UNIVERSIDAD INDUSTRIAL DE SANTANDER

MASTER'S THESIS

Reservoir porosity estimation from post-stack seismic data using a deep learning model constrained by well-porosity: A near-field application

Author:
Geo. Nicolas Cordoba Castillo

Advisor:
PhD. (c) Yesid Paul Goyes Peñafiel

Co-Advisor:
PhD. Leidy Castro Vera

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Geophysics

Universidad Industrial de Santander
Facultad de Ciencias
Escuela de Física
Maestría en Geofísica
Bucaramanga
2025

Dedicatoria

Dedico esta tesis con profundo cariño a quienes
han sido mi pilar a lo largo de este camino.
A mi mamá y mi papá, por su amor
incondicional y su apoyo constante.
A mi hermano Santiago, por creer en mí
incluso en los momentos más inciertos.
A Juliana, mi novia, por su paciencia, motivación y por
ayudarme a encontrar orden cuando más lo necesitaba.
Y a mis amigos, por estar presentes con su
compañía y aliento en cada etapa de este recorrido.

A todos ustedes, gracias por ser parte fundamental de este logro.

Estimación de la porosidad de reservorios a partir de sísmica post-apilado usando un modelo de aprendizaje profundo regularizado con datos de porosidad de pozo. Una aplicación en Near-Field Exploration

Resumen

La estimación de la distribución de porosidad resulta esencial para la exploración de hidrocarburos, captura y almacenamiento de carbono, y aprovechamiento de energía geotérmica. Los métodos tradicionales, fundamentados en inversión sísmica y ecuaciones de física de rocas, demandan extenso procesamiento, requieren ajustes paramétricos múltiples y presentan alta sensibilidad ante variaciones mineralógicas y de fluidos en la roca. Este estudio propone una metodología basada en redes neuronales (NN) que estima directamente la porosidad a partir de datos sísmicos post-apilado, superando las limitaciones mencionadas. La propuesta incorpora una función de pérdida que integra datos de porosidad obtenidos de registros de pozo para optimizar el entrenamiento. El método desarrollado superó a otras arquitecturas de aprendizaje automático en la estimación de porosidad a partir de datos sísmicos, tanto en el dominio del tiempo como de profundidad, utilizando datos sintéticos y reales. Los experimentos con datos sintéticos evidenciaron la superioridad predictiva del método propuesto frente a modelos alternativos. Adicionalmente, las pruebas con datos reales alcanzaron un $R^2 = 0,860$ en los datos de prueba, resultado corroborado mediante técnicas de validación cruzada. Estos hallazgos confirman la robusta capacidad de generalización de la red, su resistencia ante la variabilidad de datos y su estabilidad en condiciones de ruido moderado, generando resultados geológicamente interpretables. La red completó su entrenamiento en 373,7 segundos para 120.263 trazas y realizó predicciones sobre 52.910 trazas en 4 segundos, reduciendo notablemente los tiempos de procesamiento respecto a métodos convencionales y de otras redes neuronales. Los resultados subrayan el potencial de las NN para lograr estimaciones de porosidad precisas y eficientes, particularmente en el ámbito de la *Near-Field Exploration*, preservando la coherencia geológica.

Palabras clave: aprendizaje automático, redes neuronales, porosidad, sísmica post-apilado

Objetivos

Objetivo general

Estimar la porosidad en yacimientos de hidrocarburos con sísmica post-apilado mediante el uso algoritmos de inteligencia artificial.

Objetivos específicos

- Evaluar algoritmos de inteligencia artificial para establecer una relación entre la porosidad y la sísmica post-apilado usando datos de prueba.
- Implementar un modelo de inteligencia artificial incluyendo el desbalance de datos de entrenamiento para la estimación de la porosidad a partir de sísmica post-apilado con ruido.
- Evaluar la precisión de la estimación de la porosidad a partir de sísmica post-apilado con técnicas estadísticas que permitan determinar la confiabilidad del método y su replicabilidad.

Pregunta de investigación

¿Cuál es la arquitectura de inteligencia artificial y sus respectivos parámetros que permiten una estimación de baja incertidumbre de la porosidad a partir de la amplitud de sísmica post-apilado?

Reservoir porosity estimation from post-stack seismic data using a deep learning model constrained by well-porosity: A near-field application

Nicolas Cordoba-Castillo*, Paul Goyes-Peñafiel†, Leidy Castro-Vera‡

ABSTRACT

Estimating porosity distribution is essential for hydrocarbon exploration, carbon capture and storage, and geothermal energy. Traditional workflows, relying on seismic inversion and rock-physics equations, are time-intensive, parameter-dependent, and sensitive to variations in mineralogy and fluids. This study proposes a workflow that uses neural networks (NN) to directly estimate porosity from post-stack seismic, addressing the above-mentioned traditional limitations. The workflow implements a custom loss function during training that incorporates well-log-derived porosity. Our proposed workflow outperformed alternative machine learning approaches for porosity estimation from seismic data in the time and depth domain with synthetic and field data. Experiments on synthetic data demonstrated its superior predictive capability compared to other models. Moreover, evaluations on field data achieved an $R^2 = 0.860$ on the test set, an outcome reinforced by five-fold cross-validation. This performance shows the network's robust generalization, resilience to data variability, and stability under moderate noise, yielding interpretable results. The network trains efficiently in 373.7 seconds for $120,263$ traces and predicts $52,910$ traces in 4 seconds, significantly reducing the time required compared to conventional workflows. These results highlight NN's potential for accurate and efficient porosity estimation, particularly in near-field exploration, while preserving geological integrity.

INTRODUCTION

Reservoir characterization is crucial for any energy industry, either for defining prospects in hydrocarbon exploration, appraisal, and development of conventional and unconventional reservoirs or for assessing geothermal and carbon capture and storage projects. Reservoir characterization consists of estimating the target reservoir's properties to understand it better (Doyen, 2007). A reservoir characterization entails a detailed understanding of the three-dimensional arrangement and continuity of the reservoir and aquifer system, encompassing the rocks, pores, fluids, and any barriers to fluid flow (Sneider, 1990). Defining these reservoir properties, such as porosity, at the early stages of exploration enables more informed decision-making related to drilling locations, production strategies, and economic evaluations, ultimately reducing the risk and uncertainty associated with field development (Bloch and Helmold, 1995; Sneider, 1990). For instance, near-field exploration (NFE) is a field development technique wherein a new seismic survey (exploration block) is acquired adjacent to mature fields (mature block). This facilitates the extrapolation of previous analyses by capitalizing on the lateral continuity and preservation of rock properties across contiguous geological formations. Thus, accurately estimating porosity in three dimensions and understanding how it is spatially distributed is essential for obtaining a comprehensive and reliable reservoir characterization.

One of the main approaches to model the reservoir porosity is to perform a rock-physics inversion in a two-step or sequential scheme using seismic data (Doyen, 2007). In this approach, the P- and S- impedance and density are obtained and then used to estimate porosity using a local regression, a rock-physics template, or geostatistics. (Bornard et al., 2005; Bachrach, 2006; Bosch et al., 2010; Fattahi and Karimpouli, 2016). Pre-stack seismic inversion yields multiple elastic parameters that improve the discrimination of porosity, fluid saturation, and lithology (Mavko et al., 2009). Nevertheless, pre-stack seismic inversion may

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Geophysics

* **Nicolas Cordoba-Castillo (author)**. Department of Geophysics, Universidad Industrial de Santander, Colombia.

† **Paul Goyes-Peñafiel (advisor)**. Department of Computer Science, Universidad Industrial de Santander, Colombia.

‡ **Leidy Castro-Vera (co-advisor)**. Division of Earth Sciences and Geography, Aachen University, Germany.

be infeasible when the required data are unavailable. In these scenarios, reliance on post-stack data becomes the only option (Doyen, 1988; Rasmussen and Maver, 1996; Saltzer et al., 2005; Kumar et al., 2016; Yasin et al., 2020). The main disadvantage of this method is that post-stack seismic inversion provides only a single P-wave impedance estimate (Russell and Hampson, 1991), thus limiting its effectiveness in isolating these rock and fluid properties. Both approaches present technical challenges, particularly in wavelet estimation, which is highly sensitive to well-ties (Bosch et al., 2010), essential for obtaining reliable inversion results. Beyond the technical challenges, seismic inversion involves significant time investment from geophysicists, who perform the inversion steps and conduct extensive quality control (QC) processes to ensure data reliability.

Recent advancements in machine learning (ML), particularly neural network (NN), have been used for porosity estimation using post-stack seismic data (Leiphart and Hart, 2001; Iturrarán-Viveros, 2012; Feng et al., 2020; El-Dabaa et al., 2022; Vashisth and Mukerji, 2022; Jo et al., 2022). This approach leverages the universal approximation theorem (Cybenko, 1989; Hornik et al., 1989) that has been proven its effectiveness for feature extraction of seismic data (Goyes-Peñafiel et al., 2023). Some state-of-the-art neural network architectures rely on seismic attributes derived from post-stack seismic data (Leiphart and Hart, 2001; Iturrarán-Viveros, 2012; El-Dabaa et al., 2022). Furthermore, other authors have used rock-physics equations in their models (Feng et al., 2020; Vashisth and Mukerji, 2022; Jo et al., 2022) to estimate porosity directly from post-stack seismic. Feng et al. (2020) and Vashisth and Mukerji (2022) implemented a rock-physics guided method in an autoencoder architecture where they first estimate porosity from post-stack seismic using convolutional layers in the encoder part. Subsequently, the decoder leverages rock-physics equations and the convolutional seismic model, described in Russell (1988), to estimate seismic from porosity. Then, this seismic is used to calculate the misfit with the input seismic to train their models. This approach could be restrictive by selecting an appropriate rock-physics equation that matches the geological model or being oversimplistic by using a linear equation between the acoustic impedance and the porosity. Moreover, using the convolutional seismic model limits the input data domain to be exclusively on time, which is not always available. On the other hand, Jo et al. (2022) uses well-log porosity to generate synthetic porosity realizations. With rock-physics and the convolutional seismic model, they estimate multi-frequency seismic, which they use to train their model. This limits the usage of the approach by requiring a multi-frequency seismic as an input to estimate porosity, which is not commonly used. In an NFE context, leveraging previously developed models to train neural networks is possible in supervised training. Due to the lateral continuity of geological formations, the mature and exploration blocks exhibit similar rock properties, facilitating the transfer of learned features from one region to another. Consequently, supervised learning algorithms can generalize from well-characterized examples to predict properties in adjacent areas, ultimately allowing porosity estimation and overall reservoir characterization.

The main contribution of this work is to introduce a new approach for directly estimating 3D porosity from time and depth domain post-stack seismic data guided with well-derived porosity during training. The proposed method uses supervised learning of a neural network architecture design to extract the relevant features of the seismic to estimate porosity. The training is regularized with the well-derived porosity in the loss function to reduce the number of equivalent models. The input is processed as individual traces to train the neural network and then reconstructed to obtain the 3D model. The proposed method proved robustness under moderate noise conditions and maintained its generalization in performed tests. The proposed method was applied in an NFE context using field data of siliceous clastic sedimentary rocks.

METHODOLOGY

Problem Formulation

Estimating porosity from post-stack seismic data using NN is analogous to solving the inverse problem commonly employed in the hydrocarbon industry (Kim and Nakata, 2018). Both approaches share similarities, such as being ill-posed (i.e., solutions may not exist, be unique, or depend continuously on the input data) and being nonunique problems (i.e., multiple solutions can satisfy the given observations) (Kim and Nakata, 2018; Aster et al., 2018). We formulate the estimation porosity problem as a regression problem, which mathematically is

$$y = \mathcal{M}_\theta(x), \tag{1}$$

where $\mathcal{M}_\theta$ is a NN architecture with parameters θ that maps post-stack seismic data x to porosity y. During training, the parameters (weights and biases) are updated to minimize a loss function, commonly Mean Squared Error, described by the Equation 2. This minimization is achieved using the Adam optimizer, which extends Stochastic Gradient Descent with adaptive learning rates (Kingma, 2014).

$$Loss = \|\mathcal{M}_\theta(x_{obs}) - y_{obs}\|_2^2. \tag{2}$$

We propose a loss function shown in Equation 3 that extends from Equation 2 by incorporating a regularization term derived from well-log data. This integration of domain-specific information is expected to constrain the model parameters and enhance predictive performance, thereby providing a more robust framework for our analysis.

$$Loss = \frac{1}{N}\|\mathcal{M}_\theta(x_{obs}) - y_{obs}\|_2^2 + \frac{\lambda}{2}\|\mathcal{M}_\theta(x_w) - y_w\|_2^2 \ . \tag{3}$$

The proposed loss function defined in Equation 3 combines two terms based on the ℓ_2-norm. The first term measures the mean squared error between the predicted porosity $\mathcal{M}_\theta(x_{obs})$, derived from the seismic data x_{obs}, and the observed porosity y_{obs} from the pre-existing model. The second term evaluates the mean squared error between $\mathcal{M}_\theta(x_w)$, the predicted porosity from the seismic trace at the well position (x_w), and the porosity derived from well-logs (y_w). The regularization λ (set at 5) controls the relative weight of the well-data term in the loss function, balancing the influence of seismic and well-log data during training. This parameter was empirically determined through iterative experimentation and performance evaluation to identify the most effective configuration. By incorporating well-log data, this loss function ensures that the network's predictions are consistent with seismic and well-logs.

Neural Network Architecture

We proposed a NN architecture designed to approximate the functional $\mathcal{M}_\theta$, mapping seismic data (x_{obs}) to porosity (y_{obs}). In the NFE context, the seismic (x_{obs}) and porosity (y_{obs}) are the labeled data from the mature block where the NN is trained. Figure 1 describes the proposed NN architecture that integrates 1D convolutional and fully-connected layers employing an increasing filter depth while maintaining constant filter size, stride, and same padding across all convolutional layers, promoting multiscale feature extraction without diminishing spatial dimensionality. The architecture has a dropout layer of 0.1 before the fully-connected layer to mitigate overfitting, ensuring the model generalizes to unseen data. This design is particularly advantageous when seismic amplitude is the sole input, as the convolutional filters serve as artificial seismic attributes. Each filter is optimized within the network architecture to extract and map the relationship between seismic data and porosity. Moreover, the fully connected layer is strategically leveraged to augment the network's predictive capabilities. The proposed architecture was built for general purpose, allowing input seismic traces of different sample lengths, denoted by N in Figure 1.

Model Evaluation

The model's performance was evaluated during training using the determination coefficient (R^2), root mean squared error (RMSE), and mean absolute error (MAE) as defined in Equations 4, 5, and 6. These metrics are commonly employed in regression-based porosity estimation studies (Vashisth and Mukerji, 2022; Gholami et al., 2022) along variations thereof and have proven effective for model evaluation. After each epoch, these metrics were computed on the training and validation datasets to monitor the network's predictive accuracy and consistency.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}, \tag{5}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{6}$$

where y_i denotes the true (observed) porosity, $\hat{y}_i$ is the predicted porosity, and $\bar{y}$ represents the mean of the true porosity values. R^2 measures the proportion of variance in the observed data explained by the model, providing insight into overall accuracy. At the same time, RMSE quantifies the average error magnitude, and MAE highlights the average absolute deviation, offering complementary perspectives on prediction errors (Hastie, 2009).

EXPERIMENTAL SETUP

Performance Metrics

The signal-to-noise ratio (SNR) and structural similarity index measure (SSIM) were used to evaluate the network's performance under noisy conditions. SNR, defined in Equation 7, quantifies the fidelity of the reconstructed porosity predictions by comparing the signal to the noise level. SSIM, commonly used in image processing (Wang et al., 2004), defined in Equation 8, assesses the preservation of structural relationships and consistency in the reconstructed outputs. It evaluates how well the model retains spatial and structural information to determine the alignment of predicted porosity maps with geological features, even with noise. Despite not being commonly used in reservoir characterization tasks with machine learning, it can add meaningful insights and interpretation to the results obtained.

$$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\sum_{i=1}^{N} y_i^2}{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \right). \tag{7}$$

$$\text{SSIM} = \frac{(2\mu_{\hat{y}}\mu_y + C_1)(2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_{\hat{y}}^2 + C_2)}, \tag{8}$$

where, μ_y and $\mu_{\hat{y}}$ are the mean values of y and $\hat{y}$, σ_y^2 and $\sigma_{\hat{y}}^2$ are the variances of y and $\hat{y}$, σ_{xy} is the covariance between y and $\hat{y}$, and C_1 and C_2 are small constants added to stabilize the division.

Although initially designed for image mapping tasks (Isola et al., 2017), SNR and SSIM are applicable in this context as they provide a way to evaluate the network's output from an image-related perspective.

Data

Synthetic data

The synthetic dataset is a model of weakly cemented siliceous rocks, where 80% of the rock is composed of quartz and 20% of feldspar (Vashisth and Mukerji, 2022). Figure 2 shows a sample of three pairs of seismic traces with their corresponding porosity values. The seismic and

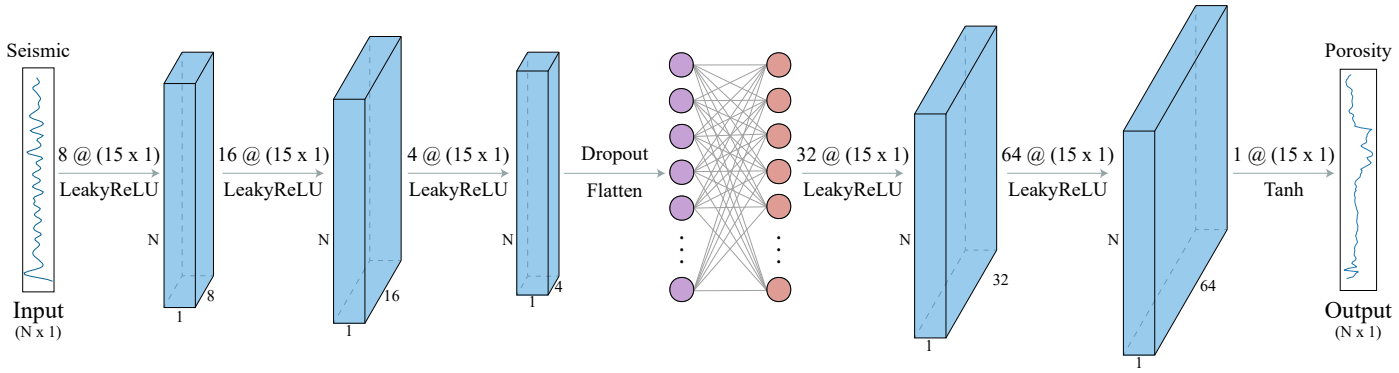


Figure 1: Scheme of the proposed neural network architecture. Key parameters include an increasing filter depth $(8, 16, 4, 32, 64)$, a filter size of 15, stride of 1, and padding set to 'same'. The input and output data are one-dimensional with N samples.

porosity range between $[-0.27, 0.26]$ and $[0.11, 0.4]$, respectively. The dataset contains 2500 pairs of seismic and porosity traces, representing a thickness of 200 m with 246 time samples. The dataset does not represent any geological structure; only high and low porosity zones were calculated using the rock-physics model. This synthetic dataset was used for controlled experiments to test various machine learning algorithms and gain insights into model performance under idealized conditions.

Field data: Illinois Basin - Decatur Project IBDP

The IBDP field data is from the Illinois Basin in the United States, acquired for a carbon capture project. The dataset includes a depth-migrated seismic survey with processing workflow (Bauer et al., 2019), 3D acoustic impedance, and a porosity model. The acoustic impedance model was used as an input for creating the porosity model using a probabilistic multi-attribute NN that selected the best six attributes for the prediction (Greenberg, 2019). Additionally, the dataset comprises four wells that target the reservoir zone. The dataset targets the Mt. Simon Formation, a reservoir composed of siliceous clastic sedimentary rocks, mainly sandstone (Greenberg, 2021). The main reservoir is the Mt. Simon Formation, which exhibits significant heterogeneity in its petrophysical properties, with the upper unit showing a typical decrease in porosity with depth, the middle unit generally exhibiting the poorest and most variable reservoir quality, and the lower unit displaying average high porosity of 22% (Leetaru and Freiburg, 2014). Regarding acquisition dimensions, the block has a sampling interval of 20 ft, with 265 inlines and 2947 crosslines at intervals of 40 and 5 ft, respectively. For this work, the dataset was cropped from inlines 53-195, crosslines 930-2140, and depth from 4700 to 6400 ft with 86 depth samples. Although the data originates from a 3D seismic volume and a corresponding 3D porosity model, the network processes each vertical trace independently, effectively treating the 3D volume as a collection of 1D inputs, with each trace mapped to its porosity label. This preserves local features along individual traces while allowing the network to handle the data in a trace-by-trace setup.

Normalization

Some ML algorithms take advantage of the normalization of the data such as NN (Feng et al., 2020). In that sense, to have comparable metrics for all the experiments, we normalized all datasets using min-max normalization in a range of $[-1, 1]$ for input and output. This approach was particularly important for handling datasets with differing units or magnitudes, such as seismic and porosity data.

Implementation Details

The network was trained for 50 epochs with an Adam Optimizer and a scheduled learning rate adjustment. The learning rate was initialized at 0.001 and reduced at 10% of its value if the R^2 of the validation set did not improve in 1 epoch. The batch used was 8 and 512 for the synthetic and field datasets, respectively. It was set up after different experiments to be approximately 0.5% of the training samples in each dataset. This was important to optimize the convergence and stability of the NN.

All experiments used Python 3.10.12, with TensorFlow 2.17.0 and scikit-learn 1.5.2 libraries for model implementation and evaluation. The training and testing processes were performed on a GPU NVIDIA GeForce GTX 1050 with 4 GB of VRAM.

EXPERIMENTS AND RESULTS

Experiment 1: synthetic data

We evaluated the proposed method with synthetic data in the first experiment by comparing its performance with several alternative ML algorithms. In this experiment $\mathcal{M}_\theta$ has $N = 246$ and $1,005,965$ trainable parameters. We compared the proposed method with models of Linear Regression (LR), Quadratic Regression (QR), Decision

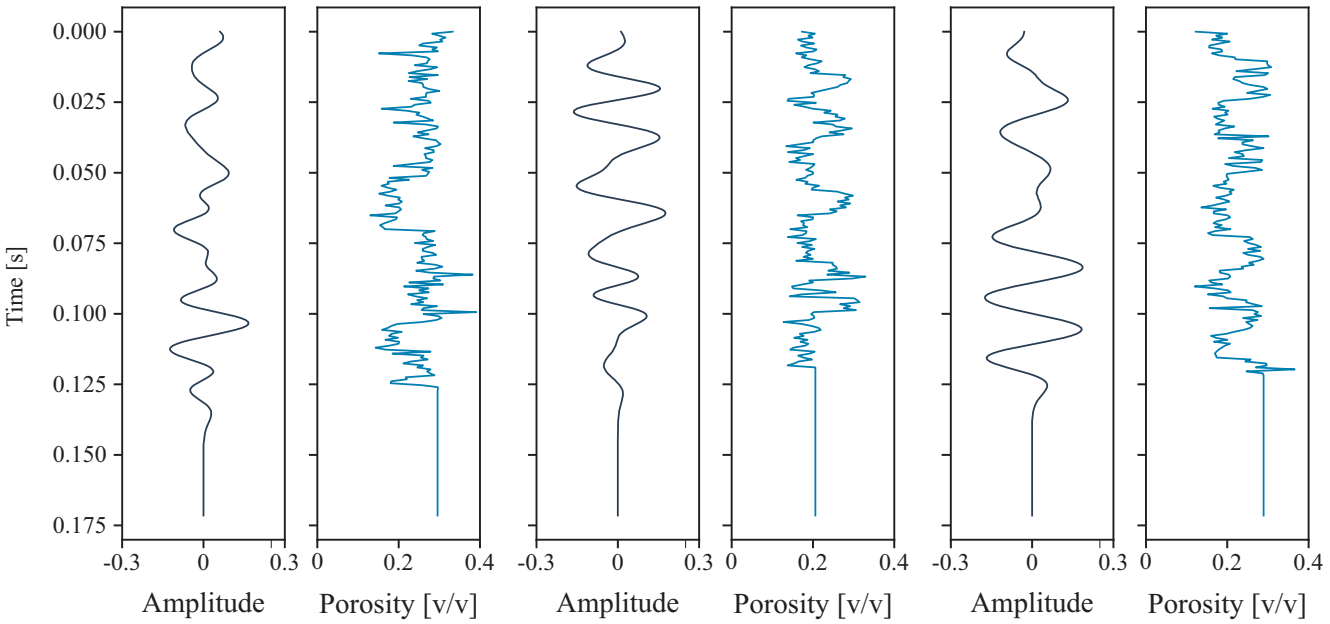


Figure 2: Example of three seismic traces with their corresponding porosity extracted from the synthetic dataset.

Trees (DT), XGBoost, MLP, and CNN. We used default hyperparameter settings for LR, QR, and DT, while the optimal hyperparameters for XGBoost were determined through a grid search as shown in Table 1. For the MLP, we designed a straightforward architecture of four layers containing 100 neurons and employing hyperbolic tangent activation functions. For the CNN implementation, we utilized the network described by Vashisth and Mukerji (2022), which has 2,128,662 trainable parameters.

Table 1: Hyperparameters used for grid search in XGBoost algorithm.

Hyperparameter	Values
Max depth	[3, 6, 9]
Learning rate	[0.01, 0.05, 0.1]
No. of estimators	[100, 300, 500, 1000]
Subsample	[0.7, 0.8, 0.9]
Col sample by tree	[0.7, 0.8, 0.9]
Booster Algorithm	[gbtree, dart]

We partitioned the original dataset for the experiments, allocating 70% for training and 30% for testing while setting aside 20% of the training portion to serve as a validation set for hyperparameter tuning. To ensure consistent and comparable error quantification across all experiments, we evaluated each model exclusively on the test data using the same metrics: R^2, RMSE, and MAE. In Figure 3 we plot the results of the predictions compared with the ground truth porosity to validate them and observe if they were consistent differentiating between low and high porosity values. All the results are summarized in Table 2 showing metrics in test data. The proposed methodology achieved optimal performance, surpassing the other ML implementations in all metrics while demonstrating improved computational efficiency by converging using fewer trainable parameters than the CNN approach.

Table 2: Performance metrics for porosity prediction models evaluated on the test set using synthetic post-stack seismic data.

Method	R^2	RMSE	MAE
LR	0.269	0.300	0.247
QR	0.015	0.349	0.295
DT	-0.752	0.465	0.372
XGBoost	0.022	0.348	0.294
MLP	0.697	0.193	0.143
CNN	0.627	0.215	0.160
Proposed	**0.729**	**0.183**	**0.134**

The experimental results in synthetic data presented in Figure 3 and Table 2 demonstrate that increasing algorithmic non-linearity in traditional machine learning approaches (LR, QR, DT, XGBoost) did not yield performance improvements. Notably, the Decision Tree implementation produced a negative R^2, indicating predictions less accurate than a baseline mean porosity estimator. In contrast, NN-based architectures outperformed conventional machine learning algorithms, achieving $R^2 > 0.6$. The visualization in Figure 3 illustrates the neural networks' capability to identify high and low porosity regions accurately. However, the CNN exhibited overfitting characteristics, while the MLP demonstrated underfitting behavior, failing to capture local data patterns.

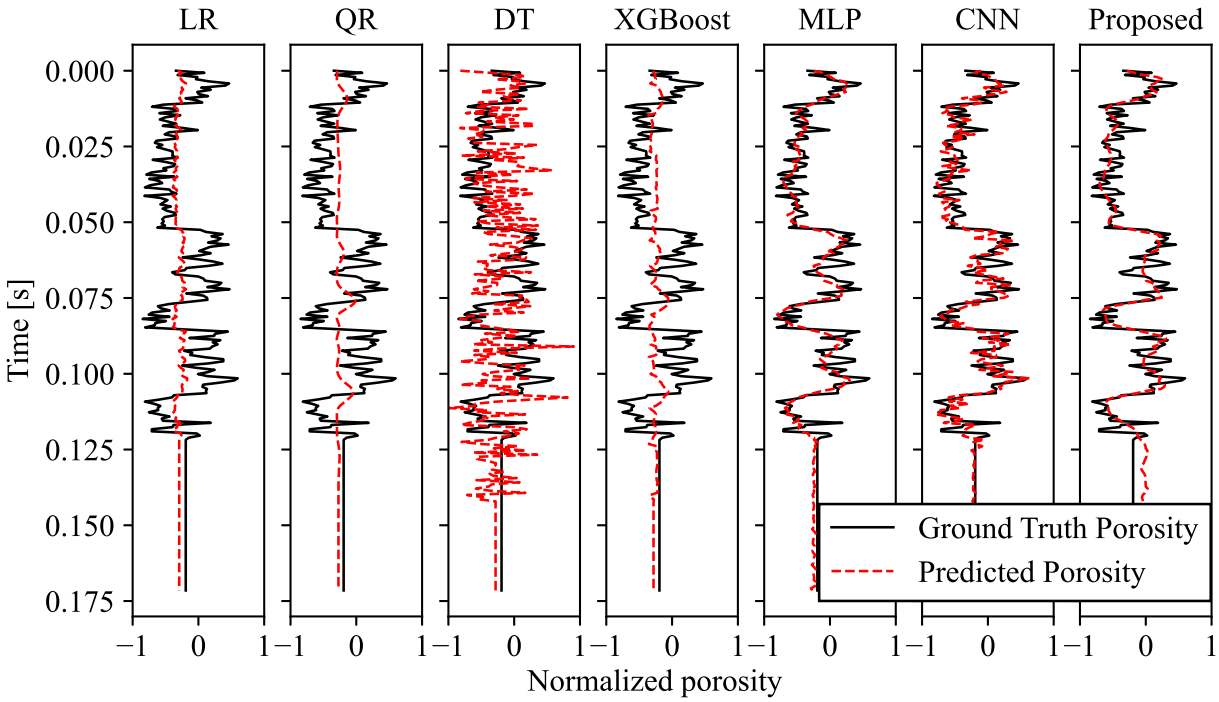


Figure 3: Predicted porosity (red) compared with the ground truth porosity (black) in synthetic test dataset using different ML algorithms (LR, QR, DT, XGBoost, MLP, CNN, and Proposed) at trace index 256.

Experiment 2: field data

In this experiment, we evaluated the methodology's predictive capabilities in a field data scenario, comparing its performance against other neural network-based algorithms. In this case, $\mathcal{M}_\theta$ has $N = 86$ and $155,405$ trainable parameters. The investigation aimed to assess algorithmic robustness in real geological formations while simulating an NFE scenario as shown in Figure 4.

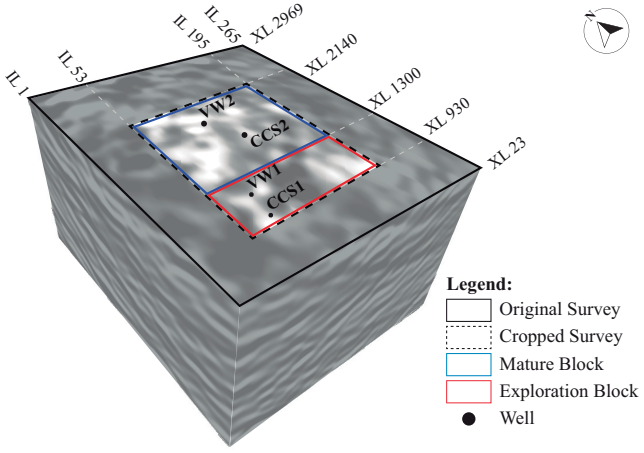


Figure 4: Schematic representation of the simulated NFE context using the IBDP field dataset. The dataset was cropped, considering the well-position. Each block has two wells to compare the results of the experiments with the well-log porosity.

The experimental design involved partitioning the dataset into two distinct regions: a mature block containing seismic and porosity data and an exploration block utilizing solely seismic data during training. For this partitioning, we selected the first 370 crosslines to represent the exploration block, with the remaining crosslines designated as the mature block. Although porosity ground-truth data existed for the exploration block, it was intentionally withheld during the training phase to simulate NFE conditions accurately. From the mature block, we allocated 70% of the data for training and 30% for testing while reserving 20% of the training portion for validation purposes. This cropping procedure ensured that each block contained two wells, enabling direct comparison between well-log-derived data and model predictions.

Table 3 presents a comparative model performance analysis across training and test datasets. The proposed model demonstrated superior generalization capabilities, achieving $R^2 = 0.860$ on unseen data, significantly outperforming both baseline architectures. While all models exhibited strong performance during training, with $R^2 > 0.9$, their test set performance revealed substantial differences in generalization ability. The MLP model, despite promising results during training with $R^2 = 0.970$, exhibited severe overfitting, as evidenced by its $R^2 = -0.392$ on the test set. The CNN architecture showed improved generalization compared to the MLP but still fell short of the proposed model's performance across all metrics. Notably, the proposed architecture achieved superior accu-

racy and demonstrated enhanced computational efficiency. With less trainable parameters, it maintains approximately one-tenth the complexity of the CNN model comprising $1,252,982$ parameters for this experiment, resulting in faster convergence during training. Specifically, the proposed method requires 5.8 minutes for training compared to 9.58 minutes for the CNN, underscoring its efficiency.

Table 3: Performance comparison of porosity prediction models on post-stack seismic field data, showing training and test set metrics.

Method	Train			Test		
	R^2	RMSE	MAE	R^2	RMSE	MAE
MLP	0.970	0.063	0.036	-0.392	0.418	0.357
CNN	0.972	0.060	**0.023**	0.799	0.159	0.118
Proposed	**0.985**	**0.044**	0.030	**0.860**	**0.132**	**0.095**

In addition to achieving good metrics, the proposed method produces porosity predictions that accurately represent geological structures, even with one-dimensional input data, as shown in Figure 5. The network effectively captures changes in porosity along seismic reflectors, identifies thin layers with distinct porosity variations, and delineates dipping structures with high fidelity. The most considerable discrepancies between predicted and ground-truth porosity occur at the upper part of the reservoir, where the NN does not map the seismic reflectors accurately and the porosity is being overestimated. Also, at the base of the reservoir, the porosity is underestimated compared with the ground truth model. Still, the predicted porosity with the proposed NN architecture remains geologically consistent, having a geological meaning, capturing the overall trend of the porosity distribution at the reservoir.

Integrating well-log-derived porosity into the loss function, as shown in Equation 3, enhances the stability and reliability of the proposed NN. In addition to improved accuracy, the proposed architecture is computationally more efficient than other methods tested in this experiment, making it a robust and practical solution for porosity estimation.

In the convolutional layers of the proposed architecture, the model learns to identify and map features from the input data. This process can be visualized in Figure 6 where the first convolutional layers have a strong seismic fingerprint in the processed signal, but in the final layers, it is noticeable how the filters are being activated by low and high porosity zones or changes at the interface. Each filter captures distinct aspects of the data, revealing the model's ability to differentiate between geological features. For example, certain filters are more activated at the top and base of the reservoir, indicating that the model is sensitive to changes in lithology. Other filters respond to areas of low porosity, highlighting the model's capacity to detect regions with specific porosity characteristics.

Additionally, the filters demonstrate varying spatial sensitivities. Some filters map broader, low-frequency features, which correspond to wider bandwidth patterns, while others are activated by thinner, high-frequency features, capturing finer geological details. This diversity in filter responses illustrates how the convolutional layers collectively extract multiscale information, enabling the network to generate geologically meaningful and structurally accurate predictions.

Experiment 3: model validation

We conducted a comprehensive validation process to evaluate the proposed model's performance and robustness. This included k-fold cross-validation to assess the model's generalization capability and a noise sensitivity analysis to evaluate its stability under varying input noise levels. These methods ensure a thorough understanding of the model's reliability across different scenarios, providing insight into its predictive accuracy and resilience to field data.

To evaluate the generalization capability of the proposed model, we performed 5-fold cross-validation. Before splitting the data into folds, the dataset was shuffled to remove its inherent spatial correlation. This step was necessary because the traces are organized sequentially, with one seismic trace following another, maintaining a spatial relationship. By mixing the data before splitting, we ensured that each fold contained a randomized distribution of samples, reducing the risk of overfitting to spatial patterns and providing a more robust evaluation of the model's predictive performance. This approach ensures the model's generalization ability, making it suitable for deployment in NFE environments.

Table 4 shows the cross-validation results demonstrating the model's strong generalization performance, with an average $R^2 = 0.966$ and a standard deviation of 0.005 across the folds. This consistency across the folds indicates the stability and reliability of the model. The R^2 on the unseen test set reached 0.973, further validating the model's predictive accuracy and potential for practical applications. These results highlight the model's robustness in generalizing new data and its suitability for porosity estimation tasks.

A series of noise sensitivity tests were conducted to evaluate the robustness of the proposed model under varying levels of noise as presented in Figure 7. These tests aimed to determine the point at which the input data degraded so that the model could no longer produce a meaningful representation of porosity. We conducted four experiments, progressively increasing the random Gaussian noise added to the seismic data until the results, despite acceptable metrics, became visually uninterpretable.

Table 5 demonstrates the model's robust performance under low to moderate noise conditions, exhibiting satisfactory metrics across all evaluation parameters. However, at higher noise levels, the model's performance deteriorates

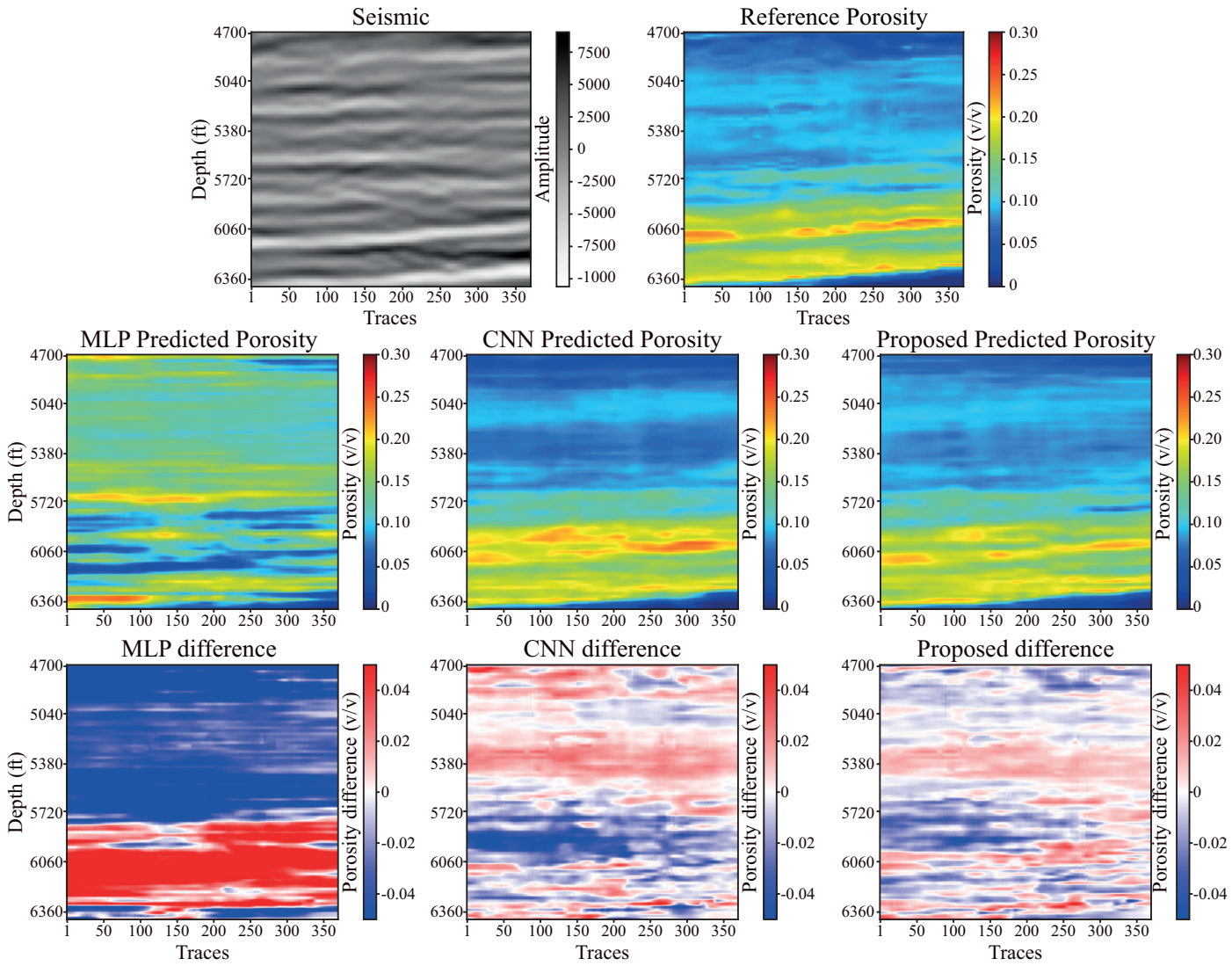


Figure 5: Comparison of porosity predictions from seismic data using neural network architectures (standard NN, CNN, and Proposed) in the exploration block within the simulated NFE context. The section corresponds to inline 83.

significantly. While the quantitative metrics at 30% noise level show only moderate degradation, as shown in Figure 7, the predicted porosity becomes geologically unreliable, with reflectors becoming poorly defined and challenging to interpret. The results show that while the model is accurate in low to moderate noise conditions, preprocessing is crucial in high-noise environments to maintain accuracy. The Gaussian noise used in this analysis represents common noise scenarios in seismic data, further emphasizing the need for robust preprocessing workflows to ensure the reliability of porosity estimates in challenging conditions.

DISCUSSION

Lateral and vertical petrophysical variations

In Figure 8, changes in the porosity were identified within the same geological formation. The Mt. Simon A unit showed the highest porosity values, while the Mt. Simon B unit showed a substantial decrease in this property. The Mt. Simon A unit is equivalent to the lower Mt. Simon formation according to core data from the VW1 well (Greenberg, 2021). The porosity model also has a good fit with the well-log porosity, validating the porosity model built with the proposed method. The porosity model obtained with the proposed method matches the edge values of porosity compared with that proposed in Greenberg (2019), as seen in Figure 8. Nevertheless, there are some differences in the mid-range porosity values. The main discrepancies occur in zones A and B at 5720 and 6280 ft from Figure 8, where the proposed method's geological structures appear inadequately mapped. Figure 9 shows that while zone B has the same structure in both methods, it is underestimated in the proposed method. However, the structure and values of the main reservoir, Mt. Simon A, are correctly mapped. Conversely, zone A lacks coher-

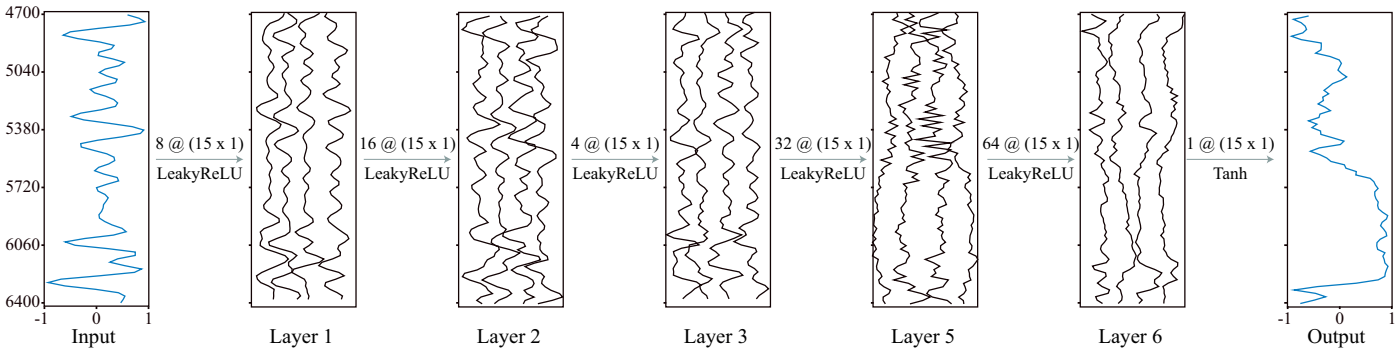


Figure 6: Feature map example of randomly selected filters from each convolutional layer from a single seismic trace.

Table 4: Results from 5-Fold Cross-Validation of the proposed porosity prediction method. Seismic traces were randomly reordered to eliminate spatial continuity, ensuring the model's performance is evaluated independently of geological spatial patterns.

Fold	R^2
Fold 1	0.960
Fold 2	0.973
Fold 3	0.969
Fold 4	0.960
Fold 5	0.968
Mean R^2	0.966
Standard Deviation	0.005

Table 5: Robustness analysis of the proposed porosity prediction method under increasing levels of Gaussian noise.

Noise (%)	SNR (dB)	SSIM	R^2
10	16.411	0.662	0.838
20	15.395	0.614	0.796
30	14.451	0.566	0.746
40	13.608	0.528	0.691

ence with the ground truth proposed by Greenberg (2019). Despite these limitations, the results from the proposed method align with the floodplain and braided depositional environment described by Leetaru and Freiburg (2014) for the Middle and Lower Mt. Simon Formation.

In addition, Figure 5 shows how the proposed method maps the seismic reflectors correctly and identifies lateral variations in porosity within the reflector. The mapping's strong numeric and geological coherence helps identify potential reservoir zones. This can be relevant for industry purposes, such as carbon storage or oil reservoirs.

The NN identifies seismic amplitude and polarity changes and maps them into porosity values through the convolutional filters as shown in Figure 6. The impact of the seismic frequency has an essential role in predicting porosity while underfitting or overfitting by using very low or high frequencies (Jo et al., 2022).

Generalization of the NN

Linear ($R^2 = 0.269$) and quadratic regressions ($R^2 = 0.015$) fail to capture the nonlinear complexity of seismic-porosity relationships, neural network models, particularly CNNs, achieve significant predictive power on test field data. These results underscore the need to use NN architectures in reservoir characterization tasks, specifically in predicting porosity with post-stack seismic data. The proposed method combined the benefits of the MLP and CNN architectures alone. This leads to obtaining the lowest RMSE, MAE, and highest R^2 in both synthetic and field data in less time. The proposed methodology showed that it could learn local spatial patterns inherent in seismic traces, patterns that might reflect subtle porosity changes.

Notably, using one-dimensional convolutional layers provided sufficient reflection continuity and coherence in our predictions, making two-dimensional convolutional layers unnecessary for achieving a satisfactory result. Comparable findings demonstrate the efficacy of 1D convolutional layers to obtain porosity from synthetic and field seismic datasets, thereby achieving robust and highly reliable outcomes (Vashisth and Mukerji, 2022; Feng et al., 2020). One likely explanation is the relatively low geological complexity of the dataset. In more challenging environments with significant lateral facies changes, two-dimensional (or even three-dimensional) convolutions may be required to capture the full extent of spatial variation.

As stated in Hunter et al. (2012), the experiments revealed that simply employing an NN algorithm does not guarantee a satisfactory solution; instead, the *design* of the network architecture is critical. Moreover, we intentionally avoided excessive parametrization to prevent overfitting and achieve faster convergence. Complex architectures such as U-Net or ResNet-like designs were also tested but did not significantly improve our proposed framework. We combined the MLP prediction capability with the spatial pattern recognition of convolutional layers, yielding a more robust predictor, as shown in the k-fold experiments.

Compared with conventional workflows, the proposed

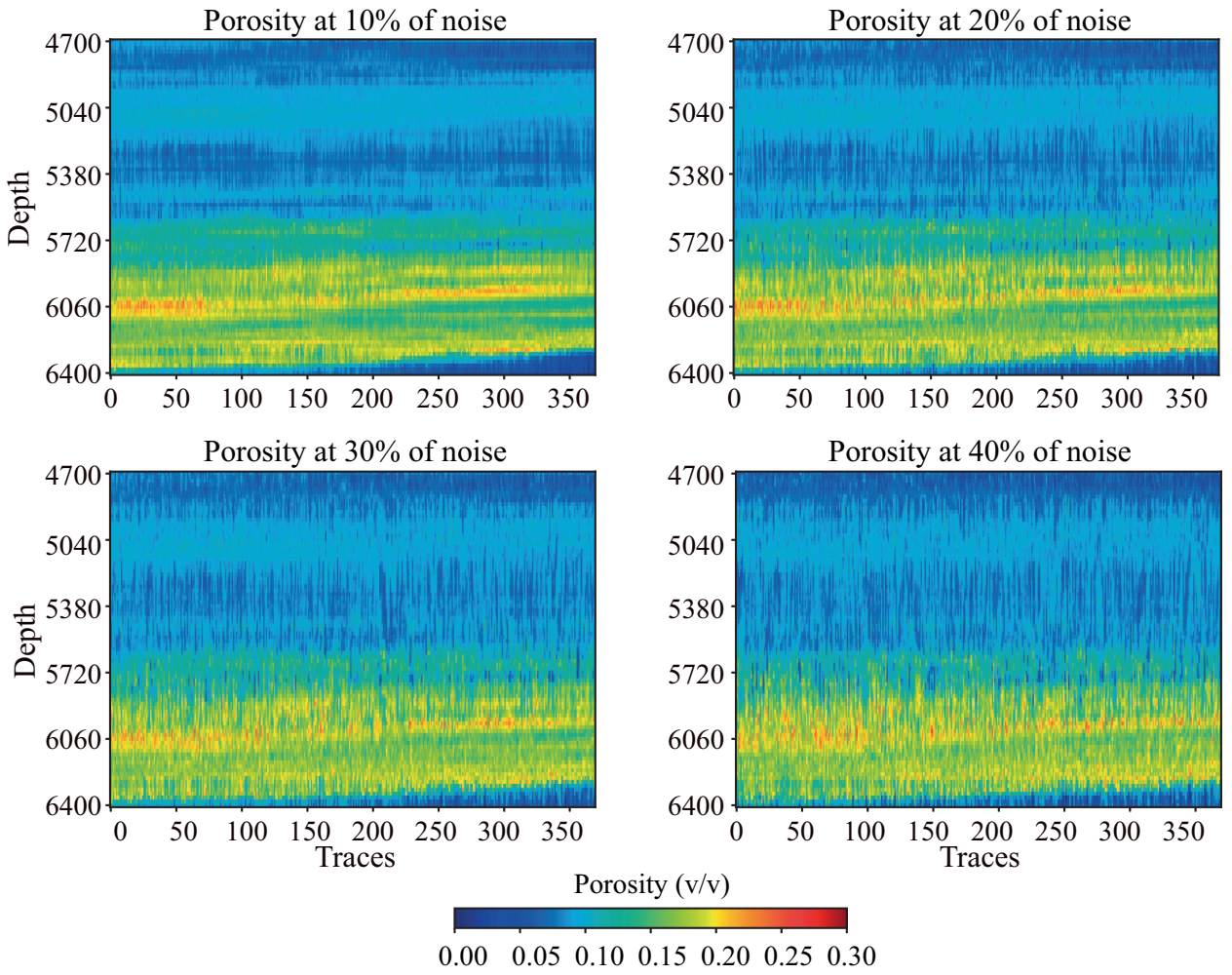
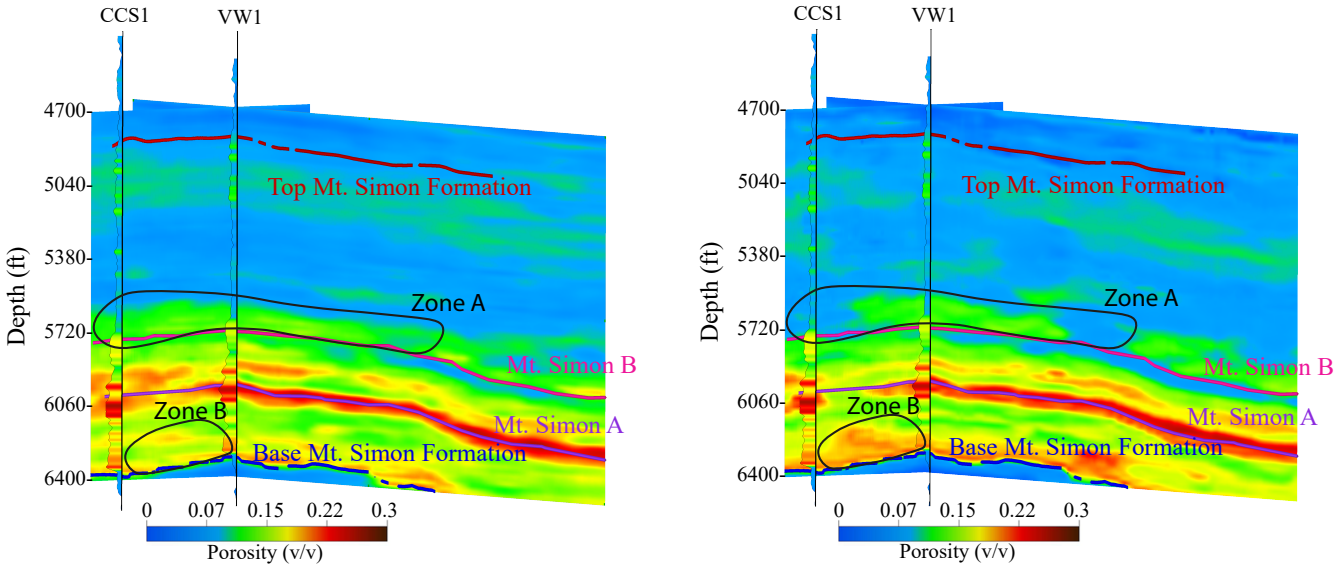


Figure 7: Noise sensitivity test of the proposed method with random Gaussian noise of 10%, 20%, 30%, and 40%, showing the result of the predicted porosity with different levels of noise in the input data. The section corresponds to inline 83.



(a) Porosity model built with the proposed method

(b) Porosity model built from seismic inversion by Greenberg (2021)

Figure 8: Comparison between (a) predicted porosity with the proposed method and (b) porosity model created from seismic inversion by Greenberg (2021). Inline 83 and Crossline 1168 with the wells CCS1 and VW1.

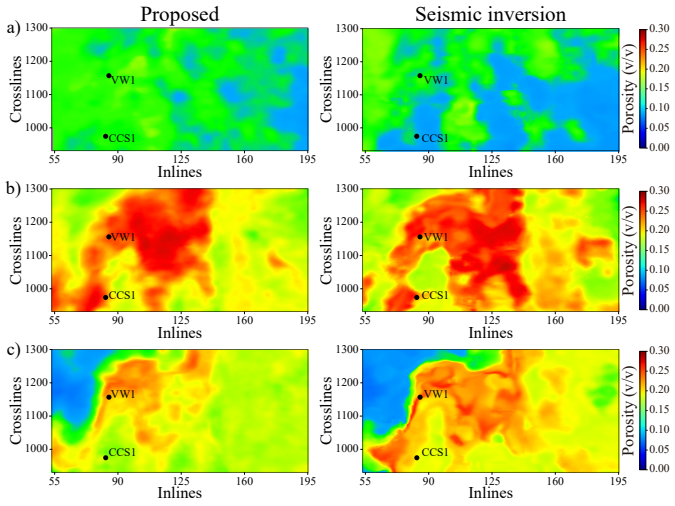


Figure 9: Z-slice of porosity from the proposed (left) and seismic inversion Greenberg (2019) (right) methods at a) $z = 5720$ ft, b) $z = 6000$ ft, and c) $z = 6280$ ft in the exploration block.

methodology for NFE substantially reduces computational overhead by requiring only the post-stack seismic dataset, a trained porosity model from the mature block, and seismic-porosity pairs at select wells. Because stratigraphic and structural interpretations are not mandatory, the interval between acquiring new seismic data in the exploration block and generating a reliable porosity model is compressed to 694.4 seconds of training for 120,263 traces and an additional 4 seconds for predicting 2,910 traces on low-end hardware. Moreover, in the neural network approach, seismic and well-data need not be in the same domain (time or depth); they must merely retain consistent dimensions (i.e., identical trace lengths). This requirement bypasses seismic–well tie operations, wavelet extraction, and other traditionally time-intensive tasks.

Data imbalance assessment

To assess the impact of data imbalance on the predictive performance of the proposed method, we conducted a series of experiments using varying proportions of training data from the mature block, ranging from 1% to 75% of the available traces, as shown in Table 6. For each proportion, the model was trained and evaluated 10 times to capture variability in performance. As expected, lower percentages of training data led to deficient performance in porosity predictions within the exploration block. When only 1% of the mature block was used, the R^2 values fluctuated significantly across iterations, with a mean of 0.538. Conversely, using higher proportions of data (e.g., 30% and above) yielded more consistent and accurate predictions, with R^2 stabilizing around 0.8 and RMSE dropping below 0.019. These results demonstrate that although the model exhibits robustness to moderate data imbalance, extremely low data availability (e.g., 1% or 2%) can obstruct learning and lead to unreliable outcomes. Nevertheless, the ability to obtain reasonable predictions with as little as 30% of the available traces highlights the efficiency of the proposed architecture in low-data regimes, a critical advantage in early exploration scenarios.

Table 6: Data imbalance assessment with different percentages of traces extracted from the mature block to train the proposed method and predict porosity. The result metrics are the mean of 10 experiments made on the exploration block.

Traces (%)	RMSE	MAE	R^2
1	0.031	0.023	0.538
2	0.027	0.020	0.642
3	0.024	0.018	0.719
4	0.023	0.018	0.729
5	0.022	0.017	0.758
10	0.022	0.017	0.769
15	0.022	0.018	0.759
20	0.021	0.016	0.788
25	0.024	0.019	0.683
30	0.020	0.015	0.805
40	0.021	0.017	0.777
50	0.021	0.016	0.778
75	0.020	0.015	0.799

Impact of data quality

According to Feng et al. (2020), normalization facilitates faster convergence. In our experiments with field data, normalizing the well-derived porosity values reduced the required training epochs by half. Furthermore, comprehensive data QC is crucial for confirming the consistency of the training data. For instance, inlines exhibiting seismic artifacts were excluded from the training and testing sets, and negative porosity values were replaced with zero. Although discarding data can limit the network's learning capacity in a supervised, data-driven setting, controlled tests introducing up to 10% Gaussian noise into the input data still produced reliable predictions, highlighting a notable degree of noise tolerance.

Impact of regularization parameter λ

The regularization term given by Equation 3 guides the prediction toward the well-log porosity. The weight of the regularization term given by λ is a hyperparameter that must be tuned according to each problem. A low λ would lead to predictions diverging from the well-log porosity domain, obtaining equivalent models that minimize the loss function but do not adequately represent the measured porosity values at the wells. On the other hand, a high λ value could lead to an accurate estimation of porosity but with imprecise spatial distribution across the block, as it may not follow the seismic reflectors. The predictions away

from the wells tend to follow flat trends. This method assumes that the well-log porosity reliably represents the actual porosity. This assumption could be assessed by calibrating the porosity models in well-logs with core data.

CONCLUSIONS

A complex relationship between porosity and post-stack seismic data cannot be found using simple machine learning algorithms such as linear or quadratic regressions or decision tree-based algorithms. This relationship must be approximated using more advanced machine learning algorithms like neural networks. The results indicate that purely data-driven approaches, such as neural networks, can estimate porosity directly from post-stack seismic, thereby overcoming the limitations of traditional input data requirements and the time-consuming steps of conventional workflows. In contrast, linear or quadratic regressions and decision tree-based solutions lack the necessary nonlinearity to predict porosity from post-stack seismic data reliably. The proposed methodology leverages a custom loss function to enhance the neural network's predictive capabilities by integrating well-log porosity. This effectively incorporates geological knowledge without relying on explicit rock-physics equations. Moreover, the methodology may provide insights into NFE projects by leveraging existing oil and gas industry data to identify new reservoirs for carbon capture or geothermal energy.

DATA AND MATERIALS AVAILABILITY

The datasets and code used in this study are publicly available for reproducibility and further research. The IBDP dataset can be accessed through the CO2DataShare platform. The synthetic dataset, initially published by Vashisth and Mukerji (2022), is available at the Rock and Wave Physics Informed Neural Networks repository. The code developed and utilized in this study for porosity prediction is also hosted on GitHub. `github.com/nicolascordoba1/MLPorosityPrediction`. The GitHub repository contains the preprocessing scripts for preparing the data, the implementations for experiments 1 and 2 as described in the study, the model validation code, and scripts for exporting the data in `.segy` format. The exported `.segy` files can be used in conventional seismic processing and interpretation software, ensuring compatibility with traditional workflows while maintaining accessibility for future research and applications.

PUBLICATIONS AND AUTHOR'S CONTRIBUTIONS

Journal Papers

1. **Córdoba-Castillo, N.**, Goyes-Peñafiel, P., Castro-Vera, L., & Khurama Velásquez, S. (2025). Reservoir porosity estimation from post-stack seismic data using a deep learning model constrained by well-porosity: A near-field application [Manuscript in *Revision*]. The Leading Edge, Special Section: Data Driven Geophysics.

International Conferences

1. **Córdoba-Castillo, N.**, Goyes-Peñafiel, P., Castro-Vera, L., & Khurama Velásquez, S. (2025). Reservoir porosity estimation from post-stack seismic data using a deep learning model constrained by well-porosity: A near-field application. EAGE ANNUAL 86th Conference and Exhibition. Toulouse, France. 2025.

2. **Córdoba-Castillo, N.,** Goyes-Peñafiel, P., & Castro-Vera, L. (2024). Estimating reservoir porosity from post-stack seismic data using machine learning techniques. SEG Latin America Virtual Student 5th Conference. Virtual.

The authors thank the 5th SEG Latin America Virtual Student Conference for their valuable suggestions, which significantly contributed to this investigation. The proposed method was honored with first place at the conference.

REFERENCES

Aster, R. C., B. Borchers, and C. H. Thurber, 2018, Parameter estimation and inverse problems: Elsevier.

Bachrach, R., 2006, Joint estimation of porosity and saturation using stochastic rock-physics modeling: Geophysics, **71**, O53–O63.

Bauer, R. A., R. Will, S. E. Greenberg, and S. G. Whittaker, 2019, Illinois basin-decatur project: Cambridge University Press.

Bloch, S., and K. Helmold, 1995, Approaches to predicting reservoir quality in sandstones: AAPG bulletin, **79**, 97–114.

Bornard, R., F. Allo, T. Coleou, Y. Freudenreich, D. H. Caldwell, and J. G. Hamman, 2005, Petrophysical seismic inversion to determine more accurate and precise reservoir properties: SPE Europec/EAGE Annual Conference, Society of Petroleum Engineers, SPE–94144–MS.

Bosch, M., T. Mukerji, and E. F. Gonzalez, 2010, Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review: Geophysics, **75**, 75A165–75A176.

Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: Mathematics of Control, Signals and Systems, **2**, 303–314.

Doyen, P. M., 1988, Porosity from seismic data: A geostatistical approach: Geophysics, **53**, 1263–1275.

———, 2007, Seismic reservoir characterization: An earth modelling perspective: EAGE.

El-Dabaa, S. A., F. I. Metwalli, A. T. Amin, and A. A. Basheer, 2022, Prediction of porosity and water saturation using a probabilistic neural network for the

bahariya formation, nader field, north western desert, egypt: Journal of African Earth Sciences, **196**, 104638.

Fattahi, H., and S. Karimpouli, 2016, Prediction of porosity and water saturation using pre-stack seismic attributes: a comparison of bayesian inversion and computational intelligence methods: Computational Geosciences, **20**, 1075–1094.

Feng, R., T. Mejer Hansen, D. Grana, and N. Balling, 2020, An unsupervised deep-learning method for porosity estimation based on poststack seismic data: Geophysics, **85**, M97–M105.

Gholami, A., M. Amirpour, H. R. Ansari, S. M. Seyedali, A. Semnani, N. Golsanami, E. Heidaryan, and M. Ostadhassan, 2022, Porosity prediction from pre-stack seismic data via committee machine with optimized parameters: Journal of Petroleum Science and Engineering, **210**, 110067.

Goyes-Peñafiel, P., E. Vargas, C. V. Correa, Y. Sun, U. S. Kamilov, B. Wohlberg, and H. Arguello, 2023, Coordinate-based seismic interpolation in irregular land survey: A deep internal learning approach: IEEE Transactions on Geoscience and Remote Sensing, **61**, 1–12.

Greenberg, S. E., 2019, Illinois state geological survey decatur 3d 2015 / macon co il: Technical report, Sterling Seismic and Reservoir Services.

———, 2021, Illinois basin - decatur project. an assessment of geological carbon sequestration options in the illinois basin: Phase iii: Technical report, Illinois State Geological Survey.

Hastie, T., 2009, The elements of statistical learning: data mining, inference, and prediction.

Hornik, K., M. Stinchcombe, and H. White, 1989, Multilayer feedforward networks are universal approximators: Neural Networks, **2**, 359–366.

Hunter, D., H. Yu, M. S. Pukish, III, J. Kolbusz, and B. M. Wilamowski, 2012, Selection of proper neural network sizes and architectures—a comparative study: IEEE Transactions on Industrial Informatics, **8**, 228–240.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, 2017, Image-to-image translation with conditional adversarial networks: Proceedings of the IEEE conference on computer vision and pattern recognition, 1125–1134.

Iturrarán-Viveros, U., 2012, Smooth regression to estimate effective porosity using seismic attributes: Journal of Applied Geophysics, **76**, 1–12.

Jo, H., Y. Cho, M. Pyrcz, H. Tang, and P. Fu, 2022, Machine-learning-based porosity estimation from multifrequency poststack seismic data: GEOPHYSICS, **87**, M217–M233.

Kim, Y., and N. Nakata, 2018, Geophysical inversion versus machine learning in inverse problems: The Leading Edge, **37**, 894–901.

Kingma, D. P., 2014, Adam: A method for stochastic optimization: arXiv preprint arXiv:1412.6980.

Kumar, R., B. Das, R. Chatterjee, and K. Sain, 2016, A methodology of porosity estimation from inversion of post-stack seismic data: Journal of Natural Gas Science and Engineering, **28**, 356–364.

Leetaru, H. E., and J. T. Freiburg, 2014, Litho-facies and reservoir characterization of the mt simon sandstone at the illinois basin–decatur project: Greenhouse Gases: Science and Technology, **4**, 580–595.

Leiphart, D. J., and B. S. Hart, 2001, Comparison of linear regression and a probabilistic neural network to predict porosity from 3-d seismic attributes in lower brushy canyon channeled sandstones, southeast new mexico: Geophysics, **66**, 1349–1358.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, The rock physics handbook: Tools for seismic analysis of porous media, 2 ed.: Cambridge University Press.

Rasmussen, K., and K. Maver, 1996, Direct inversion for porosity of post stack seismic data: SPE European 3-D Reservoir Modelling Conference, Society of Petroleum Engineers, SPE–35509.

Russell, B., and D. Hampson, 1991, Comparison of post-stack seismic inversion methods, *in* SEG technical program expanded abstracts 1991: Society of exploration geophysicists, 876–878.

Russell, B. H., 1988, 2, *in* Introduction to Seismic Inversion Methods: Society of Exploration Geophysicists.

Saltzer, R., C. Finn, and O. Burtz, 2005, Predicting vshale and porosity using cascaded seismic and rock physics inversion: The Leading Edge, **24**, 732–736.

Sneider, R. M., 1990, Reservoir description of sandstones, *in* Sandstone petroleum reservoirs: Springer, 1–3.

Vashisth, D., and T. Mukerji, 2022, Direct estimation of porosity from seismic data using rock-and wave-physics-informed neural networks: The Leading Edge, **41**, 840–846.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 2004, Image quality assessment: from error visibility to structural similarity: IEEE transactions on image processing, **13**, 600–612.

Yasin, Q., G. M. Sohail, Y. Ding, A. Ismail, and Q. Du, 2020, Estimation of petrophysical parameters from seismic inversion by combining particle swarm optimization and multilayer linear calculator: Natural Resources Research, **29**, 3291–3317.