applied sciences

MDPI

Article

Sandbody Prediction Based on Fusion of Seismic Multi-Attributes and Machine Learning Under Sedimentary Facies Constraint—A Case Study of Chenguanzhuang Area in Dongying Depression, Bohai Bay Basin

Jinshuai Liu [1,2,3,*], Chengyan Lin [1,2,3,*], Chris Elders [4] and Azhari Faris [5]

[1] National Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao 266580, China
[2] Shandong Provincial Key Laboratory of Reservoir Geology, Qingdao 266580, China
[3] School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China
[4] School of Earth and Planetary Sciences, Curtin University, Perth, WA 6102, Australia
[5] School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, QLD 4072, Australia
* Correspondence: b18010017@s.upc.edu.cn (J.L.); lincy@upc.edu.cn (C.L.)

Abstract

In complex sedimentary environments, the identification of thin sandbodies and the accurate prediction of their thickness remain challenging, particularly when relying on a single analytical approach. Taking the lower sub-member of the fourth member of the Shahejie Formation (Es_4^L) in the Chenguanzhuang area of the Dongying Depression as a case study, this study proposes a quantitative prediction method that integrates sedimentary facies constraints with machine learning-based seismic multi-attribute fusion. Based on core observations, well log data, and 3D seismic datasets, the study area is subdivided into two zones: Zone I (shallow-water delta front) and Zone II (shore–shallow lake). Sensitive attributes for each zone are optimized using Pearson correlation analysis and hierarchical clustering, and five machine learning models—SVR, Random Forest, MLP, Ridge Regression, and Lasso Regression—are systematically evaluated. The MLP model is selected for Zone I, achieving R^2 values of 0.856 and 0.936 for the training and test sets, respectively, whereas Ridge Regression combined with leave-one-out cross-validation (LOOCV) is adopted for Zone II to mitigate overfitting caused by limited well data, yielding R^2 values of 0.864 and 0.779. Compared with conventional linear regression (R^2 = 0.45), the proposed approach significantly improves the accuracy of quantitative sandbody prediction, providing a reliable geological basis for hydrocarbon exploration and an effective technical framework for similar complex sedimentary environments.

Keywords: Dongying Depression; sedimentary facies; seismic multi-attributes; machine learning; multilayer perceptron (MLP); Ridge Regression; sandbody prediction

Check for updates

Academic Editor: Nikolaos Koukouzas

Received: 5 March 2026
Revised: 23 March 2026
Accepted: 24 March 2026
Published: 30 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Seismic interpretation has long relied on seismic attributes to delineate reservoirs' geometry, lithology, and key petrophysical properties, demonstrating considerable practical effectiveness [1]. However, predictions based on single attributes are often ambiguous due to limitations in seismic data quality and resolution, particularly when targeting thin or

heterogeneous reservoirs. Consequently, multi-attribute analysis has emerged as a more reliable approach for reducing non-uniqueness and improving predictions' robustness.

In recent years, machine learning (ML) has been widely applied in engineering and geoscience fields for data-driven modeling, offering improved objectivity and efficiency while reducing reliance on expert judgment. Nevertheless, domain knowledge remains essential to ensure that ML models are physically consistent and geologically interpretable [2,3]. In petroleum exploration, ML has been increasingly employed for reservoir characterization, particularly to capture nonlinear relationships between seismic attributes and reservoir properties—an advantage over conventional linear methods [4,5]. It has also been successfully applied to predict reservoir properties using multi-dimensional seismic attributes [6–11]. Shallow-water delta systems represent important hydrocarbon exploration targets; however, their complex depositional processes and highly variable sandbody distribution pose significant challenges for accurate characterization. Existing sandstone prediction methods, including seismic sedimentology [12–14], multi-attribute analysis [15,16], seismic facies classification [17], acoustic impedance inversion [18,19], and machine learning approaches [20,21], provide effective tools for sandbody characterization [22,23], yet individual methods often lack sufficient accuracy and stability in complex or thin interbedded reservoirs. Therefore, the integration of geological, logging, and seismic data is essential for reliable sandstone identification and improved reservoir characterization, ultimately supporting optimized well placement and enhanced recovery efficiency.

To further improve sandstone reservoir prediction, this study focuses on the fourth member of the Shahejie Formation (Es_4) of the Early Eocene in the Dongying Depression and quantitatively predicts sandbody distribution. Based on core observations, well log data, and seismic datasets, shallow-water delta front deposits with well-developed subaqueous distributary channels are identified in the study area. Seismic sedimentology is applied to characterize the planar distribution of sedimentary facies, followed by zonal analysis and the optimization of seismic attributes according to facies types to identify attributes closely related to sandstone distribution. Under sedimentary facies constraints, a machine learning-based seismic multi-attribute fusion approach is employed to predict sandstone distribution and refine the spatial characterization of sandbodies. By integrating facies-controlled zonation with multi-attribute fusion, the proposed method demonstrates strong applicability in complex geological settings, significantly improving predictions' accuracy and providing a reliable basis for hydrocarbon exploration and development.

2. Geological Setting

The Bohai Bay Basin is one of the most important petroliferous basins in China. It is a large Mesozoic–Cenozoic continental rift basin located along the eastern coast of China, covering an area of approximately 200,000 km^2 [24]. The basin comprises seven sub-basins (Figure 1A), among which the Jiyang sub-basin includes the Dongying, Zhanhua, Chenzhen and Huimin depressions. The Dongying Depression is situated in the southern part of the Jiyang sub-basin (Figure 1A) and is bounded by the Qingtuozi uplift to the east, the Guangrao uplift to the southeast, the Luxi uplift to the south, the Gaoqing uplift to the west, and the Chenguanzhuang uplift to the north (Figure 1B). Structurally, the Dongying Depression can be divided into five tectonic zones from north to south: the northern steep slope zone, northern sag zone, central anticline zone, southern sag zone, and southern gentle slope zone (Figure 1C), forming a structural framework characterized by a steep fault-controlled margin in the north and a gently dipping stratigraphic overlap in the south [25]. The Dongying Depression represents one of the most hydrocarbon-rich depressions in the Bohai Bay Basin.

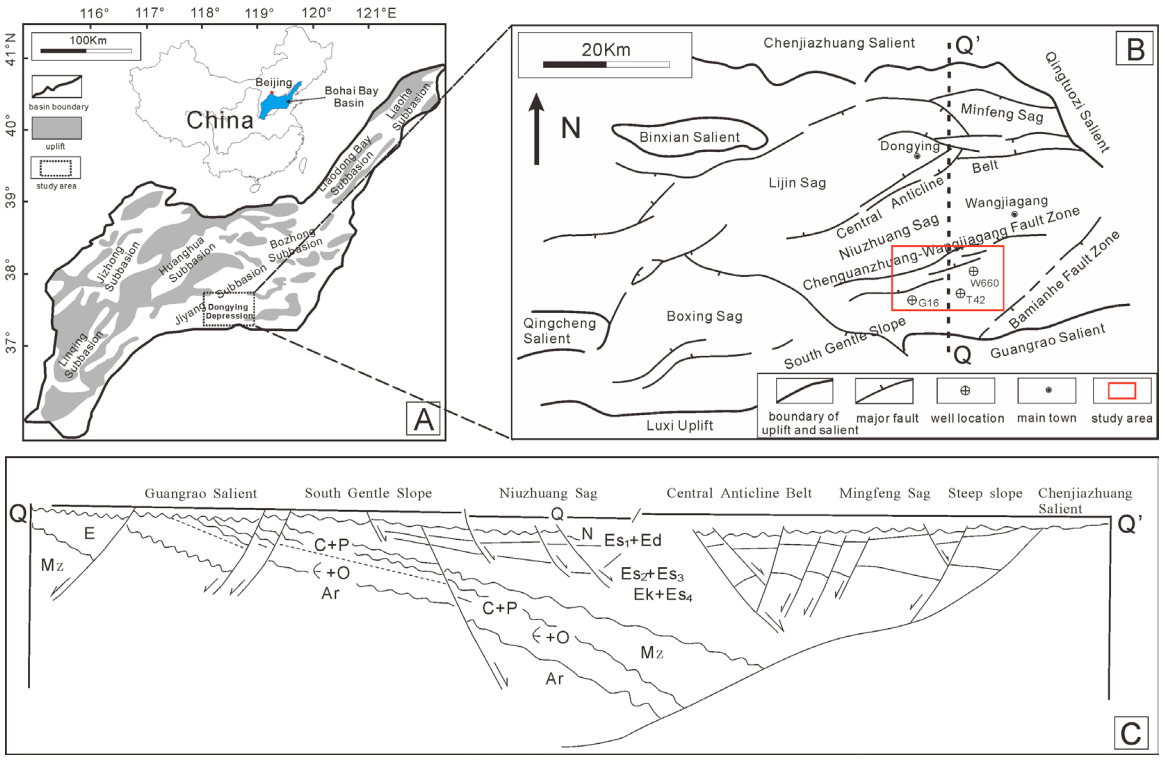


Figure 1. (**A**) Location map of the Bohai Bay Basin subtectonic units. (**B**) Structural map of the Dongying Depression; the study area is in the solid red box. (**C**) Simplified cross section (N-S) showing the internal structure and stratigraphic features (modified from Feng et al. [25]).

The tectonic evolution of the Dongying Depression during the Mesozoic–Cenozoic can be divided into two main stages: the Paleogene syn-rift stage (65.0–24.6 Ma) and the Neogene post-rift stage (24.6 Ma to present) [26]. The depression is filled with Cenozoic sediments, including the Paleogene Kongdian (Ek), Shahejie (Es), and Dongying (Ed) Formations, the Neogene Guantao (Ng) and Minghuazhen (Nm) Formations, and the Quaternary Pingyuan (Qp) Formation (Figure 2). Influenced by the Himalayan tectonic movement, these strata record three evolutionary stages of rifting: the initial rift stage (Ek and Es_4), the main rift stage (Es_3 and Es_2^L), and the late rift stage (Es_2^U, Es_1, and Ed). Together, these stages constitute a complete syn-rift sedimentary sequence, reflecting a depositional cycle from shore–shallow lake facies to deep lake facies and subsequently returning to shore–shallow lake facies [27].

The fourth member of the Shahejie Formation (Es_4) constitutes one of the principal hydrocarbon-bearing intervals in the Dongying Depression. Based on stratigraphic characteristics, Es_4 can be subdivided into an upper sub-member (Es_4^U) and a lower sub-member (Es_4^L). The deposition of this unit is strongly influenced by paleogeomorphology and lake-level fluctuations, resulting in the development of thin sandbodies. During the deposition of Es_4^L, a relatively arid paleoclimate led to elevated lake salinity, the formation of a large-scale flood delta along the southern gentle slope, and the accumulation of thick gypsum layers in the lacustrine center [27]. In contrast, during the deposition of Es_4^U, the paleoclimate gradually shifted from arid to humid, leading to decreased lake salinity and the development of extensive beach-bar sandbodies on the southern gentle slope, along with thick semi-deep to deep lacustrine mudstones and hydrocarbon source rocks in the basin's center [27].

The Chenguanzhuang area is located in the eastern part of the southern gentle slope of the Dongying Depression, between the Chenguanzhuang and Wangjiagang faults. The regional structure dips northward, and strata thicken progressively from south to north in a

step-like pattern controlled by faulting. This study focuses on the lower sub-member of Es_4 (Es_4^L), which can be further divided into two sandstone intervals, namely Es_4^{L-1} and Es_4^{L-2}, based on sedimentary characteristics and well log responses (Figure 2). The target interval of this study is Es_4^{L-1}, which consists predominantly of red beds composed of mudstone interbedded with thin sandstone layers, with individual sandstone thicknesses ranging from 0.5 to 18 m. The total sandstone thickness and net-to-gross ratio decrease from the lake margin toward the basin center. Because many of these thin sandstones fall below seismic resolution, a facies-constrained seismic multi-attribute fusion approach combined with machine learning is employed to predict sandstone thickness, thereby improving the prediction accuracy in complex geological settings.

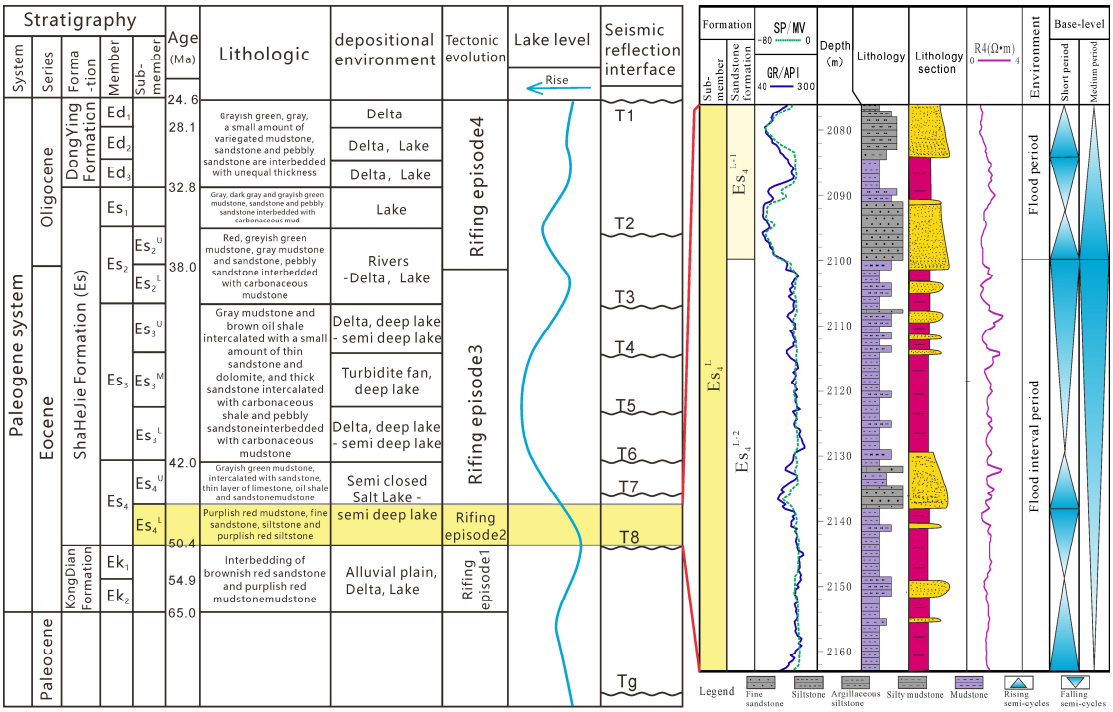


Figure 2. Stratigraphic column of the study area (modified from Feng et al. [25]).

3. Materials and Methodology

The dataset used in this study comprises core samples, well log data from 31 wells, and three-dimensional (3D) seismic data. The seismic and well log datasets were provided by the Geophysical Exploration Institute of Shengli Oilfield Company, Sinopec in Dongying City, China. The seismic dataset covers an area of approximately 210 km^2 and is displayed in normal polarity according to Society of Exploration Geophysicists (SEG) standards (SEG Polarity 1975; Special Report of the Subcommittee on Polarity Standards. Society of Exploration Geophysicists: Houston, TX, USA, 1975SEG) [28]. The dominant frequency of the Es_4 interval ranges from 24 to 30 Hz, with an estimated vertical resolution of approximately 20 m. The inline and crossline spacings are 12.5 m, and the vertical sampling interval is 1 ms. The available well log data include acoustic, compensated neutron, caliper, spontaneous potential, natural gamma ray, density, resistivity, and lithological information.

This study focuses on the fourth member of the Shahejie Formation (Es_4) of the Early Eocene in the Dongying Depression. A comprehensive workflow for sandbody prediction is established under sedimentary facies constraints, integrating seismic multi-attribute fusion and machine learning. The overall workflow follows a structured framework: data preprocessing, geological constraint, seismic attribute optimization, and machine learning-based modeling, prediction, and validation.

(1) Data preparation and seismic data preprocessing

Mixed-phase wavelet deconvolution is employed to enhance the frequency bandwidth of the seismic data, thereby improving both the dominant frequency and resolution. Diffusion filtering is subsequently applied to suppress noise and enhance the signal-to-noise ratio, achieving a balance between resolution and data fidelity.

(2) Sedimentary facies analysis and facies zonation

Sedimentary facies are identified using 90° phase rotation and stratal slicing techniques within the framework of seismic sedimentology. Based on depositional characteristics, the study area is divided into two facies zones: the shallow-water delta front (Zone I) and the shore–shallow lake (Zone II), providing the geological constraints for subsequent zoned modeling.

(3) Seismic attribute extraction and facies-constrained optimization.

A total of 30 seismic attributes related to amplitude, frequency, phase, and waveform are extracted. Under the facies constraints, sensitive attributes are optimized using Pearson correlation analysis and hierarchical clustering to reduce redundancy and improve attributes' independence. Five sensitive attributes are selected for Zone I, and two for Zone II.

(4) Machine learning modeling, prediction and accuracy validation.

In the data-rich Zone I, five machine learning algorithms (Random Forest, Lasso Regression, SVR, Ridge Regression, and MLP) are compared, and the MLP model is selected due to its superior fitting performance and generalization capability. In the data-limited Zone II, Ridge Regression combined with leave-one-out cross-validation (LOOCV) is adopted to mitigate overfitting. Model performance is evaluated using R^2, MAE, and RMSE, as well as blind well validation. Sandbody thickness is quantitatively predicted across the study area, and the results are validated against sedimentary geological constraints.

3.1. Pre-Conditioning of the Seismic Data

Seismic data quality plays a critical role in the identification of seismic reflection interfaces, seismic attribute interpretation, and reservoir characterization. The sandbodies in the Es_4^{L-1} interval within the study area are thin and deeply buried, with depths ranging from 1300 to 3500 m, corresponding to medium–deep reservoirs. In addition, the seismic data are characterized by low dominant frequency and weak reflection coefficients, making the identification of thin sandbodies challenging. Therefore, enhancing seismic data quality and improving resolution are essential for accurate reservoir characterization.

3.1.1. Frequency Bandwidth Enhancement

Mixed-phase wavelet deconvolution is employed to enhance the frequency bandwidth of the seismic data. The key procedure involves wavelet extraction and the definition of an optimal output wavelet [29]. The input wavelet is derived from the original seismic data, and an appropriate output wavelet type is selected. The amplitude spectrum of the input wavelet is adjusted with reference to the signal-to-noise ratio spectrum, while well log data are used for calibration to achieve an optimal balance between resolution and fidelity. This method effectively broadens the seismic frequency spectrum while preserving the data's fidelity. In addition, it drives the seismic signal toward a near-zero-phase condition, thereby improving vertical the resolution and enhancing the detectability of thin-layer features [30,31].

3.1.2. Diffusion Filtering Technology

Seismic data typically contain both random noise and coherent noise, such as acquisition footprints, which can significantly affect seismic attribute interpretation and lead to erroneous results. Although frequency enhancement improves resolution, it may also reduce the signal-to-noise ratio. To address this issue, diffusion filtering is applied to the frequency-enhanced seismic data to suppress noise and improve the signal-to-noise ratio. This method effectively balances noise reduction and structural preservation, enhancing the lateral continuity and improving the imaging quality of internal stratigraphic features [32]. The effectiveness of diffusion filtering is controlled by three key parameters: the data scale factor, diffusion coefficient, and number of iterations. Proper parameter selection ensures the optimal filtering performance while preserving geological features.

3.2. Seismic Sedimentology Techniques

Seismic sedimentology is an interdisciplinary approach that integrates seismic data with sedimentological analysis to investigate depositional processes and sedimentary architectures. It builds upon seismic stratigraphy and sequence stratigraphy, focusing on both model-driven seismic geomorphology and data-driven lithological interpretation [29,33]. With advances in 3D seismic data acquisition, processing, and interpretation, seismic sedimentology has become an essential tool in basin analysis, structural interpretation, hydrocarbon accumulation studies, and reservoir characterization.

3.2.1. Seismic Phase Adjustment

To improve the characterization of thin-bed lithology, a 90° (or -90°) phase rotation is applied to zero-phase seismic data. This transformation aligns the main reflection lobe with the center of the geological layer, allowing reflection events to represent lithological units directly rather than their bounding interfaces. Following phase adjustment, reflection troughs correspond to sandstone units, whereas reflection peaks correspond to mudstone intervals. Consequently, the seismic phase more accurately reflects internal lithological variations within a thickness of approximately one wavelength [34,35], providing a more direct lithological interpretation of thin sandbodies.

3.2.2. Stratal Slicing

For sedimentary bodies whose thickness is significantly smaller than their lateral extent, interpretation based solely on vertical seismic sections is often insufficient. Stratal slicing provides an effective approach for analyzing such thin-layered deposits [36,37]. Unlike conventional time slicing or horizon slicing, stratal slicing compensates for lateral variations in depositional rates and produces slices that more accurately represent isochronous sedimentary surfaces. This enables a more reliable reconstruction of paleogeomorphology and depositional evolution. In this study, the top and base of the target interval (Es_4^{L-1}) are used as reference horizons to generate forty proportionally interpolated stratal slices. Representative slices are selected for geological interpretation to characterize the planar distribution of sedimentary bodies.

3.3. Seismic Attributes Extraction and Standardization

Seismic attributes are derived through mathematical transformations of fundamental seismic data components, including time, amplitude, phase, frequency, and waveform. These attributes vary significantly as seismic waves propagate through different reflection interfaces and stratigraphic layers [38]. Variations in seismic waveforms are closely related to changes in physical properties, particularly acoustic impedance, thereby enabling the establishment of quantitative relationships between seismic attributes and lithological

interpretations derived from well log data. Strong correlations between seismic attributes and sandbody thickness provide a basis for the reliable prediction and delineation of sandbodies. More than 300 seismic attributes can be computed, of which approximately 50–60 are commonly used in practical applications. Therefore, it is essential to select optimal attributes that are sensitive to sandstone characteristics to improve prediction accuracy.

Seismic attribute values often exhibit a wide dynamic range, particularly in the form of high-amplitude variations, which can complicate interpretation. Such variability may obscure subtle features, mask local anomalies within background trends, and introduce outliers. To address this issue, min–max normalization is applied to standardize seismic attribute values, as defined in Equation (1). For a given attribute X_i, normalization (X_{nor}) is performed by subtracting the minimum value (X_{min}) and dividing by the range between the maximum (X_{max}) and minimum values. This process is applied independently to each attribute, resulting in normalized values ranging from zero to one.

$$X_{nor} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

3.4. Clustering Analysis

Clustering algorithms are fundamental unsupervised learning techniques used to identify latent structures and correlation patterns within datasets by grouping samples with similar characteristics. The objective is to maximize intra-cluster similarity while maximizing inter-cluster differences. Similarity between samples is typically measured using distance metrics or similarity indices across multiple feature dimensions. Based on their principles, clustering algorithms can be categorized into partition-based, model-based, hierarchical, density-based, graph-based, and grid-based approaches [39].

Hierarchical clustering, as a classical unsupervised learning method, organizes data into a tree-like structure without requiring a predefined number of clusters. It operates through iterative merging or splitting based on sample similarity [40]. Two main strategies are commonly used: agglomerative clustering, which follows a bottom-up approach by successively merging the most similar clusters, and divisive clustering, which adopts a top-down approach by recursively partitioning the dataset into dissimilar subsets [41].

Hierarchical clustering offers several advantages. It does not require a predefined number of clusters and can adaptively reveal data structure based on inherent distribution patterns. In addition, it provides a visual representation in the form of a dendrogram, facilitating the intuitive interpretation of the relationships among samples. Its sensitivity to outliers can also be reduced through the appropriate selection of linkage criteria and distance metrics.

3.5. Machine Learning Methods for Sandbodies Prediction

All machine learning modeling and seismic attribute analyses in this study were implemented in Python 3.8.18 using the scikit-learn framework. When constructing regression models between seismic multi-attributes and sandbody thickness, both the input attributes and target variables were normalized using Equation (1). To prevent data leakage, the normalization parameters (maximum and minimum values) were computed exclusively from the training dataset and subsequently applied to the test set and independent blind well samples. No information from the validation or test data was used during normalization, ensuring the rigor and objectivity of each model's evaluation.

Five widely used machine learning algorithms—Random Forest, Lasso Regression, Support Vector Regression (SVR), Ridge Regression, and Multilayer Perceptron (MLP)—were employed to model the relationship between seismic attributes and sandbody thickness. These algorithms have been extensively validated in petroleum geoscience applica-

tions. Comparative modeling was conducted to evaluate their predictive performance, and MLP and Ridge Regression were ultimately selected for the quantitative prediction. These two models are described in detail in the following sections.

3.5.1. Multilayer Perceptron (MLP) Artificial Neural Networks

Multilayer Perceptron (MLP) is a classical feedforward artificial neural network that has been widely applied since the 1980s [42]. It consists of at least three layers: an input layer, one or more hidden layers, and an output layer [43]. Neurons are fully connected between adjacent layers, while no connections exist within the same layer. The input layer receives feature vectors, and the hidden layers transform low-level features into higher-level representations through weighted summation, bias terms, and nonlinear activation functions [44], enabling the network to learn complex nonlinear relationships. Common activation functions, such as ReLU, Sigmoid, and Tanh, introduce nonlinearity into the model. The output layer produces the final prediction results [45].

Due to its strong generalization capability, MLP can effectively model complex nonlinear mappings and maintain a high prediction accuracy even under limited sample conditions. It has been widely applied in classification, regression, speech recognition, and other machine learning tasks [46].

3.5.2. Ridge Regression

Ordinary linear regression is highly sensitive to multicollinearity, which may lead to unstable coefficient estimation, overfitting, and poor generalization performance. Ridge Regression (L2 regularization) addresses this issue by introducing an L2 penalty term to constrain the model's coefficients, thereby stabilizing the parameter estimation and reducing the impact of multicollinearity [47].

As an effective regularization technique, Ridge Regression is particularly suitable for modeling datasets with correlated and high-dimensional features. Its performance depends on the appropriate selection of the regularization parameter λ, which balances bias and variance [48]. Ridge Regression has been widely applied in fields such as econometrics, chemometrics, engineering, and energy systems, and remains a commonly used linear modeling approach [49].

3.6. Evaluation Metric

3.6.1. Sensitivity Analysis

To effectively reduce model uncertainty, it is necessary to select input variables that have significant influence on the target variable. In this study, the Pearson correlation coefficient is used for sensitivity analysis. The linear correlation among variables can be defined by the Pearson correlation coefficient (2), which ranges from −1 to +1 within the Pearson correlation matrix and is specifically used to measure the linear relationship between two variables: a negative coefficient value indicates a negative linear relationship between variables, a positive value denotes a positive linear relationship, and a value of zero implies that there is no linear relationship between variables where r is defined as follows:

$$r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})\Big)}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}} \quad (2)$$

where N is the number of data points, x_i and y_i are the sample paired data, and $\overline{x}$ and $\overline{y}$ designate the mean of x_i and y_i, respectively.

3.6.2. Model Performance Evaluation

To comprehensively evaluate the robustness and predictive performance of the machine learning models, several commonly used evaluation metrics are adopted, including the coefficient of determination (R^2), absolute error (AE), relative error (RE), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). These metrics provide a quantitative assessment of a model's accuracy and generalization capability.

4. Results and Discussion

4.1. Results of Seismic Data Preprocessing

After applying mixed-phase wavelet deconvolution for frequency bandwidth enhancement, the quality of the seismic data is significantly improved. The dominant frequency increases from 26 Hz to 34 Hz, and the frequency bandwidth is correspondingly expanded, with the absolute bandwidth increasing from 44 Hz to 60 Hz (Figure 3). The lateral amplitude variation in the seismic sections remains relatively stable; however, termination patterns, contact relationships, and pinch-out features become more clearly defined (Figure 3). These improvements result in enhanced seismic resolution and an improved capability of thin-layer identification.

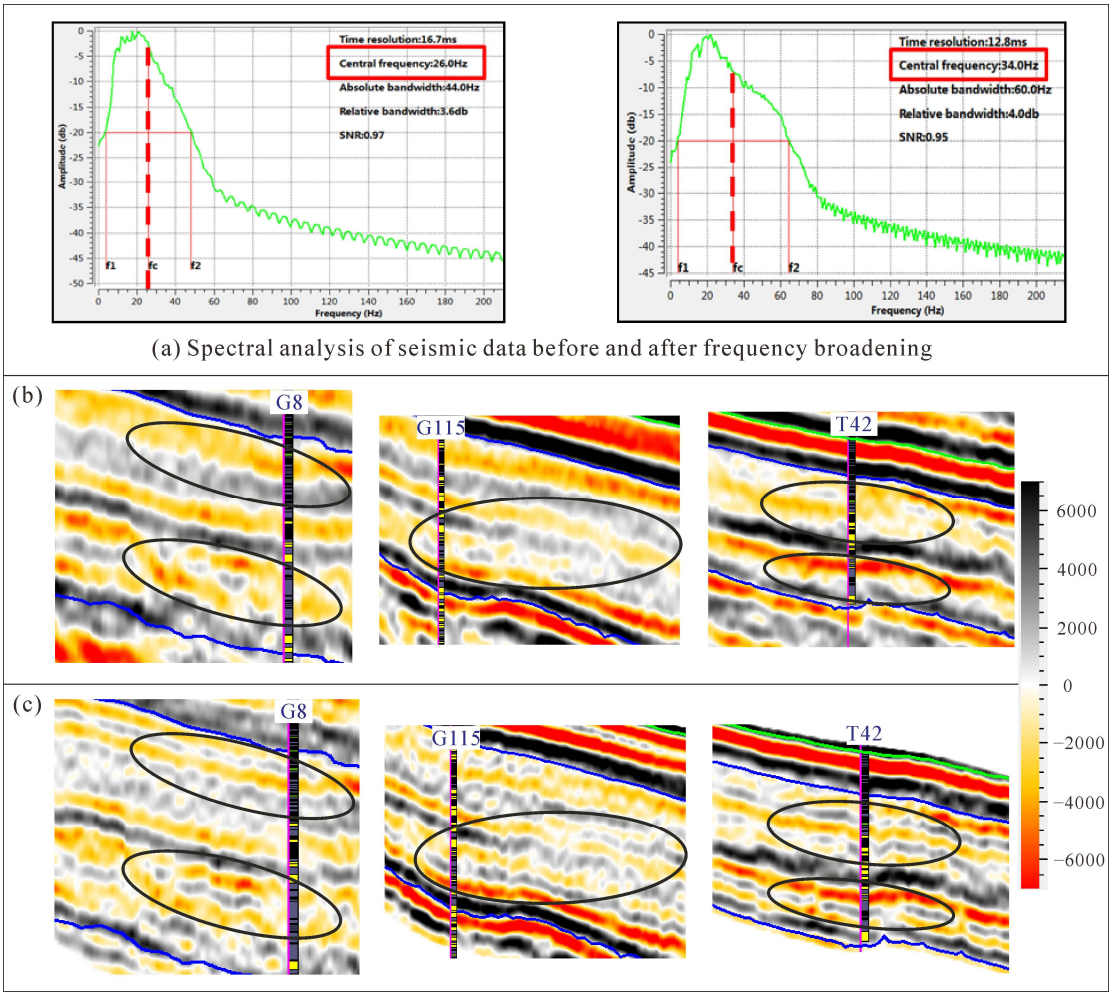


Figure 3. Seismic sections and spectrum analysis before and after frequency broadening. (**a**) Spectrum analysis before and after frequency broadening. (**b**) Seismic sections before frequency broadening. (**c**) Seismic sections after frequency broadening.

Diffusion filtering is subsequently applied to the frequency-enhanced seismic data. The noise is effectively suppressed, and the signal-to-noise ratio increases from 0.95 to 0.99.

As a result, the lateral continuity of seismic reflections is improved, leading to clearer and more interpretable seismic images (Figure 4).

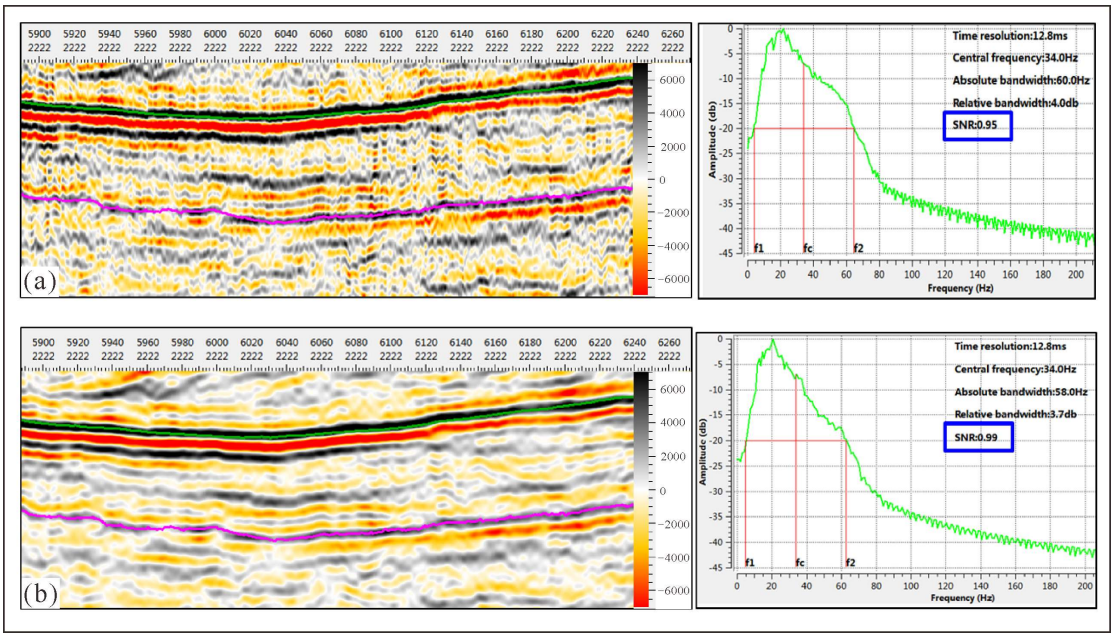


Figure 4. Comparison of seismic sections before and after diffusion filtering. (**a**) Seismic section before diffusion filtering. (**b**) Seismic section after diffusion filtering.

4.2. Sedimentary Petrological Characteristics

Mudstone color serves as an important diagnostic indicator of sedimentary environments. Purplish-red and brownish-red mudstones typically form in oxidizing shallow water environments, whereas gray and dark gray mudstones are associated with reducing deep water conditions. Green mudstones indicate suboxic to weakly reducing conditions in shallow lacustrine shoreline environments.

The Es_4^{L-1} interval in the study area is characterized by interbedded gray sandstone and purplish-red mudstone, with relatively thin sandstone layers. The identified sandstone types include siltstone, argillaceous siltstone, calcareous sandstone, conglomeratic sandstone, and fine sandstone, among which siltstone is dominant, followed by argillaceous siltstone and fine sandstone. Fine-grained sediments dominate overall, with localized occurrences of coarser-grained intervals, indicating a shallow water depositional environment with complex hydrodynamic conditions.

A variety of sedimentary structures are developed in the study area, including horizontal bedding, cross-bedding, parallel bedding, lenticular bedding, and bioturbation structures indicative of traction currents, as well as mudstone tear clasts associated with gravity flows (Figure 5). These features indicate complex and fluctuating hydrodynamic conditions during the deposition of the Es_4^{L-1} sandstones, suggesting a shallow water environment influenced by frequent hydrological variations under an overall arid climate. During high-energy flood periods, traction currents and gravity flows coexisted, whereas during lower-energy stages, traction currents dominated. In contrast, inter-flood periods were characterized by lacustrine, wave-dominated sedimentation.

Grain size characteristics further reflect hydrodynamic conditions and sediment transport mechanisms. The grain size probability curves are dominated by traction current deposits, which, together with core observations, confirm that the Es_4^{L-1} sandstone interval primarily consists of shallow-water delta front and shore–shallow lake subfacies deposits.

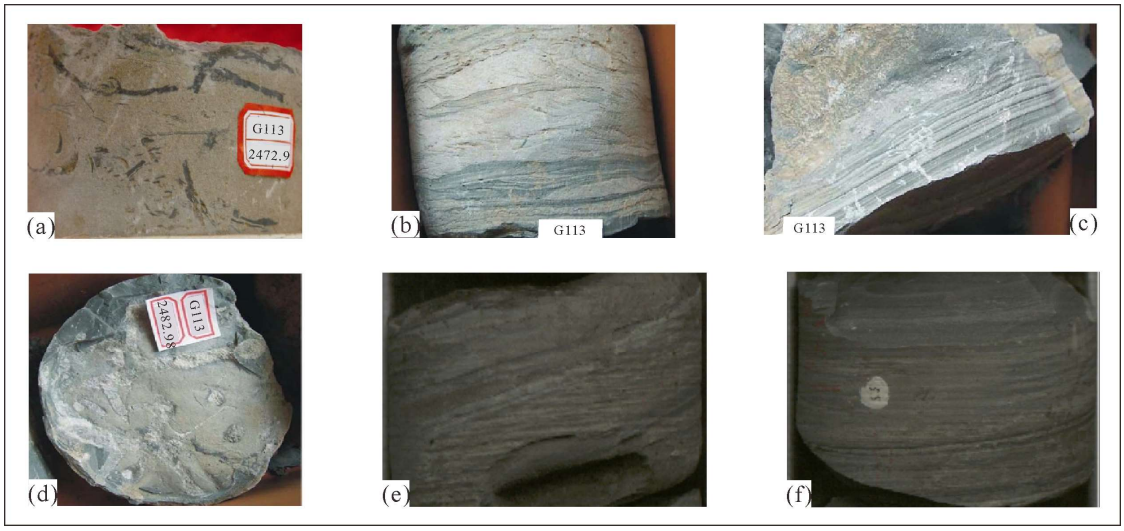


Figure 5. Typical sedimentary structures of the Es_4^{L-1} sandstone interval. (**a**) G113, 2472.9 m, gray siltstone containing dark mudstone clasts and mudstone tear clasts; (**b**) G113, 2475.95 m, gray mudstone interbedded with siltstone, with sandstone occurring in bands and displaying lenticular bedding; (**c**) G113, 2482.7 m, gray muddy siltstone exhibiting horizontal bedding; (**d**) G113, 2482.98 m, gray muddy siltstone with bioturbation structures; (**e**) G118, 3005 m, gray muddy siltstone showing cross-bedding; (**f**) G118, 3009.4 m, gray muddy siltstone showing parallel bedding.

4.3. Sedimentary Facies Characteristics

Shallow-water deltas are typically deposited in low-energy environments characterized by shallow water depth, gentle paleogeomorphology, and relatively stable tectonic conditions during sedimentation [50]. These systems are commonly associated with slow subsidence along the gentle slope of lacustrine basins and frequent lake-level fluctuations. Compared with conventional deltas, shallow-water deltas generally lack well-developed topset, foreset, and bottomset beds, but are characterized by laterally extensive distributary channel sandbodies [51]. In the study area, sediment supply is primarily derived from the southern and southeastern directions, forming a regional monoclinic structural pattern with higher elevations in the southeast and lower elevations in the northwest. The gentle paleogeomorphology and low-gradient slope provide favorable conditions for the development of shallow-water delta systems. Based on hydrodynamic processes, depositional mechanisms, and the spatial distribution of sandbodies, shallow-water delta systems can be subdivided into three subfacies: the delta plain, delta front, and prodelta. The integrated analysis of core data, well log responses, and seismic reflection characteristics indicates that the study area is dominated by shallow-water delta front subfacies. The depositional system developed within a relatively closed lacustrine environment characterized by fluctuating lake levels. Paleoclimate exerted a significant control on both sediment supply and lake-level variations. During relatively humid periods, enhanced sediment influx and elevated lake levels promoted the large-scale development of shallow-water delta front deposits. Under the combined influence of lacustrine waves and currents, sediments in the delta front were reworked and redistributed, leading to the formation of beach-bar deposits within the shore–shallow lake subfacies along the lateral margins of the delta and in local paleohigh areas. Based on this depositional model, sedimentary facies in key wells were systematically identified and interpreted (Figure 6).

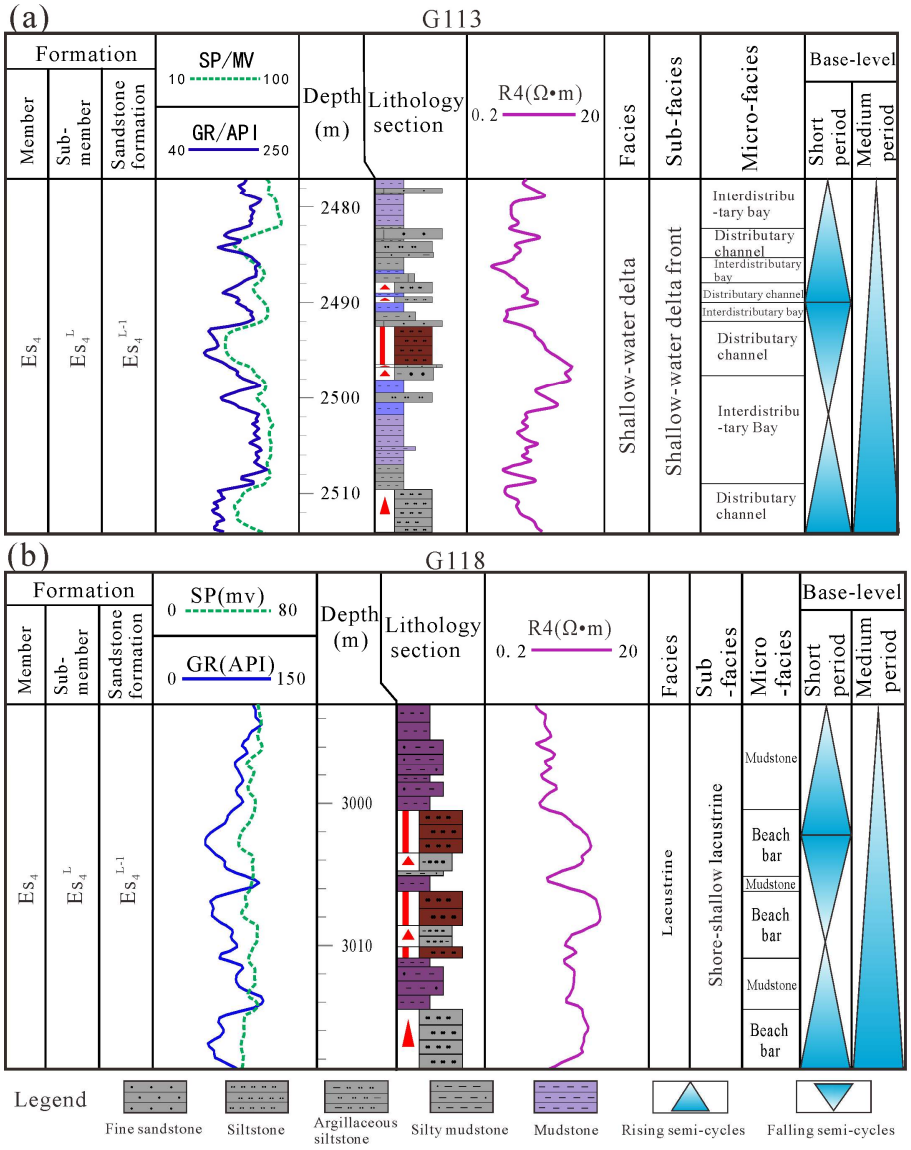


Figure 6. The sedimentary characteristics of different sedimentary facies units. (**a**) The sedimentary characteristics of well G113; (**b**) the sedimentary characteristics of well G118.

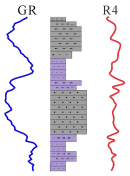
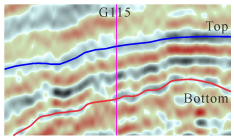
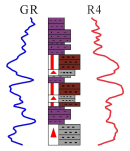
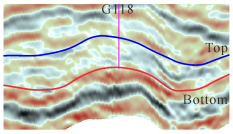
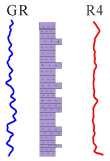
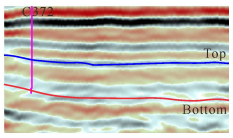
4.3.1. Shallow-Water Delta Front Subfacies

The shallow-water delta front subfacies in the study area are dominated by underwater distributary channel deposits. These sandbodies are characterized by lateral connectivity and vertical superposition, forming the principal reservoir units and extending toward the basin center. The vertical sequence commonly exhibits a coarsening upward or normal grading pattern, with thicker sandbodies in the lower part and thinner sandbodies in the upper part. The sedimentary structures observed in the sandstones include parallel bedding, lenticular bedding, bioturbation structures, and mudstone clasts, reflecting the combined influence of traction currents and episodic gravity flows (Figure 5). Due to frequent channel migration, avulsion, and erosion, mouth bars and distal bars are poorly preserved within the delta front subfacies.

The lithology of the underwater distributary channels is dominated by gray fine sandstone and siltstone, with relatively good sorting. Gamma ray (GR) logs typically exhibit box-shaped or bell-shaped patterns with low values, indicating sand-rich intervals, whereas resistivity (R4) logs show moderate values with serrated characteristics. In contrast, interdistributary areas are characterized by relatively flat GR responses and mudstone-dominated

lithology (Table 1; Figure 6a). On seismic sections, these channels are characterized by their high frequency, moderate-to-high amplitude, and moderate continuity (Table 1).

Table 1. Logging response and seismic reflection characteristics of sedimentary facies.

SUBFACIES	MICROFACIES	LOG FACIES	FEATURE DESCRIPTION	SEISMIC FACIES	REFLECTION CHARACTER
Shallow-water delta front	Subaqueous distributarychannel		Fine sandstone and siltstone, forming relatively thick sand bodies. It shows low gamma-ray (GR) values, with the GR log exhibiting a serrated bell-shaped or box-shaped motif. Resistivity(R4) values range from low to medium, and the corresponding curve displays a serrated pattern.		The seismic reflections are characterized by high frequency, medium to high amplitude, and moderate continuity.
Shore-shallow lake	Beach bar		Siltstone, relatively thin sand bodies. It shows low GR values, with the GR curve presenting a serrated funnel-shaped motif. The resistivity values(R4) are medium to high, corresponding to a funnel-shaped curve profile.		The seismic reflections are characterized by low frequency, low amplitude, and poor continuity.
	Shore-shallowlake mud		Mudstone or silty mudstone. It is identified by high GR values and low resistivity(R4) values. The corresponding curves are generally straight to slightly serrated.		The seismic reflections are characterized by high frequency, medium amplitude, and good continuity.

4.3.2. Shore–Shallow Lake Subfacies

The shore–shallow lake subfacies in the study area can be subdivided into two microfacies: beach-bar sandstones and lacustrine mudstones. Beach-bar sandstones are mainly composed of siltstone and argillaceous siltstone, with relatively thin thicknesses ranging from 1 to 5 m. The GR logs show low values with serrated funnel-shaped patterns, while resistivity (R4) logs exhibit funnel-shaped curves with moderate-to-high values. On seismic sections, beach-bar sandbodies are characterized by their high frequency, low amplitude, and relatively poor continuity.

In contrast, the lacustrine mudstone microfacies are composed of mudstone and silty mudstone, with logging responses characterized by high gamma ray values and low resistivity. Both GR and resistivity curves exhibit relatively smooth, near-baseline characteristics (Table 1; Figure 6b). Seismically, these deposits display high frequency, moderate amplitude, and good continuity (Table 1).

4.3.3. Distribution Characteristics of Well-Connected Facies

To analyze the spatial distribution and evolutionary characteristics of sedimentary facies, two cross-well sections are constructed: one along the provenance direction (SN section) and one perpendicular to it (WE section). These sections provide insight into the depositional architecture and facies distribution of the Es_4^{L-1} sandstone interval.

The cross-well facies sections indicate that two main depositional environments developed during the Es_4^{L-1} depositional period: shallow-water delta deposits and shore–shallow lake beach-bar deposits (Figure 7). During this stage, a relatively humid climate and frequent flood events led to elevated lake levels and an abundant sediment supply, resulting in the development of thick and laterally extensive sandbodies. As flood energy decreased, depositional processes gradually transitioned to lower-energy conditions, promoting the formation of both delta front deposits and beach-bar systems. Along the SN

section, stratigraphic thickness increases from well G125 to well G4, with well-developed underwater distributary channels and a decreasing sandstone content trend toward the north, indicating that the southern region served as the primary sediment source. Shore–shallow lake beach-bar deposits are mainly developed north of well G4 (Figure 7). Along the WE section, underwater distributary channels are the most prominent near well G121 in the central part of the study area. The sandstone content decreases toward both the eastern and western directions, accompanied by the thinning of individual sandbodies. The shallow-water delta front extends eastward to well W661, beyond which it transitions into shore–shallow lake deposits, with limited beach-bar development observed at well W662 (Figure 7). Overall, the Es_4^{L-1} interval is dominated by shallow-water delta front deposits, particularly underwater distributary channel sandbodies, which constitute the primary reservoir units in the study area.

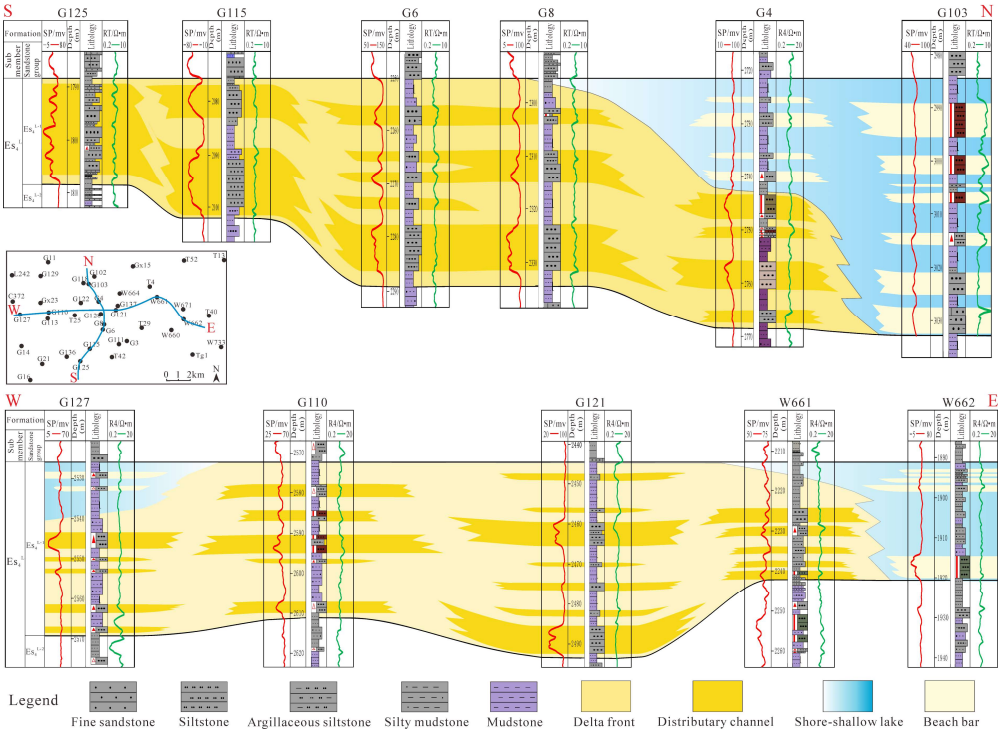


Figure 7. Sedimentary facies characteristics of the cross-well sections of the Es_4^{L-1} sandstone interval.

4.4. Seismic Sedimentary Interpretation

4.4.1. Seismic Waveform Phase Adjustment

A 90° phase adjustment is applied to enhance the geological interpretability of seismic data and improve interpretation accuracy. In this study, a multi-channel scanning method based on the Hilbert transform is used for phase adjustment [52]. The average phase of the original seismic data is approximately 2°, and therefore a phase adjustment of 88° is applied to generate a near 90° phase dataset. A comparison of the seismic sections through well G8 before and after phase adjustment is presented in Figure 8. The results show that, after phase adjustment, seismic reflection events in the 90° phase dataset correspond well with sandbody intervals identified from well logs, particularly the low-gamma ray (GR) value zones indicated by the red arrows. In this transformed dataset, seismic troughs (red) correspond to sandstone facies, whereas seismic peaks (black) correspond to mudstone facies (Figure 8). This adjustment enhances the lithological significance of seismic reflections and provides a reliable basis for identifying thin-bedded sandbodies and extracting stratal slices.

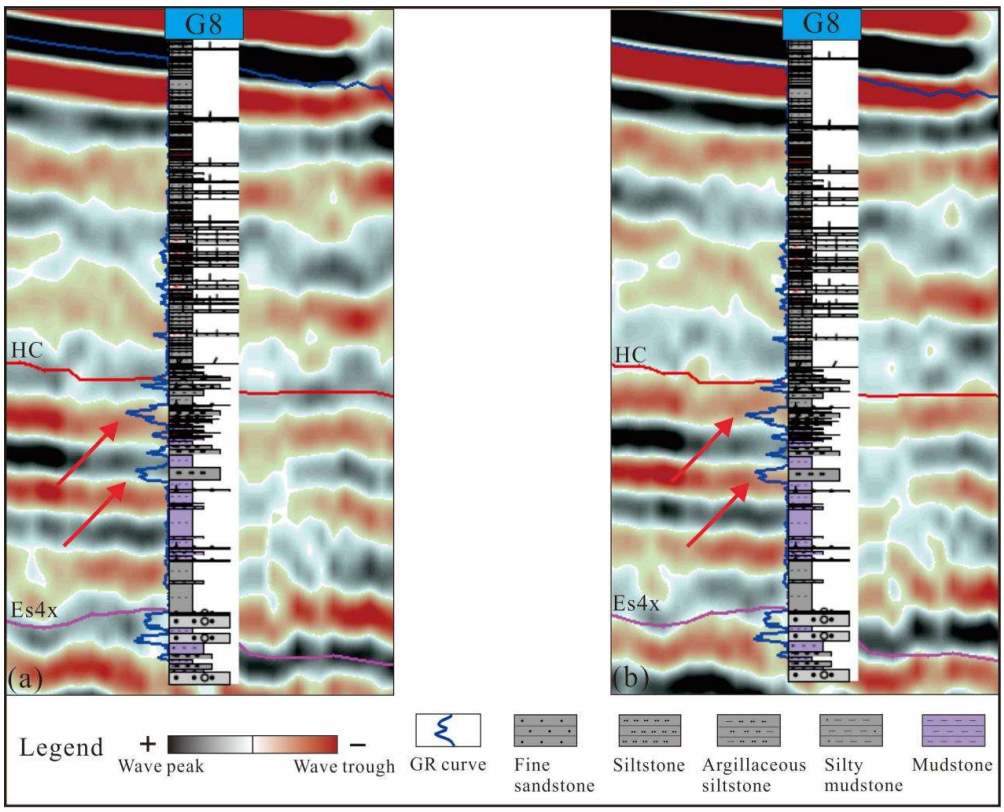


Figure 8. Comparison of seismic sections before and after 90° phase rotation at well G8. (**a**) Original seismic section; (**b**) phase-rotated seismic section.

4.4.2. Planar Distribution Characteristics of Sedimentary Facies

Using the 90° phase-adjusted seismic dataset, stratal slicing is performed based on linear interpolation. A total of 40 stratal slices are generated for the Es_4^{L-1} interval using Geoscope software (Version 4.0.0). The top and bottom horizons of the Es_4^{L-1} interval serve as reference surfaces, with a sampling interval of 1 ms in relative geological time. Representative stratal slices that best reflect the sedimentary characteristics of the Es_4^{L-1} interval are selected for facies interpretation. Through well-to-seismic calibration, the relationship between the logging facies and seismic slice attributes is established. The calibration results indicate that red–yellow negative amplitude anomalies correspond to sandstone-rich deposits, light-blue-to-blue positive amplitudes represent transitional silty mudstone, and blue positive amplitudes indicate mudstone-rich deposits (Figure 9). The lithological interpretations from wells G125, G6, G8, and G4 show good agreement with the seismic slice characteristics, while well G103 corresponds to beach-bar deposits, further validating the interpretation results. Stratal slicing significantly improves the lateral resolution of seismic data and effectively reflects paleogeomorphological features and sedimentary facies' distribution. Combined with cross-well facies correlations both parallel and perpendicular to the provenance direction, the spatial distribution of sedimentary facies within the Es_4^{L-1} interval is systematically characterized (Figure 10).

Taking wells G4 and W661 as approximate boundaries, shallow-water delta front deposits dominate the southern region, whereas shore–shallow lake beach-bar deposits are developed in the northern and eastern regions. South of well G4, stratal slices display strong red–yellow negative amplitude anomalies with discontinuous, patchy distributions, indicating well-developed underwater distributary channels. These channels are relatively well preserved due to limited reworking by lacustrine waves, resulting in large-scale channel sandbody development. In the area south of well G115, frequent channel migration and avulsion have resulted in overlapping and interconnected distribu-

tary channels, forming a quasi-planar distribution of sandbodies. In contrast, north of well G4, stratal slices exhibit weaker negative amplitude anomalies with relatively poor lateral continuity, and sandbodies are distributed as elongated, strip-like bodies. This pattern reflects the reworking of delta-front deposits by lacustrine wave action, forming beach-bar sandbodies in favorable topographic locations. Interdistributary channels and lacustrine mudstones are represented by light-blue-to-blue positive amplitude zones on stratal slices (Figure 9). Overall, the Es_4^{L-1} interval was deposited under relatively humid climatic conditions, characterized by a sufficient sediment supply, frequent lake-level fluctuations, and sustained provenance input. The combination of gentle paleotopography and continuous sediment influx facilitated the widespread development of shallow-water delta front deposits across the study area.

In summary, based on the integrated interpretation of stratal slices, single-well facies, and cross-well facies correlations, the sedimentary facies distribution of the Es_4^{L-1} interval is established. The reservoir is dominated by underwater distributary channel sandbodies of the shallow-water delta front, along with subordinate shore–shallow lake beach-bar deposits (Figure 10).

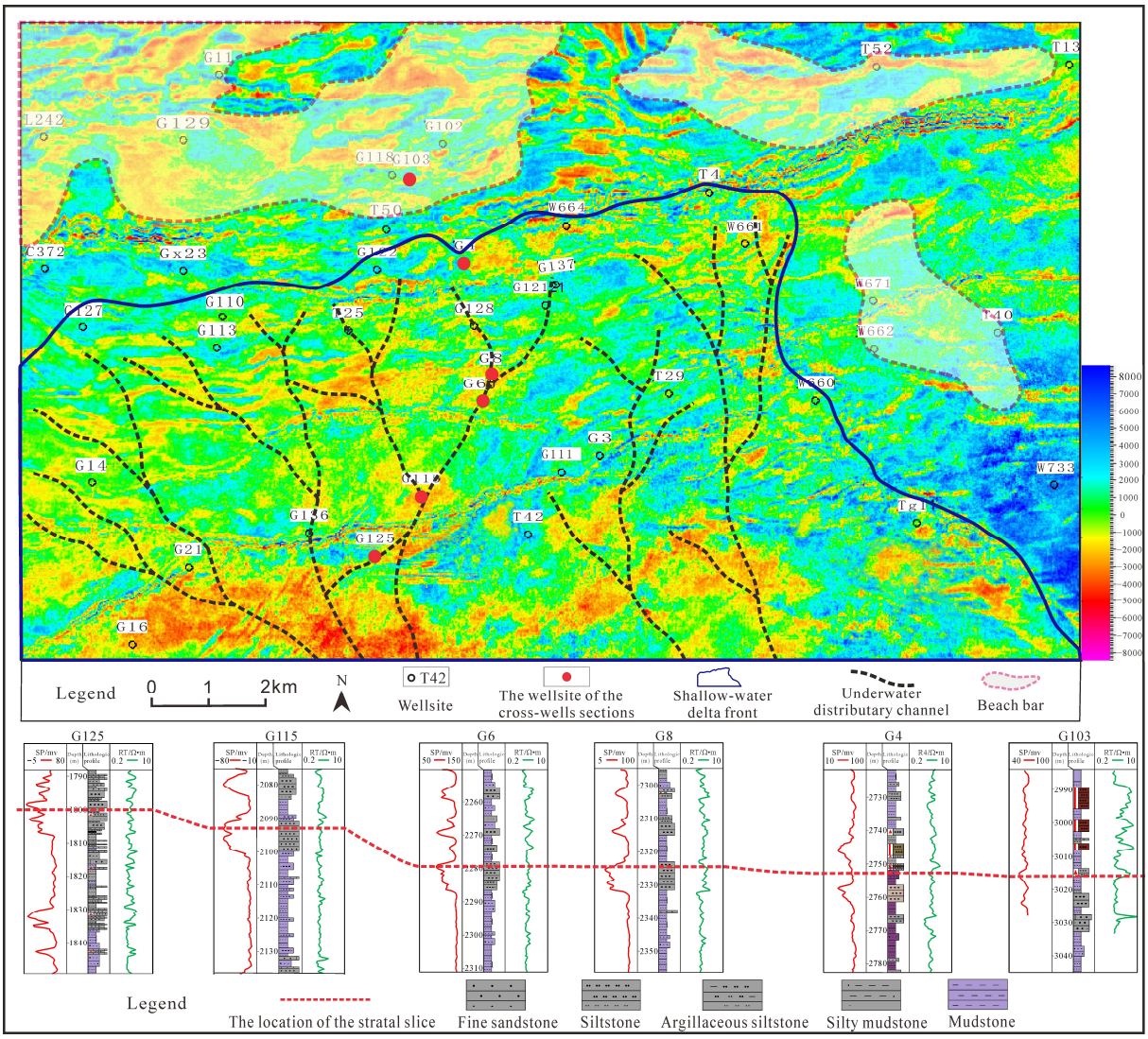


Figure 9. Sedimentary facies interpretation and lithological calibration results of amplitude stratal slices.

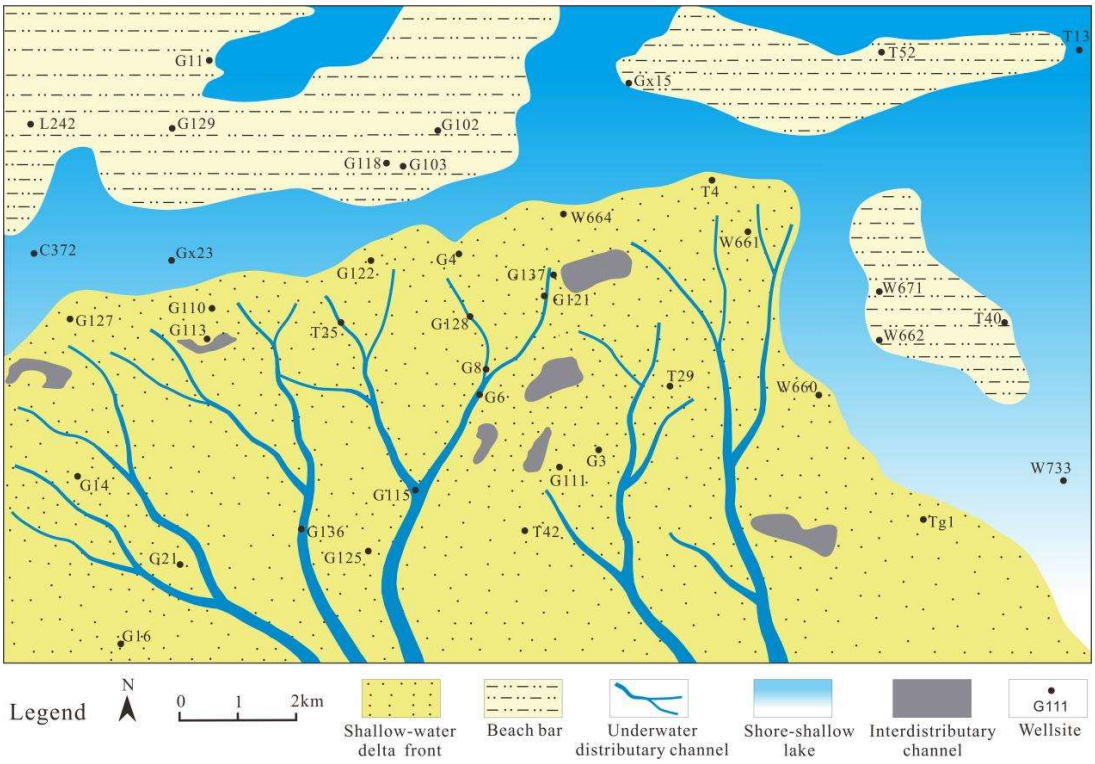


Figure 10. Sedimentary facies map of amplitude stratal slices the Es_4^{L-1} sandstone formation.

4.5. Seismic Attribute Optimization Constrained by Sedimentary Facies

Seismic data provide valuable information for characterizing reservoir properties, including lithofacies distribution and fluid-related features. Various seismic attributes reflecting reservoir characteristics can be derived through the quantitative analysis of seismic data [53]. Accurate sandbody prediction depends on the identification of sensitive seismic attributes and the establishment of robust relationships between these attributes and sandbody thickness interpreted from well log data.

Different sedimentary facies zones exhibit distinct geological characteristics and lithological associations, reflecting variations in depositional environments and processes. As a result, seismic responses—including frequency, amplitude, and continuity—vary significantly among different facies zones. Based on these differences, and in combination with the sedimentary facies analysis of the Es_4^{L-1} interval, the study area is divided into two zones: Zone I (shallow-water delta front), characterized by well-developed thick sandbodies and representing the primary sandstone-rich area, and Zone II (shore–shallow lake), where sandbodies are thinner and less developed. Accordingly, seismic attribute analysis, optimization, and sandbody prediction are conducted separately for each zone.

Seismic attributes are extracted from the Es_4^{L-1} interval using Schlumberger Petrel software (Version 2018.2), with the extraction window constrained by the top and bottom horizons of the target layer. A total of 30 commonly used seismic attributes, including amplitude-, frequency-, phase-, and waveform-related attributes, are computed to characterize the reservoir's properties from multiple perspectives. Thirty-one wells with complete logging data are selected across the study area, including twenty-two wells in Zone I and nine wells in Zone II. Based on well log interpretations, the sandbody thickness is calculated for each well. In addition, the average seismic attribute values are extracted from traces within a 25 m radius around each well and assigned as representative attribute values at the well locations.

To evaluate the relationship between seismic attributes and sandbody thickness, three representative attributes are selected for preliminary analysis: the root mean square (RMS)

amplitude, Average Instantaneous Frequency (AIF), and arc length (AL) (Figures 11–13). The RMS attribute shows relatively high values in the proximal shallow-water delta front near the sediment source and in the eastern beach-bar depositional area, indicating sandbody development, whereas lower values are observed in the distal delta front. This pattern is influenced by both the data's quality and structural factors such as faulting. The arc length attribute exhibits a similar spatial distribution. In contrast, the Average Instantaneous Frequency shows higher values in the shallow-water delta front, reflecting the presence of thin interbedded sandstone and mudstone, while displaying moderate-to-low values in other areas.

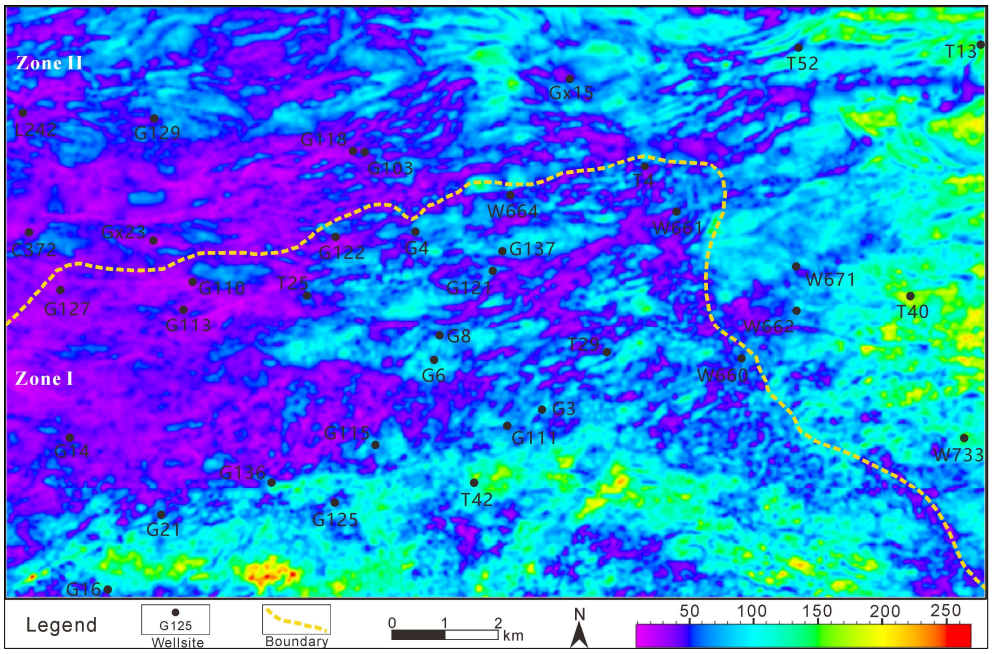


Figure 11. RMS amplitude attribute.

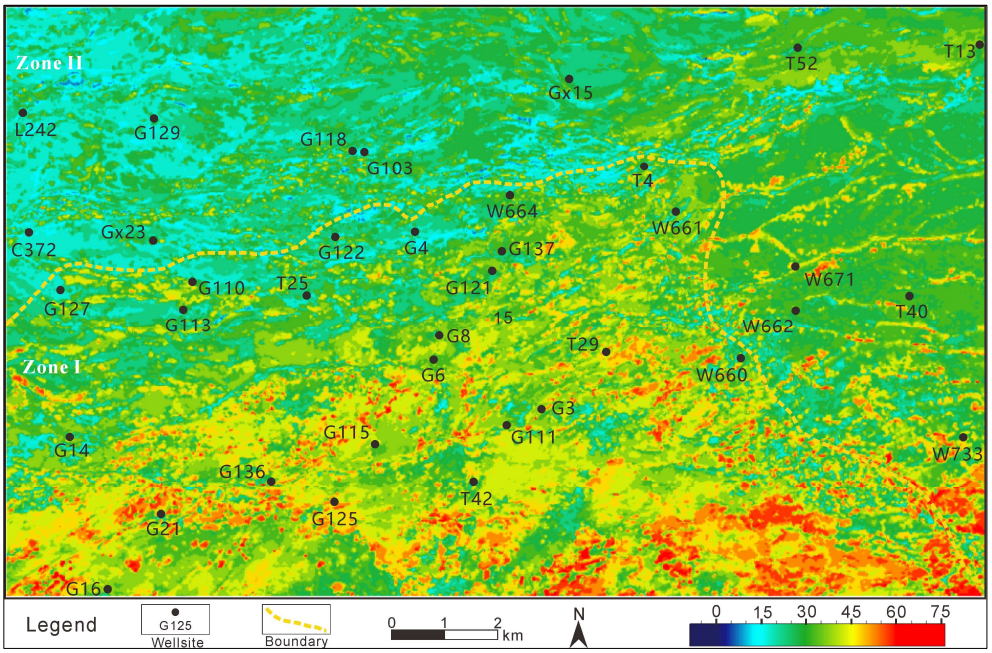


Figure 12. Average Instantaneous Frequency attribute.

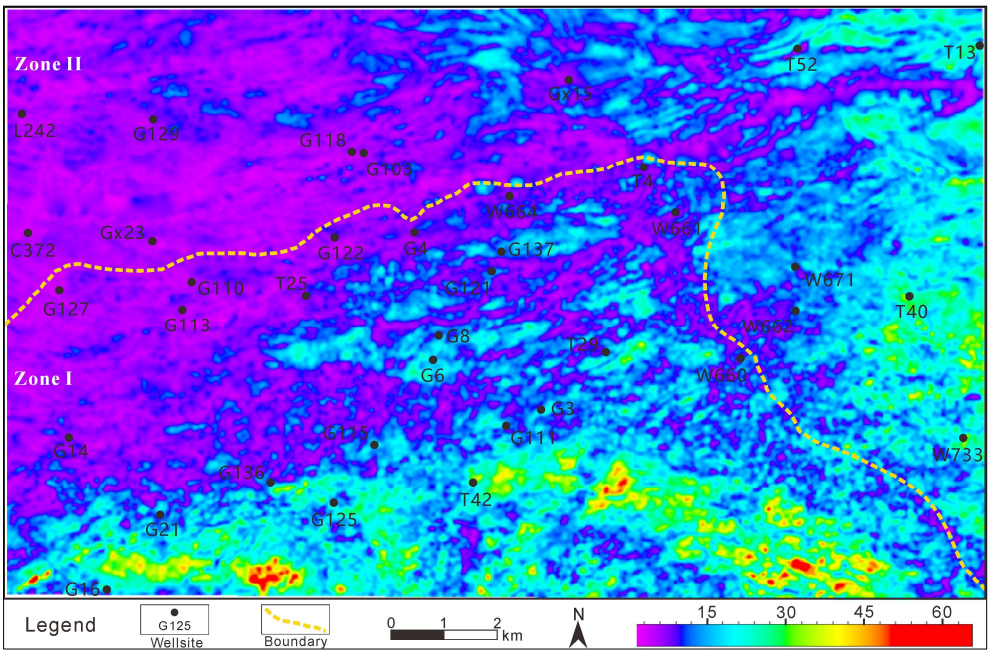


Figure 13. Arc length attribute.

Cross-plots of seismic attributes versus sandbody thickness reveal a wide dispersion of data points (Figure 14a–c), indicating weak overall correlations. Linear regression analysis is performed to quantify these relationships, using the coefficient of determination (R^2) as an evaluation metric. The R^2 values between the sandbody thickness and RMS, AIF, and AL are 0.07, 0.45, and 0.035, respectively, all indicating weak correlations and insufficient predictive capability. When analyzed separately by facies zones, the correlations improve significantly. In Zone I, the R^2 values for RMS, AIF, and AL are 0.307, 0.742, and 0.375, respectively, while in Zone II, the corresponding values are 0.688, 0.638, and 0.620 (Figure 14a'–c'). Although these results demonstrate improved relationships under facies constraints, the predictive performance remains insufficient for accurate sandbody thickness estimation. These results indicate that the variability in seismic attribute–sandbody relationships is primarily controlled by differences in sedimentary environments. The sensitivity of seismic attributes to sandbody distribution varies significantly across facies zones, and therefore a single attribute cannot be universally applied across the entire study area. Instead, facies-constrained attribute selection is required to improve prediction accuracy.

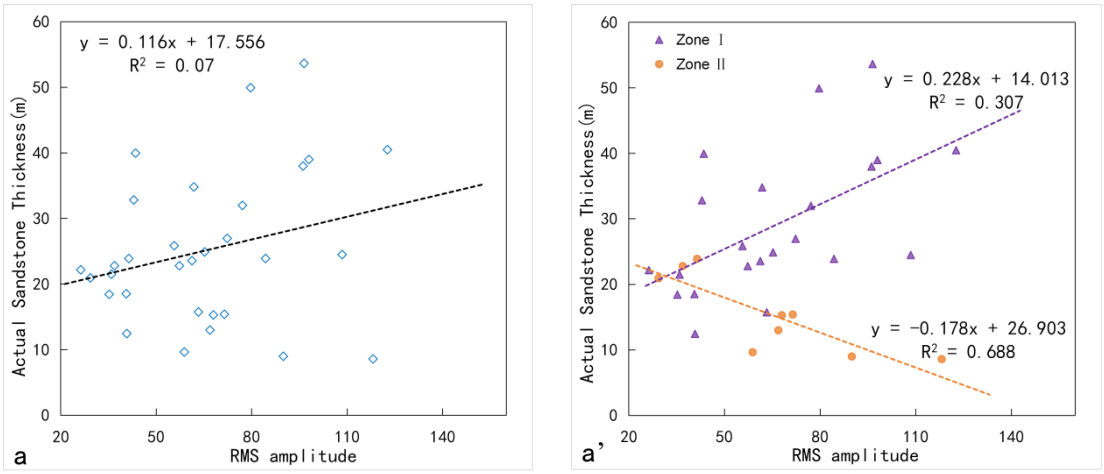


Figure 14. *Cont.*

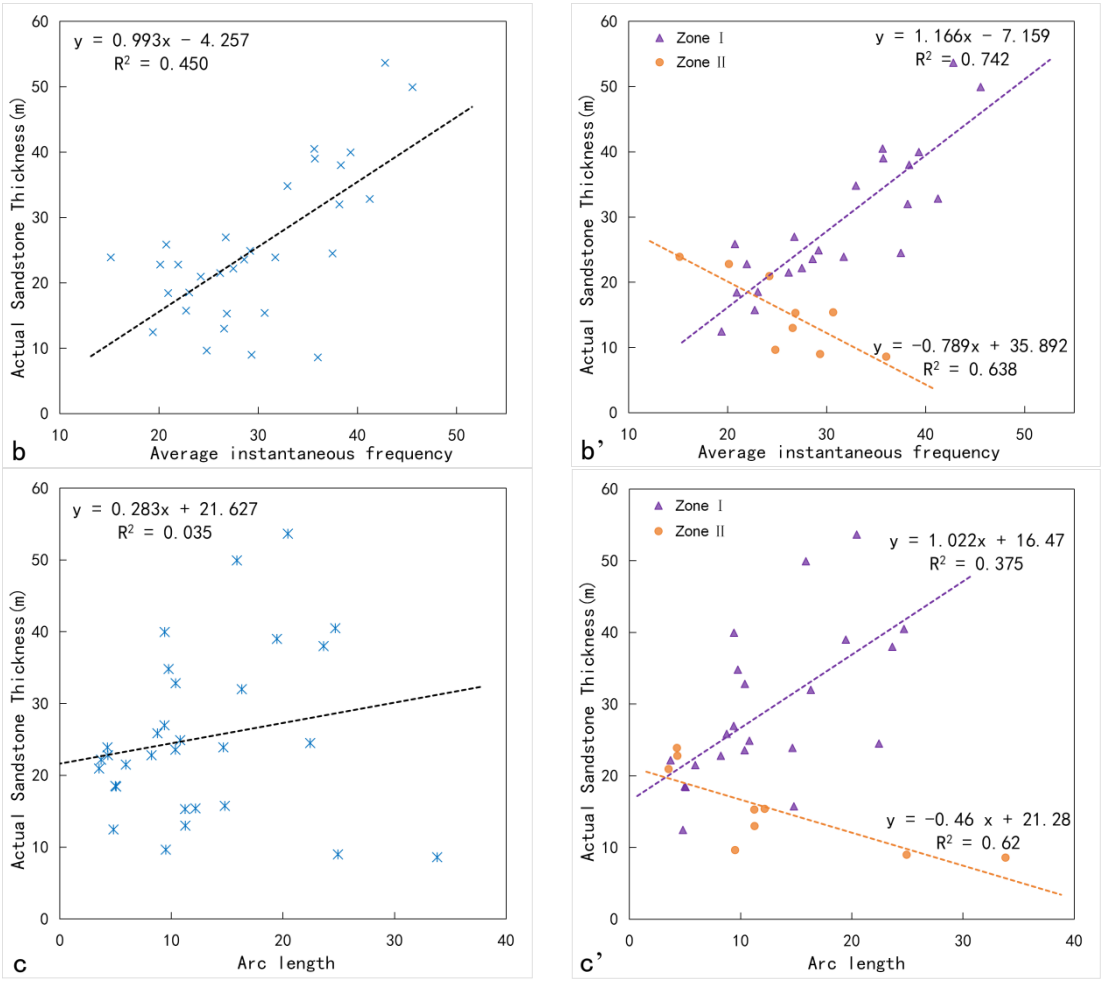


Figure 14. Correlation analysis between seismic attributes and sandstone thickness. Whole region: (**a**) RMS amplitude; (**b**) Average Iinstantaneous Frequency; and (**c**) arc length. Different regions: (**a'**) RMS amplitude; (**b'**) Average Iinstantaneous Frequency; and (**c'**) arc length.

Based on this analysis, the following principles are adopted for seismic attribute optimization: (1) the selected attributes should exhibit a strong correlation with the target variable; (2) the attributes should be as independent as possible, with minimal redundancy; and (3) the attributes with low correlation should be excluded while preserving essential information content.

4.5.1. Seismic Attribute Optimization for Zone I

An excessive number of seismic attributes may negatively affect the model's performance and prediction accuracy; therefore, it is necessary to select attributes that are sensitive to sandstone distribution. To ensure strict model validation and prevent data leakage, three blind wells are entirely excluded from the seismic attribute screening and optimization process. Feature analysis is conducted exclusively using the remaining 19 wells for model construction. In Zone I, Pearson correlation analysis is performed between sandstone thickness derived from 19 wells and 30 seismic attributes for preliminary screening. The Pearson correlation coefficient is used to quantify the relationship between the sandstone's thickness and each attribute. Attributes with an absolute correlation coefficient less than 0.5 are discarded, and the remaining 18 attributes are ranked in descending order based on their correlation strength (Figure 15). Among these, the attribute Trace Gradient–Interval Average: Maximum (TGM) exhibits the highest correlation with the sandstone thickness, with a coefficient of 0.863. These results indicate that facies-constrained attribute selection

significantly enhances the correlation between seismic attributes and sandstone thickness, while the exclusion of blind wells ensures the rigor and reliability of subsequent model validation.

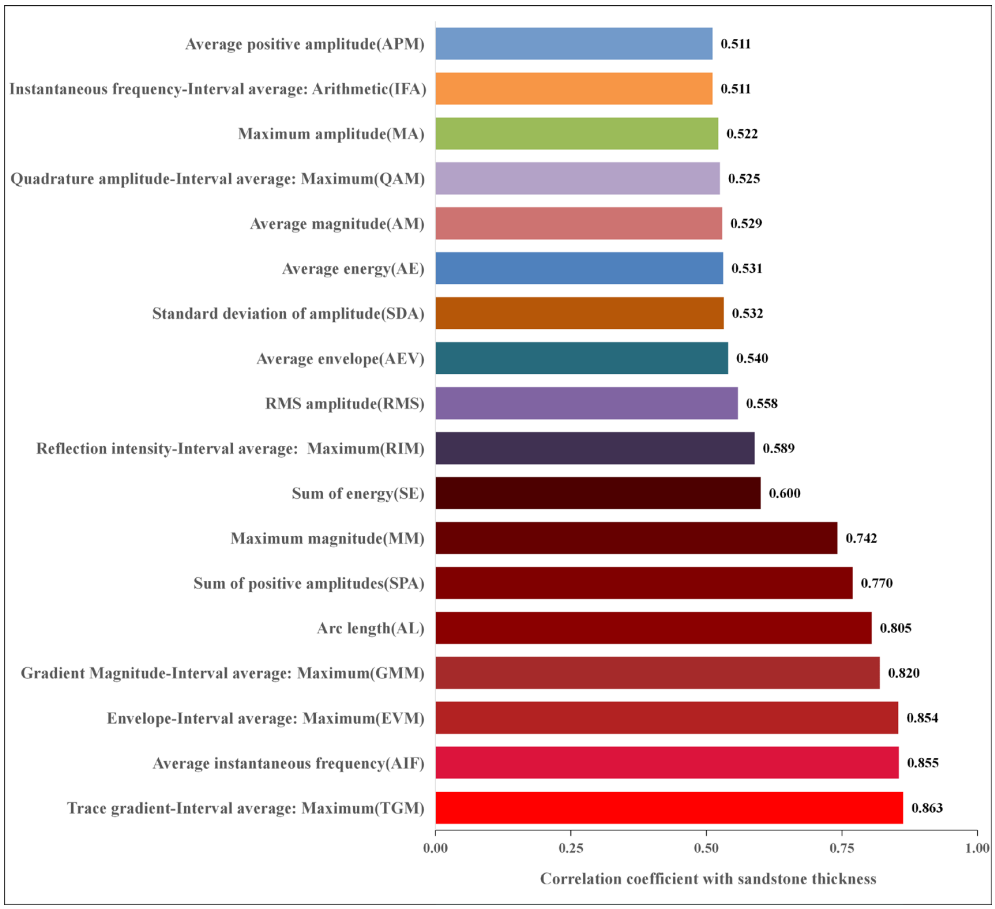


Figure 15. Ranking of correlation between seismic attributes and sandstone thickness in Zone I.

Although correlation analysis identifies attributes sensitive to sandstone, redundancy and intercorrelation among attributes remain. To improve attributes' independence and reduce redundancy, hierarchical clustering is applied to group similar attributes. The 18 selected attributes are classified into five clusters (Figure 16). Based on both the clustering results and correlation strength, five optimal attributes are selected for subsequent multi-attribute fusion: the Sum of Energy (SE), Trace Gradient–Interval Average: Maximum (TGM), Instantaneous Frequency–Interval Average: Arithmetic (IFA), Sum of Positive Amplitudes (SPA), and Average Instantaneous Frequency (AIF), with correlation coefficients of 0.600, 0.863, 0.511, 0.770, and 0.855, respectively. These attributes show a positive correlation with the sandstone's thickness and effectively reflect sandstone development. From the perspectives of energy, amplitude, and frequency, they characterize the scale, gradient, and spatial distribution of sandbodies, providing reliable inputs for subsequent machine learning modeling.

4.5.2. Seismic Attribute Optimization for Zone II

In Zone II, Pearson correlation analysis is similarly conducted between sandstone thickness from 9 wells and 30 seismic attributes. The attributes with an absolute correlation coefficient less than 0.5 are excluded, and the remaining 26 attributes are ranked according to their correlation coefficients (Figure 17). The attribute Iso-frequency component—Most of (IFCM) shows the strongest correlation with the sandstone's thickness, exhibiting a negative

correlation coefficient of −0.918, indicating a strong inverse relationship. Compared with Zone I, the correlation between seismic attributes and sandstone thickness in Zone II is significantly stronger, reflecting the more consistent seismic response of thinner sandbodies in this depositional setting.

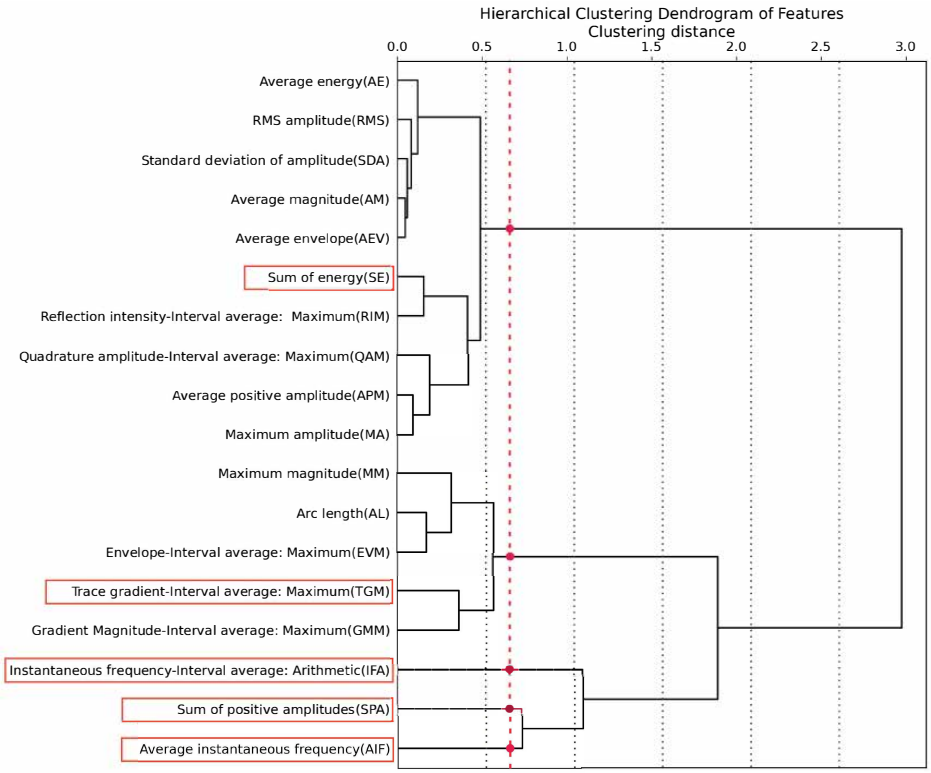


Figure 16. Unsupervised cluster analysis of seismic attributes in Zone I.

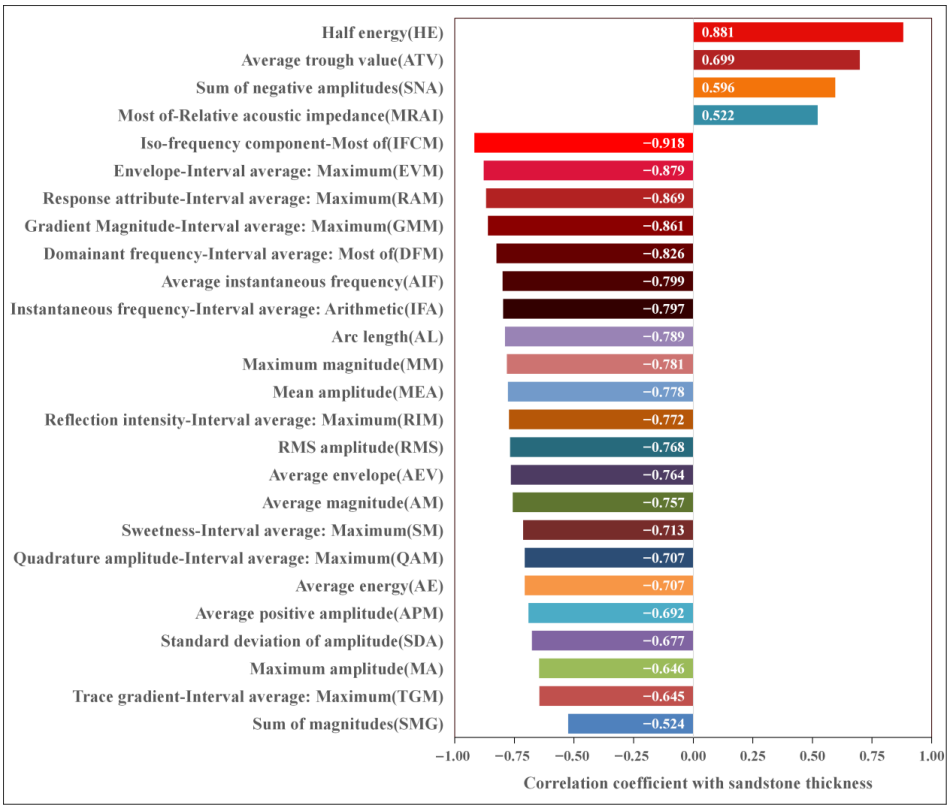


Figure 17. Ranking of correlation between seismic attributes and sandbodies thickness in Zone II.

To further reduce redundancy and ensure attributes' independence, hierarchical clustering is applied to the 26 attributes, grouping them into two clusters (Figure 18). Based on clustering results and correlation strength, two optimal attributes are selected: the Iso-frequency component—Most of (IFCM) and Envelope–Interval Average: Maximum (EVM), with correlation coefficients of −0.918 and −0.879, respectively. These attributes exhibit strong negative correlations with the sandstone's thickness, indicating that their values decrease as the sandstone's thickness increases. From the perspectives of dominant frequency and envelope amplitude, they effectively capture variations in sandstone distribution and serve as key indicators for sandbody characterization in Zone II.

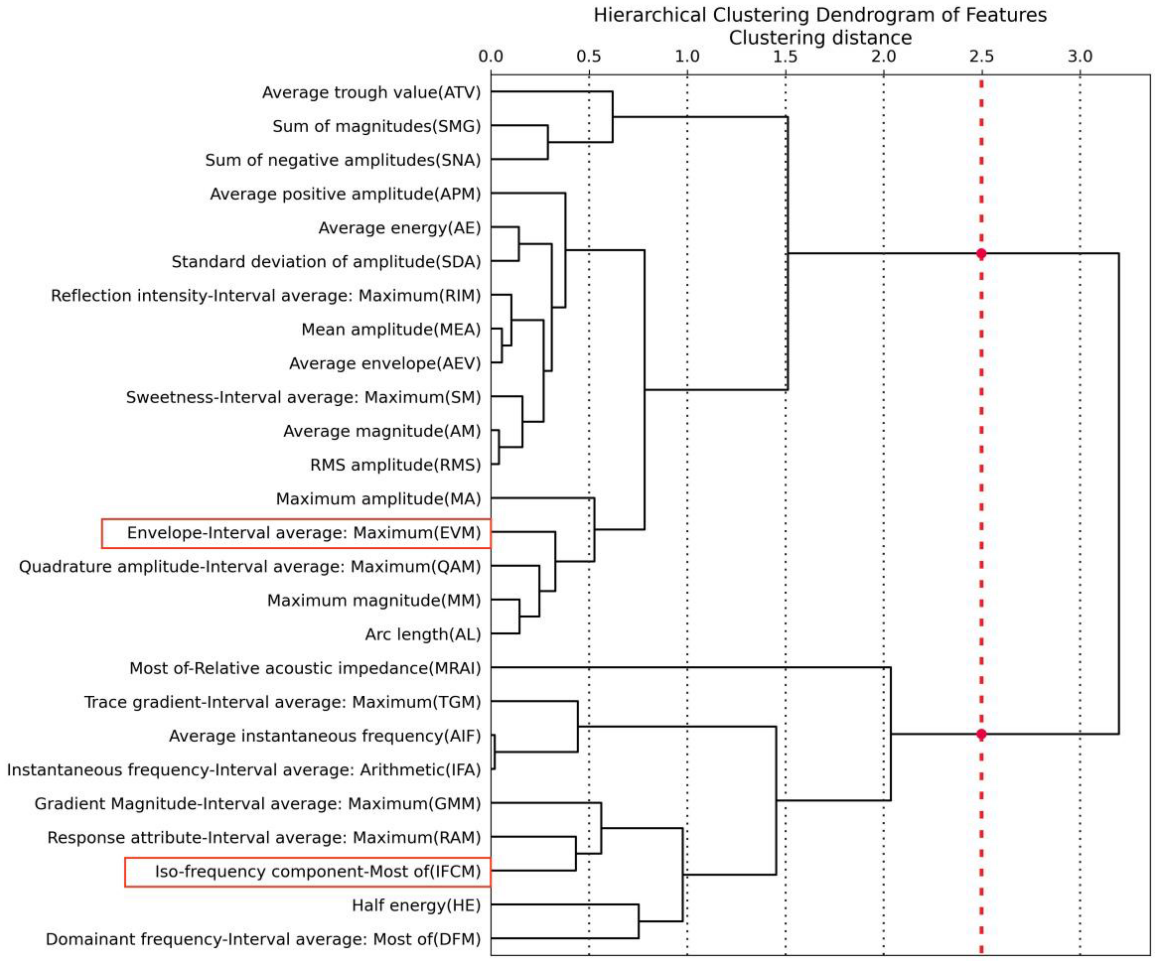


Figure 18. Unsupervised cluster analysis of seismic attributes in the Zone II.

Based on Pearson correlation analysis and hierarchical clustering, the most sensitive seismic attributes for each sedimentary facies zone are determined (Table 2). Zone I is characterized by five optimized attributes (SE, TGM, IFA, SPA, and AIF) and contains 22 wells, with the sandstone thickness ranging from 12 to 54 m (average 29.18 m), representing relatively thick sandbodies. Zone II is characterized by two optimized attributes (IFCM and EVM) and contains nine wells, with the sandstone thickness ranging from 8 to 24 m (average 15.4 m), representing relatively thin sandbodies. Based on these optimized attributes, facies-constrained multi-attribute machine learning models are subsequently applied to predict the sandbody thickness in each zone.

Table 2. Statistical table of sandstone thickness values at well points and optimized seismic attributes.

Zone I							Zone II			
Well ID	**Sandstone Thickness**	**SE**	**TGM**	**IFA**	**SPA**	**AIF**	**Well ID**	**Sandstone Thickness**	**EVM**	**IFCM**
G14	12.46	81,052.23	0.70	19.11	778.19	19.38	W733	8.6	318.9	0.83
G21	15.75	111,439.30	0.91	45.56	949.85	22.73	T40	9	263.43	0.8
G127	18.44	82,894.87	0.94	20.50	1264.05	20.92	W671	9.65	166.56	0.76
G110	18.53	113,542.80	0.76	24.03	1230.05	23.04	T52	13	160.96	0.64
G122	21.50	77,772.13	1.05	27.08	1564.62	26.16	GX15	15.3	124.07	0.57
G113	22.18	45,887.64	0.65	27.58	755.74	27.49	W662	15.4	137.87	0.75
T25	22.80	214,859.80	0.91	21.79	1794.15	21.93	G118	20.95	88.03	0.42
W660	23.57	302,309.42	2.63	28.32	839.23	28.58	L242	22.8	71.79	0.43
G121	23.90	323,793.80	1.47	33.21	1473.43	31.72	G103	23.9	72.71	0.49
G16	24.50	307,404.70	3.08	38.03	1100.67	37.49				
G137	24.90	242,065.10	1.31	29.27	1680.42	29.18				
W661	25.86	158,228.20	1.45	20.45	1029.63	20.73				
G4	26.97	332,208.60	1.12	27.36	1809.94	26.73				
G8	32.00	325,868.60	2.52	38.41	2087.01	38.18				
T29	32.83	108,270.50	1.55	43.55	1680.36	41.24	Seismic attribute name			
G136	34.81	231,154.05	2.10	29.62	1854.10	32.95	SE: Sum of energy			
G3	38.00	477,285.90	3.14	37.94	2105.08	38.33	SPA: Sum of positive amplitudes			
G6	39.00	497,438.70	2.39	35.70	2410.73	35.71	AIF: Average instantaneous frequency			
G125	39.96	73,052.72	2.97	41.54	1710.35	39.31	IFCM: Iso-frequency component - most of			
T42	40.49	539,989.40	3.22	35.74	2260.33	35.65	EVM: Envelope - Interval average: maximum			
G111	49.93	305,450.60	3.27	35.13	2451.85	45.55	TGM: Trace gradient - Interval average: maximum			
G115	53.65	410,143.20	4.17	39.79	2075.25	42.79	IFA: Instantaneous frequency - Interval average: arithmetic			

Zone I—Shallow-water delta front zone; Zone II—Shore-shallow lake zone.

4.6. Prediction Performance of Machine Learning Model and Sandbodies Prediction

4.6.1. Prediction Performance of Machine Learning Models in Zone I

Five machine learning algorithms—Random Forest, Lasso Regression, Support Vector Regression (SVR), Ridge Regression, and Multilayer Perceptron (MLP)—are employed to construct sandbody thickness prediction models. Facies-constrained seismic multi-attributes are used as input features, and all the variables are standardized prior to modeling. Three wells located at different structural positions are reserved as independent validation wells, while the remaining samples are divided into training and test sets in a ratio of approximately 75% to 25%. Sensitivity analysis is first performed to identify the key hyperparameters for each model. Subsequently, hyperparameter optimization is conducted using grid search combined with five-fold cross-validation, with mean squared error as the optimization objective. This procedure enhances models' accuracy while effectively reducing the risk of overfitting, ensuring robust generalization performance.

The MLP model adopted in this study consists of two hidden layers with ten and five neurons, respectively, enabling hierarchical feature extraction from seismic attributes. The Rectified Linear Unit (ReLU) activation function is used to improve convergence efficiency and mitigate gradient vanishing. To address the limited sample size, L2 regularization ($\alpha = 0.05$) is introduced to constrain the model's complexity. Model training is performed using mini-batch stochastic gradient descent with a batch size of four and an initial learning rate of 0.001, ensuring stable optimization. The optimized hyperparameters for all models are summarized in Table 3.

Table 3. The optimum values of hyperparameters for all models.

Model Name	Model Hyperparameters
RandomForest	n_estimators: 200; max_depth: 4; max_features: 0.8; min_samples_leaf: 2; min_samples_split: 4
Lasso	linear regression with L1 penalty; α: 0.1; selection: cyclic; max_iteration: 2000
SVR	kernel function: RBF kernel; regularization: L2 penalty; C: 150; γ: 0.2; ε: 0.8
Ridge	linear regression with L2 penalty; α: 0.1; solver: auto
MLP	input layer: 5 neurons; hidden layer 1: 10 neurons; hidden layer 2: 5 neurons; output layer: 1 neuron; activation: ReLU; optimizer: Adam; α: 0.05; learning rate: 0.001; max iteration: 1000; batch_size: 4

The performance comparison of the five models is presented in Table 4. On the training set, all models achieve R^2 values ranging from 0.814 to 0.856, indicating that they can effectively capture the relationship between seismic attributes and sandbody thickness. On the test set, models' performances show clear differentiation. Random Forest yields a relatively lower accuracy (R^2 = 0.762; RMSE = 6.721), while Lasso Regression achieves R^2 = 0.902, demonstrating a stable linear fitting performance. SVR and Ridge Regression further improve prediction accuracy, with R^2 values of 0.929 and 0.933, respectively. Among all the models, the MLP model exhibits the best overall performance. The MLP model achieves R^2 values of 0.856 and 0.936 for the training and test sets, respectively, with MAE = 3.003 and RMSE = 3.474. The small difference between training and test performance (R^2_difference = 0.080), along with consistent error metrics, indicates that the model does not suffer from overfitting. The predicted values are highly concentrated along the y = x reference line (Figure 19), further demonstrating the reliability of the model.

Table 4. Statistical parameters of the performance of all models for different training/test set in predicting sandstone thickness.

Model Name	R^2_Training Set	R^2_Test Set	R^2_Difference	MAE_Test Set	RMSE_Test Set
RandomForest	0.814	0.762	0.052	5.303	6.721
Lasso	0.839	0.902	0.063	3.779	4.308
SVR	0.855	0.929	0.074	3.354	3.661
Ridge	0.845	0.933	0.088	3.262	3.574
MLP	0.856	0.936	0.080	3.003	3.474

Note: R^2_difference = $|R^2$_Training set − R^2_Test set$|$.

To further validate the reliability and generalization capability of the primary multilayer perceptron (MLP) model, this study employs Ridge Regression as an auxiliary verification method. By comparing the prediction trends of these nonlinear and linear approaches, the stability of sandbody thickness prediction patterns is systematically assessed. On the test set, the Pearson correlation coefficient between MLP and Ridge predictions reaches 0.9995, with the predicted values distributed almost entirely along the y = x diagonal (Figure 20a). This remarkable consistency indicates that the complex patterns captured by the nonlinear MLP model align closely with the overall trends revealed by the linear Ridge model. Furthermore, radar chart comparisons (Figure 20b) demonstrate comparable performance between the two models across key metrics, including the test set R^2, RMSE, and MAE. Notably, both the models maintain low R^2_difference values (defined as the absolute difference in R^2 between the training and test sets), indicating negligible overfitting. These cross-validation results robustly confirm that the MLP model delivers predictions with excellent geological plausibility and statistical stability, establishing it as a reliable tool for sandbody thickness prediction in the study area.

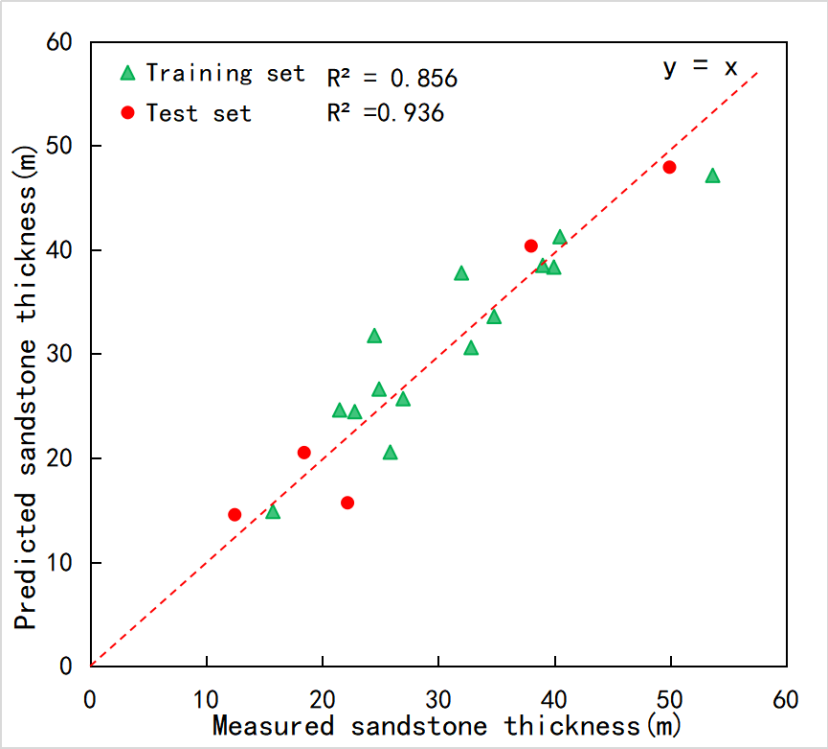


Figure 19. Relationship between the predicted sandstone thickness from the MLP model and the sandstone thickness from well.

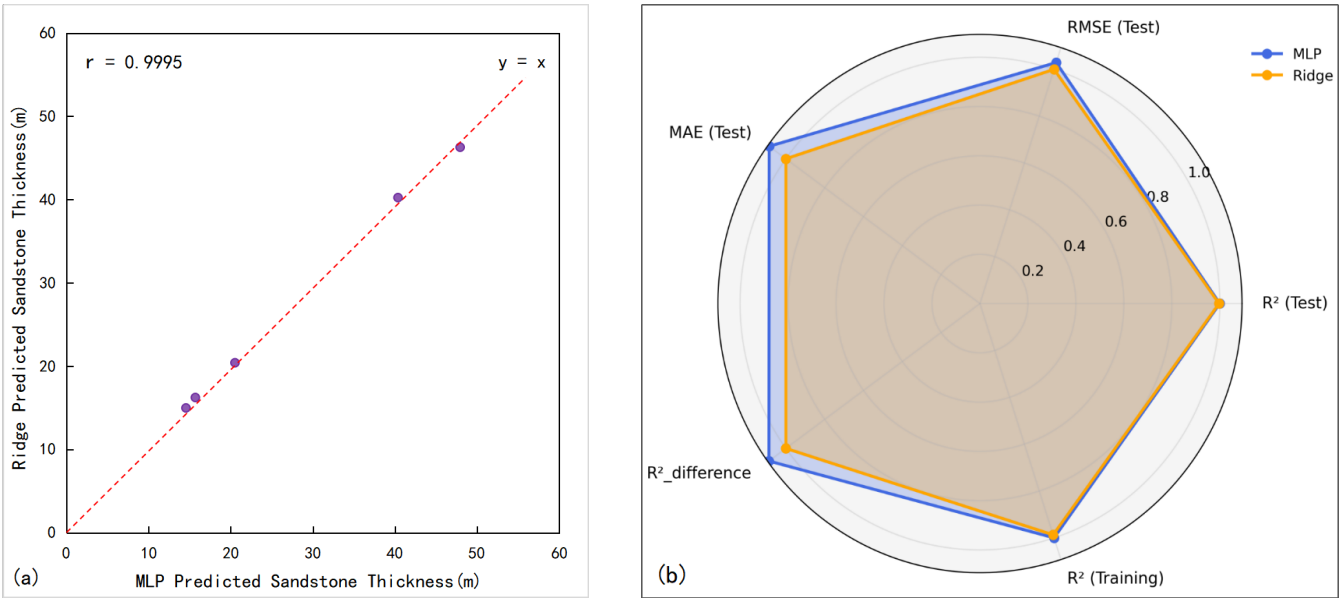


Figure 20. Comparison between MLP and Ridge model performance. (**a**) Scatter plot of model predicted values; (**b**) radar chart of model performance metrics.

Permutation importance analysis is conducted to evaluate the contribution of individual seismic attributes (Figure 21). The results show that the Sum of Positive Amplitudes (SPA) and Trace Gradient–Interval Average: Maximum (TGM) are the most influential features, indicating that strong amplitude responses and gradient variations are closely associated with well-developed sandbodies. Variations in Sum of Energy (SE)also demonstrate a moderate correlation with the development of sandbodies, while frequency-related attributes (IFA and AIF) exhibit relatively weaker contributions, reflecting their sensitivity to thin interbedded lithologies.

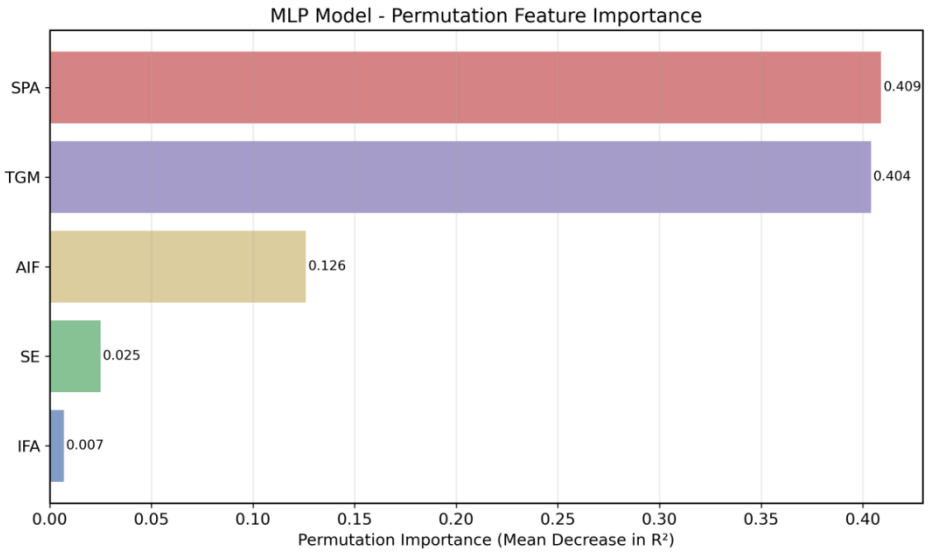


Figure 21. Permutation feature importance plot of MLP model.

The predictive performance of the MLP model is further validated using three independent wells. The predicted sandbody thickness values show good agreement with measured data, with the absolute errors ranging from 0.7 m to 2.0 m and the relative errors between 3.8% and 8.5% (Figure 22; Table 5). The low MAE, MSE, and RMSE values confirm the high predictive accuracy and reliability of the model in practical applications.

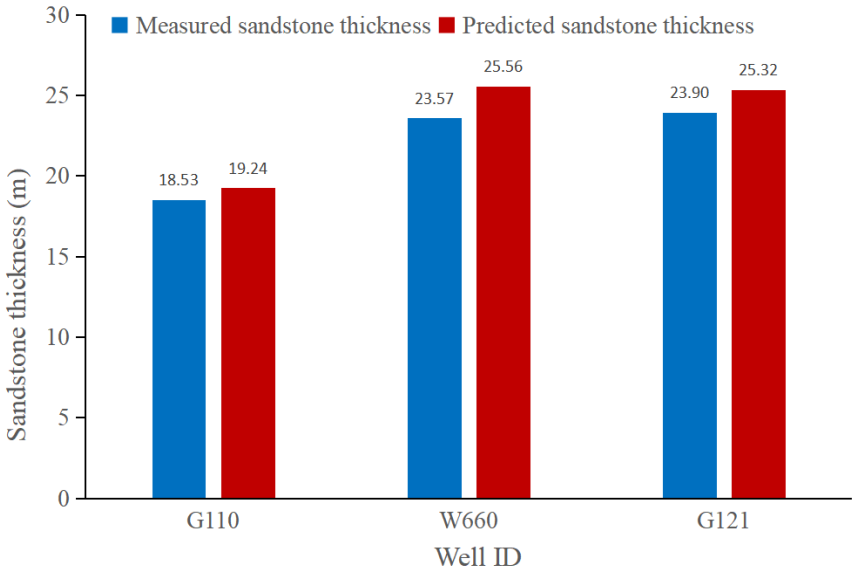


Figure 22. Comparison of the measured and predicted sandbodies thickness in validation wells.

Table 5. Calculated statistical parameters for the verification wells.

Fitted Model	Well ID	Measured Sandstone Thickness (m)	Predicted Sandstone Thickness (m)	Absolute Error (m)	Relative_ Error (%)	MAE (m)	MSE	RMSE
MLP	G110	18.53	19.24	0.708	3.818	1.375	3.205	1.790
	W660	23.57	25.56	1.425	5.962			
	G121	23.90	25.32	1.992	8.450			

In summary, the MLP model demonstrates strong generalization capability and predictive reliability. The consistency between training and test performance, agreement with

independent validation wells, and consistency with Ridge Regression results collectively confirm the robustness and applicability of the proposed method for sandbody thickness prediction in the study area.

4.6.2. Prediction Performance of Ridge Regression Model in Zone II

Due to the limited availability of the only nine wells in Zone II, leave-one-out cross-validation (LOOCV) combined with Ridge Regression is adopted to improve models' stability and generalization performance. A weighted Ridge Regression model is constructed to quantify the relationship between seismic attributes and sandbody thickness.

Two optimized seismic attributes, Iso-frequency component—Most of (IFCM) and Envelope–Interval Average: Maximum (EVM), are selected as input features. These attributes are weighted according to the absolute values of their correlation coefficients with sandbody thickness. After feature standardization, grid search is applied to determine the optimal regularization parameter of the Ridge Regression model. The model is evaluated through nine LOOCV iterations, in which eight wells are used for training and one well is used to test in each iteration. This strategy effectively mitigates the overfitting risks associated with small-sample datasets. The results show that the predicted values are generally aligned with the y = x reference line in both the training and test datasets (Figure 23), indicating consistent prediction trends with the measured values. The model achieves an R^2 of 0.864 for the training set and 0.779 for the test set, demonstrating good predictive capability under limited data conditions. The small difference between the training and test performance (R^2_difference = 0.085) further indicates that overfitting is effectively controlled. The error metrics, including the MAE, MSE, and RMSE, remain at relatively low levels and show no significant divergence between the training and test datasets (Table 6), confirming the robustness and stability of the model.

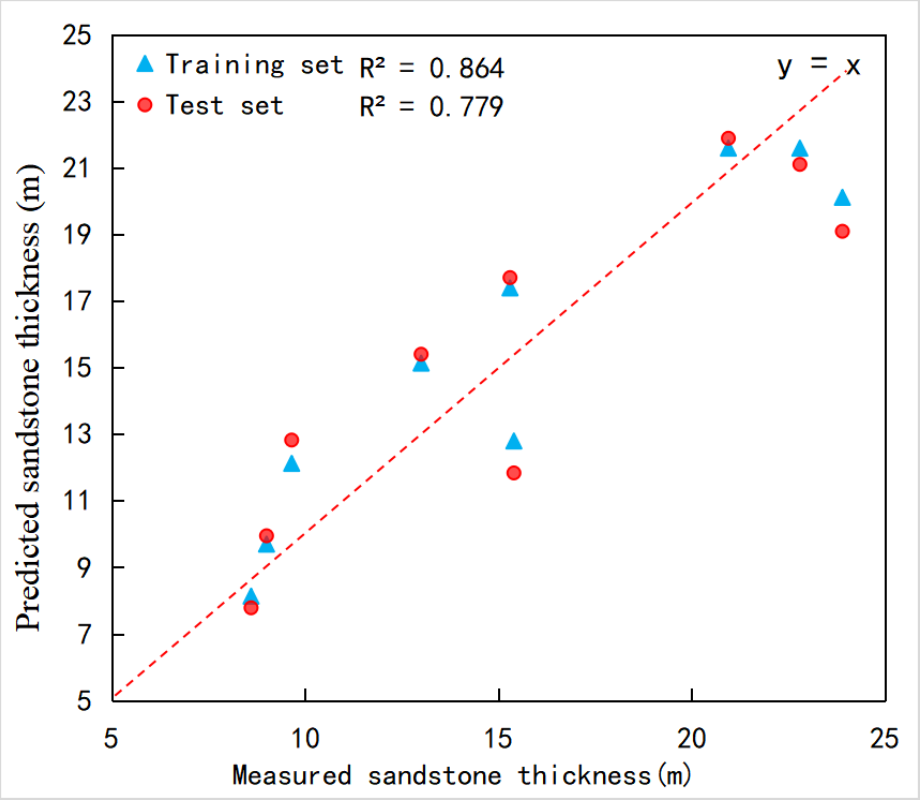


Figure 23. Relationship between predicted sandstone thickness from the Ridge Regression model and sandstone thickness from well.

Table 6. Calculated statistical parameters for the Ridge Regression model.

Data Set	R^2	MSE	MAE	RMSE	MAPE
Training set	0.864	4.302	1.789	2.074	12.23%
Test set	0.779	6.968	2.306	2.639	15.81%
Absolute difference	0.085	2.666	0.517	0.565	3.59%

4.6.3. Sandbodies Quantitative Prediction Results

Based on the trained models, sandbody thickness is predicted across the study area. The MLP model is applied to Zone I, while the Ridge Regression model is used for Zone II, and the results are integrated to generate a unified sandbody thickness distribution map (Figures 24–26). In Zone I, the southern shallow-water delta front exhibits a continuous, sheet-like distribution of thick sandbodies. High-value zones (>40 m) are observed around wells G111, T42, and G115, corresponding well with the measured data and indicating the development of large-scale distributary channel systems. In contrast, the northern region is characterized by thinner sandbodies (<22 m), displaying discontinuous, patchy distributions that reflect the distal delta front environment. In Zone II, sandbodies are generally thin and discontinuous, with sandstone thickness mostly below 36 m. No large contiguous high-value zones are observed; instead, localized increases in thickness occur near individual wells. This distribution pattern is consistent with the depositional characteristics of shore–shallow lake subfacies, which are dominated by mudstone with scattered beach-bar sandbodies. These results demonstrate the capability of the Ridge Regression model to capture weak and discontinuous sandbody signals under limited data conditions.

Overall, the study area exhibits a clear spatial pattern of thick, continuous sandbodies in the south and thin, discontinuous sandbodies in the north. The distribution of high-value zones aligns well with the inferred paleocurrent direction and the development of underwater distributary channels, indicating strong consistency with sedimentary facies interpretations. This consistency provides independent geological validation of the model's predictions, beyond statistical evaluation alone.

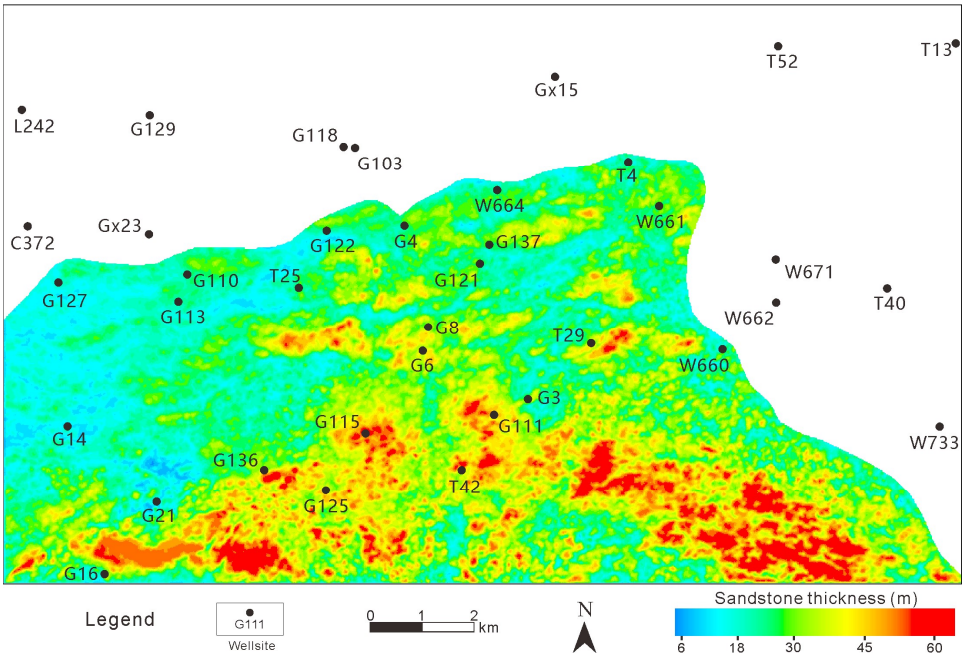


Figure 24. Predicted sandbodies distribution of Zone I.

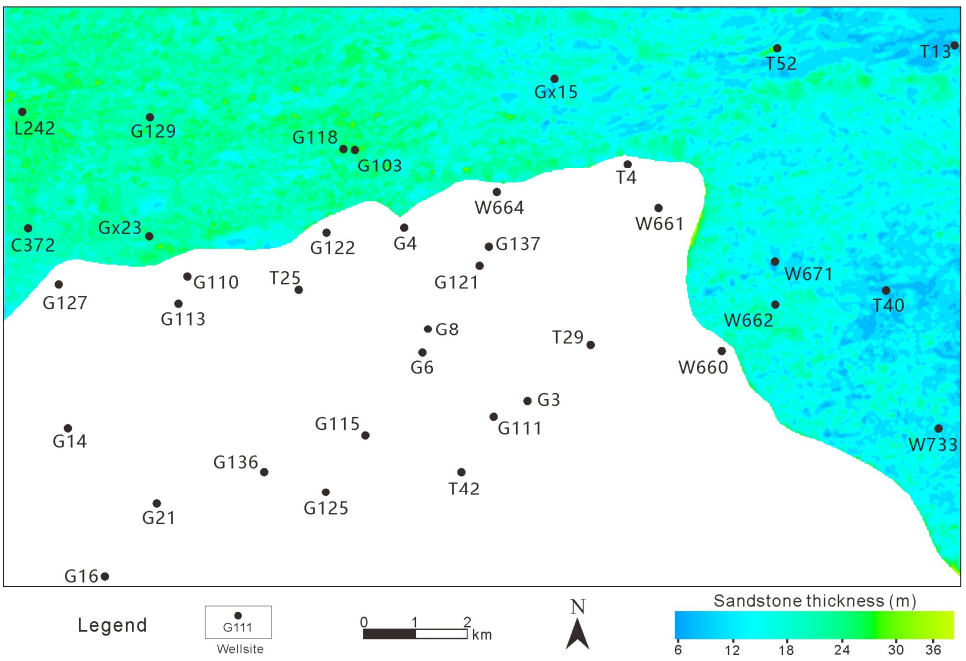


Figure 25. Predicted sandbodies distribution of Zone II.

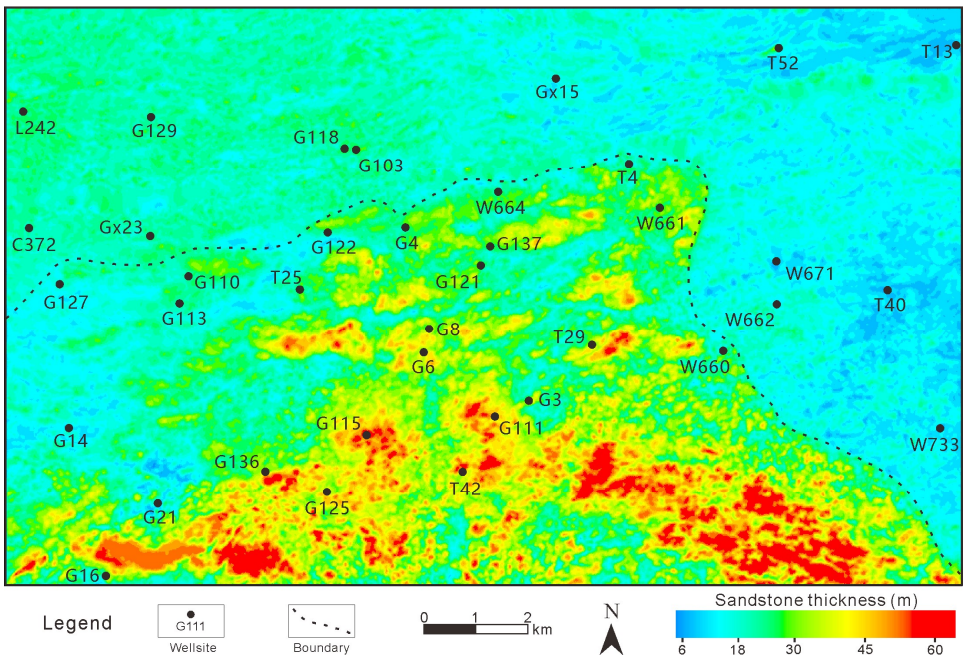


Figure 26. Predicted sandbodies distribution of the entire study area.

To further assess the effectiveness of the proposed method, a comparative analysis is conducted using a conventional linear regression model based on the most correlated seismic attribute (Average Instantaneous Frequency). Although this attribute shows the highest correlation (R = 0.67), the resulting prediction exhibits significant deviations from measured values in several areas, particularly around wells T40 and T52 (Figure 27). The quantitative comparison of model performance (Table 7) demonstrates that the proposed facies-constrained multi-attribute machine learning approach significantly outperforms the conventional linear regression method in both R^2 and error metrics. The predicted sandbody thickness distribution shows strong agreement with sedimentary facies patterns, further confirming the effectiveness and reliability of facies-controlled modeling.

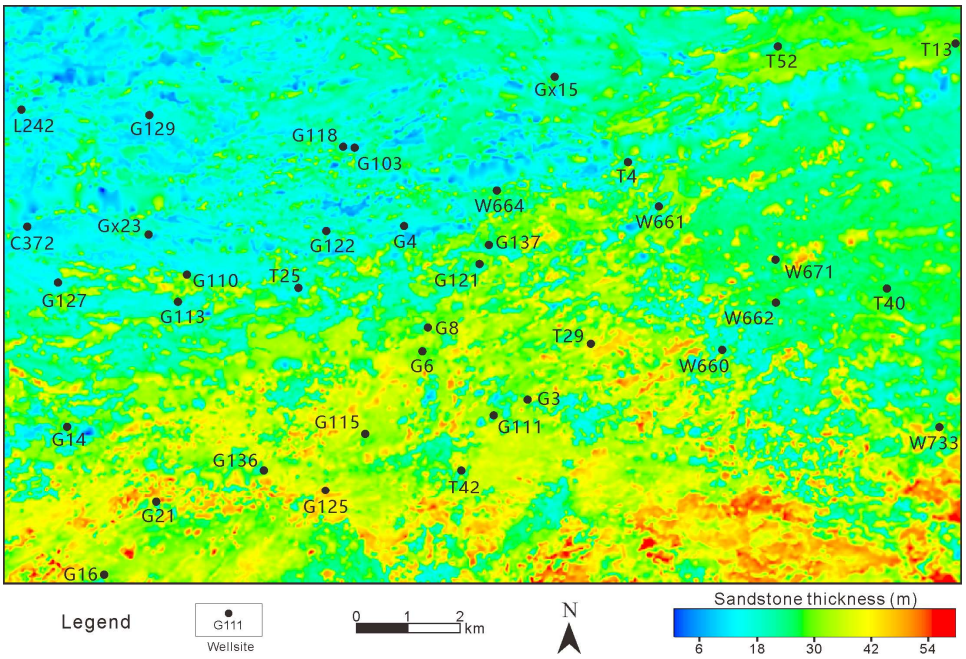


Figure 27. The sandbody predicted by the simple linear regression model.

Table 7. Comparison of statistical parameters of different sandstone prediction methods.

Type	Prediction Methods	R^2	MAE	MRE	RMSE
①	Simple linear regression	0.450	6.479	37.47%	8.375
②	MLP and Ridge regression	0.858	2.571	11.77%	3.181

4.6.4. Prediction Limitations and Corresponding Optimization Strategies

In this study, sandbody thickness prediction is conducted under sedimentary facies constraints by integrating seismic attribute optimization and machine learning models, based on a relatively limited dataset of 31 wells (22 in Zone I and 9 in Zone II). Although the proposed workflow effectively addresses variations in sandbody identification across different sedimentary environments, several limitations remain, primarily related to sample size, well distribution, and seismic resolution.

First, well control across the study area is limited. Although the 22 wells in Zone I allow for the division of training, test, and validation datasets to support neural network modeling, they are still insufficient to fully capture the spatial heterogeneity of sandbody distribution. In Zone II, only nine wells are available, representing a small-sample scenario. Under such conditions, conventional dataset partitioning becomes less effective, and models' performance may be influenced by individual well data, thereby constraining generalization capability.

Second, the spatial distribution of wells is uneven. Wells are densely distributed in the southern region, where thick sandbodies are developed, but are sparse in the northern region, where sandbodies are thinner. This uneven distribution leads to biased learning, with the model achieving higher accuracy in well-controlled areas while exhibiting reduced prediction reliability in sparsely drilled regions.

Third, the vertical resolution of seismic data imposes inherent limitations. The vertical resolution in the study area is approximately 20 m. The thickness of a single sandbody layer ranges from 0.5 to 18 m, which is lower than the seismic resolution. However, the prediction in this study is based on cumulative sandbody thickness,

with sandbody thickness ranging from 8 to 54 m. Most major sandbodies exceed the resolution limit and can be reliably identified; however, thinner sandbodies (<20 m), particularly in the shore–shallow lake beach-bar zones in the northern region, fall below the seismic resolution threshold. These thin layers exhibit weak reflection responses, making accurate characterization challenging.

To address these limitations, several mitigation strategies are implemented. These include facies-constrained attribute optimization, L2 regularization, leave-one-out cross-validation (LOOCV), independent blind well validation, consistency verification between linear and nonlinear models, and geological validation through sedimentary facies analysis. These approaches collectively reduce the overfitting risk and enhance the model's reliability. In addition, multi-attribute fusion combined with nonlinear machine learning models enables the integration of seismic information related to reflection energy, gradient, and frequency. This approach helps capture the composite response of sub-resolution sandbodies and reduces the ambiguity associated with a limited seismic resolution.

Overall, the limitations related to well distribution and seismic resolution primarily affect the prediction of thin sandbodies in the northern region, while having minimal impact on the prediction of thick, laterally continuous sandbodies in the southern region. Multiple validation results demonstrate that the model's predictions are consistent with geological patterns and maintain acceptable error levels, satisfying the requirements for sandbody distribution mapping and reservoir evaluation in the study area. Future improvements may include increasing well density, acquiring higher-resolution seismic data, and incorporating multi-model fusion strategies to further enhance predictions' accuracy, particularly for thin sandbodies and sparsely sampled regions.

5. Conclusions

This study investigates the quantitative prediction of thin sandbodies in the lower sub-member of the fourth member of the Shahejie Formation (Es4) in the Chenguanzhuang area of the Dongying Depression. By integrating sedimentary facies analysis, seismic attribute optimization, and machine learning, a facies-constrained prediction framework is established to address the challenges associated with low seismic resolution and high prediction uncertainty in complex sedimentary environments. The results demonstrate that the proposed approach effectively improves prediction accuracy and reliably captures the spatial distribution of sandbodies. The main conclusions are summarized as follows:

(1) Sedimentary facies characterization provides a fundamental geological constraint for modeling. Based on integrated analysis of core data, well logs, and 3D seismic data, two subfacies are identified within the $Es_4^{L\text{-}1}$ interval: the shallow-water delta front (Zone I) and the shore–shallow lake (Zone II). Zone I is characterized by thick, laterally continuous underwater distributary channel sandbodies, whereas Zone II is dominated by thin, discontinuous beach-bar sandbodies. This facies-based zonation provides a reliable geological basis for targeted modeling.

(2) Seismic data preprocessing and attribute optimization significantly enhance prediction reliability. Frequency enhancement and diffusion filtering improve seismic resolution and data quality, facilitating thin sandbody identification. Facies-constrained attribute optimization using Pearson correlation and hierarchical clustering selects five optimal attributes for Zone I and two for Zone II, with maximum correlation coefficients of 0.863 and −0.918, respectively. This process reduces redundancy and strengthens the relationship between seismic attributes and sandbody thickness.

(3) A facies-differentiated modeling strategy improves predictions' accuracy and robustness. In the data-rich Zone I (22 wells), the MLP model achieves the best performance among the five algorithms, with R^2 values of 0.856 (training) and 0.936 (test), and an

RMSE of 3.474. Cross-validation with Ridge Regression confirms the model's stability. In the data-limited Zone II (nine wells), Ridge Regression combined with LOOCV effectively mitigates overfitting, yielding R^2 values of 0.864 (training) and 0.779 (test). Compared with conventional linear regression (R^2 = 0.450), the proposed method demonstrates superior accuracy and robustness under varying geological conditions.

(4) The predicted sandbody distribution is consistent with sedimentary patterns and provides practical exploration guidance. The predicted results reveal a clear spatial pattern, with thick, continuous sandbodies developed in the southern region and thin, discontinuous sandbodies distributed in the northern region. This distribution is highly consistent with sedimentary facies evolution and paleogeomorphological controls, providing a reliable geological basis for hydrocarbon exploration and development planning in the study area.

Author Contributions: Conceptualization, C.L. and J.L.; methodology, J.L.; software C.L. and C.E.; validation, A.F.; formal analysis, C.L.; investigation, J.L.; resources, C.L.; data curation, C.L.; writing—original draft preparation, J.L.; writing—review and editing, C.L. and C.E.; visualization, J.L. and A.F.; supervision, C.L. and C.E.; project administration, C.L.; and funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the National Natural Science Foundation of China (No. 42372156), the Key R&D Program of Shandong Province, China (Grant No. 2022CXPT048), and the Research Contract "Seismic Sedimentology Interpretation Methods Testing" (Contract No. 30200000-19-FW2099-0047), and the China Scholarship Council (grant no. 202106450061).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Acknowledgments: We acknowledge the Geophysical Research Institute of Shengli Oilfield Company, SINOPEC, for providing core samples, well log data, and geological data for this study. Special thanks to Fenglei Li of Shandong Institute of Petroleum and Chemical Technology and Guoyin Zhang, Yuqi Wu, Xianguo Zhang, Lihua Ren, Yuan Zhan, Pingshan Ma, Huafeng Liu and Guiang Li of China University of Petroleum (East China) for their constructive suggestions and for revising the manuscript. The discussion with them has played a positive role in our research and article writing. We also sincerely thanks Beijing Knoxda Petroleum Technology Co., Ltd. for providing the Geoscope software (Version 4.0.0) and Curtin University for providing the Petrel software (Version 2018.2), which provided important support for the smooth implementation of this study. Special thanks are extended to the editor and anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Zhao, L.; Zhu, X.; Zhao, X.; You, Y.; Xu, M.; Wang, T.; Geng, J. Deep Carbonate Reservoir Characterization Using Multiseismic Attributes: A Comparison of Unsupervised Machine-Learning Approaches. *Geophysics* **2024**, *89*, B65–B82. [CrossRef]
2. Piryonesi, S.M.; El-Diraby, T.E. Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *J. Transp. Eng. Part B Pavements* **2020**, *146*, 04020022. [CrossRef]
3. Isah, A.; Tariq, Z.; Mustafa, A.; Mahmoud, M.; Okoroafor, E.R. A Review of Data-Driven Machine Learning Applications in Reservoir Petrophysics. *Arab. J. Sci. Eng.* **2025**, *50*, 20343–20377. [CrossRef]
4. Li, G.; Lin, C.; Wu, Y.; Ma, P.; Tahmasebi, P.; Dong, C.; Liu, W.; Du, X.; Zhao, Z. Machine Learning Applications in Tight Porous Media: Challenges, Advances, and Future Directions. *Earth-Sci. Rev.* **2025**, *271*, 105306. [CrossRef]

5. Qin, D.; Li, Y.; Li, J.; Peng, L.; Liu, Q.; Liu, X. Multi-Parameter Neural Network Deep Learning with a Novel Amplitude-versus-Offset Attribute for Hydrocarbon Detection in the X Gas Field of the Xihu Sag. *J. Geophys. Eng.* **2025**, *22*, 1863–1875. [CrossRef]
6. Gholami, A.; Ansari, H.R. Estimation of Porosity from Seismic Attributes Using a Committee Model with Bat-Inspired Optimization Algorithm. *J. Pet. Sci. Eng.* **2017**, *152*, 238–249. [CrossRef]
7. Liu, L.; Li, W.; Du, Y.; Yue, D.; Zhang, X.; Hou, J. Reservoir prediction method of fusing frequency-decomposedseismic attributes using Stacking ensemble learning. *Oil Geophys. Prospect.* **2024**, *59*, 12–22. [CrossRef]
8. Otchere, D.A.; Arbi Ganat, T.O.; Gholami, R.; Ridha, S. Application of Supervised Machine Learning Paradigms in the Prediction of Petroleum Reservoir Properties: Comparative Analysis of ANN and SVM Models. *J. Pet. Sci. Eng.* **2021**, *200*, 108182. [CrossRef]
9. Wang, Q.; Wang, Z.; Gao, D.; Gao, Z.; Jia, J.; Zhu, J.; Gao, J. Seismic Attribute Analysis with a Combination of Convolutional Autoencoder and Random Forest in a Turbidite Reservoir. *Geophysics* **2024**, *89*, WA207–WA217. [CrossRef]
10. Zou, C.; Zhao, L.; Xu, M.; Chen, Y.; Geng, J. Porosity Prediction With Uncertainty Quantification From Multiple Seismic Attributes Using Random Forest. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2021JB021826. [CrossRef]
11. Ntibahanana, M.; Jianguo, S.; Luemba, M.; Tondozi, K.; Imani, G.; Mohamed, B. Ensemble of Neural Networks Utilizing Seismic Attributes for Rock-Property Inversion With Uncertainty Estimation. *Earth Space Sci.* **2024**, *11*, e2023EA003101. [CrossRef]
12. Huang, D.; Song, X.; Wang, Y.; Qin, G. Advanced Seismic Sedimentology Techniques for Characterizing Shallow-Water Fan Deltas: Reservoir Architecture and Sedimentary Evolution of the Upper Karamay Formation, Bai21 Area, NW Junggar Basin, China. *Appl. Sci.* **2025**, *15*, 2306. [CrossRef]
13. Lou, M.; Cai, H.; He, X.; Liu, Y.; Huang, X.; Zhang, X.; Liu, H. Application of Seismic Sedimentology in Characterization of Fluvial-Deltaic Reservoirs in Xihu Sag, East China Sea Shelf Basin. *Pet. Explor. Dev.* **2023**, *50*, 138–151. [CrossRef]
14. Zhang, X.; Lin, C.; Zhang, T. Seismic Sedimentology and Its Application in Shallow Sea Area, Gentle Slope Belt of Chengning Uplift. *J. Earth Sci.* **2010**, *21*, 471–479. [CrossRef]
15. Li, N.; Zhang, J.; Matsushima, J.; Song, C.; Luan, X.; Dou, M.; Chen, T.; Wang, L. Sensitivity Analysis of Seismic Attributes and Oil Reservoir Predictions Based on Jointing Wells and Seismic Data—A Case Study in the Taoerhe Sag, China. *Front. Mar. Sci.* **2022**, *9*, 1011770. [CrossRef]
16. Sun, L.-P.; Zheng, X.-D.; Shou, H.; Li, J.-S.; Li, Y.-D. Quantitative Prediction of Channel Sand Bodies Based on Seismic Peak Attributes in the Frequency Domain and Its Application. *Appl. Geophys.* **2010**, *7*, 10–17. [CrossRef]
17. Li, H.; Gao, R.; Wang, Y. Predicting the Thickness of Sand Strata in a Sand-Shale Interbed Reservoir Based on Seismic Facies Analysis. *J. Geophys. Eng.* **2020**, *17*, 592–601. [CrossRef]
18. Anees, A.; Zhang, H.; Ashraf, U.; Wang, R.; Thanh, H.V.; Radwan, A.E.; Ullah, J.; Abbasi, G.R.; Iqbal, I.; Ali, N.; et al. Sand-Ratio Distribution in an Unconventional Tight Sandstone Reservoir of Hangjinqi Area, Ordos Basin: Acoustic Impedance Inversion-Based Reservoir Quality Prediction. *Front. Earth Sci.* **2022**, *10*, 1018105. [CrossRef]
19. Zrelli, A.; Amiri, A.; Sebei, K.; Abidi, O.; Barhoumi, N.; Inoubli, M.H. Reservoir Quality Assessment Based on Seismic Inversion and Petrophysical Modeling: A Case Study of the Silurian Acacus Formation, Ghadames Basin, Southern Tunisia. *Acta Geophys.* **2026**, *74*, 106. [CrossRef]
20. Manzoor, U.; Ehsan, M.; Radwan, A.E.; Hussain, M.; Iftikhar, M.K.; Arshad, F. Seismic Driven Reservoir Classification Using Advanced Machine Learning Algorithms: A Case Study from the Lower Ranikot/Khadro Sandstone Gas Reservoir, Kirthar Fold Belt, Lower Indus Basin, Pakistan. *Geoenergy Sci. Eng.* **2023**, *222*, 211451. [CrossRef]
21. Zhao, X.; Chen, X.; Chen, W.; Liu, M.; Yao, Y.; Wang, H.; Zhang, H.; Yao, G. Quantitative Classification and Prediction of Diagenetic Facies in Tight Gas Sandstone Reservoirs via Unsupervised and Supervised Machine Learning Models: Ledong Area, Yinggehai Basin. *Nat. Resour. Res.* **2023**, *32*, 2685–2710. [CrossRef]
22. Huang, W.; Wang, J.; Chen, H.; Yang, J.; Wu, J.; Ma, N.; Meng, Z.; Jing, Y.; Xu, F.; Ning, B.; et al. Application of Multi-Wave and Multi-Component Seismic Data in the Description on Shallow-Buried Unconsolidated Sand Bodies: Example of Block J of the Orinoco Heavy Oil Belt in Venezuela. *J. Pet. Sci. Eng.* **2021**, *205*, 108786. [CrossRef]
23. Liu, Y.; Xie, P.; Duan, D.; Yao, Y.; Liu, P.; Li, W.; Chen, B.; Wang, W.; Hou, J. Multiple Seismic Attributes Fusion Approach with Support Vector Regression and Forward Simulation for Sand Body Prediction and Sedimentary Facies Interpretation—A Case of the X Gas Field in Xihu Sag. *Interpretation* **2022**, *11*, SA33–SA45. [CrossRef]
24. Liu, C.; Wang, P. The Role of Algal Blooms in the Formation of Lacustrine Petroleum Source Rocks—Evidence from Jiyang Depression, Bohai Gulf Rift Basin, Eastern China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2013**, *388*, 15–22. [CrossRef]
25. Feng, Y.; Jiang, S.; Hu, S.; Li, S.; Lin, C.; Xie, X. Sequence Stratigraphy and Importance of Syndepositional Structural Slope-Break for Architecture of Paleogene Syn-Rift Lacustrine Strata, Bohai Bay Basin, E. China. *Mar. Pet. Geol.* **2016**, *69*, 183–204. [CrossRef]

26. Zahid, M.A.; Chunmei, D.; Lin, C.; Gluyas, J.; Jones, S.; Zhang, X.; Munawar, M.J.; Ma, C. Sequence Stratigraphy, Sedimentary Facies and Reservoir Quality of Es4s, Southern Slope of Dongying Depression, Bohai Bay Basin, East China. *Mar. Pet. Geol.* **2016**, *77*, 448–470. [CrossRef]
27. Wang, J.; Pang, Y.; Cao, Y.; Peng, J.; Liu, K.; Liu, H. Sedimentary Environment Constraints on the Diagenetic Evolution of Clastic Reservoirs: Examples from the Eocene "Red-Bed" and "Gray-Bed" in the Dongying Depression, China. *Mar. Pet. Geol.* **2021**, *131*, 105153. [CrossRef]
28. Thigpen, B.B.; Dalby, A.E.; Landrum, R. Special report of the Subcommittee on polarity standards. *Geophysics* **1975**, *40*, 694–699. [CrossRef]
29. Zeng, H.; Hentz, T.F. High-Frequency Sequence Stratigraphy from Seismic Sedimentology: Applied to Miocene, Vermilion Block 50, Tiger Shoal Area, Offshore Louisiana. *Bulletin* **2004**, *88*, 153–174. [CrossRef]
30. Caers, J.; Avseth, P.; Mukerji, T. Geostatistical Integration of Rock Physics, Seismic Amplitudes, and Geologic Models in North Sea Turbidite Systems. *Lead. Edge* **2001**, *20*, 308–312. [CrossRef]
31. Zeng, H.; Kerans, C. Seismic Frequency Control on Carbonate Seismic Stratigraphy: A Case Study of the Kingdom Abo Sequence, West Texas. *Bulletin* **2003**, *87*, 273–293. [CrossRef]
32. Sun, X.; Du, S.; Tang, L. Coherent-enhancing anisotropic diffusion filtering technique. *Oil Geophys. Prospect.* **2004**, *39*, 651–655. [CrossRef]
33. Zeng, H.; Zhao, W.; Xu, Z.; Fu, Q.; Hu, S.; Wang, Z.; Li, B. Carbonate Seismic Sedimentology: A Case Study of Cambrian Longwangmiao Formation, Gaoshiti-Moxi Area, Sichuan Basin, China. *Pet. Explor. Dev.* **2018**, *45*, 830–839. [CrossRef]
34. Zeng, H.; Backus, M.M. Interpretive Advantages of 90°-Phase Wavelets: Part 1—Modeling. *Geophysics* **2005**, *70*, C7–C15. [CrossRef]
35. Zeng, H.; Backus, M.M. Interpretive Advantages of 90°-Phase Wavelets: Part 2—Seismic Applications. *Geophysics* **2005**, *70*, C17–C24. [CrossRef]
36. Zeng, H. Stratal Slicing: Benefits and Challenges. *Lead. Edge* **2010**, *29*, 1040–1047. [CrossRef]
37. Zeng, H.; Ambrose, W.A.; Villalta, E. Seismic Sedimentology and Regional Depositional Systems in Mioceno Norte, Lake Maracaibo, Venezuela. *Lead. Edge* **2001**, *20*, 1260–1269. [CrossRef]
38. Chopra, S.; Marfurt, K.J. Seismic Attributes—A Historical Perspective. *Geophysics* **2005**, *70*, 3SO–28SO. [CrossRef]
39. Ezugwu, A.E.; Ikotun, A.M.; Oyelade, O.O.; Abualigah, L.; Agushaka, J.O.; Eke, C.I.; Akinyelu, A.A. A Comprehensive Survey of Clustering Algorithms: State-of-the-Art Machine Learning Applications, Taxonomy, Challenges, and Future Research Prospects. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104743. [CrossRef]
40. Saxena, A.; Prasad, M.; Gupta, A.; Bharill, N.; Patel, O.P.; Tiwari, A.; Er, M.J.; Ding, W.; Lin, C.-T. A Review of Clustering Techniques and Developments. *Neurocomputing* **2017**, *267*, 664–681. [CrossRef]
41. Govender, P.; Sivakumar, V. Application of *k*-Means and Hierarchical Clustering Techniques for Analysis of Air Pollution: A Review (1980–2019). *Atmos. Pollut. Res.* **2020**, *11*, 40–56. [CrossRef]
42. Wasserman, P.D.; Schwartz, T. Neural Networks. II. What Are They and Why Is Everybody so Interested in Them Now? *IEEE Expert* **1988**, *3*, 10–15. [CrossRef]
43. Lashkarbolooki, M.; Hezave, A.Z.; Ayatollahi, S. Artificial Neural Network as an Applicable Tool to Predict the Binary Heat Capacity of Mixtures Containing Ionic Liquids. *Fluid Phase Equilibria* **2012**, *324*, 102–107. [CrossRef]
44. Mohammadi, M.-R.; Hemmati-Sarapardeh, A.; Schaffie, M.; Husein, M.M.; Ranjbar, M. Application of Cascade Forward Neural Network and Group Method of Data Handling to Modeling Crude Oil Pyrolysis during Thermal Enhanced Oil Recovery. *J. Pet. Sci. Eng.* **2021**, *205*, 108836. [CrossRef]
45. Qian, X.; Wang, B.; Chen, J.; Fan, Y.; Mo, R.; Xu, C.; Liu, W.; Liu, J.; Zhong, P. An Explainable Ensemble Deep Learning Model for Long-Term Streamflow Forecasting under Multiple Uncertainties. *J. Hydrol.* **2025**, *662*, 133968. [CrossRef]
46. Gardner, M.W.; Dorling, S.R. Artificial Neural Networks (the Multilayer Perceptron)—A Review of Applications in the Atmospheric Sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [CrossRef]
47. Smith, G.; Campbell, F. A Critique of Some Ridge Regression Methods. *J. Am. Stat. Assoc.* **1980**, *75*, 74–81. [CrossRef]
48. Hoerl, R.W. Ridge Regression: A Historical Context. *Technometrics* **2020**, *62*, 420–425. [CrossRef]
49. Zhao, K.; Hu, X.; Cai, J.; Yang, S.; Yu, Z.; Jiang, Y.; Chen, L. Unveiling Hydrothermal Dynamics: Ridge Regression-Driven Prediction of Water-Rock Interaction Fractured Aquifer in Coal Mine Gobs during Thermal Energy Storage. *J. Energy Storage* **2026**, *147*, 120121. [CrossRef]
50. Zhu, X.; Pan, R.; Zhao, D.; Liu, F.; Wu, D.; Li, Y.; Wang, R. Formation and development of shallow-water deltasin lacustrine basin and typical case analyses. *J. China Univ. Pet.* **2013**, *37*, 7–14.
51. Zhu, X.; Zeng, H.; Li, S.; Dong, Y.; Zhu, S.; Zhao, D.; Huang, W. Sedimentary Characteristics and Seismic Geomorphologic Responses of a Shallow-Water Delta in the Qingshankou Formation from the Songliao Basin, China. *Mar. Pet. Geol.* **2017**, *79*, 131–148. [CrossRef]

52. Zhifeng, S.; Chengyan, L.; Dongxing, D.; Haisheng, B.; Huaiqiang, R. Application of Seismic Architecture Interpretation in Enhancing Oil Recovery in Late Development Stage—Taking Meandering River Reservoir in Dongying Depression as an Example. *J. Pet. Sci. Eng.* **2020**, *187*, 106769. [CrossRef]
53. Yu, J.; Jiang, X. Application of Seismic Attribute Optimization in Reservoir Prediction. *Oil Gas Geol.* **2003**, *24*, 291–295.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.