

Desarrollo de una red neuronal físicamente informada (PINN) usando kernels NTK para la solución del problema directo e inverso de propagación sísmica acústica

Juan Diego Bohórquez López

Trabajo de Grado para Optar al Título de Magister en Geofísica

Directora

Ana Beatriz Ramirez Silva

PhD Ingenieria Electrónica

Codirector

Sergio Alberto Abreo Carrillo

Dr.-Ing Ingenieria Electrónica

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Física

Maestría en Geofísica

Bucaramanga

2026

### **Dedicatoria**

*Dedico esta tesis a mis padres y hermanos, por estar a mi lado durante toda mi formación y apoyarme siempre a nivel personal.*

*A mis amigos, de quienes he aprendido mucho y quienes hicieron de este recorrido una experiencia más agradable; gracias por todo su apoyo.*

### **Agradecimientos**

Agradezco a mis directores, la profesora Ana Ramirez y el profesor Sergio Abreo, por sus valiosos aportes y enseñanzas como docentes e investigadores, sin los cuales no hubiera sido posible llevar a cabo este trabajo y vivir gratificantes experiencias relacionadas con la divulgación tanto dentro como fuera del país. Además, agradezco su enorme calidad humana, que hizo todo el proceso más enriquecedor.

Agradezco al Grupo de Investigación en Conectividad y Procesamiento de Señales (CPS) de la Universidad Industrial de Santander por su apoyo al proporcionar recursos de hardware y software para este trabajo. También agradezco a la Agencia Nacional de Hidrocarburos (ANH); al Ministerio de Ciencia, Tecnología e Innovación (Minciencias); y al convenio ANH-Minciencias 0166-2021.

## Tabla de Contenido

<b>Introducción</b>	<b>13</b>
<b>1 Objetivos</b>	<b>17</b>
1.1 Objetivo general	17
1.2 Objetivos específicos	17
<b>2 Metodología</b>	<b>18</b>
2.1 Formulación del problema	18
2.1.1 Inversión sísmica de velocidad	18
2.1.2 Propagación acústica	19
2.2 Red neuronal físicamente informada	21
2.2.1 Arquitectura	21
2.2.2 Función de costo	23
2.2.3 Kernels NTK	25
2.3 Evaluación del modelo	27
<b>3 Configuración experimental</b>	<b>28</b>
3.1 Datos Sintéticos	28
3.1.1 Submuestreo	30
3.1.2 Fuente sísmica	30
3.2 Normalización	31
3.3 Detalles de implementación	32
<b>4 Resultados</b>	<b>33</b>
4.1 Experimento 1: Modelo pozo cruzado	33
4.1.1 Entrenamiento	34
4.1.2 Modelo directo	34

4.1.3	Modelo inverso . . . . .	35
4.1.4	Datos observados . . . . .	36
4.2	Experimento 2: Modelo Sismológico . . . . .	37
4.2.1	Entrenamiento . . . . .	38
4.2.2	Modelo directo . . . . .	38
4.2.3	Modelo inverso . . . . .	39
4.2.4	Datos observados . . . . .	39
4.3	Experimento 3: Modelo de cubrimiento imperfecto de receptores . . . . .	39
4.3.1	Distribución de receptores . . . . .	41
4.3.2	Entrenamiento . . . . .	42
4.3.3	Modelo Directo . . . . .	43
4.3.4	Modelo Inverso . . . . .	43
4.3.5	Datos observados . . . . .	44
4.4	Experimento 4: Cambio de geometría . . . . .	44
4.4.1	Configuraciones evaluadas . . . . .	46
4.4.2	Entrenamiento . . . . .	48
4.4.3	Modelo directo . . . . .	49
4.4.4	Modelo inverso . . . . .	50
4.4.5	Datos observados . . . . .	50
<b>5</b>	<b>Discusión . . . . .</b>	<b>51</b>
5.1	Implementación de los kernels NTK . . . . .	51
5.2	Modelado acústico . . . . .	53
5.3	Inversión de velocidad . . . . .	54
5.4	Entrenamiento . . . . .	54
5.5	Efecto de la variación en la distribución de receptores . . . . .	56
5.6	Efecto de las variaciones en la distribución de receptores . . . . .	57

RED PINN PARA LA SOLUCIÓN DEL PROBLEMA SÍSMICO ACÚSTICO	6
5.7 Geometría de las redes feed-forward . . . . .	58
<b>6 Conclusiones . . . . .</b>	<b>60</b>
<b>Referencias . . . . .</b>	<b>62</b>

## Lista de Tablas

Tabla 1	Arquitectura de las redes que conforman la red PINN propuesta . . . . .	23
Tabla 2	Términos residuales de la red PINN . . . . .	24
Tabla 3	Errores del modelo directo en el experimento 1 . . . . .	35
Tabla 4	Errores del modelo directo en el experimento 2 . . . . .	39
Tabla 5	Errores del modelo directo en el experimento 3 . . . . .	44
Tabla 6	Arquitecturas utilizadas en el experimento 4 . . . . .	47

## Lista de Figuras

Figura 1	Arquitectura de red PINN. . . . .	20
Figura 2	Representación esquemática del dominio computacional de la red PINN para el campo de presión. . . . .	29
Figura 3	Ondícula de referencia para el modelamiento de los datos sintéticos. . . . .	31
Figura 4	Diagramas de evolución del entrenamiento de la red PINN en el entrenamiento de pozo cruzado . . . . .	34
Figura 5	Comparación del modelo directo en escenario de pozo cruzado para los campos $\phi, v_x, v_z$ . . . . .	35
Figura 6	Comparación entre el modelo real de velocidad de onda P y el parámetro físico $\alpha$ invertido por la red PINN (cuarta red <i>feed-forward</i> ) en el escenario de pozo cruzado. . . . .	36
Figura 7	Resultados de la predicción de datos observados para el modelo sísmico de pozo cruzado. . . . .	37
Figura 8	Diagramas de evolución del entrenamiento de la red PINN en el entrenamiento del modelo sísmológico . . . . .	38
Figura 9	Comparación del modelo directo en escenario sísmológico para los campos $\phi, v_x, v_z$ . . . . .	40
Figura 10	Comparación entre el modelo real de velocidad de onda P y el parámetro físico $\alpha$ invertido por la red PINN (cuarta red <i>feed-forward</i> ) para el escenario sísmológico. . . . .	41
Figura 11	Resultados de la predicción de datos observados para el modelo sísmológico. . . . .	42
Figura 12	Diagramas de evolución del entrenamiento de la red PINN en los entrenamientos de cambio de disposición de receptores . . . . .	43

Figura 13	Comparación de los resultados de propagación para el experimento 3 en los casos de cambio de receptores para los campos $\phi, v_x, v_z$ . . . . .	45
Figura 14	Comparación entre el modelo real de velocidad de onda P y el parámetro físico $\alpha$ invertido por la red PINN (cuarta red <i>feed-forward</i> ) para los escenarios con variación de receptores. . . . .	46
Figura 15	Comparación entre los resultados de predicción de datos observados para los escenarios con variación de receptores. . . . .	46
Figura 16	Métricas de los entrenamientos en el experimento de cambio de geometría. . . . .	48

### **Lista de Apéndices**

- Anexo A: Experimento 4 - Inferencias de campos de onda
- Anexo B: Experimento 4 - Inversiones de  $\alpha$
- Anexo C: Experimento 4 - Evolución de la función de costo
- Anexo D: Experimento 4 - Inferencias de datos observados

## Resumen

**Título:** Desarrollo de una red neuronal físicamente informada (PINN) usando kernels NTK para la solución del problema directo e inverso de propagación sísmica acústica\*

**Autor:** Juan Diego Bohórquez López\*\*

**Palabras Clave:** PINN, problema inverso, propagación sísmica acústica, redes neuronales

**Descripción:** La inversión sísmica y el modelado de la propagación de ondas son herramientas fundamentales para la caracterización de las propiedades físicas del subsuelo. Sin embargo, los métodos numéricos convencionales, desde el modelado acústico por diferencias finitas hasta la inversión de onda completa (FWI), requieren discretizaciones espaciales estrictas y presentan elevados costos computacionales. En este trabajo se desarrolla una Red Neuronal Físicamente Informada (PINN), formulada a partir de la ecuación acústica de primer orden, orientada a resolver de manera simultánea la propagación sísmica acústica y la estimación de modelos de velocidad, en escenarios sintéticos bidimensionales. La red se optimiza mediante la incorporación de kernels NTK (Neural Tangent Kernel), que permiten calcular un ajuste adaptativo para los pesos del entrenamiento, logrando reducir el contraste entre los términos de la función de costo de 32 a 1 orden de magnitud y equilibrar las escalas de aprendizaje entre los distintos componentes físicos. Los datos de entrenamiento y validación fueron generados a partir de un propagador numérico externo, utilizado como método de propagación de referencia. Los resultados demuestran que la red PINN reproduce los campos de presión y velocidad con errores relativos del orden de  $10^{-1}$ . Estos hallazgos confirman el potencial de las PINNs con NTK como herramientas eficientes para la simulación e inversión sísmica en entornos sintéticos.

---

\*Trabajo de grado

\*\*Facultad de Ciencias. Escuela de Física. Maestría en Geofísica. Directora: PhD. Ana Ramírez Silva. Codirector: PhD. Sergio Abreo Carrillo

### Abstract

**Title:** Development of a Physics-Informed Neural Network (PINN) using Neural Tangent Kernels for solving the forward and inverse acoustic seismic propagation problem\*

**Author:** Juan Diego Bohórquez López\*\*

**Key Words:** PINN, inverse problem, acoustic seismic propagation, neural networks

**Description:** Seismic inversion and wave propagation modeling are fundamental tools for characterizing subsurface physical properties. However, conventional numerical methods, ranging from finite-difference acoustic modeling to full waveform inversion (FWI), require strict spatial discretizations and entail high computational costs. In this work, a Physics-Informed Neural Network (PINN) is developed based on the first-order acoustic wave equation, aimed at simultaneously solving the forward problem of acoustic wave propagation and the inverse problem of velocity model estimation in two-dimensional synthetic scenarios. The network is optimized through the incorporation of Neural Tangent Kernel (NTK)-based weighting, enabling an adaptive adjustment of the training loss components. This approach reduces the imbalance between loss terms from 32 to 1 order of magnitude, effectively balancing the learning scales among the different physical constraints. Training and validation datasets were generated using an external numerical solver, employed as a reference wave propagation method. Results demonstrate that the PINN accurately reproduces pressure and velocity fields, achieving relative errors on the order of  $10^{-1}$ . These findings highlight the potential of NTK-enhanced PINNs as efficient tools for seismic simulation and inversion in controlled synthetic environments.

---

\*Master Thesis

\*\*Facultad de Ciencias. Escuela de Física. Maestría en Geofísica. Directora: PhD. Ana Ramírez Silva. Codirector: PhD. Sergio Abreo Carrillo

## Introducción

La inversión sísmica y la simulación de la propagación de ondas constituyen herramientas fundamentales para la caracterización de las propiedades físicas del subsuelo, al permitir inferir heterogeneidades y estructuras geológicas a partir de mediciones sísmicas (AlAli & Anifowose, 2022; Mondol, 2010). El problema inverso busca estimar parámetros físicos distribuidos en profundidad mediante el ajuste de modelos numéricos a datos observados (X. Wang, 2017; Yang & Ma, 2019), mientras que el problema directo describe la propagación de las ondas sísmicas en un modelo geológico dado, siendo la base para la generación de imágenes sísmicas y metodologías como la inversión de onda completa (FWI, *Full Waveform Inversion*) (Virieux et al., 2009). La construcción de modelos de velocidad requiere un modelo inicial, el cual puede obtenerse mediante enfoques basados en rayos o en la ecuación de onda (X. Wang, 2017). Aunque los métodos de FWI han sido aplicados con éxito gracias al desarrollo de la computación de alto desempeño, persisten desafíos relacionados con la construcción de modelos iniciales confiables, la adquisición de datos de baja frecuencia y la alta complejidad algorítmica (Tromp, 2020).

Los avances recientes en aprendizaje profundo han impulsado el desarrollo de soluciones en geociencias, como la detección de sismos o la estimación de mecanismos focales (Kuang et al., 2021). Sin embargo, la aplicación directa de estos enfoques a la inversión sísmica presenta limitaciones asociadas a la escasez de datos etiquetados. En el contexto de la inversión de onda completa (FWI), un “dato etiquetado” se entiende como un par entrada–salida donde las observaciones sísmicas (por ejemplo, sismogramas) están asociadas a un modelo de propiedades del subsuelo conocido (como el campo de velocidad). Este tipo de información es difícil de obtener en escenarios reales, ya que el modelo verdadero del subsuelo no es directamente observable y solo puede inferirse de manera indirecta mediante mediciones que no siempre están disponibles debido a su alto costo. Con el fin de mitigar esta limitación, se han propuesto enfoques híbridos que integran conocimiento físico en el proceso de aprendizaje, entre los que se destacan las redes

neuronales físicamente informadas.

Las redes neuronales físicamente informadas, introducidas por Raissi et al. (2017) bajo la sigla PINN (*Physics-Informed Neural Networks*), incorporan explícitamente las ecuaciones diferenciales parciales que gobiernan el fenómeno físico dentro de la función de costo. En este marco, la física actúa como un mecanismo de regularización, ya que restringe el espacio de soluciones admisibles al imponer que la red neuronal no solo ajuste los datos observados, sino que también satisfaga las leyes físicas del sistema. Este enfoque reduce la dependencia de grandes volúmenes de datos etiquetados y favorece una mayor capacidad de generalización en escenarios con información limitada (Karniadakis et al., 2021). En el contexto de la sismica, las PINN han sido empleadas para la generación de modelos de velocidad mediante la ecuación eikonal (Waheed et al., 2021), la ecuación acústica de segundo orden (Rasht-Behesht et al., 2022; Xu et al., 2019; Yang & Ma, 2019), así como para la simulación directa de la propagación de ondas (Ren et al., 2024; Zou et al., 2024). Una de las principales ventajas de estos enfoques es la reducción significativa del costo computacional una vez finalizado el entrenamiento de la red neuronal.

En el contexto del aprendizaje profundo, existen diversas arquitecturas de redes neuronales artificiales (ANN), cada una diseñada para capturar distintos tipos de patrones en los datos. Entre las más comunes se encuentran las redes completamente conectadas (*fully connected neural networks*, FCNN), en las cuales la información se propaga de manera unidireccional desde la capa de entrada, a través de múltiples capas ocultas, hasta la capa de salida. Por otro lado, las redes neuronales recurrentes (RNN) incorporan conexiones retroalimentadas que permiten modelar dependencias temporales, mientras que las redes neuronales convolucionales (CNN) están diseñadas para explotar estructuras espaciales en los datos, siendo ampliamente utilizadas en problemas de visión por computadora (Montesinos López et al., 2022).

En el caso de las redes neuronales físicamente informadas (PINN), la arquitectura más utilizada corresponde a redes completamente conectadas, debido a su capacidad para aproximar funciones continuas en dominios espacio-temporales. No obstante, a pesar de su potencial, estas archi-

tecturas presentan limitaciones asociadas con la inestabilidad del entrenamiento. Esta inestabilidad se origina principalmente en las diferencias de escala entre los distintos términos que componen la función de costo, lo que conduce a una dinámica de aprendizaje desbalanceada y dificulta la convergencia (S. Wang et al., 2022). Una estrategia para mitigar estos problemas consiste en introducir pesos adaptativos  $\lambda$  que permitan equilibrar la contribución de cada término durante el proceso de optimización. En este contexto, basándose en la formulación del *Neural Tangent Kernel* (NTK) desarrollada por Jacot et al. (2018), S. Wang et al. (2022) propusieron una metodología que ajusta automáticamente dichos pesos a partir de la traza de la matriz del NTK. Conceptualmente, la traza del NTK actúa como un indicador de la influencia relativa de cada término de la función de costo en la actualización de los parámetros de la red. De esta manera, al reescalar los términos de la pérdida en función de esta magnitud, se logra equilibrar sus tasas de aprendizaje efectivas, lo que contribuye a estabilizar el proceso de entrenamiento y mejorar la convergencia.

Mientras que Rasht-Behesht et al. (2022) formulaban el problema acústico mediante la ecuación de onda de segundo orden, en este trabajo se propone una alternativa basada en la ecuación acústica de primer orden, una formulación ampliamente utilizada que describe explícitamente las componentes de presión, velocidad y densidad del medio.

Desde la perspectiva de los problemas inversos geofísicos, aspectos como la estabilidad del proceso de inversión, la robustez frente a limitaciones en los datos y la capacidad de resolución del modelo recuperado constituyen elementos centrales para la interpretación de los resultados. En este sentido, las redes neuronales físicamente informadas ofrecen un marco de optimización en el que las ecuaciones diferenciales que gobiernan la propagación de ondas actúan como una forma de regularización física del problema inverso, restringiendo el espacio de soluciones a aquellas que satisfacen las leyes físicas del sistema. Esto contribuye a mejorar la estabilidad del proceso de inversión y a obtener modelos físicamente consistentes incluso en escenarios con cobertura de datos limitada. Adicionalmente, la incorporación de estrategias de ponderación adaptativa basadas en el análisis del *Neural Tangent Kernel* permite equilibrar la dinámica de entrenamiento entre los

distintos términos de la función de costo, favoreciendo una optimización más estable y robusta. En conjunto, estos elementos proporcionan un marco metodológico que busca mejorar la capacidad del modelo para recuperar estructuras del subsuelo a partir de información sísmica parcial, una condición frecuente en aplicaciones sismológicas reales.

En este contexto, el presente trabajo contribuye al avance del estado del arte en redes PINN aplicadas al modelado e inversión de ondas sísmicas acústicas al:

1. Adoptar una formulación acústica de primer orden para guiar el entrenamiento
2. Integrar una estrategia de pesos dinámicos basada en el cálculo del kernel NTK (S. Wang et al., 2022) para mejorar la estabilidad y la convergencia del modelo
3. Evaluar el desempeño del método propuesto en escenarios sintéticos bidimensionales que representan configuraciones sísmicas (experimento 1) y sismológicas (experimentos 2, 3 y 4) controladas.

## 1 Objetivos

### 1.1 Objetivo general

Desarrollar una red neuronal físicamente informada (PINN) orientada a la solución del problema directo de propagación acústica de una fuente y de la solución del problema inverso de estimación de modelos de velocidad, en medios sintéticos 2D.

### 1.2 Objetivos específicos

- Diseñar e implementar una red neuronal físicamente informada (PINN) en lenguaje Python, formulada a partir de la ecuación acústica de primer orden e implementada mediante una arquitectura totalmente conectada, orientada a la solución de los problemas directo e inverso en medios bidimensionales.
- Entrenar la red neuronal físicamente informada (PINN) sobre escenarios sintéticos bidimensionales, utilizando como datos de entrada la formulación física de la propagación sísmica acústica y observaciones simuladas de los campos de presión y velocidad de las partículas.
- Optimizar el proceso de entrenamiento de la red mediante la incorporación de kernels NTK (Neural Tangent Kernel), permitiendo el ajuste adaptativo de los pesos de la función de costo en cada uno de sus términos, con el objetivo de mejorar la estabilidad y la convergencia.
- Validar los resultados obtenidos por la red PINN en ambos problemas, comparando las simulaciones directas con las soluciones numéricas obtenidas mediante métodos de diferencias finitas, y evaluando las inversiones con respecto a los modelos de referencia sintéticos (ground truth).

## 2 Metodología

### 2.1 Formulación del problema

#### 2.1.1 Inversión sísmica de velocidad

La inversión sísmica tiene como objetivo reconstruir un modelo cuantitativo del subsuelo a partir de mediciones sísmicas. En este proceso se estiman parámetros físicos espacialmente variables (como la velocidad de onda P), que controlan la propagación de las ondas en el medio (X. Wang, 2017).

La inversión sísmica constituye un caso particular de los problemas inversos mal condicionados, caracterizados por su no unicidad y no linealidad (Aster et al., 2019; Y. Wang, 2016). En la práctica, estos problemas se abordan comúnmente mediante esquemas iterativos que minimizan una función de costo definida a partir del desajuste entre los datos observados y los datos sintéticos, empleando algoritmos de optimización y estrategias de regularización orientadas a estabilizar la solución (Aster et al., 2019).

De manera general, un problema de estimación de parámetros puede expresarse como un sistema de ecuaciones de la forma:

$$G(m) = d, \quad (1)$$

donde  $m$  representa el conjunto de parámetros físicos que definen el modelo,  $d$  corresponde a las observaciones disponibles, y  $G$  es el operador directo que encapsula la física del problema (como la ecuación de onda) y establece la relación entre el dominio del modelo y el dominio de los datos (Aster et al., 2019).

En el contexto de la inversión sísmica acústica, la ecuación (1) se interpreta considerando  $m$  como el modelo de velocidades de onda P,  $d$  como los datos sísmicos registrados en superficie o en profundidad, y  $G$  como el operador de propagación de ondas acústicas. Resolver el problema inverso consiste en estimar  $m$  a partir de  $d$ , lo cual requiere técnicas capaces de mitigar las dificul-

tades inherentes al carácter mal condicionado del operador  $G$ . Entre estas estrategias se incluyen la incorporación de términos de regularización, como la regularización de Tikhonov (Y. Wang et al., 2012), así como el uso de esquemas multiescala en frecuencia en la FWI y la construcción de modelos iniciales o de fondo (*background*) cercanos a la solución, que reduzcan la no linealidad del problema (Virieux & Operto, 2009).

A diferencia de los enfoques tradicionales basados en esquemas de inversión iterativa, en este trabajo el modelo de velocidad se obtiene como una función aprendida por una red neuronal, evaluable de manera continua en cualquier punto del dominio espacial. Este enfoque permite reformular el problema inverso dentro de un marco de aprendizaje profundo con restricciones físicas, evitando la necesidad de un proceso iterativo externo de actualización del modelo.

### 2.1.2 Propagación acústica

La propagación de ondas acústicas en un medio bidimensional puede expresarse mediante la ecuación acústica de segundo orden para el potencial escalar  $\phi$ , bajo la suposición de ausencia de fuerzas de cuerpo adicionales y con densidad constante:

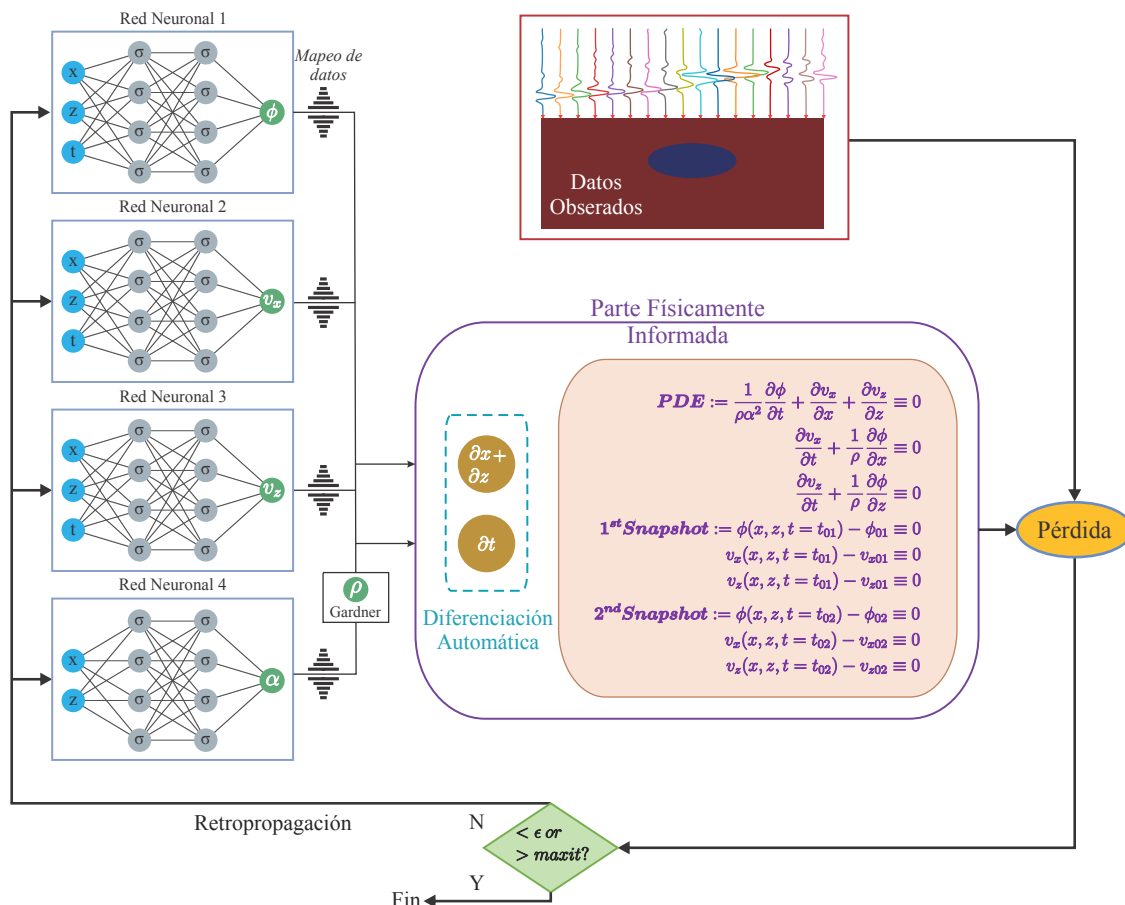
$$\alpha^2 \nabla^2 \phi + f = \frac{\partial^2 \phi}{\partial t^2}, \quad (2)$$

en donde  $\nabla \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$  es el operador Laplaciano definido en coordenadas cartesianas  $(x, z)$  del sistema,  $f$  es una fuerza externa (o función de fuente) y  $\alpha$  caracteriza la velocidad de onda acústica que depende de las propiedades del medio a través de la relación  $\alpha = \sqrt{K/\rho}$ , con  $K$  y  $\rho$  siendo el módulo de incompresibilidad (módulo de bulk) y la densidad, respectivamente.

La suposición de densidad constante permite simplificar la formulación de la ecuación de onda, desacoplando los efectos de la densidad y evitando la aparición de términos adicionales asociados a gradientes espaciales de  $\rho$ . No obstante, esta hipótesis no se cumple en medios geológicos reales, donde la densidad puede variar significativamente y afectar la amplitud y el comportamiento de las ondas propagadas. En consecuencia, el modelo acústico considerado representa una aproximación que captura adecuadamente la cinemática de las ondas, pero puede limitar la precisión en

**Figura 1**

*Arquitectura de red PINN.*



*Nota.* Izquierda: Redes neuronales completamente conectadas, la salida se aproxima a la solución tanto del problema directo como del problema inverso. Derecha: Leyes físicas y datos observados (sismogramas) utilizados para optimizar los parámetros de la PINN. la descripción de efectos dinámicos asociados a contrastes de impedancia.

Una formulación equivalente consiste en expresar la ecuación (2) como un sistema de ecuaciones de primer orden, introduciendo explícitamente los campos de velocidad de las partículas  $v_x, v_z$  y la densidad del medio  $\rho$ , expresada como:

$$\begin{aligned}
\frac{1}{\rho\alpha^2} \frac{\partial\phi}{\partial t} &= -\frac{\partial v_x}{\partial x} - \frac{\partial v_z}{\partial z} + f \\
\frac{\partial v_x}{\partial t} &= -\frac{1}{\rho} \frac{\partial\phi}{\partial x} \\
\frac{\partial v_z}{\partial t} &= -\frac{1}{\rho} \frac{\partial\phi}{\partial z}.
\end{aligned} \tag{3}$$

Esta formulación de primer orden proporciona directamente los campos de presión  $\phi$  y velocidades  $v_x, v_z$ , permitiendo obtener un cubo espacio–temporal que describe la propagación de ondas en el medio.

## 2.2 Red neuronal físicamente informada

Las Redes Neuronales Físicamente Informadas (PINN) combinan la información recopilada a partir de mediciones de un fenómeno físico con las ecuaciones diferenciales parciales (EDP) que describen parcial o totalmente la ley física subyacente a estos datos (Raissi et al., 2019). Esta información se incorpora directamente en la función de pérdida como un término penalizador durante el proceso de entrenamiento de una red neuronal artificial (ANN, por sus siglas en inglés *Artificial Neural Network*). En este contexto, un término penalizador corresponde a una contribución adicional en la función de costo que cuantifica el grado de incumplimiento de la ecuación diferencial gobernante; es decir, penaliza aquellas soluciones que no satisfacen la física del problema. De esta manera, la red no solo se ajusta a los datos observados, sino que también es forzada a aproximar soluciones físicamente consistentes. Las EDP pueden ser ecuaciones de orden entero, ecuaciones integro-diferenciales, ecuaciones fraccionales o ecuaciones estocásticas (Karniadakis et al., 2021).

### 2.2.1 Arquitectura

Las arquitecturas para las redes PINN han estado tradicionalmente basadas en redes *feed-forward* completamente conectadas (Rasht-Behesht et al., 2022). Siguiendo este enfoque, la arquitectura utilizada en esta investigación mantiene esta configuración.

El modelo propuesto se compone de tres redes neuronales independientes (figura 1) para aproximar los campos de onda  $\phi, v_x, v_z$ , utilizando como variables de entrada las coordenadas físicas  $x, z, t$ . Cada una presenta la misma configuración: una capa de entrada con 3 neuronas, tres

capas ocultas de 30 neuronas cada una y una capa de salida con 1 neurona (tabla 1), que representa el campo correspondiente. La selección de esta arquitectura se basó en configuraciones reportadas en la literatura de PINNs, particularmente en Raissi et al. (2017), donde se obtienen buenos resultados con entre 1 y 8 capas ocultas y entre 10 y 40 neuronas por capa. En este trabajo, la configuración adoptada responde a un compromiso entre capacidad de representación y costo computacional. Adicionalmente, su sensibilidad fue evaluada en el experimento 4, analizando la influencia del número de capas y neuronas en la convergencia y precisión.

Para estimar la velocidad de onda P ( $\alpha$ ), se incorpora una cuarta red neuronal *feedforward* con una arquitectura distinta, que recibe como entrada únicamente las coordenadas espaciales ( $x, z$ ), dado que  $\alpha$  es independiente del tiempo. Esta selección se basa en configuraciones reportadas para problemas de inversión con PINNs (Rasht-Behesht et al., 2022), donde se emplean redes más profundas para aproximar parámetros del medio. En este contexto, se adopta una arquitectura más profunda que la de los campos de onda, con el fin de capturar variaciones espaciales del modelo de velocidad. La densidad ( $\rho$ ) se aproxima mediante la relación empírica de Gardner ( $\rho = 0.31\alpha^{0.25}$ ), utilizando como entrada la salida de la red que estima  $\alpha$ . Esta aproximación reduce el número de variables independientes en el problema de inversión. No obstante, la relación de Gardner fue derivada para rocas sedimentarias consolidadas (Gardner et al., 1974), por lo que su aplicabilidad en otros materiales puede ser limitada y requerir relaciones de física de rocas más adecuadas.

Los valores obtenidos de cada red se integran en la parte físicamente informada del modelo (lado derecho de la figura 1), donde se definen las ecuaciones diferenciales que restringen el espacio de soluciones de la PINN. Además, se incorpora un término residual basado en los datos observados para construir la función de costo minimizada durante la optimización.

Los elementos de la ecuación diferencial (ecuación (3)) se obtienen mediante diferenciación automática de los campos de onda respecto a las variables de entrada, utilizando TensorFlow (Abadi et al., 2016). La diferenciación automática permite calcular derivadas exactas de funciones representadas computacionalmente mediante la aplicación sistemática de la regla de la cadena

(Paszke et al., 2017). En PINNs, esto permite evaluar de forma eficiente y precisa los términos diferenciales sin recurrir a aproximaciones como diferencias finitas. Se seleccionó como función de activación  $\tanh()$  para mantener consistencia con Rasht-Behesht et al. (2022), donde se reporta buen desempeño en arquitecturas similares.

**Tabla 1**

*Arquitectura de las redes que conforman la red PINN propuesta*

	$\phi$	$v_x$	$v_z$	$\alpha$
Neuronas de entrada	3	3	3	2
Capas ocultas	3	3	3	5
Neuronas por capa	30	30	30	20
Neuronas de salida	1	1	1	1

### 2.2.2 Función de costo

La función de costo que guía el entrenamiento de la red PINN integra los distintos términos necesarios para imponer simultáneamente la física del problema y la información proveniente de los datos observados y modelados (ecuación (4)). Dicha función se define como una combinación ponderada de errores cuadráticos medios (MSE), donde cada término se evalúa sobre dominios específicos del espacio y del tiempo, de acuerdo con la naturaleza de la restricción que representa.

En particular, la función de costo incluye tres componentes principales: (i) el término asociado al residuo de la ecuación diferencial parcial (EDP), que garantiza el cumplimiento de la ecuación acústica gobernante dentro del dominio; (ii) el término correspondiente a los sismogramas sintéticos, que impone el ajuste entre las predicciones de la red y los datos observados en las posiciones de los receptores; y (iii) los términos asociados a instantáneas de los campos de onda  $\phi$ ,  $v_x$  y  $v_z$ , evaluadas en dos instantes temporales distintos, los cuales proporcionan información adicional sobre la evolución dinámica del campo de ondas.

**Tabla 2***Términos residuales de la red PINN*

Residual	Expresión	Descripción
$R_{EDP} :=$	$\frac{1}{\rho\alpha^2} \frac{\partial\phi}{\partial t} + \frac{\partial\phi}{\partial x} + \frac{\partial\phi}{\partial z} \equiv 0$ $\frac{\partial v_x}{\partial t} + \frac{1}{\rho_x} \frac{\partial\phi}{\partial x} \equiv 0$ $\frac{\partial v_z}{\partial t} + \frac{1}{\rho_z} \frac{\partial\phi}{\partial z} \equiv 0$	EDP
$R_{\phi_1} :=$	$\phi(x, z, t_{01}) - \phi_{01} \equiv 0$	Primer instantánea de tiempo (01)
$R_{v_{x1}} :=$	$v_x(x, z, t_{01}) - v_{x01} \equiv 0$	
$R_{v_{z1}} :=$	$v_z(x, z, t_{01}) - v_{z01} \equiv 0$	
$R_{\phi_2} :=$	$\phi(x, z, t_{02}) - \phi_{02} \equiv 0$	Segunda instantánea de tiempo (02)
$R_{v_{x2}} :=$	$v_x(x, z, t_{02}) - v_{x02} \equiv 0$	
$R_{v_{z2}} :=$	$v_z(x, z, t_{02}) - v_{z02} \equiv 0$	
$R_{obs} :=$	$v_x(x_r, z_r, t) - EHE(x_r, z_r, t) \equiv 0$ $v_z(x_r, z_r, t) - EHZ(x_r, z_r, t) \equiv 0$	*Datos observados en las posiciones de los receptores

\*Los términos  $EHE$  y  $EHZ$  representan los datos de velocidad de partícula observados en los receptores sísmicos. En particular,  $EHE(x_r, z_r, t)$  corresponde al componente horizontal de la velocidad de partícula ( $v_x$ ), mientras que  $EHZ(x_r, z_r, t)$  representa el componente vertical ( $v_z$ ).

$$\begin{aligned}
\mathcal{L}(\Theta) = & \lambda_1 MSE_{EDP} + \lambda_2 MSE_{Obs} + \lambda_3 MSE_{\phi(x,z,t_{01})} + \\
& \lambda_4 MSE_{v_x(x,z,t_{01})} + \lambda_5 MSE_{v_z(x,z,t_{01})} + \\
& \lambda_6 MSE_{\phi(x,z,t_{02})} + \lambda_7 MSE_{v_x(x,z,t_{02})} + \\
& \lambda_8 MSE_{v_z(x,z,t_{02})}.
\end{aligned} \tag{4}$$

Aquí,  $\Theta$  representa el conjunto de pesos y sesgos de todas las redes neuronales (NNs, por sus siglas en inglés *Neural Networks*) evaluadas en los puntos de entrenamiento, mientras que los hiperparámetros  $\lambda_i$  son los coeficientes que controlan la contribución relativa de cada término en la función de costo. Aunque estos pesos pueden fijarse de manera heurística, esta práctica puede conducir a inestabilidad numérica debido a las diferencias en las magnitudes y tasas de variación de los distintos términos de la función de costo (naturaleza multiescala), lo que genera desbalances en los gradientes durante el entrenamiento (S. Wang, Wang & Perdikaris, 2021). Para mitigar

este efecto, se adoptó un esquema de pesos dinámicos basado en el cálculo de la traza del kernel NTK (*Neural Tangent Kernel*), siguiendo la metodología propuesta por S. Wang et al. (2022). Esta estrategia permite ajustar los  $\lambda_i$  en cada época de entrenamiento en función de la escala de los gradientes, favoreciendo una convergencia más estable y equilibrada al reducir la discrepancia entre las tasas de aprendizaje efectivas de los distintos términos.

### 2.2.3 *Kernels NTK*

Cada uno de los coeficientes  $\lambda_i$  que componen la función de costo (ecuación (4)) está asociado a un término residual definido en la tabla 2. Estos términos representan los objetivos del proceso de optimización y corresponden a los residuos de la EDP, los datos observados y las instantáneas de los campos  $\phi, v_x, v_z$ . En el caso de los datos observados y las instantáneas (*snapshots*), el residual se define como la diferencia entre la salida de la red neuronal y los datos generados mediante el propagador numérico. Por su parte, el residual de la EDP se construye evaluando la ecuación gobernante utilizando las salidas de las redes *feedforward*, de modo que cuantifica el grado de incumplimiento de la ley física en cada punto del dominio.

Todos los términos residuales se reducen mediante el error cuadrático medio (MSE), definido de manera general como:

$$MSE_{R_k} = \frac{1}{N_{R_k}} \sum_{i=1}^{N_{R_k}} (R_k(x_i, z_i, t_i))^2 \quad (5)$$

donde  $R_k$  denota cualquier residual incluido en la función de costo y  $N_{R_k}$  es el número de puntos muestreados en su respectivo subdominio de entrenamiento.

Los kernels NTK se obtienen a partir de los jacobianos de cada una de las variables de salida involucradas en los términos residuales. De manera compacta, el kernel total  $K_T(t)$  puede escribirse como:

$$K_T(t) = \begin{bmatrix} J_{EDP} \\ J_{v_x} \\ J_{v_z} \\ J_{\phi} \end{bmatrix} \begin{bmatrix} J_{EDP}^T, J_{v_x}^T, J_{v_z}^T, J_{\phi}^T \end{bmatrix} \quad (6)$$

o, similarmente:

$$K_T(t) = \begin{bmatrix} K_{EE}(t) & K_{Ev_x}(t) & K_{Ev_z}(t) & K_{E\phi}(t) \\ K_{v_xE}(t) & K_{v_xv_x}(t) & K_{v_xv_z}(t) & K_{v_x\phi}(t) \\ K_{v_zE}(t) & K_{v_zv_x}(t) & K_{v_zv_z}(t) & K_{v_z\phi}(t) \\ K_{\phi E}(t) & K_{\phi v_x}(t) & K_{\phi v_z}(t) & K_{\phi\phi}(t) \end{bmatrix}. \quad (7)$$

Cada término diagonal se define como:

$$K_{R_i R_j}(t) = J_{R_i}(t) J_{R_j}^T(t) \quad (8)$$

donde  $J_{R_i}(t) = \partial R_i / \partial \Theta$  corresponde al jacobiano del residual  $R_i$  con respecto a los parámetros de las redes *feedforward* ( $\Theta$ ). De manera general, el kernel  $K_{R_i R_j}(t)$  se calcula para todos los términos que componen la función de costo (ecuación (4)), considerando tanto el dominio espacial en el que se evalúan como la red neuronal asociada a cada variable.

Siguiendo el esquema de S. Wang et al. (2022), los pesos adaptativos se obtienen a partir de la traza de la matriz NTK:

$$\lambda_{R_k} = \frac{Tr(K_T(t))}{Tr(K_{R_k R_k}(t))} \quad (9)$$

donde  $Tr(\cdot)$  representa la traza del kernel correspondiente. Dado que la traza solo involucra los términos diagonales, la matriz completa puede reducirse a:

$$K_T(t) = \begin{bmatrix} K_{EE}(t) & & & \\ & K_{v_x v_x}(t) & & \\ & & K_{v_z v_z}(t) & \\ & & & K_{\phi\phi}(t) \end{bmatrix}. \quad (10)$$

### 2.3 Evaluación del modelo

Las inferencias realizadas con la red entrenada fueron evaluadas en los distintos escenarios definidos mediante las normas  $L1$  y  $L2$ , utilizadas para calcular los errores relativos de predicción (ecuaciones (11) y (12)). En este contexto,  $x \in \mathbb{R}^n$  representa el vector de datos sintéticos (referencia) y  $\bar{x} \in \mathbb{R}^n$  el vector de valores predichos por la red PINN.

$$r_{L1} = \frac{\|x - \bar{x}\|_1}{\|x\|_1} \quad (11)$$

$$r_{L2} = \frac{\|x - \bar{x}\|_2}{\|x\|_2}. \quad (12)$$

Por otra parte, para evaluar el ajuste entre los datos de referencia y las predicciones del modelo, se utilizó el coeficiente de determinación ( $R^2$ ), el cual cuantifica la proporción de la varianza explicada por el modelo. Su valor máximo es 1 (ajuste perfecto) y puede tomar valores negativos cuando el modelo presenta un desempeño inferior al de un modelo basado en la media de los datos. Si  $\hat{y}_i$  es el valor predicho de la muestra  $i$  y  $y_i$  el valor de referencia correspondiente, para un total de  $n$  muestras,  $R^2$  se define como:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

donde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  y  $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \epsilon_i^2$ .

### 3 Configuración experimental

#### 3.1 Datos Sintéticos

En el contexto de esta red PINN, los datos sintéticos corresponden a información generada numéricamente que permite restringir el espacio de soluciones en un entorno controlado, mediante la minimización de los términos residuales asociados a la física de propagación acústica dentro de la función de costo. Estos datos actúan como referencia para guiar el aprendizaje de la red, asegurando consistencia con el comportamiento físico esperado.

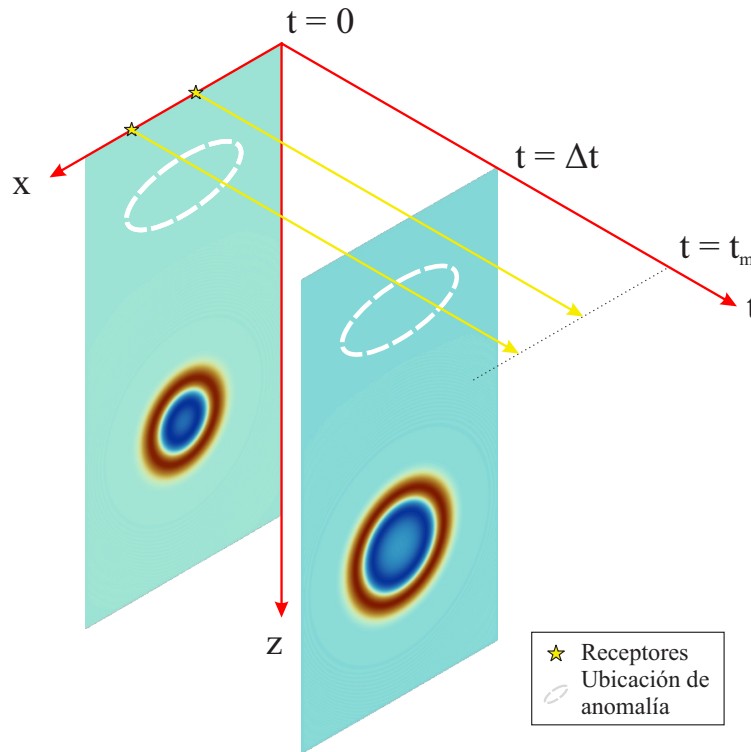
Para abordar el problema sísmico acústico, se implementó una estrategia de entrenamiento basada en instantáneas (*snapshots*), que utiliza dos registros temporales de los campos de onda  $(\phi, v_x, v_z)$  separados por un intervalo  $\Delta t$ . Este intervalo fue seleccionado de manera heurística, garantizando que ambos instantes de tiempo se ubicaran antes de la interacción del frente de onda con la anomalía de velocidad, sin ningún otro criterio para su selección. De esta forma, se limita la información disponible sobre la propagación, de modo que las instantáneas representen un comportamiento cercano al de un medio homogéneo isotrópico, mientras que los sismogramas sintéticos constituyen la principal fuente de información sobre la interacción con el modelo de velocidad completo, de forma análoga a lo que ocurre en la realidad. En este esquema, la primera instantánea de tiempo captura la forma del frente de onda, mientras que la segunda impone la dirección de propagación en el tiempo  $t$ .

Adicionalmente, las trazas sintéticas de velocidad en las componentes  $x$  y  $z$  se utilizan como datos observados en posiciones discretas del dominio. Estos sismogramas se obtienen a partir de receptores sintéticos ubicados en coordenadas predefinidas  $x, z$ , como se muestra en la figura 2, para todo el intervalo temporal de propagación. Su inclusión en la función de pérdida se realiza mediante los términos residuales correspondientes para cada componente, proporcionando restricciones adicionales al modelo para reproducir la respuesta observada del sistema.

La generación de todos los datos sintéticos se realizó mediante un propagador numérico

**Figura 2**

*Representación esquemática del dominio computacional de la red PINN para el campo de presión.*



desarrollado por el grupo de investigación CPS, UIS. Este propagador, escrito en lenguaje C, resuelve la ecuación acústica de primer orden en medios heterogéneos ( $\alpha$  y  $\rho$  variables), utilizando un esquema de diferencias finitas de octavo orden de precisión. Todos los experimentos definidos en este trabajo se configuraron considerando el número de Courant–Friedrichs–Lewy (CFL) máximo permitido, con el fin de seleccionar pasos espaciales y tamaño de malla de modo que se garantizara la estabilidad numérica de la simulación. Para integrarlo en el flujo de trabajo, se desarrolló una interfaz en Python encargada de ejecutar el propagador, gestionar sus parámetros de entrada y procesar las salidas requeridas para el entrenamiento de la red PINN. Esta integración permitió un control sobre las condiciones numéricas y la reproducibilidad de los escenarios sintéticos empleados.

### 3.1.1 *Submuestreo*

Para llevar a cabo el entrenamiento de la red PINN, los datos simulados se submuestran previamente con el objetivo de reducir la densidad de puntos, lo que permite disminuir la carga computacional manteniendo una calidad suficiente de la información utilizada. Para garantizar una cobertura adecuada del dominio físico, la selección de los puntos se lleva a cabo de manera uniforme.

1. Los datos etiquetados en este esquema de aprendizaje corresponden a las instantáneas de tiempo de la propagación de los campos  $\phi, v_x, v_z$ . Esta información excluye las regiones absorbentes (CPML), las cuales no se consideran parte del dominio de entrenamiento, ya que en estas zonas, debido a su implementación, la PDE (ecuación (3)) no es aplicable.
2. Los datos observados en este esquema de aprendizaje corresponden a los sismogramas que capturan la variación temporal de los campos  $v_x, v_z$  dentro del modelo. Para reducir la cantidad de muestras que procesa la red, se submuestra cada vector en un patrón regular equiespaciado.

### 3.1.2 *Fuente sísmica*

La fuente sísmica que se usó para la generación de los datos sintéticos de entrenamiento fue una fuente gaussiana, definida matemáticamente por la ecuación (14). Esta elección se debe a su suavidad y simplicidad espectral, lo cual resulta conveniente para el entrenamiento de la red PINN, al reducir la presencia de componentes de alta frecuencia.

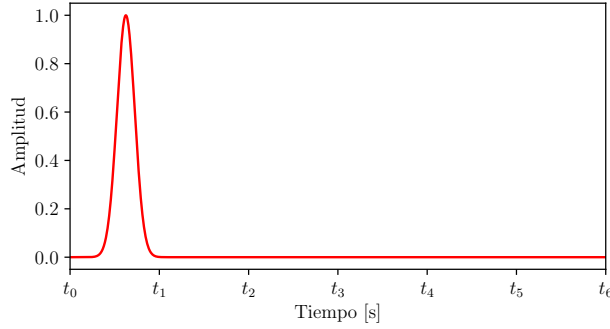
Para el experimento 1 se empleó una frecuencia dominante  $f_q$  de 16  $Hz$ , mientras para los experimentos 2, 3 y 4 una  $f_q$  de 1.6  $Hz$ , asegurando una estabilidad numérica en el modelamiento. La forma de esta ondícula para el modelado directo puede observarse en la figura 3 (gráfico de referencia).

$$f = e^{-\zeta t^2} \quad (14)$$

siendo  $\zeta = 2\pi^2 f_q^2$ .

### Figura 3

*Ondícula de referencia para el modelamiento de los datos sintéticos.*



### 3.2 Normalización

Con el fin de garantizar una convergencia estable durante el entrenamiento de las redes neuronales, se implementaron estrategias de normalización tanto en las variables de entrada como en las salidas de las redes *feed-forward*. Estas normalizaciones permiten equilibrar las escalas de los distintos términos de la función de costo y mejorar el comportamiento numérico del proceso de optimización.

Los pasos de normalización implementados fueron:

1. Para los datos de entrada de cada red *feed-forward*, las coordenadas espacio-temporales  $(x, z, t)$  se mapearon al intervalo  $[-1, 1]$  siguiendo la ecuación (15). Este reescalamiento se realizó utilizando las dimensiones del dominio computacional en cada dirección, con el objetivo de homogenizar la escala de las entradas y facilitar el aprendizaje de las dependencias espaciales y temporales por parte de la red neuronal.

$$\mathbf{H} = 2 \left( \frac{\mathbf{X}}{\mathbf{u}_b} \right) - 1 \quad (15)$$

donde  $\mathbf{H}$  representa las coordenadas normalizadas,  $\mathbf{X} = (x, z, t)$  corresponde al vector de entrada de la red *feed-forward* ( $\mathbf{X} = (x, z)$  para la red de  $\alpha$ ), y  $\mathbf{u}_b = (a_x, a_z, t_f - t_s)$  son los

límites del dominio computacional en cada dimensión.

2. Para asegurar que la ecuación diferencial parcial se encontrara adecuadamente balanceada y que los campos de onda y los parámetros físicos aproximados satisficieran de manera consistente la ecuación gobernante, se desnormalizaron las salidas de las redes neuronales. Para ello, se aplicó una estrategia de reescalamiento que transformó las salidas normalizadas de cada red a un dominio donde existe la solución física de los campos de onda y de los parámetros del medio.

Matemáticamente, cada salida  $\hat{y}$  producida por la red neuronal se reescaló a su correspondiente valor físico  $y$  mediante una transformación afín de la forma:

$$y = a\hat{y} + b \quad (16)$$

donde los coeficientes  $a$  y  $b$  fueron seleccionados de manera heurística para garantizar que las soluciones obtenidas fueran consistentes con los valores esperados en el medio de propagación.

Este mapeo se aplicó específicamente a los parámetros de velocidad de onda ( $\alpha$ ), el potencial escalar ( $\phi$ ) y los componentes de velocidad de partícula ( $v_i$ ), asegurando que los valores obtenidos mantuvieran coherencia con el comportamiento físico esperado en la solución de la EDP. Por su parte, la densidad ( $\rho$ ) fue calculada a partir de la salida escalada de  $\alpha$ .

### 3.3 Detalles de implementación

Utilizando la función de costo definida en la ecuación (4), los parámetros de la red  $\Theta$  pueden estimarse resolviendo un problema de optimización, el cual consiste en minimizar la ecuación (17).

$$\Theta^* = \operatorname{argmin}\{MSE(\Theta)\}. \quad (17)$$

De manera convencional, las redes PINN se entrenan mediante la optimización simultánea de la función de costo total (ecuación (4)), utilizando métodos estándar de descenso de gradiente estocástico (Raissi et al., 2017; Xinquan Huang & T. Alkhalifah, 2023), como el optimizador Adam (Kingma & Ba, 2014). Bajo este enfoque pueden presentarse limitaciones, entre ellas el riesgo

de que la optimización converja a mínimos locales subóptimos. Por esta razón, recientemente se han propuesto metodologías de optimización multiobjetivo aplicadas a redes PINN, con resultados favorables (Lu et al., 2023), así como el uso de optimizadores de segundo orden como L-BFGS en problemas de inversión de velocidades sísmicas. No obstante, en este último caso no se ha evidenciado una mejora significativa en el proceso de entrenamiento (Rasht-Behesht et al., 2022). En este trabajo se adoptó el enfoque más común y de menor costo computacional, implementando el entrenamiento mediante el optimizador Adam en la librería TensorFlow, con los hiperparámetros:  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1e - 08$ , y una tasa de aprendizaje de 0.0001.

El presente trabajo fue implementado sobre un clúster de cómputo compuesto por nodos equipados con GPU Tesla V100 SXM2 (16 GB de memoria, 640 NVIDIA Tensor Cores y 5 120 NVIDIA CUDA Cores). Cada entrenamiento se ejecutó de manera individual en un nodo, utilizando TensorFlow versión 1.0 con aceleración por GPU, lo que permitió optimizar los tiempos de cómputo.

## 4 Resultados

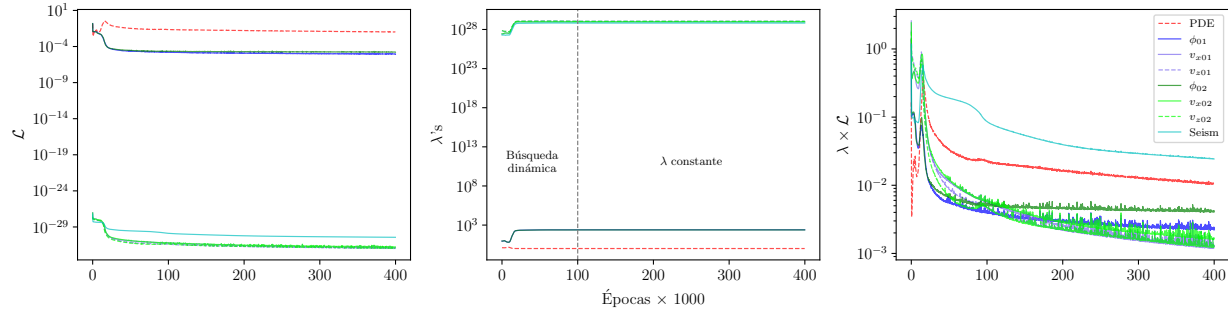
### 4.1 Experimento 1: Modelo pozo cruzado

El modelo utilizado en el escenario sísmico es un pozo cruzado con una anomalía elíptica de baja velocidad. Las dimensiones del dominio espacial son  $x = 1.23 \text{ km}$ ,  $z = 0.43 \text{ km}$ . En este contexto, se simuló la propagación de una onda acústica con origen en  $x = 0.43 \text{ km}$ ,  $z = 0.25 \text{ km}$ , a la izquierda de la anomalía de velocidad ( $2.5 \text{ km}$ , en un modelo de fondo de  $3 \text{ km}$ ). La evolución temporal de los campos escalares  $\phi$ ,  $v_x$ ,  $v_z$  se registró mediante 20 receptores sintéticos distribuidos en profundidad, con un espaciamiento de  $22.07 \text{ m}$ , ubicados al lado derecho del modelo (denotados como triángulos amarillos, ver figura 6).

Los receptores capturan la información sobre los campos de velocidad  $v_x$ ,  $v_z$  durante todo el tiempo de la propagación. En este escenario, se modelaron 950 pasos temporales, espaciados

**Figura 4**

*Diagramas de evolución del entrenamiento de la red PINN en el entrenamiento de pozo cruzado*



cada  $0.0006 s$ , lo que resulta en una propagación completa de  $0.57 s$ . Los dos instantes temporales seleccionados para extraer los datos etiquetados destinados al entrenamiento de la red neuronal son  $t_1 = 0.1 s$  y  $t_2 = 0.115 s$ .

El submuestreo de este experimento para los *snapshots* de entrenamiento de cada campo corresponde a una cuadrícula de  $50 \times 50$ , y para cada dato observado, una tasa de submuestreo temporal de  $1/2$ .

#### 4.1.1 Entrenamiento

Este modelo fue entrenado durante un total de 400 000 épocas, con una duración de 3 días, 16 horas y 29 minutos. La figura 4 muestra la evolución del entrenamiento, teniendo en cuenta la convergencia de los términos de la función de costo, el ajuste de los pesos adaptativos ( $\lambda$ 's) calculados y los términos de pérdida ponderados. De izquierda a derecha: 1) Evolución de los términos de pérdida durante el entrenamiento. 2) Evolución de los pesos ( $\lambda$ 's) obtenidos de la implementación de los kernels NTK. 3) Términos de pérdida ponderados por sus correspondientes  $\lambda$ , permitiendo la convergencia del modelo.

#### 4.1.2 Modelo directo

Las tres redes neuronales *feed-forward* entrenadas, correspondientes a los campos de onda  $\phi, v_x, v_z$ , predicen la distribución espacial del campo de onda real para cualquier tiempo dentro del dominio de la red PINN, resolviendo el problema directo en el modelo sísmico de pozo cruzado.

La inferencia de estos campos de onda en el tiempo de la segunda instantánea ( $t = 0.12$  s) puede verse en la figura 5 y sus respectivos errores de norma L1 y L2 en la tabla 3.

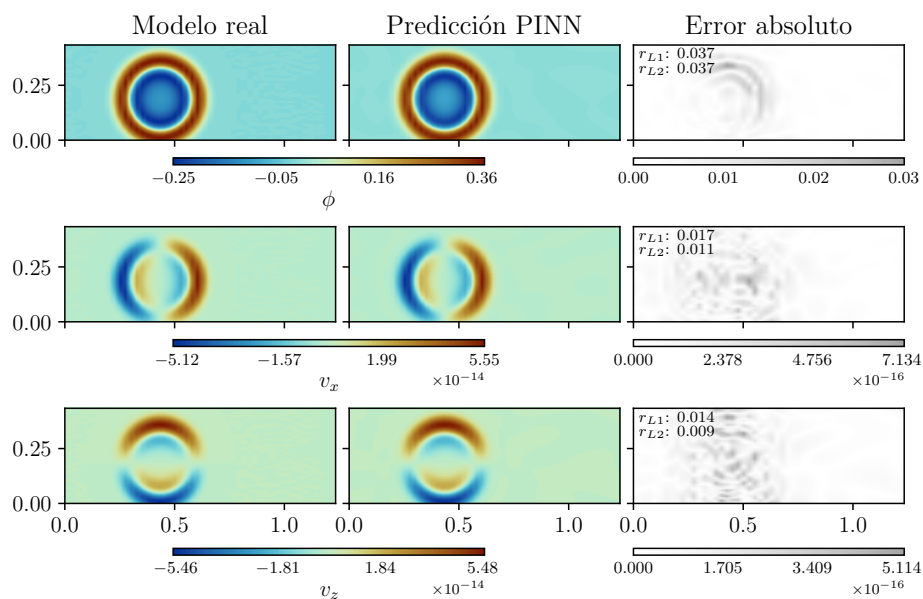
**Tabla 3**

*Errores del modelo directo en el experimento 1*

Campo	Error L1	Error L2
$\phi$	0.038	0.037
$v_x$	0.017	0.011
$v_z$	0.014	0.009

**Figura 5**

*Comparación del modelo directo en escenario de pozo cruzado para los campos  $\phi$ ,  $v_x$ ,  $v_z$ .*



*Nota.* Tiempo de predicción  $t = 0.12$ s,  $r_{L1}$  y  $r_{L2}$  corresponden al error relativo para la norma L1 y la norma L2, respectivamente.

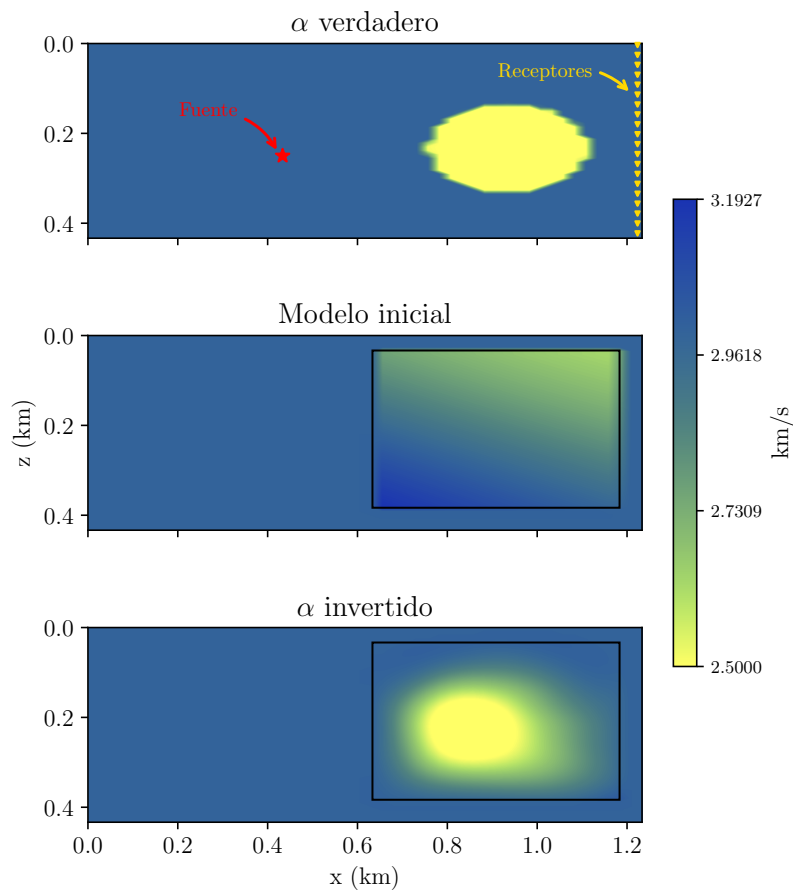
#### 4.1.3 Modelo inverso

El resultado de la cuarta red neuronal *feed-forward* del parámetro físico  $\alpha$  después del entrenamiento puede predecir la velocidad sísmica real para cualquier coordenada espacial dentro del

dominio de la red PINN, resolviendo así el problema inverso de este experimento. Este resultado, y su comparación con el modelo teórico de velocidad de onda P puede verse en la figura 6.

### Figura 6

*Comparación entre el modelo real de velocidad de onda P y el parámetro físico  $\alpha$  invertido por la red PINN (cuarta red feed-forward) en el escenario de pozo cruzado.*



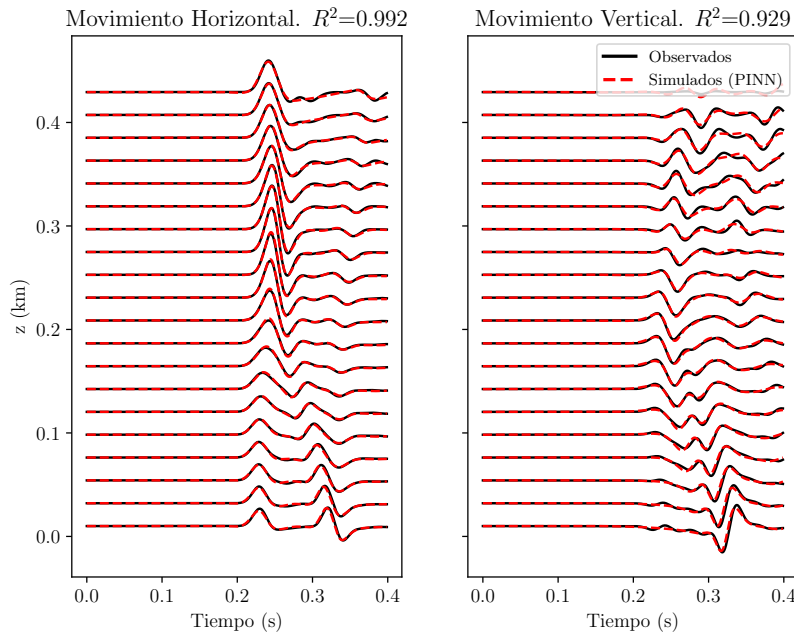
#### 4.1.4 Datos observados

Los datos observados pueden obtenerse como subproductos de las redes neuronales *feed-forward* de los campos de onda  $v_x, v_z$ , utilizando como datos de entrada las posiciones espaciales de los receptores dentro del dominio de la red PINN (representados a la derecha de la anomalía en la figura 6), y un vector de tiempo dentro de la propagación. Este resultado, y su comparación con los datos originales pueden verse en la figura 7, en donde se observa un coeficiente de determinación

$R^2$  de 0.992 para el campo  $v_x$  (movimiento horizontal) y de 0.929 para el campo  $v_z$  (movimiento vertical).

### Figura 7

*Resultados de la predicción de datos observados para el modelo sísmico de pozo cruzado.*



## 4.2 Experimento 2: Modelo Sismológico

El modelo utilizado en el escenario sismológico es una variación del modelo sísmico, caracterizada en este caso por una mayor dimensión en  $z$  que en  $x$ . Las dimensiones del dominio espacial son  $x = 6.8 \text{ km}$ ,  $z = 14.16 \text{ km}$ . En este contexto, la fuente sísmica se ubicó en las coordenadas  $x = 3.38 \text{ km}$ ,  $z = 9.77 \text{ km}$ . La anomalía, de forma elíptica y de baja velocidad ( $2.5 \text{ km}$ , en un modelo de fondo de  $3 \text{ km}$ ), se encuentra próxima a la superficie. Se simuló la propagación de una onda acústica desde este origen y se registró la evolución temporal de los campos escalares  $\phi$ ,  $v_x$ ,  $v_z$  mediante 20 receptores distribuidos en superficie, con un espaciamiento de  $357.59 \text{ m}$ , ubicados en el límite superior del dominio (denotados como triángulos amarillos, ver figura 10).

Los receptores capturan la información sobre los campos de velocidad  $v_x$ ,  $v_z$  durante todo el tiempo de la propagación. En este escenario, se modelaron 1100 pasos temporales, espaciados

cada 0.006  $s$ , lo que resulta en una propagación completa de 6.6  $s$ . Los dos instantes temporales seleccionados para extraer los datos etiquetados destinados al entrenamiento de la red neuronal son  $t_1 = 1 s$  y  $t_2 = 1.15 s$ .

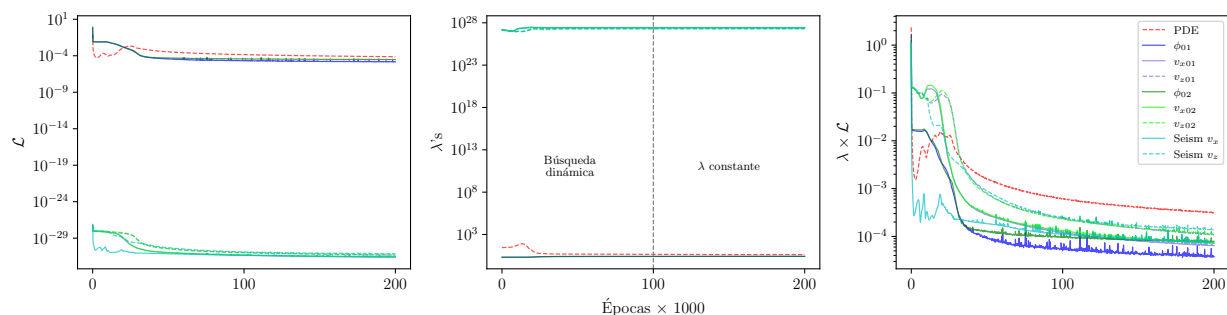
El submuestreo de este experimento para los *snapshots* de entrenamiento de cada campo corresponde a una cuadrícula de  $55 \times 55$ , y para cada dato observado, una tasa de submuestreo temporal de 1/2.

#### 4.2.1 Entrenamiento

Este modelo fue entrenado durante un total de 200 000 épocas, con una duración de 23 horas y 52 minutos. La figura 8 muestra la evolución del entrenamiento, teniendo en cuenta la convergencia de los términos de la función de costo, el ajuste de los pesos adaptativos ( $\lambda$ 's) calculados y los términos de pérdida ponderados. De izquierda a derecha: 1) Evolución de los términos de pérdida durante el entrenamiento. 2) Evolución de los pesos ( $\lambda$ 's) obtenidos de la implementación de los kernels NTK. 3) Términos de pérdida ponderados por sus correspondientes  $\lambda$ , permitiendo la convergencia del modelo.

### Figura 8

Diagramas de evolución del entrenamiento de la red PINN en el entrenamiento del modelo sísmico



#### 4.2.2 Modelo directo

Las tres redes neuronales *feed-forward* entrenadas, correspondientes a los campos de onda  $\phi$ ,  $v_x$ ,  $v_z$ , predicen la distribución espacial del campo de onda real para cualquier tiempo dentro

del dominio de la red PINN, resolviendo así el problema directo en el modelo sismológico. Un ejemplo de la inferencia de estos campos de onda en un tiempo posterior a la segunda instantánea ( $t = 1.2$  s tomando como referencia el tiempo total de la propagación) puede verse en la figura 9 y sus respectivos errores de norma L1 y L2 en la tabla 4.

**Tabla 4**

*Errores del modelo directo en el experimento 2*

Campo	Error L1	Error L2
$\phi$	0.121	0.107
$v_x$	0.099	0.066
$v_z$	0.134	0.086

#### 4.2.3 *Modelo inverso*

El resultado de la cuarta red neuronal *feed-forward* del parámetro físico  $\alpha$  después del entrenamiento puede predecir la velocidad sísmica real para cualquier coordenada espacial dentro del dominio de la red PINN, resolviendo así el problema inverso en el modelo sismológico. Este resultado, y su comparación con el modelo teórico de velocidad de onda P puede verse en la figura 10.

#### 4.2.4 *Datos observados*

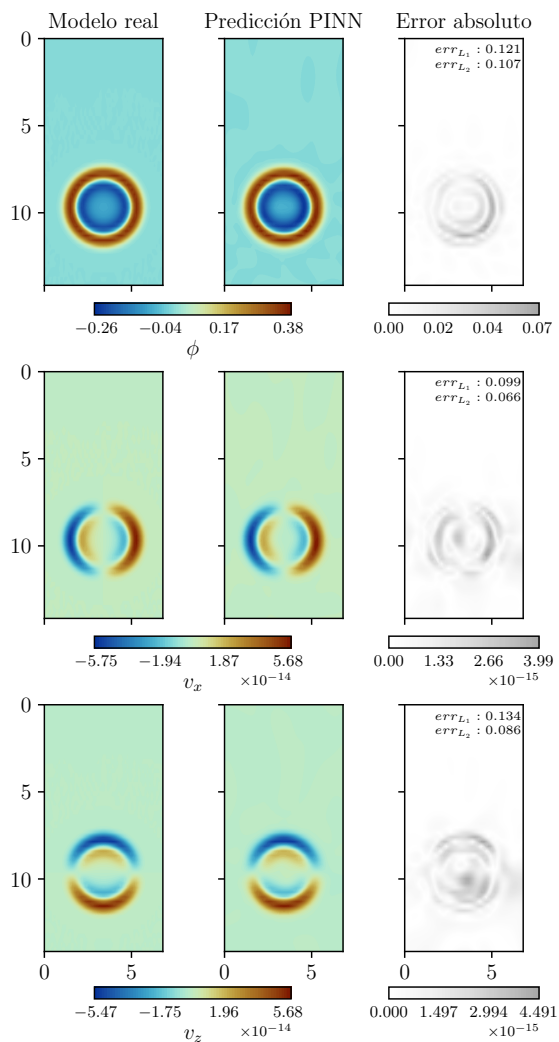
Los datos observados pueden obtenerse como subproducto de las redes neuronales *feed-forward* de los campos de onda  $v_x, v_z$ , utilizando el valor de cada campo y su evolución temporal en las coordenadas espaciales de los receptores dentro del dominio de la red PINN. Este resultado, y su comparación con los datos originales puede verse en la figura 11, en donde se observa un coeficiente de determinación  $R^2$  de 0.654 para el campo  $v_x$  (movimiento horizontal) y de 0.999 para el campo  $v_z$  (movimiento vertical).

### 4.3 Experimento 3: Modelo de cubrimiento imperfecto de receptores

Para evaluar de forma preliminar el impacto de la distribución de los receptores en superficie, considerando un esquema no uniforme, ni denso ni equiespaciado, se realizó una variación

**Figura 9**

Comparación del modelo directo en escenario sísmológico para los campos  $\phi$ ,  $v_x$ ,  $v_z$ .



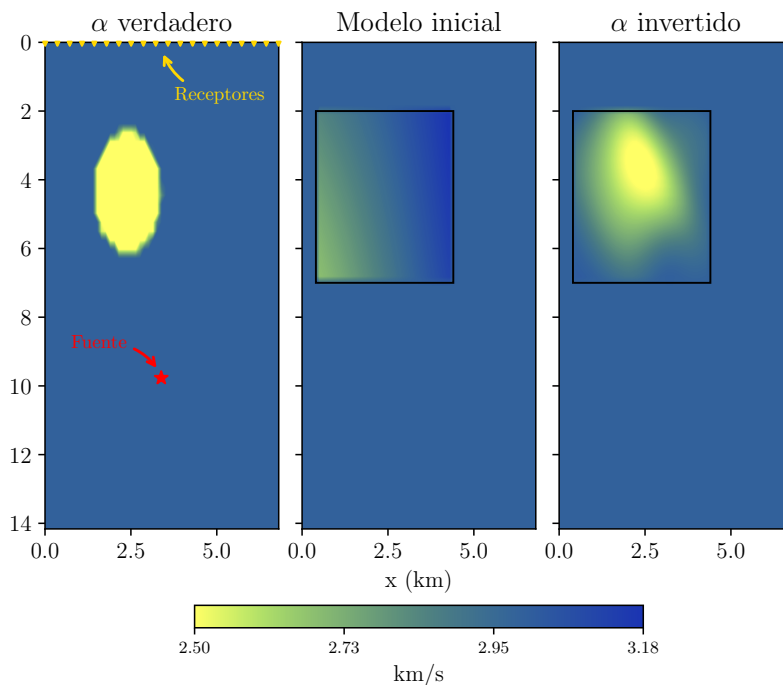
*Nota.* Tiempo de predicción  $t = 1.2s$ ,  $r_{L_1}$  y  $r_{L_2}$  corresponden al error relativo para la norma L1 y la norma L2, respectivamente.

del despliegue de receptores empleado en el modelo del experimento 2. Este tipo de configuración busca representar condiciones más cercanas a escenarios reales de adquisición sísmica, donde pueden existir restricciones socioambientales, cuerpos de agua o zonas de difícil acceso que limitan la ubicación de los receptores en superficie.

La intención de este análisis no es establecer una métrica formal de sensibilidad del modelo

**Figura 10**

*Comparación entre el modelo real de velocidad de onda P y el parámetro físico  $\alpha$  invertido por la red PINN (cuarta red feed-forward) para el escenario sísmológico.*



frente a la distribución de los datos observados, sino generar una primera aproximación cualitativa y comparar los resultados con los obtenidos en el caso más controlado.

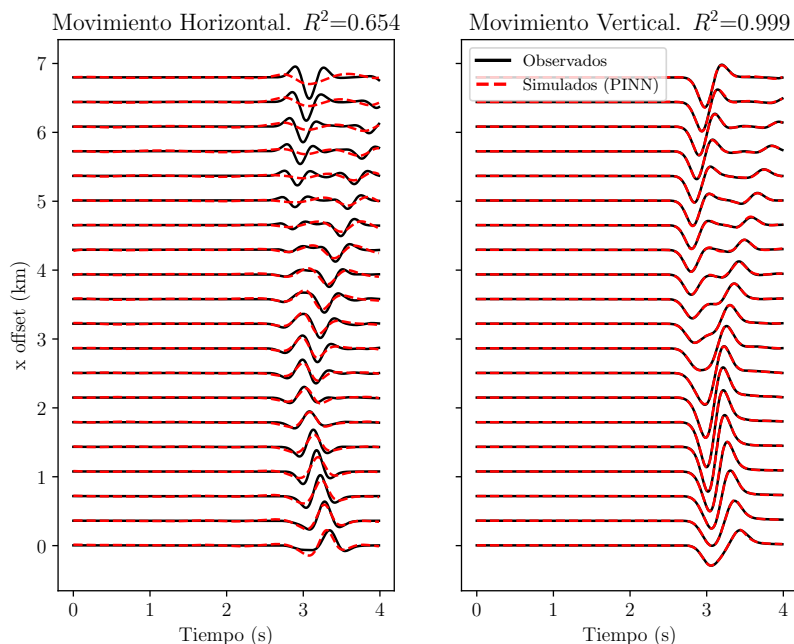
#### **4.3.1 Distribución de receptores**

Con el fin de analizar el efecto de la geometría de adquisición sobre el desempeño de la red PINN, se consideraron dos configuraciones distintas de distribución de receptores en superficie, orientadas a representar un escenario con menor densidad de datos y otro con una disposición espacial no equidistante.

- **Caso 1:** se reduce el número total de receptores de 20 a 15 (-25 %), manteniendo una distribución regular y equiespaciada a lo largo de la superficie del dominio, con un espaciamiento constante de 485.31 m entre receptores (figura 14a).
- **Caso 2:** se conserva el número total de receptores (20), pero se introduce una distribución no

**Figura 11**

*Resultados de la predicción de datos observados para el modelo sismológico.*



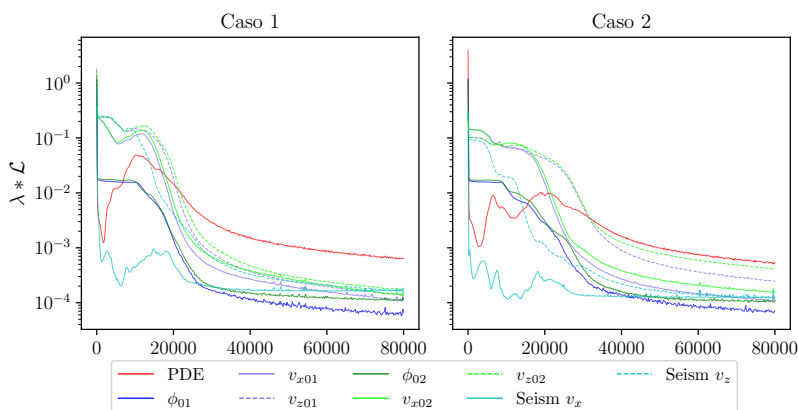
equidistante en superficie. Los receptores se organizan en tres clústeres con espaciamentos diferenciados: seis receptores separados 522.64 m, nueve receptores separados 174.21 m y cinco receptores separados 348.42 m, dispuestos de izquierda a derecha a lo largo de la superficie del dominio (figura 14b).

**4.3.2 Entrenamiento**

Cada uno de los escenarios con variación en la distribución de receptores fue entrenado durante un total de 200 000 épocas. El tiempo total de entrenamiento fue de 22 horas y 3 minutos para el caso 1, y de 22 horas y 36 minutos para el caso 2. La evolución del entrenamiento, considerando la convergencia de los términos de la función de costo ponderados mediante los pesos adaptativos ( $\lambda$ ) calculados en cada época, se presenta en la figura 12. En ambos escenarios se observa un comportamiento estable del proceso de optimización.

**Figura 12**

*Diagramas de evolución del entrenamiento de la red PINN en los entrenamientos de cambio de disposición de receptores*

**4.3.3 Modelo Directo**

Ambos escenarios con variación en la distribución de receptores mostraron un comportamiento adecuado en la aproximación de los campos de onda  $\phi$ ,  $v_x$  y  $v_z$ . En términos generales, los errores relativos se mantuvieron en rangos comparables para ambos casos, dentro del rango de 0.098 hasta 0.121 para la norma  $L1$ , y el rango de 0.068 hasta 0.109, rangos similares si se toman en cuenta los errores obtenidos en el experimento 2 (tabla 4). Se observa una ligera mejora en el caso 1 para el campo  $v_z$ , mientras que para los campos  $\phi$ ,  $v_x$  los menores errores se obtuvieron en el caso 2. Un ejemplo de las inferencias espaciales de los campos de onda en un tiempo posterior al de la segunda instantánea temporal ( $t = 1.2$  s, tomando como referencia el tiempo total de propagación) se presenta en la figura 13 y sus respectivos errores de norma  $L1$  y  $L2$  en la tabla 5.

**4.3.4 Modelo Inverso**

Para los dos escenarios evaluados se observa un comportamiento contrastante en el resultado de la red neuronal de  $\alpha$ . En el caso 2, la red logra recuperar de manera adecuada tanto la magnitud como la distribución espacial del modelo de velocidad, reproduciendo de forma consistente la geometría de la anomalía de baja velocidad (figura 14b). En contraste, en el caso 1, aunque la red aproxima valores de velocidad coherentes con el modelo teórico, no logra reconstruir correc-

**Tabla 5***Errores del modelo directo en el experimento 3*

Caso	Campo	Error L1	Error L2
1	$\phi$	0.118	0.109
	$vx$	0.098	0.073
	$vz$	0.105	0.074
2	$\phi$	0.109	0.104
	$vx$	0.107	0.068
	$vz$	0.121	0.074

tamente la distribución espacial del parámetro  $\alpha$ , presentando una representación más suavizada y con pérdida de la geometría de la anomalía (figura 14a).

#### 4.3.5 Datos observados

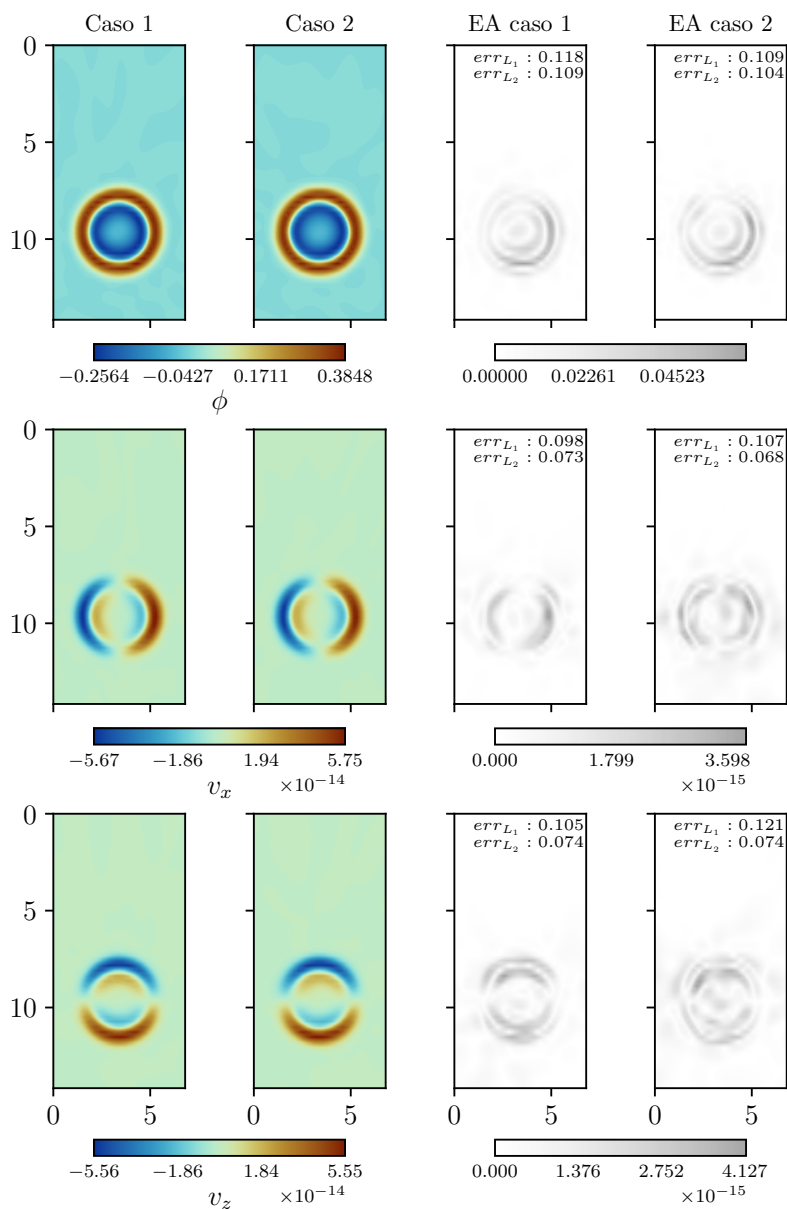
El análisis de los datos observados mediante la métrica  $R^2$  evidencia un comportamiento diferenciado entre las componentes de velocidad. Para el campo  $v_z$ , se obtiene un ajuste prácticamente perfecto en ambos escenarios, con valores de  $R^2 \approx 0.99$ . En contraste, para el campo  $v_x$  ninguno de los casos presenta un ajuste adecuado, reflejado en valores negativos de  $R^2$ , lo que indica una ausencia de correlación entre los datos predichos por la red y los datos sintéticos observados. La comparación de los sismogramas observados y predichos para ambos escenarios se muestra en la figura 15.

#### 4.4 Experimento 4: Cambio de geometría

Con el objetivo de evaluar la sensibilidad del enfoque PINN frente a la geometría de las redes neuronales utilizadas, se definió una serie de experimentos de entrenamiento en los que se modificó sistemáticamente la arquitectura de las redes *feed-forward*. En particular, se analizaron variaciones en el número de capas ocultas y en la cantidad de neuronas por capa, con el fin de identificar su influencia sobre la precisión del modelo y su capacidad de ajuste en el entorno sísmológico sintético utilizado en el experimento 2 (figura 10).

**Figura 13**

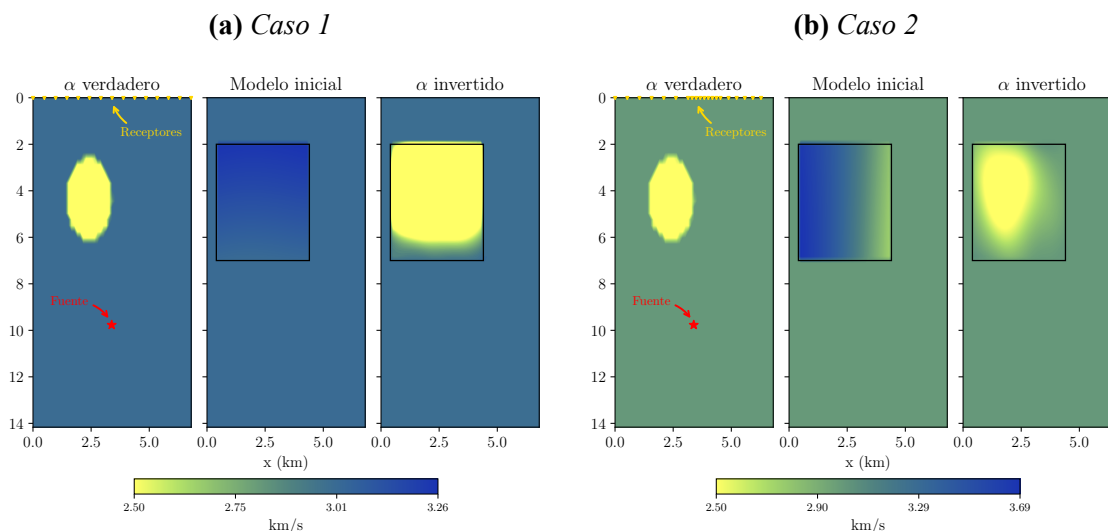
*Comparación de los resultados de propagación para el experimento 3 en los casos de cambio de receptores para los campos  $\phi$ ,  $v_x$ ,  $v_z$ .*



*Nota.* Tiempo de predicción  $t = 1.2s$ ,  $r_{L1}$  y  $r_{L2}$  corresponden al error relativo para la norma L1 y la norma L2, respectivamente.

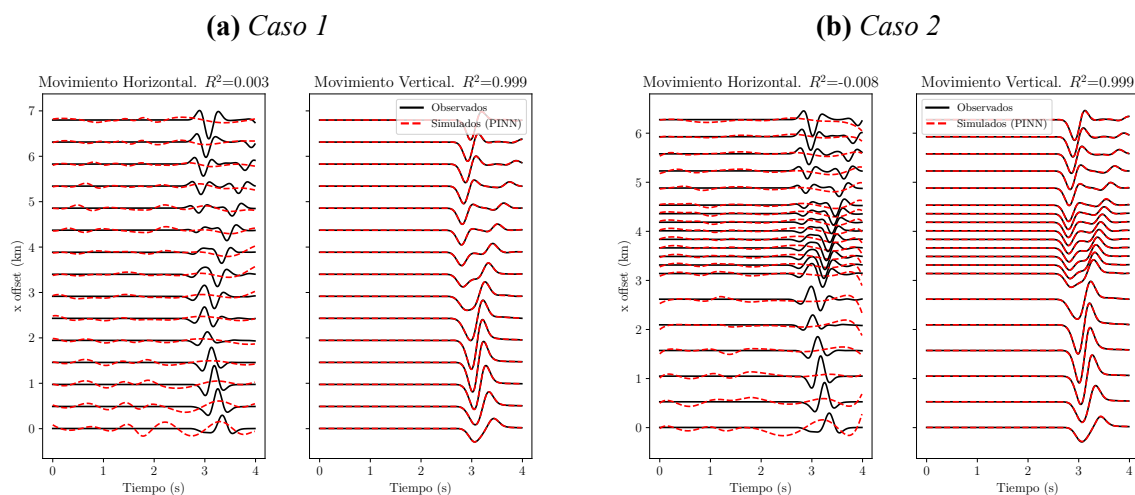
**Figura 14**

Comparación entre el modelo real de velocidad de onda P y el parámetro físico  $\alpha$  invertido por la red PINN (cuarta red feed-forward) para los escenarios con variación de receptores.



**Figura 15**

Comparación entre los resultados de predicción de datos observados para los escenarios con variación de receptores.



#### 4.4.1 Configuraciones evaluadas

Se evaluaron nueve configuraciones distintas de redes completamente conectadas para la aproximación de los campos de onda  $v_x$ ,  $v_z$  y  $\phi$  (redes 1, 2 y 3 en la figura 1). En todos los casos,

se mantuvo constante la arquitectura de la red asociada al parámetro físico  $\alpha$  (tabla 6), con el fin de aislar y analizar la influencia de la geometría de las redes encargadas de aproximar los campos de onda sobre el desempeño del entrenamiento. Las configuraciones consideradas abarcaron desde redes de capacidad limitada (2 capas ocultas con 15 neuronas por capa) hasta arquitecturas más profundas y densas (4 capas ocultas con 45 neuronas por capa). En todos los experimentos se utilizó la función de activación  $\tanh(\cdot)$ , en concordancia con los experimentos previos.

El desempeño de cada configuración se evaluó a partir de las siguientes métricas:

- Errores relativos ( $L_1, L_2$ ) en la predicción de los campos de onda  $\phi, v_x, v_z$  y del parámetro físico  $\alpha$ .
- Coeficiente de determinación ( $R^2$ ) para cuantificar el ajuste de los datos observados.
- Error cuadrático medio (MSE) obtenido de todas las funciones de costo en la última época del entrenamiento.

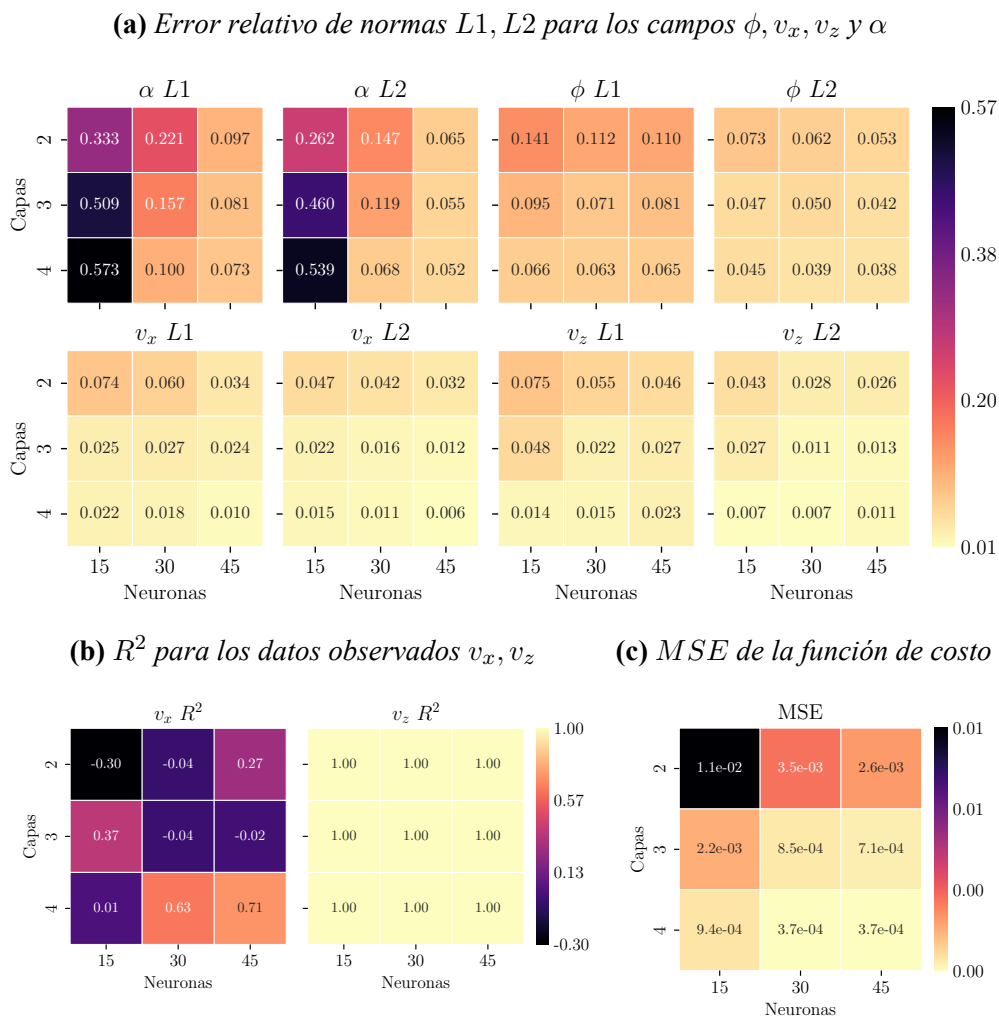
**Tabla 6**

*Arquitecturas utilizadas en el experimento 4*

Exp	Redes $v_x, v_z, \phi$		Red $\alpha$	
	Capas	Neuronas	Capas	Neuronas
1	2	15	5	20
2	2	30	5	20
3	2	45	5	20
4	3	15	5	20
5	3	30	5	20
6	3	45	5	20
7	4	15	5	20
8	4	30	5	20
9	4	45	5	20
			Épocas	200 000

**Figura 16**

*Métricas de los entrenamientos en el experimento de cambio de geometría.*



### 4.4.2 Entrenamiento

Las nueve configuraciones de red fueron entrenadas durante 200 000 épocas utilizando el optimizador Adam, con una tasa de aprendizaje de 0.0001. En todos los casos se empleó el mismo esquema de cálculo de pesos  $\lambda_i$  mediante el kernel NTK, consistente con los experimentos anteriores, aplicado de manera independiente en cada entrenamiento. La evolución temporal de la función de costo y de sus términos ponderados se presenta en el anexo C, mientras que en esta sección se reportan únicamente los valores finales de las métricas globales de error.

El error cuadrático medio (MSE) total, resumido en la figura 16c, se encuentra en el rango de  $1.1 \times 10^{-2}$  a  $3.7 \times 10^{-4}$ . Se observa una disminución sistemática del error al incrementar tanto la profundidad de la red como el número de neuronas por capa, siendo la profundidad el factor con mayor impacto en la reducción del error global. Este comportamiento sugiere una mayor capacidad de representación de funciones complejas en arquitecturas más profundas.

Adicionalmente, se evidencia que las arquitecturas más simples presentan dificultades para alcanzar un régimen de convergencia adecuado, mostrando un estancamiento temprano durante el entrenamiento. Este comportamiento puede asociarse a limitaciones inherentes de redes totalmente conectadas, tales como el sesgo espectral y la dificultad para capturar componentes de alta frecuencia de la solución (Rahaman et al., 2018). En este contexto, el uso de ponderación basada en NTK resulta menos efectivo en redes de baja capacidad, dado que persisten discrepancias en las tasas de convergencia de los distintos términos de la función de costo. En contraste, las arquitecturas más profundas y anchas permiten una mejor alineación entre los diferentes componentes del error, reduciendo las diferencias de escala y favoreciendo un proceso de optimización más estable.

#### 4.4.3 *Modelo directo*

Las nueve configuraciones evaluadas mostraron un comportamiento adecuado en la aproximación de los campos de onda  $\phi$ ,  $v_x$  y  $v_z$ . En términos generales, los errores relativos disminuyen a medida que aumenta la capacidad de la red, tanto en número de capas como en número de neuronas por capa, como se evidencia en la figura 16a.

1. Para el campo  $\phi$ , los errores relativos en norma  $L_1$  y  $L_2$  varían desde valores de 0.141 y 0.073, respectivamente, en la configuración de menor capacidad (2 capas y 15 neuronas por capa), hasta valores mínimos de 0.064 y 0.038 en las configuraciones más profundas y densas (4 capas con 30 y 45 neuronas por capa).
2. Para el campo  $v_x$ , se observa una reducción más marcada de los errores, con valores de  $L_1$  y  $L_2$  que disminuyen desde 0.074 y 0.047, respectivamente, en la configuración de menor capacidad (2 capas y 15 neuronas por capa) hasta 0.010 y 0.006, respectivamente, alcanzados

en la configuración de 4 capas con 45 neuronas por capa.

3. Para el campo  $v_z$ , los errores relativos presentan un comportamiento similar, reduciéndose desde 0.075 y 0.043 en la configuración más simple (2 capas y 15 neuronas por capa) hasta valores mínimos del orden de 0.014 y 0.007 en configuraciones de 4 capas con 30 y 45 neuronas por capa.

Las inferencias espaciales de los campos de onda para todas las configuraciones evaluadas se presentan en el anexo A.

#### **4.4.4 Modelo inverso**

El comportamiento de las nueve configuraciones en la aproximación del parámetro físico  $\alpha$  muestra una mayor sensibilidad a la geometría de la red. En particular, se observa una reducción significativa de los errores relativos al aumentar el número de capas ocultas, siempre que la red cuente con al menos 30 neuronas por capa. En contraste, las configuraciones con solo 15 neuronas presentan errores considerablemente mayores, incluso al incrementar la profundidad de la red.

Los errores relativos en norma  $L_1$  para  $\alpha$  varían desde valores del orden de 0.57 en la configuración de 4 capas con 15 neuronas por capa, hasta un mínimo de 0.07 en la configuración de 4 capas con 45 neuronas por capa. De manera consistente, los errores en norma  $L_2$  disminuyen desde 0.54 hasta 0.05 en estas mismas configuraciones (figura 16a).

Las distribuciones espaciales del parámetro de velocidad invertido con sus errores asociados se presentan en el anexo anexo B.

#### **4.4.5 Datos observados**

En el caso de los datos observados, la métrica  $R^2$  evidencia un comportamiento diferenciado entre las componentes de velocidad, reflejando una capacidad de ajuste no uniforme de la red frente a los distintos campos físicos involucrados. Para la componente vertical  $v_z$ , se obtiene un ajuste prácticamente perfecto en todas las configuraciones evaluadas, con valores de  $R^2 = 1.0$ , lo que indica una adecuada reconstrucción de esta variable a partir de los datos observados.

En contraste, la componente horizontal  $v_x$  presenta un desempeño considerablemente infe-

rior, con valores de  $R^2$  que varían desde valores negativos en las arquitecturas correspondientes a los escenarios 1, 2, 5 y 6 (tabla 6), hasta un valor máximo de  $R^2 = 0.71$  en la configuración de mayor capacidad (4 capas y 45 neuronas por capa), como se muestra en la figura 16b. Este comportamiento sugiere que la representación adecuada del campo  $v_x$  en la superficie libre requiere una mayor capacidad expresiva de la red, tanto en términos de profundidad como de número de neuronas por capa. En particular, los resultados indican que configuraciones con al menos 4 capas y 30 neuronas por capa constituyen un umbral mínimo para lograr un ajuste significativo. Esto permite explicar el bajo desempeño observado en arquitecturas más simples, como las empleadas en los experimentos 2 y 3.

La representación de los datos observados por parte de la red PINN en cada experimento se muestra en el anexo anexo D.

## 5 Discusión

### 5.1 Implementación de los kernels NTK

La implementación de los kernels NTK en la función de costo mostró un impacto positivo en la estabilidad y convergencia del entrenamiento de la red PINN en todos los experimentos realizados. Este resultado aborda directamente una problemática documentada en el estado del arte sobre PINNs: el mal condicionamiento del problema de optimización. Estudios como el de Krishnapriyan et al. (2021) han demostrado que las PINNs pueden presentar fallas sistemáticas de entrenamiento asociadas a la rigidez del problema diferencial y a un paisaje de pérdida extremadamente difícil de optimizar, más que a una falta de expresividad en la arquitectura de la red. En particular, estos autores evidencian que el proceso de optimización tiende a quedar atrapado en mínimos locales que satisfacen parcialmente la ecuación, pero fallan en capturar la dinámica física subyacente.

De forma complementaria, S. Wang, Teng y Perdikaris (2021) identificaron que esta falla se debe a una “patología de gradientes” causada por la rigidez numérica (*stiffness*) en la dinámica

del flujo de gradientes. Esta condición genera un desbalance en las magnitudes de los gradientes retropropagados provenientes de los distintos términos de la función de pérdida (datos observados, condiciones iniciales/contorno y el residuo de la PDE), lo que sesga el entrenamiento y puede impedir la convergencia.

En este contexto, en los experimentos realizados se observó que, sin la estrategia de ponderación adaptativa basada en NTK, los términos de la función de costo presentaban diferencias de hasta 32 órdenes de magnitud (figuras 4 y 8). Los términos asociados a la PDE y a las dos instantáneas del campo  $\phi$  mostraron los mayores residuos, coincidiendo correspondientemente con los valores más pequeños de  $\lambda$  asignados durante el ajuste. Estas diferencias son representativas del problema de entrenamiento multiescala descrito por S. Wang et al. (2022), que surge al intentar optimizar simultáneamente todos los términos residuales en una red PINN.

Con la incorporación de los pesos dinámicos basados en la traza del NTK, estas diferencias se redujeron significativamente, alcanzando discrepancias de aproximadamente un orden de magnitud entre los distintos términos de la función de costo. Esto permitió lograr un balance más adecuado entre las diferentes componentes del problema y favoreció un proceso de entrenamiento en el que el ajuste considera simultáneamente todos los términos de la función de pérdida. Desde la perspectiva de los problemas inversos sísmicos, este balance resulta particularmente relevante debido a la posible presencia de múltiples mínimos locales en el proceso de inversión. En este sentido, un entrenamiento más equilibrado puede contribuir a evitar el estancamiento en soluciones que, aunque satisfacen parcialmente las restricciones físicas (particularmente aquellas cuyos gradientes dominan el proceso de optimización), correspondan a configuraciones del modelo del subsuelo poco consistentes desde el punto de vista geofísico.

Los resultados de los experimentos muestran que la evolución de los pesos  $\lambda$  permite ajustar la contribución de cada término en función de su escala relativa, evitando que el entrenamiento sea dominado por un único término residual. En los experimentos realizados se observó que, con la ponderación adaptativa, el cálculo de los términos  $\lambda_i$  alcanzó valores estables en aproximadamente

100 000 épocas, por lo que en las iteraciones posteriores se optó por mantener estos pesos constantes y detener su actualización. Esta consideración es específica para las geometrías y configuraciones evaluadas en este trabajo, ya que la dinámica de entrenamiento puede variar dependiendo del problema físico, la arquitectura de la red y la distribución de los datos utilizados durante el entrenamiento.

## 5.2 Modelado acústico

En los experimentos 1 y 2, se consiguió que la red aprendiera el modelado directo (salida de las redes 1, 2 y 3, correspondientes a los campos  $\phi, v_x, v_z$ , figuras 5 y 9) obteniendo un ajuste tanto en los modelos de propagación como en los datos observados generados por la PINN. Los errores de las normas  $L_1, L_2$  fueron menores en el experimento de pozo cruzado, sin embargo, se debe tener en cuenta que fue un experimento con mayor cantidad de épocas de entrenamiento y una mayor densidad de datos de muestreo.

El escenario de pozo cruzado en su dominio espacial de entrenamiento (quitando las fronteras absorbentes) se submuestreó con una cuadrícula de  $50 \times 50$  que representa un espacio de  $1.23 \text{ km} \times 0.43 \text{ km}$ , lo que corresponde a una densidad de  $4726.79 \text{ } d_{obs}/\text{km}^2$  en cada matriz para todos los campos. Por otra parte, en el escenario sismológico, se tiene un dominio espacial de entrenamiento submuestreado con una cuadrícula de  $55 \times 55$  que representa un espacio de  $6.802 \text{ km} \times 14.162 \text{ km}$ , lo que corresponde a una densidad de  $31.4 \text{ } d_{obs}/\text{km}^2$  en cada matriz para todos los campos. Esto demuestra que la densidad de los datos sintéticos con los cuales se entrena la red PINN puede reducirse cuando el modelo computacional crece, al igual que reducir el número de épocas de entrenamiento, y aún así lograr un entrenamiento satisfactorio para el modelado de los campos  $(\phi, v_x, v_z)$ . El tiempo que tarda la red, una vez entrenada, en producir un *snapshot* de  $55 \times 55$  para uno de los campos es de  $0.1 \text{ s}$ , y el tiempo que toma en reconstruir los 3 cubos de propagación es de  $4 \text{ s}$ .

Para los datos observados también se obtuvo una diferencia entre el experimento de pozo cruzado y el sismológico. Para el primer caso, el resultado tuvo el mejor ajuste según el coeficiente

de determinación  $R^2$  (0.992 para el campo  $v_x$  y 0.929 para el campo  $v_z$ ). En el segundo caso, se observó una disminución del ajuste del campo  $v_x$  con respecto al campo  $v_z$  (figura 11), principalmente en los sismogramas más alejados de la anomalía, lo cual podría indicar una atenuación más rápida que la del campo  $v_z$ , así como el hecho de que en la frontera de rebote del frente de onda (superficie libre), las propagaciones empleadas en este trabajo presentan una diferencia de al menos un orden de magnitud en la amplitud del campo  $v_x$  en comparación con las del campo  $v_z$ .

### 5.3 Inversión de velocidad

En los experimentos 1 y 2, se consiguió que la red aproximara el modelado inverso (salida de la red 4, correspondiente al campo de velocidades acústicas  $\alpha$ ), obteniendo en ambos casos un resultado cercano en magnitud, ubicación y geometría, sin embargo, en el modelo sismológico mostró un resultado menos fiel.

La diferencia entre los resultados de ambos experimentos se puede explicar nuevamente en función de la diferencia en la cantidad de datos sintéticos utilizados para aproximar los *snapshots*, lo que podría producir resultados menos precisos luego del entrenamiento, además de la diferencia en las épocas de entrenamiento entre un caso y el otro.

Este resultado, en todos los casos, puede interpretarse como un modelo tomográfico, el cual representa una aproximación que no define los bordes de la anomalía, pero es adecuado y podría usarse como modelo inicial dentro de un proceso de inversión de mayor resolución (como FWI). El tiempo que tarda la red, una vez entrenada, en producir un resultado de  $55 \times 55$  del modelo de velocidades es de 0.01 s.

### 5.4 Entrenamiento

En la fase experimental, se determinó que la arquitectura propuesta requiere un régimen de entrenamiento extenso, aproximadamente de 200 000 épocas, para alcanzar una convergencia asintótica en las cuatro redes neuronales *feed-forward*. En este punto, el modelo logra una reconstrucción consistente de los campos de presión y una estimación de parámetros alineada con el *ground-truth* (ver figura 10). Se observó que prolongar el entrenamiento más allá de este umbral

refina la precisión de la inversión, especialmente en la configuración del caso sísmico de pozo cruzado (figura 6), sin embargo, el tiempo de cómputo aumenta en un factor de 3.7 (de 23.52 horas en un entrenamiento de 200 000 épocas a 88.5 horas en uno de 400 000 épocas, en promedio). Este requerimiento computacional sugiere que la superficie de error del problema multiobjetivo (PDE y datos) posee una topografía compleja, donde la optimización demanda un número elevado de iteraciones para equilibrar la magnitud de los gradientes provenientes de los distintos términos de la función de pérdida.

Desde la perspectiva de la teoría de problemas inversos, este comportamiento se relaciona con el carácter mal condicionado de la inversión sísmica. En este contexto, la función de costo no se limita al ajuste de datos (*misfit*), sino que incorpora la ecuación acústica como una forma de regularización física (Raissi et al., 2019). Esta restricción reduce el espacio de soluciones admisibles al favorecer aquellas que satisfacen, al menos aproximadamente, la ecuación diferencial que gobierna el fenómeno físico, contribuyendo así a disminuir la ambigüedad asociada al problema inverso. No obstante, la interacción entre los términos de datos y los términos físicos puede introducir una dinámica de optimización rígida, lo que podría explicar la necesidad de ciclos de entrenamiento prolongados para estabilizar la trayectoria del descenso de gradiente.

Finalmente, se destaca la capacidad de las PINNs para representar el modelo del subsuelo como una función continua del espacio, lo que permite aproximar los campos físicos sin depender de una discretización estricta sobre una malla numérica (Rasht-Behesht et al., 2022). Esta representación implícita puede facilitar la recuperación de estructuras a distintas escalas espaciales. Sin embargo, persiste el desafío del sesgo espectral, una tendencia inherente de las redes neuronales totalmente conectadas a priorizar el aprendizaje de componentes de baja frecuencia (estructuras globales) antes que detalles de alta frecuencia (heterogeneidades locales) (Rahaman et al., 2018). En este sentido, los resultados sugieren que, si bien el enfoque basado en NTK contribuye a estabilizar el proceso de entrenamiento, la resolución final del modelo invertido continúa condicionada por la capacidad de la red para representar gradientes de velocidad abruptos, lo que define en la

práctica el límite de resolución del método propuesto.

### 5.5 Efecto de la variación en la distribución de receptores

Los experimentos realizados con variaciones en la distribución de receptores permitieron analizar de manera preliminar la sensibilidad de la red PINN frente a geometrías de adquisición no idealizadas. Se observó que la modificación en la densidad y el espaciamiento de los receptores tuvo un impacto diferenciado sobre el problema directo, el problema inverso y el ajuste de los datos observados.

En el caso del modelo directo, los resultados muestran una alta robustez frente a los cambios en la distribución de receptores. Las predicciones de los campos de onda  $\phi$ ,  $v_x$  y  $v_z$  mantienen errores relativos bajos y comparables entre ambos escenarios evaluados, incluso cuando se reduce el número total de receptores o se introduce una distribución no equidistante. Este comportamiento sugiere que la formulación físicamente informada de la red, basada en la ecuación de onda, proporciona una regularización suficiente para compensar una menor o irregular cobertura de datos en superficie, al menos en el contexto sintético considerado.

En contraste, el problema inverso asociado a la reconstrucción del parámetro de velocidad  $\alpha$  muestra una mayor dependencia de la geometría de adquisición. Mientras que en el escenario con distribución no equidistante (caso 2) la red logra recuperar adecuadamente tanto la magnitud como la distribución espacial de la anomalía de baja velocidad, en el escenario con menor densidad de receptores (caso 1) la inversión presenta una anomalía más suavizada y una pérdida significativa de su geometría. Este mejor desempeño puede asociarse a la presencia de un clúster de mayor densidad de receptores en la región superior, lo que incrementa la información disponible en esta zona del dominio. Estos resultados indican que, si bien la información física impuesta por la PINN contribuye a estabilizar la inversión, la cobertura espacial y la distribución de los datos siguen siendo factores que determinan la capacidad de resolver adecuadamente la estructura del subsuelo.

## 5.6 Efecto de las variaciones en la distribución de receptores

Los experimentos realizados con variaciones en la distribución de receptores permitieron analizar de manera preliminar la sensibilidad de la red PINN frente a geometrías de adquisición no idealizadas. Se observó que la modificación en la densidad y el espaciamiento de los receptores tuvo un impacto diferenciado sobre el problema directo, el problema inverso y el ajuste de los datos observados.

En el caso del modelo directo, los resultados muestran una alta robustez frente a los cambios en la distribución de receptores. Las predicciones de los campos de onda  $\phi$ ,  $v_x$  y  $v_z$  mantienen errores relativos bajos y comparables entre ambos escenarios evaluados, incluso cuando se reduce el número total de receptores o se introduce una distribución no equidistante. Este comportamiento sugiere que la formulación físicamente informada de la red, basada en la ecuación de onda, proporciona una regularización suficiente para compensar una menor o irregular cobertura de datos en superficie, al menos en el contexto sintético considerado.

En contraste, el problema inverso asociado a la reconstrucción del parámetro de velocidad  $\alpha$  muestra una mayor dependencia de la geometría de adquisición. Mientras que en el escenario con distribución no equidistante (caso 2) la red logra recuperar adecuadamente tanto la magnitud como la distribución espacial de la anomalía de baja velocidad, en el escenario con menor densidad de receptores (caso 1) la inversión presenta una anomalía más suavizada y una pérdida significativa de su geometría. Este mejor desempeño puede asociarse a la presencia de un clúster de mayor densidad de receptores en la región superior cercana a la fuente sísmica, lo que incrementa la información disponible en esta zona del dominio. Estos resultados indican que, si bien la información física impuesta por la PINN contribuye a estabilizar la inversión, la cobertura espacial y la distribución de los datos siguen siendo factores que determinan la capacidad de resolver adecuadamente la estructura del subsuelo.

El análisis de los datos observados muestra un ajuste prácticamente perfecto para la componente vertical de la velocidad  $v_z$  en ambos escenarios, en contraste con el bajo desempeño en

la componente horizontal  $v_x$ , la cual presenta valores negativos del coeficiente  $R^2$ . Esta asimetría sugiere una sensibilidad diferenciada entre la magnitud de las componentes  $v_x$  y  $v_z$ , posiblemente debido al aplanamiento de la onda incidente, lo cual es consistente con el comportamiento observado en el experimento 4.

En conjunto, estos resultados evidencian que, aunque las redes PINN mantienen un desempeño robusto en el problema directo bajo configuraciones de adquisición menos controladas, la calidad de la inversión sísmica puede depender tanto de la cantidad de receptores como de su distribución espacial. En este sentido, resulta necesario desarrollar experimentos adicionales que permitan aislar estos factores y evaluar su influencia de manera independiente.

En conjunto, estos resultados evidencian que, aunque las redes PINN mantienen un desempeño robusto en el problema directo bajo configuraciones de adquisición menos controladas, la calidad de la inversión sísmica puede depender tanto de la cantidad de receptores como de su distribución espacial. En este sentido, resulta necesario desarrollar experimentos adicionales que permitan aislar estos factores y evaluar su influencia de manera independiente.

### **5.7 Geometría de las redes feed-forward**

El diseño de la arquitectura de una red neuronal físicamente informada (PINN), incluyendo el número de capas y neuronas por capa en sus redes independientes, es un factor que afecta su capacidad para aprender soluciones a ecuaciones diferenciales parciales (EDP) e identificar sus parámetros asociados, como se describe en Raissi et al., 2019. En dicho estudio se observa que, de manera general, un incremento en el número de capas y neuronas por capa tiende a mejorar la capacidad de representación de la red, lo cual puede conducir a soluciones más precisas tanto para el problema directo como para el inverso abordado por una red PINN.

Para este trabajo, las nueve configuraciones evaluadas muestran una clara influencia de la geometría de la red sobre la convergencia. El descenso pronunciado del MSE y de los errores relativos L1 y L2 en redes más profundas sugiere que un mayor número de capas mejora la capacidad de representación. Sin embargo, la falta de un patrón claro al aumentar el número de neuronas indica

que la capacidad de la red no escala linealmente con el número de parámetros.

Las redes con 30 neuronas por capa presentan el menor error  $L1$  y  $L2$  para el campo  $v_z$ , así como el menor error  $L1$  para el campo  $\phi$ , lo que sugiere la existencia de un equilibrio entre complejidad y capacidad de representación. En este sentido, aumentar el número de neuronas no necesariamente conduce a mejoras en el ajuste, lo que indica la posible existencia de un umbral a partir del cual el incremento de parámetros no aporta beneficios significativos. La reducción del error no se asocia a un proceso de sobreajuste (*overfitting*), dado que el entrenamiento de la red PINN no se basa en la minimización de un conjunto de datos etiquetados independiente, sino en la imposición simultánea de restricciones físicas (residuo de la EDP) y datos observados dentro de una misma función de costo. En este contexto, los distintos términos de la función de pérdida actúan como mecanismos de regularización física. Adicionalmente, la consistencia entre la disminución del MSE global y la mejora en la reconstrucción de los campos de onda y del modelo de velocidad sugiere que el modelo captura adecuadamente la solución física del problema, en lugar de ajustar ruido.

El modelado directo refleja que el aumento del número de capas mejora la calidad de las predicciones de los campos de onda  $\phi, v_x, v_z$ . Sin embargo, la componente  $\phi$  muestra persistencia de errores hasta que se alcanza una profundidad de 4 capas, lo que muestra que este campo requiere una mayor capacidad de representación de la red.

En la predicción de datos observados, el ajuste para  $v_z$  es consistentemente alto ( $R^2 > 0.98$ ), mientras que para  $v_x$  el ajuste es pobre, especialmente en redes con 2 capas ocultas.

Esta asimetría puede explicarse, en parte, por diferencias en la magnitud de los datos observados entre ambos campos en las posiciones de los receptores, similar a lo observado en el experimento 2, donde se presentan variaciones de hasta un orden de magnitud. Esta disparidad favorece un ajuste más rápido en el campo de mayor magnitud, mientras que el otro converge más lentamente.

Adicionalmente, este comportamiento puede estar asociado a la geometría de propagación

del campo de onda. Dado que la fuente se encuentra en profundidad, el frente de onda incide en los receptores con una componente predominantemente horizontal, lo que implica que la derivada espacial  $\partial\phi/\partial x$  es cercana a cero en dichas posiciones. Esto reduce la sensibilidad del modelo respecto al campo  $v_x$ , dificultando su ajuste en comparación con  $v_z$ .

## 6 Conclusiones

En este trabajo se desarrolló e implementó una Red Neuronal Físicamente Informada (PINN) para la inversión sísmica acústica de modelos de velocidad de onda P, evaluada en escenarios sintéticos. Asimismo, se analizó la influencia de la distribución de datos observados y de la geometría de las redes internas en el desempeño del modelo. Los resultados muestran que la metodología PINN permite reconstruir dentro de un error aceptable los campos de onda y el modelo de velocidad en entornos sintéticos controlados, lo que evidencia una adecuada incorporación de las leyes físicas que gobiernan el problema de propagación de ondas.

La inclusión de pesos dinámicos en la función de costo permitió estabilizar el proceso de entrenamiento y equilibrar automáticamente los distintos términos físicos y de datos, reduciendo diferencias iniciales de varios órdenes de magnitud sin requerir ajustes manuales o una asignación heurística de pesos. Esta estrategia facilitó la convergencia del modelo y permitió abordar de forma conjunta los problemas directo e inverso dentro del mismo marco de optimización.

Se observó que la arquitectura de la red influye significativamente en su capacidad de generalización, siendo la profundidad un factor más determinante que el número de neuronas por capa. Asimismo, los experimentos evidenciaron que la distribución espacial de los receptores afecta la calidad de la inversión: mientras que el problema directo mostró una alta robustez frente a variaciones en la geometría de adquisición, la reconstrucción del modelo de velocidad resultó más sensible a la densidad y al espaciamiento de los receptores, afectando principalmente la recuperación de la distribución espacial de las anomalías del subsuelo. En este sentido, resulta necesario aislar de

manera sistemática los efectos de la densidad y el espaciamiento de los receptores, con el fin de cuantificar el aporte individual de cada factor en la reconstrucción del modelo de velocidad.

No obstante, el alcance de los resultados obtenidos se encuentra condicionado por algunas limitaciones inherentes al enfoque implementado. En primer lugar, los experimentos se realizaron en escenarios sintéticos controlados, por lo que el comportamiento del método frente a datos sísmicos reales, particularmente en presencia de ruido y errores de medición, requiere ser evaluado en trabajos futuros. De igual forma, aunque la incorporación de la ecuación acústica dentro de la función de costo introduce una forma de regularización física que contribuye a estabilizar el proceso de inversión, la propagación real de las ondas sísmicas está gobernada por un comportamiento elástico del medio. En este sentido, el uso de una formulación acústica constituye una simplificación del fenómeno físico, lo que puede limitar la capacidad del modelo para representar adecuadamente efectos asociados a la propagación elástica, particularmente en medios complejos o altamente heterogéneos.

Adicionalmente, la capacidad de resolución espacial del método depende tanto de la arquitectura de la red como de la cobertura de los datos disponibles. En particular, la recuperación de heterogeneidades de pequeña escala puede verse limitada por factores como el sesgo espectral de las redes neuronales totalmente conectadas y por la distribución espacial de los receptores utilizados durante el entrenamiento. Finalmente, el proceso de entrenamiento requiere un número elevado de iteraciones para alcanzar una convergencia estable, lo que plantea desafíos de escalabilidad computacional cuando se consideran dominios de mayor tamaño o configuraciones en tres dimensiones.

En conjunto, los resultados obtenidos confirman el potencial de las PINNs como una herramienta prometedora para la inversión sísmica en entornos sintéticos, y proporcionan una base metodológica para continuar explorando la integración de aprendizaje profundo con restricciones físicas en problemas inversos geofísicos.

### Referencias

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. <https://doi.org/10.48550/ARXIV.1605.08695>
- AlAli, A., & Anifowose, F. (2022). Seismic velocity modeling in the digital transformation era: A review of the role of machine learning. *Journal of Petroleum Exploration and Production Technology*, 12(1), 21-34. <https://doi.org/10.1007/s13202-021-01304-0>
- Aster, R. C., Borchers, B., & Thurber, C. H. (2019). *Parameter estimation and inverse problems* (3.<sup>a</sup> ed.). Elsevier. <https://doi.org/10.1016/C2015-0-02458-3>
- Gardner, G. H. F., Gardner, L. W., & Gregory, A. R. (1974). Formation velocity and density; the diagnostic basics for stratigraphic traps. *Geophysics*, 39(6), 770-780. <https://doi.org/10.1190/1.1440465>
- Jacot, A., Gabriel, F., & Hongler, C. (2018). Neural Tangent Kernel: Convergence and Generalization in Neural Networks [Publisher: arXiv Version Number: 4]. <https://doi.org/10.48550/ARXIV.1806.07572>
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422-440. <https://doi.org/10.1038/s42254-021-00314-5>
- Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/ARXIV.1412.6980>
- Krishnapriyan, A. S., Gholami, A., Zhe, S., Kirby, R. M., & Mahoney, M. W. (2021). Characterizing possible failure modes in physics-informed neural networks. <https://doi.org/10.48550/ARXIV.2109.01050>

- Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications*, 12(1), 1432. <https://doi.org/10.1038/s41467-021-21670-x>
- Lu, B., Moya, C., & Lin, G. (2023). NSGA-PINN: A multi-objective optimization method for physics-informed neural network training. *Algorithms*, 16(4), 194. <https://doi.org/10.3390/a16040194>
- Mondol, N. H. (2010). Seismic Exploration. En *Petroleum Geoscience: From Sedimentary Environments to Rock Physics* (pp. 375-402, Vol. 1). Springer Berlin Heidelberg.
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Fundamentals of artificial neural networks and deep learning. En *Multivariate statistical machine learning methods for genomic prediction* (pp. 379-425). Springer International Publishing. [https://doi.org/10.1007/978-3-030-89010-0\\_10](https://doi.org/10.1007/978-3-030-89010-0_10)
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in PyTorch. <https://api.semanticscholar.org/CorpusID:40027675>
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F. A., Bengio, Y., & Courville, A. (2018). On the Spectral Bias of Neural Networks. <https://doi.org/10.48550/ARXIV.1806.08734>
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. <https://doi.org/10.48550/ARXIV.1711.10566>
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>

- Rasht-Behesht, M., Huber, C., Shukla, K., & Karniadakis, G. E. (2022). Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions. *Journal of Geophysical Research: Solid Earth*, 127(5), e2021JB023120. <https://doi.org/10.1029/2021JB023120>
- Ren, P., Rao, C., Chen, S., Wang, J.-X., Sun, H., & Liu, Y. (2024). SeismicNet: Physics-informed neural networks for seismic wave modeling in semi-infinite domain. *Computer Physics Communications*, 295, 109010. <https://doi.org/10.1016/j.cpc.2023.109010>
- Tromp, J. (2020). Seismic wavefield imaging of earth's interior across scales. *Nature Reviews Earth & Environment*, 1(1), 40-53. <https://doi.org/10.1038/s43017-019-0003-8>
- Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1-WCC26. <https://doi.org/10.1190/1.3238367>
- Virieux, J., Operto, S., Ben-Hadj-Ali, H., Brossier, R., Etienne, V., Sourbier, F., Giraud, L., & Haidar, A. (2009). Seismic wave modeling for seismic imaging. *The Leading Edge*, 28(5), 538-544. <https://doi.org/10.1190/1.3124928>
- Waheed, U. B., Haghghat, E., Alkhalifah, T., Song, C., & Hao, Q. (2021). PINNeik: Eikonal solution using physics-informed neural networks. *Computers & Geosciences*, 155, 104833. <https://doi.org/10.1016/j.cageo.2021.104833>
- Wang, S., Teng, Y., & Perdikaris, P. (2021). Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5), A3055-A3081. <https://doi.org/10.1137/20M1318043>
- Wang, S., Wang, H., & Perdikaris, P. (2021). On the eigenvector bias of fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 384, 113938. <https://doi.org/10.1016/j.cma.2021.113938>

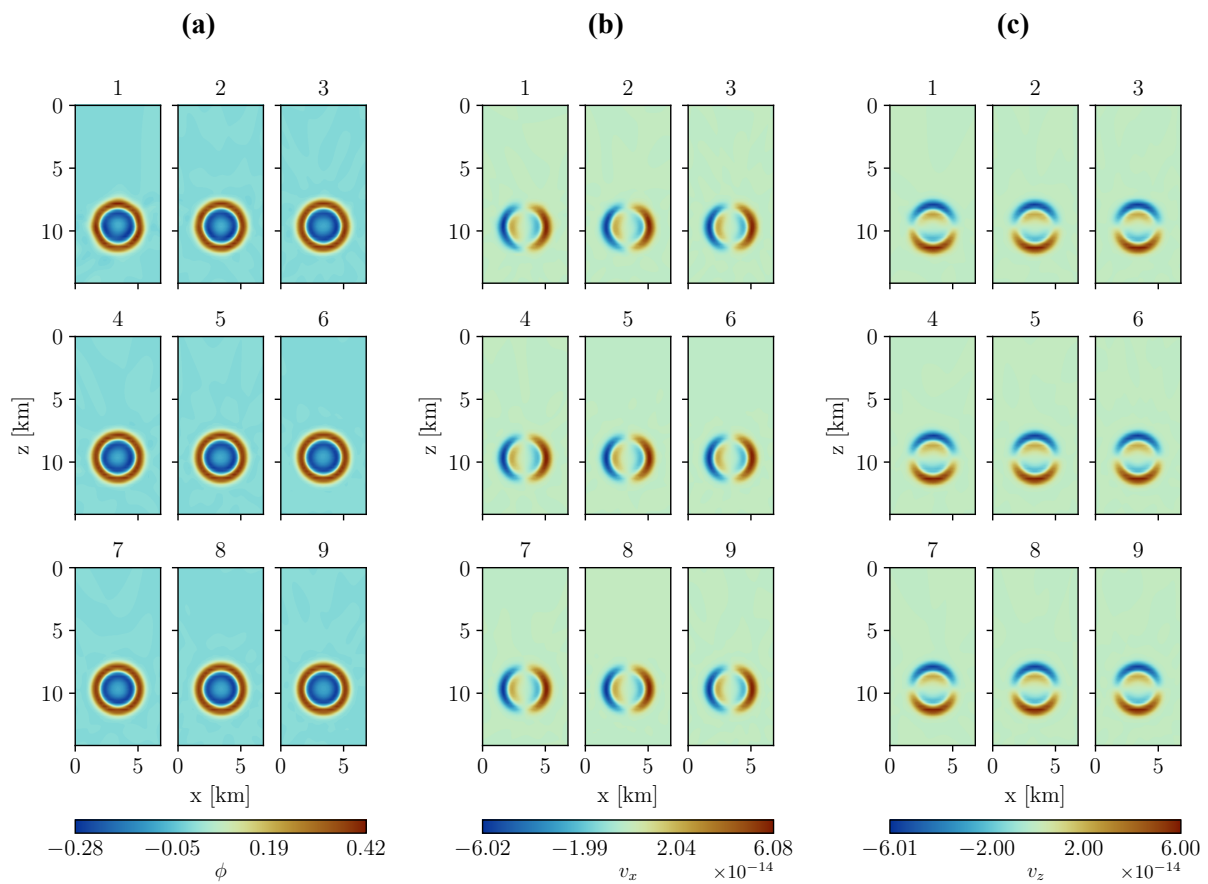
- Wang, S., Yu, X., & Perdikaris, P. (2022). When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449, 110768. <https://doi.org/10.1016/j.jcp.2021.110768>
- Wang, X. (2017). Study and application of velocity model building method for the areas with complicated structures. En *Relative fidelity processing of seismic data* (1.<sup>a</sup> ed., pp. 117-182). Wiley. <https://doi.org/10.1002/9781119052913.ch4>
- Wang, Y., Yang, C., & Cao, J. (2012). ON TIKHONOV REGULARIZATION AND COMPRESSIVE SENSING FOR SEISMIC SIGNAL PROCESSING. *Mathematical Models and Methods in Applied Sciences*, 22(2), 1150008. <https://doi.org/10.1142/S0218202511500084>
- Wang, Y. (Ed.). (2016). *Seismic inversion: Theory and applications* (1.<sup>a</sup> ed.). Wiley. <https://doi.org/10.1002/9781119258032>
- Xinquan Huang & T. Alkhalifah. (2023). Microseismic source imaging using physics-informed neural networks with hard constraints.
- Xu, Y., Li, J., & Chen, X. (2019). Physics informed neural networks for velocity inversion. *SEG Technical Program Expanded Abstracts 2019*, 2584-2588. <https://doi.org/10.1190/segam2019-3216823.1>
- Yang, F., & Ma, J. (2019). Deep-learning inversion: A next-generation seismic velocity model building method. *GEOPHYSICS*, 84(4), R583-R599. <https://doi.org/10.1190/geo2018-0249.1>
- Zou, J., Liu, C., Zhao, P., & Song, C. (2024). Seismic Wavefields Modeling With Variable Horizontally Layered Velocity Models via Velocity-Encoded PINN. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-11. <https://doi.org/10.1109/TGRS.2024.3411472>

## Apéndices

## Anexo A: Experimento 4 - Inferencias de campos de onda

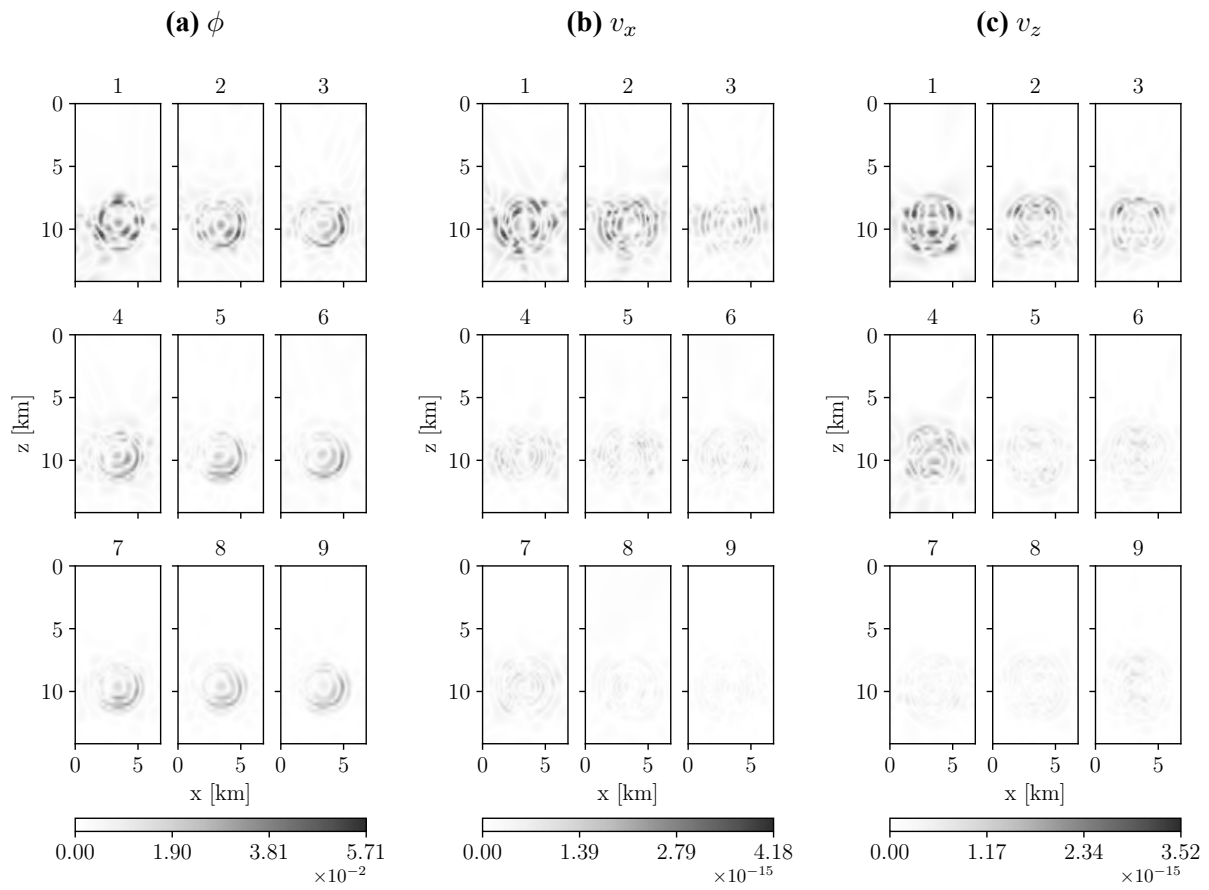
Figura A-1

Resultados del modelamiento directo en los experimentos de cambio de geometría ( $t = 1.15$  s)



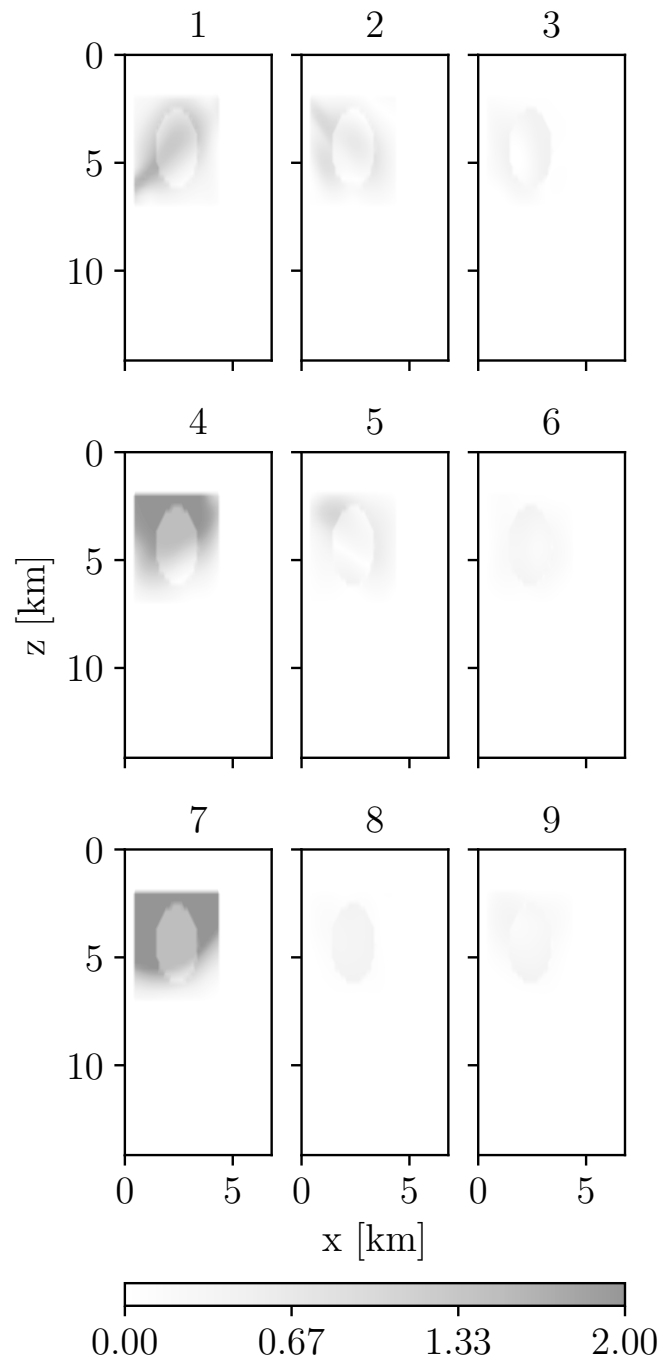
**Figura A-2**

*Errores absolutos de los campos generados en los experimentos de cambio de geometría ( $t = 1.15$  s)*



**Anexo B: Experimento 4 - Inversiones de  $\alpha$** **Figura B-1**

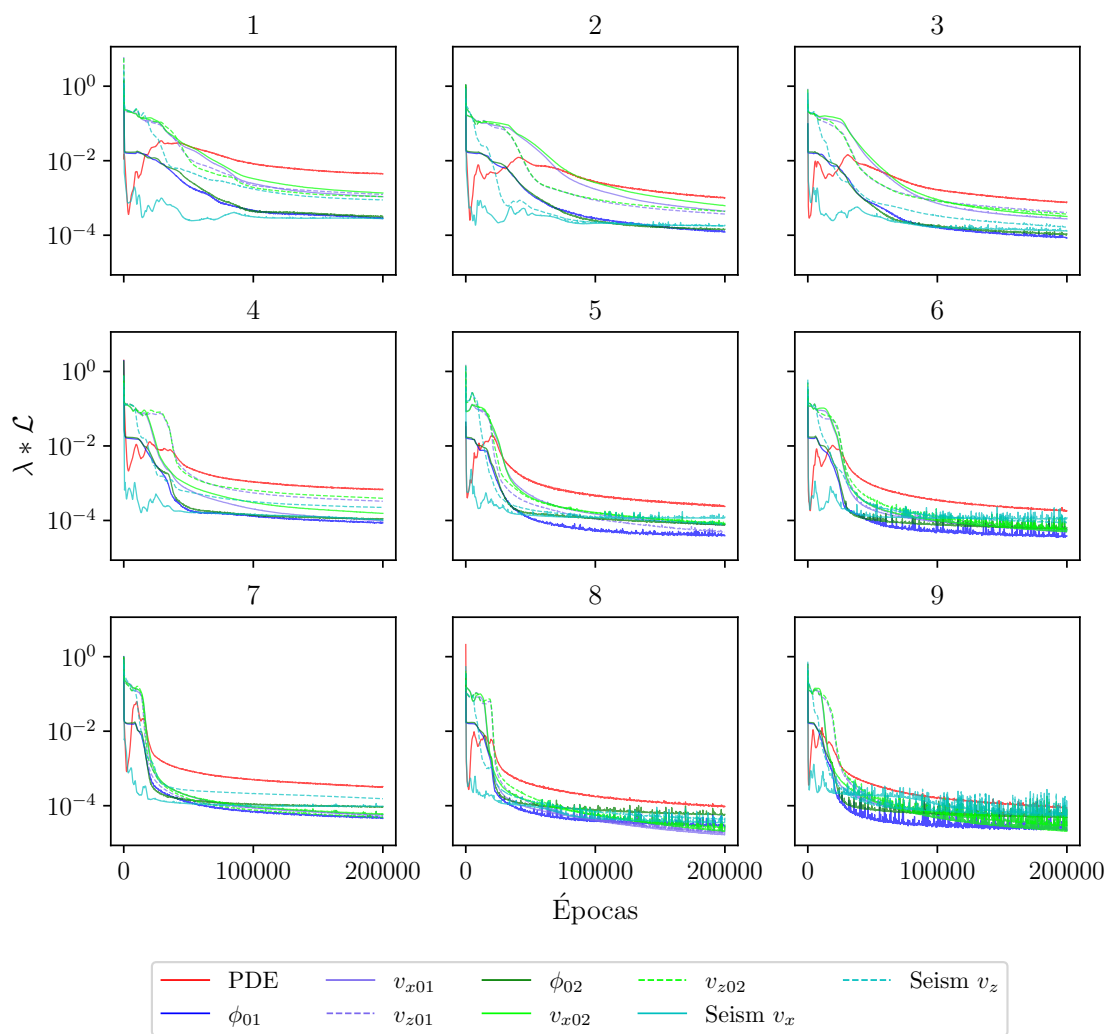
*Misfit de la inversión de velocidad de onda P en los experimentos de cambio de geometría*



**Anexo C: Experimento 4 - Evolución de la función de costo**

**Figura C-1**

*Evolución de los entrenamientos en los experimentos de cambio de geometría*



**Anexo D: Experimento 4 - Inferencias de datos observados****Figura D-1***Resultados de la predicción de datos observados en los experimentos de cambio de geometría*