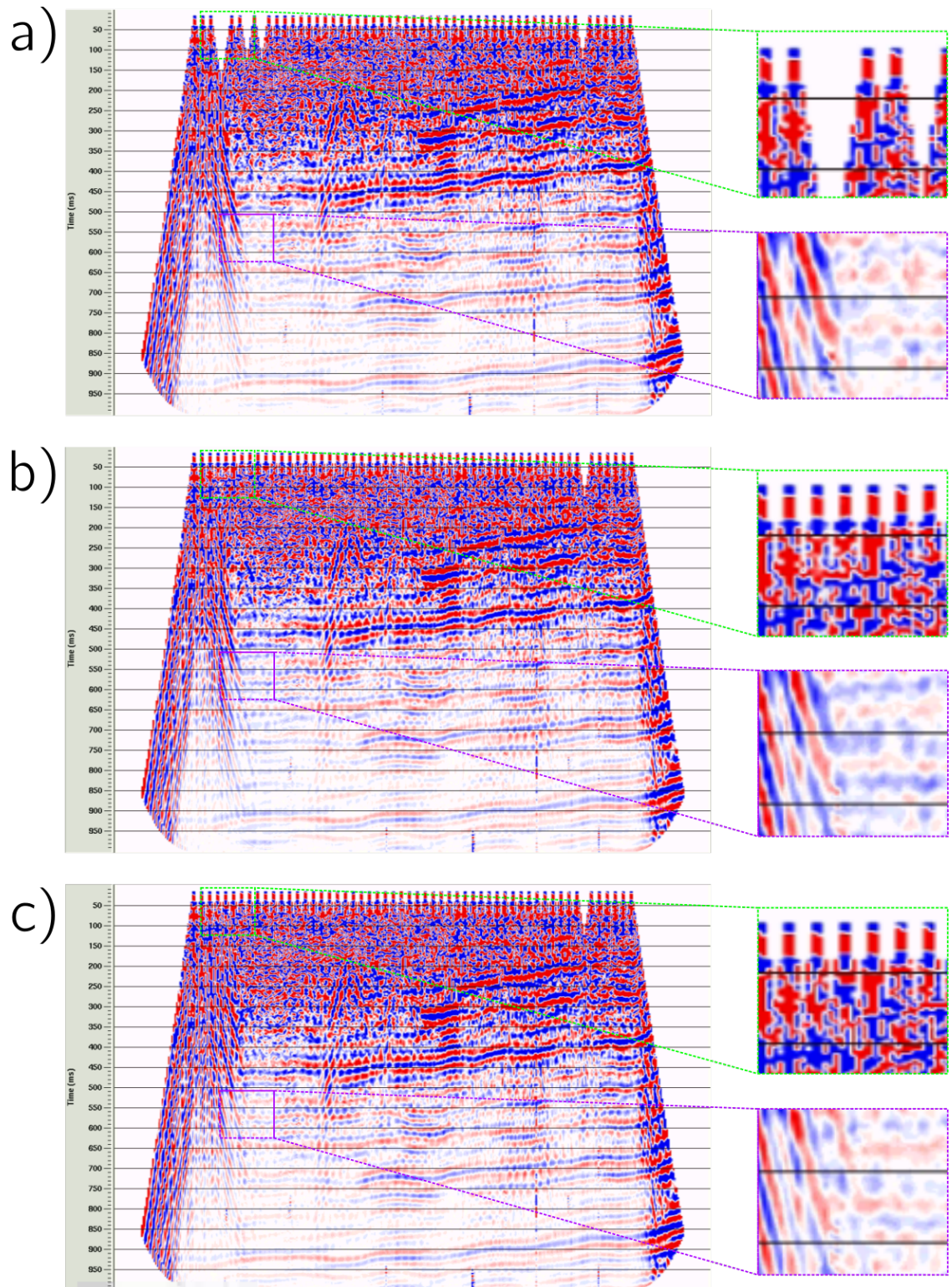


Figura 21. Imágenes de apilamiento bruto obtenidas con el software PROMAX del a) cubo submuestreado, b) el cubo con disparos reocnstruidos y c) el cubo original completo.

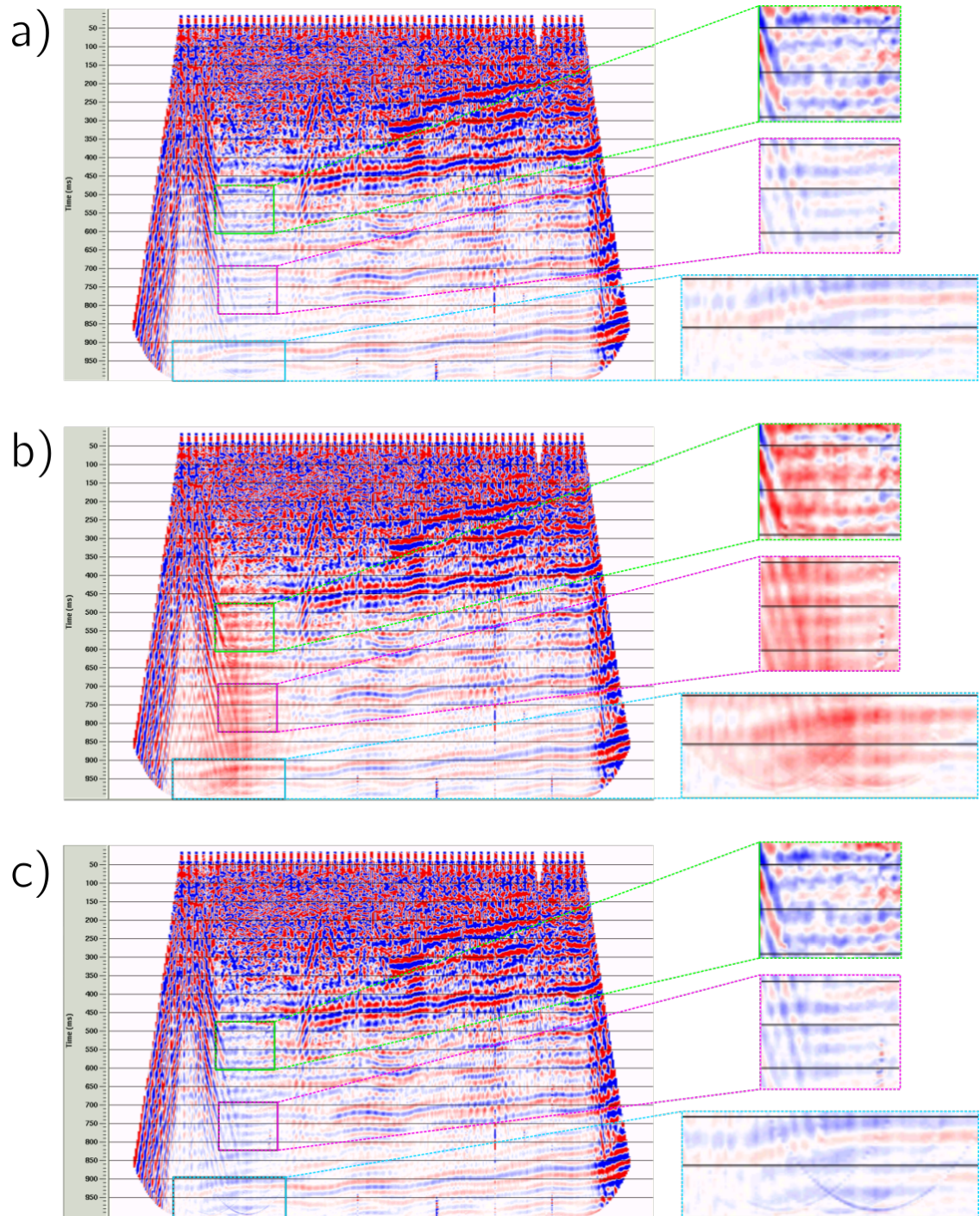


Figura 22. Imágenes de apilamiento bruto obtenidas con el software PROMAX del cubo con disparos reconstruidos con a) el método propuesto, b) Aprendizaje externo c) Aprendizaje Interno.

6. CONCLUSIONES

En este trabajo se propuso una metodología de reconstrucción de datos sísmicos adquiridos mediante geometrías compresivas usando un enfoque de aprendizaje profundo interno y externo. Este enfoque permite utilizar datos sísmicos externos y posteriormente realizar un acondicionamiento a la complejidad del subsuelo utilizando los mismos datos incompletos.

Se realizó la formulación matemática del modelo de adquisición de datos sísmicos utilizando geometrías compresivas mediante notación vectorial para el dominio $2D$ y *cross-spread*. Este modelo realiza un muestreo de disparos sísmicos con el fin de simular su posterior proceso de reconstrucción, apreciándose la naturaleza mal condicionada del problema inverso. De esta manera, se da por cumplido el objetivo 1.

Se realizó una búsqueda de las bases de datos sísmicas disponibles asegurando que las geometrías de adquisición fueran 2D o 3D, específicamente en el último caso fueron geometrías 3D ortogonales. En ese orden de ideas, se obtuvieron los conjuntos de datos de **SEAM phase II** y **Stratton** que corresponden a datos sísmicos adquiridos bajo geometrías 3D ortogonales en el dominio *cross-spread* y, además, el conjunto de datos de **Alaska** que corresponde a un cubo de datos sísmicos muestreado bajo geometrías 2D, específicamente con la geometría *split-spread*. Realizada esta selección de bases de datos se da por cumplido el objetivo 2.

Posteriormente, se planteó un esquema que combina algoritmos de aprendizaje profundo externo e interno para abordar la problemática planteada. Este enfoque se eligió debido a que permite utilizar los pocos datos sísmicos disponibles para entrenar

un algoritmo y adicionalmente promueve el acondicionamiento de los parámetros de la red a la complejidad particular del dato a reconstruir a través del aprendizaje interno. Este esquema mejora el desempeño en la tarea de reconstruir disparos en un cubo sísmico en comparación con el uso de cada esquema de aprendizaje por separado. Con este planteamiento, se da por cumplido el objetivo 3.

En la evaluación del desempeño del algoritmo propuesto para la recuperación de disparos sísmicos en conjuntos de datos adquiridos con geometrías 3D ortogonales en el dominio cross-spread, se obtuvieron resultados sobresalientes en las métricas de evaluación PSNR y SSIM, con ganancias de hasta $12dB$ y $0,47$ en comparación con los métodos del estado del arte. Además, se observó que el rendimiento obtenido con el algoritmo propuesto superó a los valores obtenidos al utilizar métodos de aprendizaje externo e interno por separado. Al analizar los resultados visuales, se pudo apreciar que el método propuesto logró mejores aproximaciones a los datos reales, con una mayor continuidad en los eventos del disparo sísmico. A través del estudio de ablación se concluyó que el algoritmo propuesto superó a los demás métodos del estado del arte en la mayoría de los casos, incluso cuando se removieron hasta el 35% y el 50% de los disparos en los conjuntos de datos **SEAM Phase II** y **Stratton** respectivamente. Además, se notó que el rendimiento del algoritmo no varió significativamente al cambiar el tipo de muestreo utilizado para remover los disparos sísmicos. En resumen, se cumplió el objetivo de evaluar el desempeño del algoritmo propuesto para la recuperación de disparos sísmicos. Con esta evaluación se da por cumplido el objetivo 4.

Finalmente, se compararon los resultados obtenidos con el método propuesto, los datos reales y los datos submuestreados sobre el conjunto de datos muestreados **Alaska** haciendo uso del software de procesamiento sísmico PROMAX. Las imá-

genes de apilamiento bruto obtenidas por medio de PROMAX muestran que efectivamente el método propuesto ofrece una mejoría sobre las mismas al permitir un análisis de regiones no muestreadas que preservan las características de los datos reales al conseguir una continuidad de los eventos sísmicos. Además, el método propuesto consigue ofrecer imágenes apiladas con menos artefactos y eventos sísmicos mejor definidos que los demás métodos de comparación. Con esta comparación se da por cumplido el objetivo 5.

7. TRABAJO FUTURO

Para trabajos futuros, se plantea validar la efectividad del algoritmo en otros tipos de geometrías de adquisición utilizadas en la exploración sísmica. Además, se propone un estudio para determinar qué posiciones son más favorables para la eliminación de disparos con el algoritmo propuesto. Por último, se planea incorporar ruido a los datos de entrada para obtener un algoritmo capaz de eliminar ruido y estudiar la influencia de este en las reconstrucciones ofrecidas por el método propuesto.

BIBLIOGRAFÍA

- Abadi, Martín et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015 (vid. pág. 47).
- Angarita, Edgar Eduardo Yáñez et al. "Rapid screening and probabilistic estimation of the potential for CO₂-EOR and associated geological CO₂ storage in Colombian petroleum basins". En: *Petroleum Geoscience* 28.1 (2022), petgeo2020-110 (vid. pág. 15).
- Beck, Amir y Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". En: *SIAM journal on imaging sciences* 2.1 (2009), págs. 183-202 (vid. pág. 46).
- Bioucas-Dias, José M y Mário AT Figueiredo. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration". En: *IEEE Transactions on Image processing* 16.12 (2007), págs. 2992-3004 (vid. pág. 46).
- Boyd, Stephen et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122 (vid. pág. 46).
- Chen, Haibo et al. "Artistic style transfer with internal-external learning and contrastive learning". En: *Advances in Neural Information Processing Systems* 34 (2021), págs. 26561-26573 (vid. pág. 17).

- Chen, Yangkang et al. "An open-source Matlab code package for improved rank-reduction 3D seismic data denoising and reconstruction". En: *Computers & Geosciences* 95 (2016), págs. 59-66 (vid. pág. 45).
- Dabov, Kostadin et al. "Image denoising with block-matching and 3D filtering". En: *Image processing: algorithms and systems, neural networks, and machine learning*. Vol. 6064. SPIE. 2006, págs. 354-365 (vid. pág. 46).
- Goyes-Peñafiel, Paul et al. "A Consensus Equilibrium Approach for 3-D Land Seismic Shots Recovery". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), págs. 1-5 (vid. págs. 14, 16, 30, 45).
- Gülünay, Necati. "Seismic trace interpolation in the Fourier transform domain". En: *Geophysics* 68.1 (2003), págs. 355-369 (vid. pág. 29).
- Kaur, Harpreet, Nam Pham y Sergey Fomel. "Seismic data interpolation using CycleGAN". En: *SEG technical program expanded abstracts 2019*. Society of Exploration Geophysicists, 2019, págs. 2202-2206 (vid. pág. 29).
- Kingma, Diederik P y Jimmy Ba. "Adam: A method for stochastic optimization". En: *arXiv preprint arXiv:1412.6980* (2014) (vid. pág. 37).
- Li, Yuanming et al. "Seismic data augmentation based on conditional generative adversarial networks". En: *Sensors* 20.23 (2020), pág. 6850 (vid. pág. 14).
- Liao, Xuejun, Hui Li y Lawrence Carin. "Generalized alternating projection for weighted-2,1 minimization with applications to model-based compressive sensing". En: *SIAM Journal on Imaging Sciences* 7.2 (2014), págs. 797-823 (vid. pág. 46).

- López, Kareth León et al. "Regular multi-shot subsampling and reconstruction on 3d orthogonal symmetric seismic grids via compressive sensing". En: *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE. 2019, págs. 1-5 (vid. págs. 14, 16, 17, 30).
- Mandelli, Sara et al. "Interpolation and denoising of seismic data using convolutional neural networks". En: *arXiv preprint arXiv:1901.07927* (2019) (vid. pág. 29).
- Marques, Elaine Crespo et al. "A review of sparse recovery algorithms". En: *IEEE access* 7 (2018), págs. 1300-1322 (vid. pág. 46).
- Mikhailiuk, A y A Faul. "Deep learning applied to seismic data interpolation". En: *80th EAGE Conference and Exhibition 2018*. Vol. 2018. 1. European Association of Geoscientists & Engineers. 2018, págs. 1-5 (vid. pág. 16).
- Mora, John y Mauricio Rocha. "Implementación del dominio del CROSS SPREAD en procesamiento sísmico 3D en cuencas colombianas". En: *11th Simposio Bolivariano- Exploracion Petrolera en las Cuencas Subandinas*. European Association of Geoscientists & Engineers. 2012, cp-330 (vid. pág. 16).
- Naghizadeh, Mostafa y Mauricio D Sacchi. "f-x adaptive seismic-trace interpolation". En: *Geophysics* 74.1 (2009), págs. V9-V16 (vid. pág. 29).
- Ortiz, Iván et al. *Algoritmo de reconstrucción de datos sísmicos faltantes adquiridos mediante geometrías compresivas en el dominio cross-spread empleando optimización convexa con aproximación de bajo rango*. <https://noesis.uis.edu.co/items/a4965267-3def-4837-8a2f-141cc5c06041>. 2022 (vid. pág. 47).

- Osher, Stanley et al. "An iterative regularization method for total variation-based image restoration". En: *Multiscale Modeling & Simulation* 4.2 (2005), págs. 460-489 (vid. pág. 47).
- Ronneberger, Olaf, Philipp Fischer y Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". En: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, págs. 234-241 (vid. pág. 30).
- SEAM Phase II Land Seismic Datasets*. <https://seg.org/News-Resources/Research-Data/Land-Seismic-Data>. (Accessed on 11/30/2022) (vid. pág. 32).
- Spitz, Simon. "Seismic trace interpolation in the FX domain". En: *Geophysics* 56.6 (1991), págs. 785-794 (vid. pág. 29).
- Tang, Shuhang et al. "Reconstruction of sparsely sampled seismic data via residual U-Net". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2020), págs. 1-5 (vid. págs. 14, 30).
- Ulyanov, Dmitry, Andrea Vedaldi y Victor Lempitsky. "Deep image prior". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 9446-9454 (vid. págs. 15, 26).
- Villarreal, Ofelia et al. "Seismic source reconstruction in an orthogonal geometry based on local and non-local information in the time slice domain". En: *Journal of Applied Geophysics* 170 (2019), pág. 103846 (vid. págs. 15, 16).
- Villarreal, Ofelia P et al. "Compressive sensing seismic acquisition by using regular sampling in an orthogonal grid". En: *2017 IEEE 7th International Workshop on*

- Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, págs. 1-5 (vid. págs. 14, 16, 30).
- Wang, Benfeng et al. "Intelligent missing shots' reconstruction using the spatial reciprocity of Green's function based on deep learning". En: *IEEE Transactions on Geoscience and Remote Sensing* 58.3 (2019), págs. 1587-1597 (vid. págs. 14, 30).
- Wang, Qin et al. "Seismic data interpolation using deep internal learning". En: *Exploration Geophysics* 51.6 (2020), págs. 683-697 (vid. págs. 30, 45).
- Wang, Tengfei, Hao Ouyang y Qifeng Chen. "Image inpainting with external-internal learning and monochromic bottleneck". En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, págs. 5120-5129 (vid. págs. 17, 27).
- Wang, Yanghua. "Seismic trace interpolation in the fxy domain". En: *Geophysics* 67.4 (2002), págs. 1232-1239 (vid. págs. 17, 29).
- Wapenaar, CPA. "3-D migration of cross-spread data: Resolution and amplitude aspects". En: *Geophysics* 62.4 (1997), págs. 1220-1225 (vid. pág. 16).
- Zhang, Tao et al. "Hyperspectral image reconstruction using deep external and internal learning". En: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, págs. 8559-8568 (vid. pág. 17).