Available online at:
Al-Fiziya: Journal of Materials Science, Geophysics,
Instrumentation and Theoretical Physics
Website: http://journal.uinjkt.ac.id/index.php/al-fiziya
Vol. 8 No. I 2025, 27-44

Harnessing Machine Learning for Caprock Capacity and Seal Integrity Assessment in CCS: A Multi-Attribute Seismic Inversion Study from the F3 Block, North Sea

Fadhlur Rahman[1*], Abdul Haris[1], Dwandari Ralanarko[2], Humbang Purba[2], Wrahaspati Rulandoko[2], Praditio Riyadi[3], Muhammad Nafian[3]

[1]Department of Physics, Universitas Indonesia, Indonesia
[2]PHE OSES, RDTX Building, South Jakarta, Jakarta, Indonesia
[3]Department of Physics, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

fadhrahman1911@gmail.com

Submitted: June; Revised: July; Approved: August; Available Online: August.

Abstracts. Accurate characterization of subsurface petrophysical properties is a critical prerequisite for evaluating the suitability of geological formations for carbon capture and storage (CCS), particularly in identifying high-capacity reservoirs and effective sealing intervals. This study explores the use of a machine learning approach, Random Forest (RF) regression, for multi-attribute seismic inversion to predict porosity and acoustic impedance in the F3 Block, offshore Netherlands. The integration of ten seismic attributes with two well log datasets enables the construction of predictive models capable of resolving complex lithological variations within deltaic settings. The RF algorithm's robustness against geological noise and its ability to model nonlinear relationships offer significant advantages over conventional inversion workflows, especially in heterogeneous and interbedded formations. The results demonstrate that RF-based inversion produces petrophysical volumes with improved spatial continuity and alignment with depositional patterns, offering a promising avenue for CCS site screening and reservoir-seal evaluation. The method's ability to capture subtle textural and facies changes also enhances understanding of potential CO_2 migration pathways and trap integrity. This research underscores the potential of data-driven inversion frameworks in supporting geoscientific decision-making for CCS development, particularly in data-limited or geologically complex offshore regions.

Keywords: CCS, F3, Inversion, Machine Learning, Petrophysics

DOI : 10.15408/fiziya.v8i1.47182

INTRODUCTION

The accurate prediction of subsurface petrophysical properties, particularly porosity and acoustic impedance (AI), serves as a fundamental pillar in evaluating the feasibility of geological formations for carbon storage. In the context of Carbon Capture and Storage (CCS), where the safe and permanent sequestration of CO_2 is paramount, these properties govern not only the storage capacity and injectivity of the reservoir but also the sealing effectiveness of overlying caprocks. Porosity dictates the available pore volume that can host injected CO_2, while AI provides critical insights into lithology, fluid content, and mechanical continuity. These parameters are essential in assessing both the suitability of the reservoir and the integrity of the caprock system, which must remain impermeable to ensure containment over geologic timescales. Traditionally, seismic inversion techniques—such as deterministic, stochastic, or geostatistical approaches—have been widely applied to convert seismic reflection data into quantitative models of the subsurface. However, these conventional approaches often rely on simplifying assumptions, such as linearity and stationarity, which may be invalid in the presence of heterogeneous sedimentary sequences or structurally deformed caprocks. The effectiveness of such methods is further constrained by their strong dependence on a priori models and well log data, which are frequently sparse or spatially limited in CCS candidate sites. As a result, the inversion outputs can become biased or unstable, particularly in areas where seismic amplitudes are influenced by overlapping effects of stratigraphy, diagenesis, and fluid variability. These limitations pose significant challenges in CCS-focused settings, where accurate delineation of both storage and seal units is critical to risk mitigation and long-term performance assessment. In light of these challenges, data-driven methods have emerged as transformative alternatives that offer improved flexibility and adaptability. In particular, machine learning (ML) approaches provide a framework capable of capturing nonlinear relationships and high-dimensional patterns embedded in seismic datasets without requiring explicit physical assumptions [1].

Within the growing suite of machine learning (ML) tools, Random Forest (RF) has gained considerable traction as a robust, interpretable, and practically efficient algorithm, particularly under the geophysical and geological uncertainties typical of Carbon Capture and Storage (CCS) reservoirs. As an ensemble learning method, RF aggregates the outcomes of multiple decision trees trained on diverse data subsets and feature combinations, thereby reducing prediction variance and guarding against overfitting, which is an essential advantage when working with seismic data that are often noisy, spatially incomplete, or affected by acquisition footprint. In CCS-specific contexts, including caprock characterization, this robustness becomes critical for detecting subtle variations in acoustic impedance or lateral seal continuity that may influence long-term containment integrity. A further strength of RF lies in its ability to rank feature importance, allowing geoscientists to identify which seismic attributes most influence inversion outcomes, thus enhancing trust in model transparency. Zou et al. [1] demonstrated the efficacy of RF in predicting porosity from multiple seismic attributes, reporting strong predictive performance and the ability to quantify uncertainty through ensemble dispersion metrics. Such capabilities are especially valuable in CCS site screening and development phases, where uncertainty analysis informs risk-based decision making for storage integrity and monitoring strategy design. Compared to more complex ML models such as deep neural networks, RF offers a more interpretable yet competitive alternative, although it may be less effective in capturing deep hierarchical or temporal patterns. More broadly, ML methods, including RF, remain dependent on representative training datasets and may struggle when extrapolated to geologies outside their training domain. Despite these limitations, RF provides a compelling balance between accuracy, explainability, and resilience to noisy input, making it a strong candidate for practical deployment in CCS inversion workflows, particularly in challenging environments where traditional physics-based methods fall short.

One of the most significant advantages of RF in this context lies in its integration with multi-attribute inversion techniques. By incorporating multiple seismic-derived attributes—each with distinct physical sensitivity to lithological and fluid variations—into a supervised learning model, RF can offer enhanced predictions that reflect the complex realities of subsurface systems. For instance, attributes such as instantaneous amplitude, Hilbert envelope, and RMS amplitude may each respond differently to porosity variations, fluid substitution effects, or thin-bed stratigraphy. When used collectively, these attributes allow for a richer and more discriminative inversion process, improving the delineation of CO_2 reservoir zones, internal heterogeneities, and potential migration pathways. Topór and Sowiżdżał [2] applied RF in conjunction with other ML tools to estimate porosity and saturation types in Miocene sandstone formations, finding that ensemble methods could successfully resolve lithofacies variability and fluid distribution that conventional inversion often overlooks. Their work highlights how ML-based multi-attribute inversion facilitates a more spatially continuous, geologically coherent, and physically plausible prediction of reservoir properties—critical traits for CCS site characterization, where decisions must balance subsurface uncertainty with operational feasibility. In addition, this approach allows researchers to iteratively refine attribute selection and model architecture based on feedback from uncertainty maps and validation data, further boosting confidence in the interpreted results.

The Dutch sector of the North Sea—specifically the F3 Block—presents an ideal natural laboratory to test and deploy RF-based seismic inversion workflows under CCS-relevant conditions. The F3 Block hosts a publicly available, high-resolution seismic dataset alongside two key well logs, enabling method development and benchmarking in a geologically diverse environment. The block's subsurface is typified by deltaic siliciclastic systems, comprising interbedded sandstone–shale units, sinuous channel deposits, and locally deformed caprock analogs. Such features are representative of many storage complexes currently under consideration in Europe's North Sea and other continental shelf basins, making the site a meaningful proxy for broader CCS applications. Safari et al. [3] previously emphasized how attribute continuity and structural smoothing significantly improved seismic interpretation in this area, aiding fault detection and stratigraphic delineation. Nevertheless, accurately predicting porosity and AI remains challenging due to overlapping seismic signatures, weak reflectors in low-contrast zones, and the fine layering of deltaic facies. By applying a Random Forest-based multi-attribute inversion strategy to the F3 Block, this study aims to bridge these interpretational gaps, delivering spatially refined maps of porosity and impedance that honor well log constraints and reveal internal heterogeneities of potential CO_2 storage intervals. In so doing, the method demonstrates how machine learning can enhance our ability to screen, characterize, and monitor CCS sites using conventional seismic datasets—thus offering a scalable and data-efficient pathway for subsurface decarbonization planning.

RESEARCH METHODS

Machine Learning

Machine learning (ML) refers to a suite of data-driven algorithms that enable computational models to identify patterns and make data-based predictions without being explicitly programmed. In the context of Carbon Capture and Storage (CCS), where accurate estimation of porosity, acoustic impedance, and seal integrity is crucial for injectivity and containment assessment, ML provides a flexible framework for seismic inversion and petrophysical property prediction. Unlike traditional deterministic inversion methods, which often rely on linear assumptions and prior model constraints, ML can model nonlinear, high-dimensional relationships between seismic attributes and rock properties, making it particularly advantageous in thinly bedded, heterogeneous, or poorly constrained formations. As highlighted by Bishop, ML is grounded in probabilistic modeling and statistical inference, with learning defined as the ability to generalize from past observations [4]. This foundation has enabled ML models to surpass traditional methods in resolution and adaptability, particularly in complex geologies common to CCS sites. The rapid growth in geophysical data availability, combined with increased computational capacity, has catalyzed the use of ML across Earth science domains, including CCS reservoir monitoring and seal characterization [5].

Recent applications underscore ML's capacity to enhance seismic inversion outcomes. For instance, Anjom et al. applied a data-driven AVO inversion using ML to resolve subtle gas-bearing sand signatures, a task where conventional inversion often fails due to thin-bed interference and amplitude attenuation [6]. Although their study focused on offshore Egypt, the challenges addressed mirror those found in depleted gas fields considered for CO_2 injection. Similarly, Wu et al. employed Random Forest and other ensemble models to fuse seismic velocities, resistivity, and borehole data for lithological classification, demonstrating improved delineation of reservoir facies and internal baffles [7]. These examples show that ML models, and RF in particular, can integrate disparate geophysical inputs to resolve petrophysical targets with higher sensitivity and spatial continuity than deterministic methods. In CCS workflows, this enables more accurate identification of injectivity zones, barriers to flow, and potential leakage pathways—insights that are critical for site feasibility and risk mitigation.

Moreover, the evolution of physics-guided machine learning frameworks has further improved model reliability and physical consistency. By incorporating geophysical constraints such as conservation laws and domain-specific knowledge, these hybrid approaches reduce overfitting and enhance generalization—key benefits in CCS where well control is often sparse. Karpatne et al. argued for embedding physical principles into ML architectures to ensure geoscientific validity, laying the foundation for physics-informed learning [8]. This paradigm has been successfully applied in CCS-relevant tasks, including the hybrid CNN framework proposed by Hu et al. to predict porosity and sandbody thickness in the Bohai Bay Basin, which outperformed conventional inversion techniques [9]. In addition, ML is gaining ground in microseismic monitoring for CO_2 injection, where real-time classification and spatial localization of induced events play a crucial role in operational safety [10]. Collectively, these advances underscore the growing importance of ML as not merely a computational add-on, but as a core inversion strategy capable of resolving the nonlinear, uncertain, and stratigraphically complex nature of CO_2 storage formations.

Vol.8 No. I 2025, 27-44
P-ISSN: 2621-0215, E-ISSN: 2621-489X

Random Forest in Seismic Inversion

Random Forest (RF) is a tree-based ensemble algorithm that relies on stochastic sampling to construct multiple decision trees, each trained on different bootstrapped subsets of the data and a random selection of features [11]. While the ensemble framework inherently improves model generalization, its performance remains sensitive to key hyperparameters, particularly in high-dimensional geophysical applications such as seismic inversion for CO_2 storage assessment. In this study, the RF model was carefully tuned to balance predictive accuracy, computational efficiency, and resistance to overfitting factors especially critical in heterogeneous and data-limited CCS reservoirs. The number of trees (n_estimators) was set to 100 to ensure sufficient diversity across the ensemble while maintaining tractable runtime during full-section inference. The maximum depth (max_depth) was limited to 12, a value selected based on iterative testing to allow each tree to learn meaningful nonlinear interactions between multi-attribute inputs and petrophysical outputs, without fragmenting the feature space excessively or responding to spurious variations in seismic amplitude. Shallower depths led to underfitting of log–seismic relationships, while deeper trees increased variance without improving correlation at the well locations.

Although the parameters controlling node purity such as min_samples_split and min_samples_leaf were kept at their default values, their regularizing effects were enhanced by the use of a ±3 sample windowing scheme during training data generation. This windowed approach increased the density of training examples derived from each log curve, while embedding local temporal context that helps the model learn transitions in seismic response across thin-bed and vertically heterogeneous facies. Each training instance comprised a 10-dimensional feature vector built from seismic-derived attributes, including amplitude, gradient, Hilbert envelope, RMS, relative acoustic impedance, low-pass and curvature components. These features were selected for their sensitivity to lithological contrasts, fluid-related effects, and stratigraphic geometry, all of which govern CO_2 injectivity and caprock integrity in CCS systems. The input attributes were standardized prior to training using z-score normalization to ensure fair contribution across features and reduce model sensitivity to scale disparities.

Independent RF models were trained for porosity and acoustic impedance using calibration data from two wells (Well 34 and Well 61) strategically positioned within the F3 Block. Once trained, the models were deployed across the full seismic section to predict spatial distributions of AI and porosity, producing continuous property volumes that were evaluated against well log measurements. The resulting correlation coefficients ranged from 0.90 to 0.94 at the calibration wells, indicating strong alignment between predicted and actual log values. These results suggest that the selected RF configuration offers an effective balance between model complexity and stability, capturing the nonlinear seismic–log relationships required for reliable petrophysical inversion in CCS reservoir settings. As emphasized by Breiman [11], the strength of the Random Forest algorithm lies in its ability to reduce variance through ensemble averaging, which in this case proved essential for maintaining prediction consistency across a geologically variable and stratigraphically complex setting.

Multi-attribute

Vol.8 No. I 2025, 27-44
P-ISSN: 2621-0215, E-ISSN: 2621-489X

Multi-attribute seismic inversion offers significant value in Carbon Capture and Storage (CCS) site evaluation by integrating multiple seismic-derived attributes to refine the prediction of key subsurface properties critical to CO_2 storage feasibility. Each attribute carries unique geological sensitivities: amplitude aids in identifying impedance contrasts associated with reservoir-caprock interfaces; phase reveals reflector continuity relevant for trap integrity; frequency attributes relate to bed thickness crucial for vertical sealing; and similarity enhances the delineation of structural boundaries, which can influence potential leakage pathways [18]. In the F3 Block of the Dutch North Sea, a well-documented CCS analogue, seismic attributes such as similarity, dip, and curvature have shown effectiveness in highlighting buried fluvial channels and deltaic stratigraphy—features that critically impact both injectivity and containment reliability [19].

Kabaca emphasized that attributes like instantaneous phase, raw amplitude, and RMS amplitude were instrumental in delineating stratigraphic layering and depositional trends across the F3 Block [20]. These characteristics are not only relevant for hydrocarbon systems but are equally essential in evaluating the continuity of potential storage zones and the lateral extent of sealing facies in CCS settings. Additionally, the utility of these attributes in horizon flattening and channel boundary mapping supports the interpretation of baffle geometries and sedimentary heterogeneities, which may act as either migration barriers or vertical leakage conduits within storage formations [21].

In CCS-focused inversion workflows, multi-attribute inputs are often integrated with machine learning algorithms like Random Forest to enhance the prediction of porosity and acoustic impedance—both of which are foundational to assessing reservoir capacity and seal integrity. Zou et al. applied a multi-attribute Random Forest framework and successfully improved porosity prediction accuracy while providing ensemble-based uncertainty metrics, which are essential in risk-informed CCS site screening [22]. In a parallel study within the same F3 Block, the inclusion of attributes such as seismic energy and envelope amplitude within inversion modeling enabled better lithological differentiation, helping to distinguish porous injection zones from low-permeability barriers [23]. Furthermore, advanced attribute integration techniques like 3D Wheeler transformation, when combined with curvature and dip, allowed for the reconstruction of paleo-geomorphological trends and stratigraphic architecture—information pivotal in identifying compartmentalization risks and migration pathways in a CCS scenario [24].

Collectively, these studies underscore that multi-attribute seismic inversion, originally developed for hydrocarbon exploration, holds substantial potential for CCS site assessment. By capturing geological complexity and enhancing interpretability, this method supports a more informed evaluation of both storage capacity and containment security—two pillars of any viable CCS operation..

Seismic Inversion

Seismic inversion remains a cornerstone in geophysical reservoir characterization, providing a structured methodology for transforming reflection-based seismic amplitudes into spatially resolved subsurface property volumes. Among its outputs, acoustic impedance (AI) stands as a fundamental parameter due to its sensitivity to lithological contrasts and fluid content, offering a critical linkage between seismic response and petrophysical interpretation [25]. In clastic-dominated settings such as deltaic or fluvial reservoirs, AI serves not only as a lithology proxy but also as a first-order predictor for properties like porosity and permeability. In the F3 Block of the Dutch North Sea well-documented CCS analogue characterized by interbedded sandstone–shale sequences model-based post-stack inversion has been successfully applied to derive AI volumes that illuminate sedimentary architectures, enabling the identification of reservoir-prone intervals and internal facies transitions [25]. However, traditional inversion approaches remain bounded by intrinsic limitations, including reliance on low-frequency initial models, assumption of linear relationships between reflectivity and impedance, and sensitivity to seismic noise and vertical resolution loss due to wavelet bandwidth constraints. These challenges are especially pronounced in CCS projects, where subtle contrasts between seal and reservoir units and thin-bed heterogeneity can govern the long-term containment of injected CO_2.

To address these constraints, recent developments have introduced data-driven methodologies that augment or replace conventional workflows with machine learning (ML)-based models trained on seismic-derived attributes. As discussed in the previous section, multi-attribute seismic analysis allows for the extraction of amplitude, phase, curvature, and frequency-domain information each with specific geological sensitivities that together provide a richer input set for supervised inversion. These attributes are subsequently fed into inversion frameworks such as Random Forest or deep neural networks, which can capture nonlinear relationships between seismic patterns and rock properties with greater flexibility than physics-based transforms alone. For instance, Jo et al. [26] presented a deep learning framework that integrates prestack seismic data to predict porosity directly, showing improved adaptability to heterogeneous lithologies where conventional AI-to-porosity transformations falter. Similarly, Mojeddifar et al. [27] proposed a pseudo-forward modeling strategy that employs similarity-based attributes to estimate porosity without relying on a fixed physical kernel, thus mitigating the rigidity of traditional inversion algorithms.

Despite these advancements, ML-based inversion remains susceptible to challenges such as data quality dependence, overfitting in sparse training conditions, and limited generalizability beyond the training domain. Moreover, while these methods can reproduce high-frequency variations in predicted properties, their effective vertical resolution remains fundamentally limited by the seismic wavelet's bandwidth and signal-to-noise ratio factors that must be carefully considered in any CCS feasibility study. To enhance reliability, hybrid strategies that embed physical constraints into ML architectures are increasingly being explored, allowing geoscientists to harness the pattern recognition capabilities of data-driven models while maintaining consistency with geophysical principles. Collectively, these trends suggest a paradigm shift where multi-attribute inversion evolves from a deterministic signal-matching process into a probabilistic learning framework, offering improved robustness, better handling of uncertainty, and a deeper integration of geological complexity all of which are essential for risk-informed decision-making in CO_2 storage site evaluation.

Methodology

Following the upsampling of the seismic data to a 2 ms sampling interval, a diverse set of ten seismic attributes was extracted from each trace to enhance subsurface interpretability. These attributes were selected for their capacity to highlight various aspects of geological variability, ranging from amplitude-based contrasts to frequency and geometry-related patterns. The attribute set included raw amplitude, RMS amplitude, Hilbert envelope, instantaneous phase, amplitude gradient, curvature, relative acoustic impedance, the derivative of the envelope, and two Gaussian-filtered amplitude components. The rationale behind this selection was grounded in both geophysical theory and geological applicability. Attributes such as RMS and envelope are known to respond to energy concentrations that often correlate with porous or gas-charged zones. Curvature and gradient tend to accentuate subtle structural features like channel flanks or compaction-related flexures that could compromise seal integrity. The inclusion of smoothed and derivative forms was intended to capture multi-scale signal variations, providing sensitivity to both fine layering and broader stratigraphic architecture. Post-model analysis confirmed that envelope, curvature, and RMS amplitude were consistently the most influential attributes for both porosity and acoustic impedance prediction, aligning with their expected geological importance and reinforcing the reliability of the attribute framework.

Rather than relying on a conventional model-based inversion that begins with an assumed starting model and forward modeling, this study adopted a data-guided approach in which the relationships between seismic attributes and petrophysical properties were learned from the available well control. Once trained using the calibration from Well 34 and Well 61, the method was applied trace by trace across the entire seismic section to produce continuous volumes of predicted porosity and acoustic impedance. These volumes revealed features consistent with known deltaic stratigraphy in the F3 Block, including lateral facies transitions, multi-story channel fills, and vertically persistent sealing units. The impedance volumes captured the broad internal architecture of the reservoir system, while the porosity predictions helped distinguish between cleaner sand bodies and heterolithic intervals with limited injectivity. This mapping was especially useful in identifying potential injection zones and the continuity of caprock analogs, both of which are essential for evaluating containment integrity in CO_2 storage scenarios. The results not only aligned well with the interpreted geological framework but also extended the predictive capacity into areas lacking direct borehole data, offering a practical tool for reservoir screening and early-phase CCS planning.

To address the need for confidence in the predictions, the inversion outputs were accompanied by a basic uncertainty evaluation. This was achieved by quantifying the variability among individual predictions within the model ensemble for each sample location. The resulting ensemble spread served as a proxy for uncertainty, highlighting intervals where the model exhibited low internal agreement. As expected, higher uncertainty tended to cluster near facies transitions, thinly bedded sequences, and low-reflectivity zones where seismic signals are inherently ambiguous. Conversely, thick, well-bounded units such as clean sand packages and continuous shales showed relatively low prediction spread. These spatial patterns were further validated against well log comparisons, where residual analysis confirmed that the largest deviations between predicted and observed values occurred near lithological boundaries or in zones of acoustic impedance overlap. Although overall model performance remained strong—with correlation coefficients between 0.90 and 0.94 across both wells—the inclusion of uncertainty diagnostics added another layer of interpretive value. This information is particularly relevant for CCS feasibility studies, where understanding the degree of confidence in reservoir and seal predictions can inform both risk management and monitoring strategy design.

Together, the integration of carefully selected seismic attributes, data-driven property prediction, and ensemble-based uncertainty screening provides a robust and transferable framework for subsurface evaluation. The method is well suited to clastic environments with limited well control and strong lateral variability, such as the F3 Block and other deltaic reservoirs under consideration for carbon storage. By moving beyond deterministic workflows and embracing a probabilistic understanding of geophysical response, this approach supports more informed decisions about injectivity, seal reliability, and long-term containment performance in CO_2 storage projects.

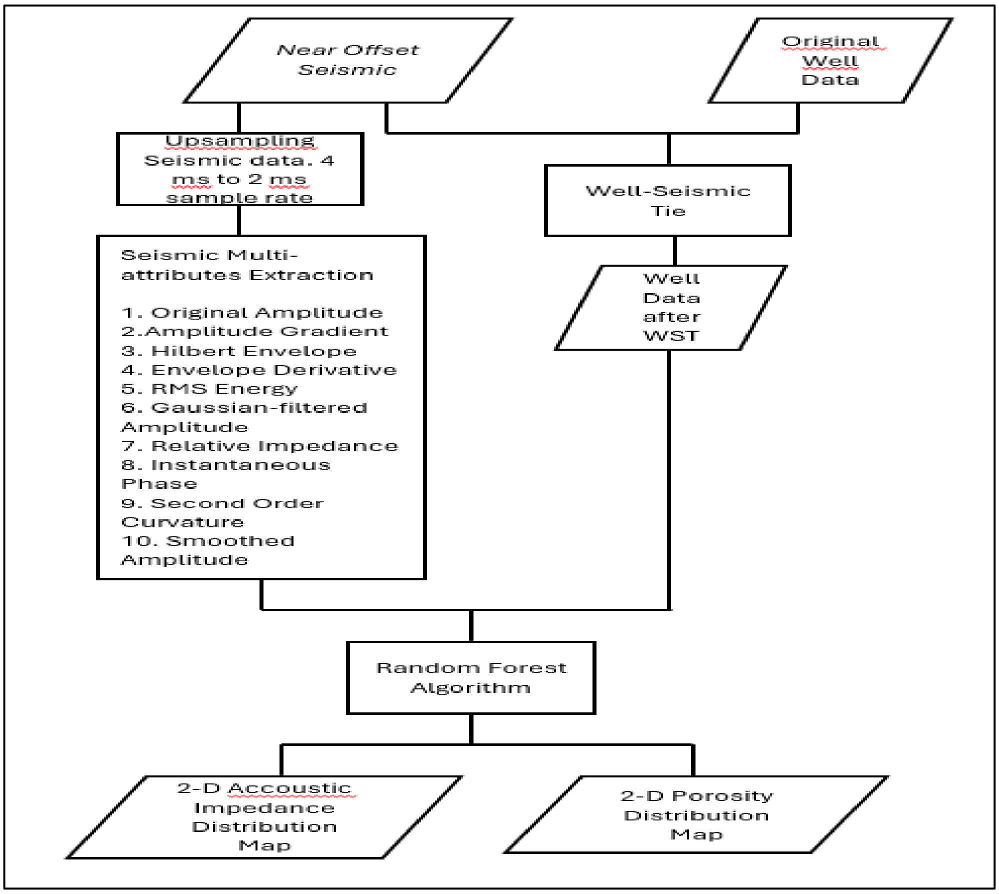

Figure 1. Random Forest Workflow

RESULTS AND DISCUSSION

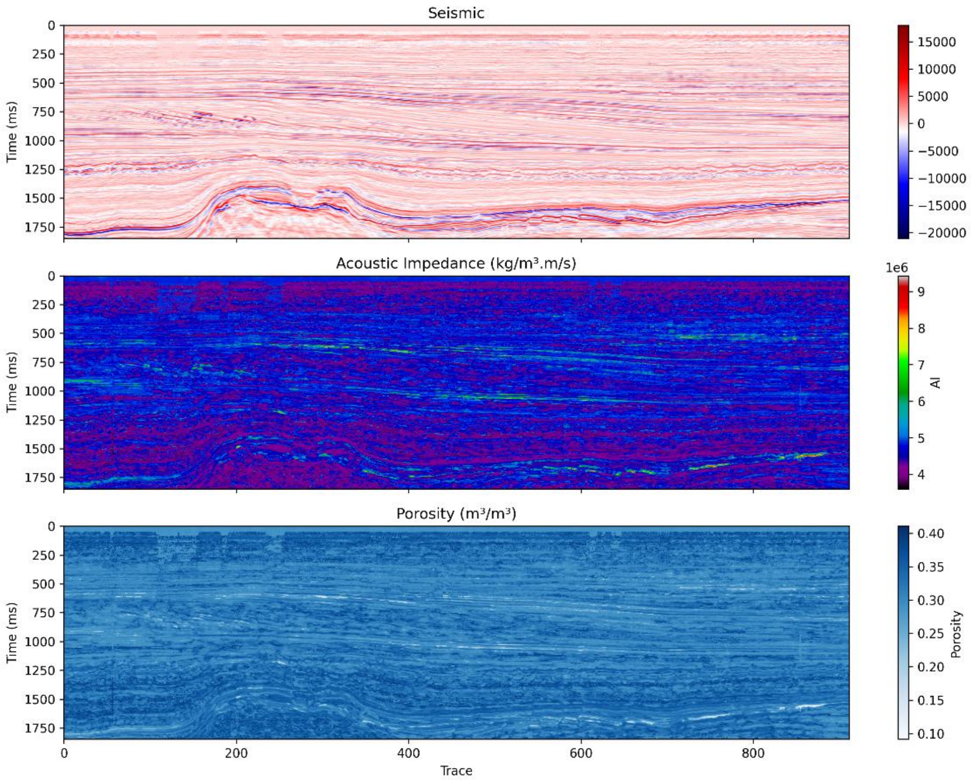

Figure 2. 2-D Inversion AI and Porosity Section

Figure 2 provides insight into the subsurface geological architecture of the F3 Block, as revealed by the spatial distribution of porosity and acoustic impedance. The alignment of high-porosity zones with low acoustic impedance values suggests the presence of sand-rich depositional environments, likely distributary channels or mouth bars, which are characteristic of deltaic systems [20]. These sand-dominated facies are not only relevant in hydrocarbon plays, but also critically important as potential CO_2 injection intervals in a CCS framework, due to their higher storage capacity and injectivity. The curvilinear geometry of these features, along with abrupt changes in impedance boundaries, further supports the interpretation of channel complexes bordered by shale-rich levees or floodplain deposits [18]. Such geometries often define natural flow baffles or barriers, which may enhance long-term CO_2 containment by restricting lateral plume migration. These lateral facies variations are typical in delta lobe switching and multi-phase progradation observed in the F3 setting [3].

The seismic expression of possible minor faulting or compactional deformation can also be inferred from the discontinuities and impedance breaks across the map. These features, while subtle, may play an important role in compartmentalizing the reservoir and controlling lateral fluid flow. In CCS contexts, even minor structural discontinuities can influence pressure build-up or act as potential leak pathways, making their identification essential for long-term risk mitigation. The clarity of such features in the AI map demonstrates the effectiveness of multi-attribute modeling in enhancing subsurface interpretation [25]. Additionally, the coherence between predicted petrophysical properties and geological structures illustrates the advantage of using ensemble learning techniques for complex stratigraphic frameworks [1].

Another important geological interpretation from Figure 2 involves the identification of channel stacking and possible stratigraphic traps. The vertical and lateral continuity of low-impedance, high-porosity bodies suggests multi-generational channel systems, indicating phases of deposition followed by abandonment and avulsion, typical in wave- and tide-influenced deltaic environments [18]. These repeated depositional events can lead to vertically stacked reservoir units that are favorable not only for hydrocarbon accumulation but also for CO_2 storage layering, offering potential for plume stratification and enhanced storage security. The spatial separation of impedance lows, juxtaposed with sealing facies, may indicate locations with high reservoir potential that warrant further evaluation. This ability to delineate reservoir geometry and potential flow barriers from seismic-derived attributes is crucial in early exploration and development planning [3], and particularly essential when evaluating a formation's capacity to act as a secure CO_2 sink.

In addition to these observations, the consistent alignment between the geometry of impedance anomalies and expected depositional patterns in deltaic regimes highlights the model's capability in preserving geological realism. Subtle textural differences, which may correspond to grain-size variations or diagenetic overprints, appear to be captured in the AI and porosity maps—an encouraging outcome given the limitations of post-stack data. From a CCS perspective, such fidelity is critical for modeling CO_2 injectivity and assessing the extent of heterogeneities that could influence plume dynamics. Moreover, these attribute-based predictions complement structural interpretation by revealing gentle deformation zones or paleotopographic lows where sediments preferentially accumulated, possibly forming natural traps for injected CO_2. This comprehensive view reinforces the value of multi-attribute and machine learning workflows in reducing interpretational ambiguity and providing a data-driven lens into the depositional history of complex clastic systems such as the F3 Block—while also informing key risk and capacity metrics relevant to carbon storage.

Porosity Correlation

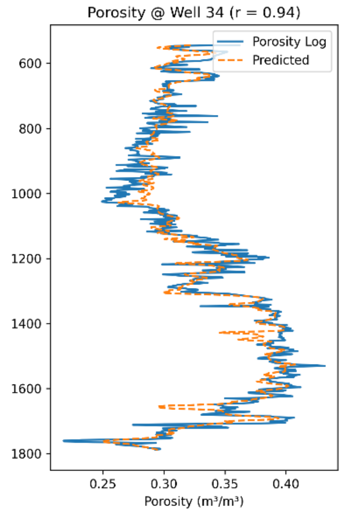

Figure 3. Porosity Correlation Well 34

Figure 3 shows the correlation between predicted and log-derived porosity values at Well 34, demonstrating a strong linear agreement throughout the interval. This high correlation is attributed to the presence of thick, clean sandstones that dominate the lithology around this well location, which typically show strong and predictable relationships with seismic attributes [2]. From a CCS standpoint, such clean and laterally continuous sandstone bodies are highly desirable injection targets, as they offer high porosity, permeability, and volume capacity needed for efficient and secure CO_2 sequestration. The accuracy of the Random Forest model in capturing these variations validates the suitability of machine learning for porosity inversion in well-calibrated and geologically homogeneous settings [1]. This performance is further enhanced by the use of carefully selected seismic attributes, such as RMS amplitude and envelope, which are sensitive to lithological and fluid contrasts [18]—making them particularly effective in delineating high-capacity reservoir zones for potential CO_2 storage.

The reliable prediction at Well 34 also highlights the benefits of having high-quality training data and minimal geological noise. These conditions mirror ideal CCS injection scenarios, where reservoir predictability directly correlates with long-term storage integrity and plume migration control. The clean signal and simple facies distribution allow the Random Forest algorithm to effectively generalize and extrapolate porosity patterns across the seismic section [25]. In deltaic environments like F3, such conditions are often encountered within distributary channels that are laterally extensive and have minimal shale intercalations, reducing potential barriers to injectivity and improving storage uniformity. Therefore, the results at Well 34 serve not only as a reference case for evaluating inversion performance across the field, but also as a conceptual model for identifying prime CO_2 injection candidates in similar clastic reservoirs.

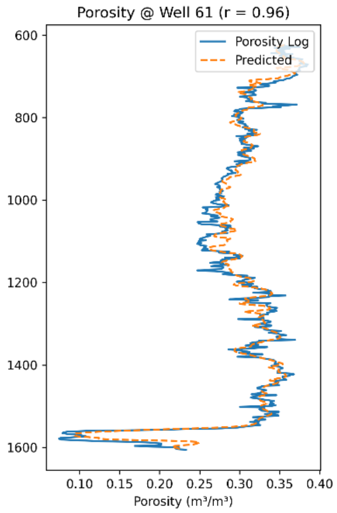

Figure 4. Porosity Correlation Well 61

In contrast to Well 34, the porosity correlation at Well 61 exhibits greater scatter, particularly at shallower and deeper depths. This increased dispersion is likely caused by complex interbedding of sand and shale layers, which introduce seismic interference and reduce the fidelity of attribute-based predictions [20]. The facies at this location may include more heterolithic or distal deposits, such as levee or overbank sands, which complicate the relationship between porosity and seismic response [18]. From a CCS perspective, such variability introduces greater uncertainty in assessing both reservoir capacity and injectivity, as the presence of interbedded shale may hinder vertical plume migration and affect long-term storage behavior. Nevertheless, the overall trend of the predicted values still aligns with the general porosity profile, indicating the model's resilience [1].

The Random Forest model, despite encountering more complex geological variability, was able to maintain a reasonable prediction across the well interval. This highlights the importance of ensemble methods in handling non-linear and noisy data environments [25], particularly when evaluating candidate storage sites where heterogeneity may obscure capacity estimation. The performance at Well 61 emphasizes the need for improved attribute selection or local model tuning when applying machine learning in geologically heterogeneous zones. Nonetheless, the result remains within acceptable limits and offers a useful approximation for initial reservoir characterization, especially in early-stage screening of CCS storage intervals where rapid assessments are required.

AI Correlation

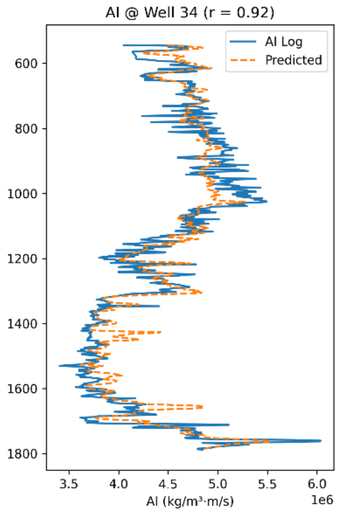

Figure 5. AI Correlation Well 34

The acoustic impedance (AI) correlation at Well 34 presents a nearly linear relationship between predicted and actual log values, especially in intervals dominated by clean sandstone. These sandstones exhibit low density and velocity, resulting in low AI values that are easier to capture using seismic attributes such as amplitude and frequency-derived measures [1]. The high correlation reflects the Random Forest model's strong capacity to learn attribute–impedance relationships in geologically straightforward contexts [25]. In deltaic deposits, such homogeneity is often encountered in the axis of distributary channels [20]. For the purpose of CCS, this level of predictability becomes vital, as such sand-prone units can act as effective CO_2 storage reservoirs due to their high injectivity and predictable sealing boundaries.

Moreover, the absence of structural complexities and thin layering at this well location minimizes distortion of the seismic signal, leading to high-quality input features (Mazloum, 2020). These favorable geological conditions enable accurate prediction and demonstrate the strength of machine learning approaches in enhancing post-stack impedance inversion. From a carbon storage standpoint, such well-behaved reservoir units offer a baseline for evaluating seal–reservoir pairs, enabling early-stage estimation of storage volume and injectivity. The consistent results at Well 34 reinforce the reliability of the model in well-characterized zones and provide a benchmark for validation in assessing storage site quality across the broader CCS field deployment area.

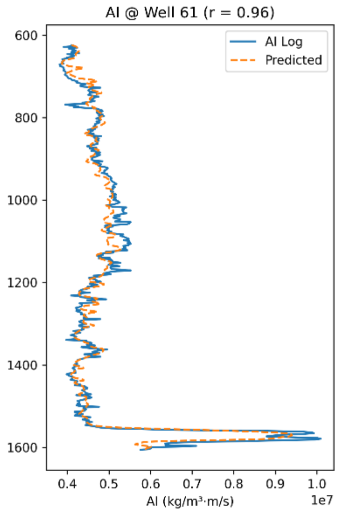

Figure 6. AI Correlation Well 61

Figure 6 reveals that the AI prediction at Well 61 is moderately correlated with the actual log values but shows a broader spread than that observed at Well 34. This deviation suggests that Well 61 is located in a more geologically complex setting, possibly near facies transitions or thin interbedded sequences [18]. In such environments, the seismic response becomes less distinct, and the attributes carry mixed signals that obscure true impedance contrasts [20]. Despite these complications, the overall AI trend remains intact and within an acceptable error range (Zou et al., 2021). These conditions—marked by heterogeneity and subtle stratigraphic mixing—are often found in the marginal or distal portions of potential CO_2 storage reservoirs, where injectivity and seal effectiveness must be assessed with higher scrutiny.

The capability of the Random Forest algorithm to capture impedance variations even in noisy zones speaks to its robustness and adaptability. However, this figure also illustrates the limitations of applying a uniform model across varied facies without localized adjustments [25]. For CCS site characterization, this underscores the necessity of tailoring inversion workflows based on depositional environment and seal–reservoir architecture to improve predictions of caprock integrity and injection capacity. The moderate correlation at Well 61 suggests that future models may benefit from stratified training or facies-dependent attribute weighting to refine impedance estimation in more variable lithologies. Still, the results demonstrate the potential of data-driven inversion techniques in extracting valuable impedance information from post-stack seismic data, especially in support of CCS screening where accurate property distribution is crucial for risk reduction.

CONCLUSION AND RECOMMENDATION

Conclusion

This study underscores the value of Random Forest (RF)-based multi-attribute seismic inversion as a powerful and interpretable machine learning approach for predicting key petrophysical properties—specifically porosity and acoustic impedance (AI)—with direct relevance to CCS site evaluation. By leveraging ten seismic attributes that capture a broad range of structural and stratigraphic information, the RF model was able to generate high-resolution 3D property volumes across the F3 Block. These outputs are not only consistent with well-log data but also preserve geological continuity, such as channel geometries and lithological transitions, that are critical in determining both reservoir capacity and seal performance for CO_2 storage.

The accuracy observed at Well 34 reflects the model's strong generalization in clean sandstone-dominated settings with minimal heterogeneity—indicating high confidence for predicting reservoir quality in primary injection targets. Meanwhile, the model's resilience at Well 61, despite lithological noise and facies complexity, highlights its adaptability across less favorable CCS contexts, where uncertainties often impede decision-making. Importantly, the AI results also delineate potential caprock intervals and baffle zones that could restrict vertical migration, thereby informing both reservoir storage efficiency and containment integrity assessments.

Furthermore, the use of ensemble learning allows for both prediction and uncertainty estimation, making RF particularly advantageous for pre-injection risk analysis and regulatory reporting. When integrated with conventional static modeling, RF-based inversion offers a scalable and transferable method to screen large datasets for CO_2 storage suitability. As the global energy transition increasingly prioritizes decarbonization, such tools will be essential for accelerating the safe, effective, and scientifically credible deployment of CCS technologies.

Recommendation

Drawing upon the results obtained and the constraints identified throughout this investigation, several forward-looking suggestions are put forth to inform both subsequent academic inquiries and prospective practical implementations:

a. Incorporate additional borehole data, core measurements, and reservoir pressure information to enhance validation accuracy and extend model generalizability across the entire storage formation.
b. Couple the inverted porosity and impedance outputs with stress modeling and mineralogical sensitivity analyses to assess caprock sealing behavior under CO_2 plume migration and storage conditions.
c. Divide training data based on stratigraphic or depositional facies to improve model resolution in heterogeneous or thinly interbedded zones often present in CO_2 storage complexes.

REFERENCES

[1] C. Zou, L. Zhao, M. Xu, Y. Chen, and J. Geng, "Porosity prediction with uncertainty quantification from multiple seismic attributes using random forest," *J. Geophys. Res. Solid Earth*, vol. 126, no. 7, p. e2021JB021826, Jul. 2021.

[2] T. Topór and K. Sowiżdżał, "Application of machine learning tools for seismic reservoir characterization study of porosity and saturation type," *Nafta-Gaz*, vol. 78, no. 3, pp. 165–175, Mar. 2022.

[3] M. R. Safari, K. Taheri, H. Hashemi, and A. Hadadi, "Structural smoothing on mixed instantaneous phase energy for automatic fault and horizon picking: Case study on F3 North Sea," *J. Pet. Explor. Prod. Technol.*, vol. 13, pp. 775–785, 2023.

[4] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Science, vol. 521, no. 7553, pp. 436–444, 2015.

[6] K. Anjom, H. A. Hafez, and G. Khalaf, "Machine learning and AVO Class II workflow for hydrocarbon exploration," Scientific Reports, vol. 13, Art. no. 4509, 2023.

[7] W. Wu, R. Zhu, Y. Zhou, and Z. Zhang, "Joint interpretation of geophysical data: Applying machine learning to combine seismic velocities, electrical resistivity, well-logs, and lithological columns," Physics of the Earth and Planetary Interiors, vol. 312, 2021.

[8] A. Karpatne, G. Atluri, J. H. Faghmous, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2318–2331, 2017.

[9] X. Hu, J. Li, Y. Wang, and B. Zhang, "A novel seismic inversion method based on multiple attributes and machine learning for hydrocarbon reservoir prediction in Bohai Bay Basin, Eastern China," Front. Earth Sci., vol. 12, 2024.

[10] A. Sinha, A. Tiwari, J. F. Shragge, and M. M. Haney, "3D microseismic monitoring using machine learning," J. Geophys. Res. Solid Earth, vol. 125, no. 3, 2020.

[11] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32.

[12] A. B. Kassa and M. T. Dugda, "Implementation of Machine Learning Algorithms for Seismic Events Classification," arXiv preprint arXiv:2502.05197, 2024.

[13] S. S. Shukla et al., "An Ensemble Random Forest Model for Seismic Energy Forecast," Nat. Hazards Earth Syst. Sci. Discuss., preprint, 2024.

[14] C. Puryear, R. Tharimela, and V. Egorov, "Spectral Extrapolation and Random Forest for High Resolution Prediction of Subsurface Properties," in SEG Int. Mtg. Appl. Geoscience & Energy, 2021.

[15] S. S. Shukla, J. Dhanya, P. Kumar, P. Priyanka, and V. Dutt, "An Ensemble Random Forest Model for Seismic Energy Forecast," Nat. Hazards Earth Syst. Sci. Discuss., preprint, 2024.

[16] P. Domel, C. Hibert, V. Schlindwein, and A. Plaza-Faverola, "Event Recognition in Marine Seismological Data Using Random Forest Machine Learning Classifier," Geophys. J. Int., vol. 235, pp. 589–609, 2023.

[17] L. Izquierdo-Horna, J. Zevallos, and Y. Yepez, "An Integrated Approach to Seismic Risk Assessment Using Random Forest and Hierarchical Analysis: Pisco, Peru," Heliyon, vol. 8, e10926, 2022.

[18] D. Mazloum, "Advanced Seismic Reservoir Characterization of the F3 Block," M2 Thesis, Université Grenoble Alpes, 2024.

[19] M. A. Ishak, M. A. Islam, M. R. Shalaby, and N. Hasan, "The Application of Seismic Attributes and Wheeler Transformations for the Geomorphological Interpretation of Stratigraphic Surfaces: A Case Study of the F3 Block, Dutch Offshore Sector, North Sea," Geosciences, vol. 8, no. 3, p. 79, 2018.

[20] E. Kabaca, "Seismic Stratigraphic Analysis Using Multiple Seismic Attributes: A Case Study from the F3 Block, North Sea," M.S. Thesis, University of Oklahoma, 2019.

[21] A. M. El-Din, M. A. Abdelrahman, and M. M. El-Gohary, "Unsupervised machine learning-based multi-attributes analysis for gas channel detection and seismic facies classification," Arabian J. of Geosciences, vol. 17, 2024.

[22] C. Zou, L. Zhao, M. Xu, Y. Chen, and J. Geng, "Porosity prediction with uncertainty quantification from multiple seismic attributes using Random Forest," J. Geophys. Res. Solid Earth, vol. 126, no. 6, 2021.

[23] H. H. Hassan, K. El-Maghawry, and W. M. Hassan, "Model-Based Inversion in North Sea F3-Block Dutch Sector," in Proc. of the 6th Int. Conf. on Petroleum Engineering, 2015.

[24] Current Science Editorial Board, "Fracture characterization using seismic attributes and machine learning," Current Science, vol. 122, no. 5, pp. 623–629, 2022.

[25] P. K. Kushwaha, S. P. Maurya, P. Rai, and N. P. Singh, "Porosity prediction from offshore seismic data of F3 Block, the Netherlands using multi-layer feed-forward neural network," Current Science, vol. 119, no. 10, pp. 1652–1662, Nov. 2020.

[26] H. Jo, D. Lee, and S. Park, "Porosity prediction from prestack seismic data via deep learning," arXiv preprint arXiv:2111.13581, 2021.

[27] M. Mojeddifar, M. K. Ilkhchi, and S. Rezaee, "Porosity prediction from seismic inversion of a similarity attribute using a pseudo-forward equation," Petroleum Science, vol. 12, no. 4, pp. 665–675, Dec. 2015.