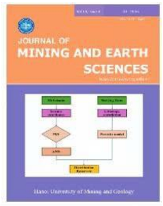
Journal of Mining and Earth Sciences

Website: https://jmes.humg.edu.vn



Application of principal component analysis on seismic attributes to predict the distribution of early miocene reservoirs in the northeast Bach Ho

Muoi Duy Nguyen *, Anh Ngoc Le, Ngan Thi Bui, Hong Minh Thi Nguyen, Hang Thu Thi Nguyen

Hanoi University of Mining and Geology, Hanoi, Vietnam

ARTICLE INFO

Article history:
Received 15th Mar. 2025
Revised 18th June 2025
Accepted 29th June 2025

Keywords:
Artificial neural network,
Bach Ho field,
Cuu Long basin,
PCA,
Seismic attribute,
UNN.

ABSTRACT

This study addresses the challenge of predicting the distribution of early Miocene reservoir rocks in the Northeast Bach Ho field, located in the Cuu Long basin, which has been a significant site for oil and gas exploration. The study aims to apply advanced seismic attribute analysis combined with the Principal Component Analysis (PCA) method to enhance the accuracy of reservoir distribution predictions in an area with limited wells data. The seismic attributes analyzed include Seismicity, RMS amplitude, Instantaneous frequency, Coverage, RAI, Instantaneous phase, Sweetness, and t attenuation, which were used to determine seismic facies classification.*

Four principal components, PC0, PC1, PC2, and PC3, were selected after analyzing the seismic attributes, accounting for 85.44% of the total variance. These components were then used as inputs for training an Unsupervised Neural Network (UNN), a method particularly useful when well data are sparse or unavailable. The output of the trained network, combined with facies analysis and porosity data from well logs, was employed to predict reservoir distribution in the study area.

The results showed that the sandstone deposits in the early Miocene sediments serve as potential reservoir rocks. These potential reservoirs are primarily located around the central part of the study area, with additional deposits in the eastern, northeastern, and western regions. The reservoirs are associated with river and deltaic sedimentary environments. The combination of PCA and UNN effectively reduced noise in the seismic data, leading to clearer identification of seismic facies classification and improved reservoir prediction. This integrated method is proven to be highly effective in improving oil and gas exploration strategies, particularly in regions with limited or no well data, and can support further assessment of the oil and gas potential of the Northeast Bach Ho field and the broader Cuu Long basin.

**Corresponding author*
E - mail: nguyenduymuoi@humg.edu.vn
DOI: 10.46326/JMES.2025.66(4).01

1. Introduction

The Bach Ho oil field, located in the Cuu Long basin, is one of the most important oil and gas exploration regions on Vietnam's continental shelf (Figure 1). With over 30 years of production, the field has shown significant potential, especially in the fractured granitoid basement and Miocene sedimentary reservoirs. Despite extensive exploration efforts, predicting the distribution of reservoir rocks in this area remains difficult due to limited well data and complex geological structures.

The main goal of this study is to enhance the prediction of reservoir distribution by using seismic attribute analysis combined with Principal Component Analysis (PCA) and Unsupervised Neural Network (UNN) methods. These approaches provide a more effective solution in areas with limited well control, which is the case in the Northeast Bach Ho area. Applying seismic attributes, such as seismicity, RMS amplitude, instantaneous frequency, and others, together with PCA, helps to reduce noise in seismic data and makes seismic facies easier to identify.

The study uses Principal Component Analysis (PCA) to identify the key principal components and applies the UNN method for seismic facies classification. This approach effectively predicts the distribution of potential reservoir rocks, focusing on early Miocene sandstones. The results highlight areas with high reservoir potential, mainly in the central, eastern, northeastern, and western parts of the study area, which are associated with fluvial and deltaic sedimentary environments.

This innovative integration of PCA and UNN markedly improves the accuracy of reservoir prediction, offering a robust tool for oil and gas exploration in the Cuu Long Basin. The approach proves particularly effective in areas with limited well data, enabling more efficient exploration strategies and optimizing oil and gas exploitation potential. Furthermore, the methodology demonstrates its versatility, showing promise for application in other sedimentary basins with comparable geological characteristics.

Bach Ho field has been an Oil and gas deposits in fractured granitoid basement rocks, as well as in lower Miocene and Oligocene crystalline sandstone-sandstone sediments, with the basement rock layer being the main oil reservoir, with large reserves and high productivity. Cuu

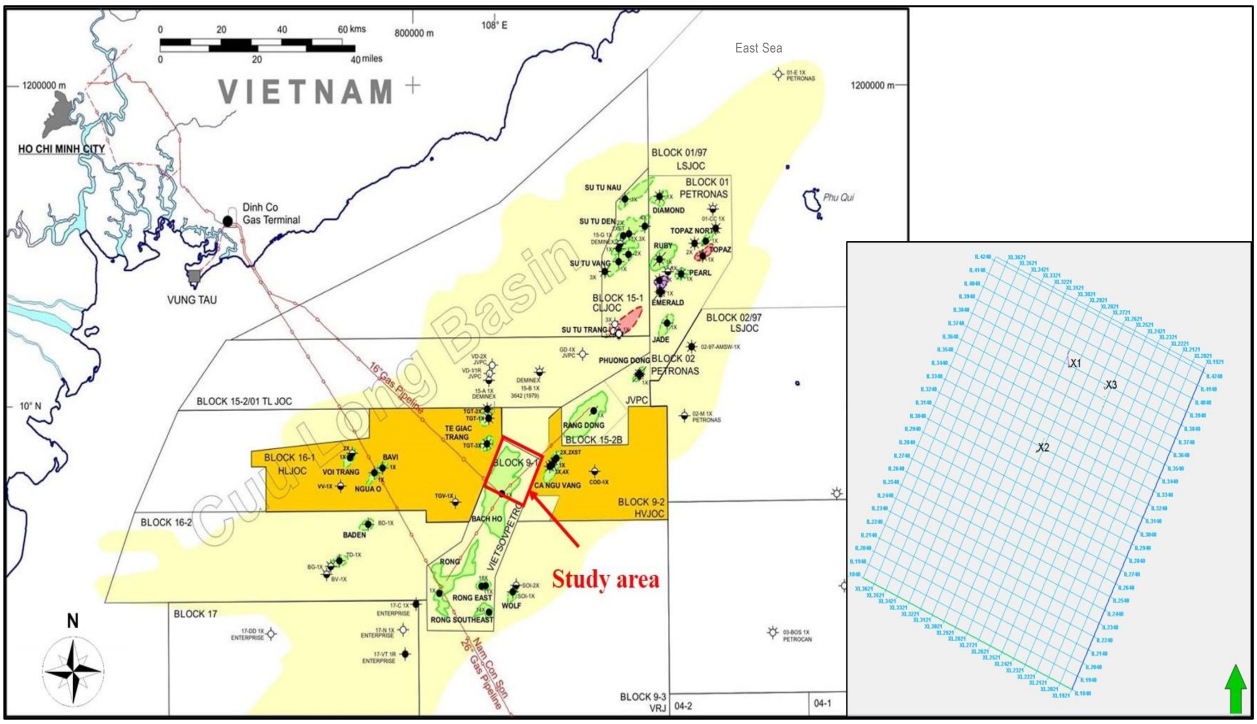


Figure 1. Location map of study area and wells in the seismic survey.

Long sedimentary basin was formed during the phases of India-Asia continental collision, rifting and sagging of the earth's crust, which occurred in the late Mesozoic to Cenozoic period. Extension and subsidence activities created a series of grabens, half-grabens, and horsts. The basin underwent three stages: pre-rift (before middle Eocene), syn-rift stage (middle Eocene- early Miocene), and post-rift (middle Miocene to present day) (Nguyen, 2009). The sedimentary stratigraphy and depositional environment of the Bach Ho field are consistent with the regional stratigraphic framework of the Cuu Long basin. The depositional environment in the Lower Miocene was fluvial, coastal plain, and shallow lacustrine/ marine conditions.

2. Database and Methodology

2.1. Database

The data used in this study include seismic and well data. Seismic used 3D-PSDM seismic with an area of 80 km^2. In general, the seismic data have high resolution and good quality, allowing interpretation and analysis of seismic properties. Interpretation and analysis of seismic properties were performed by the Petrel software. The early Miocene sedimentary suite in this study is defined by two reflectors, SH5 and SH7 (Figure 2). Well data, including 3 wells X1, X2, and X3 were used for facies analysis and porosity calculation (Figure 1). In addition, data collected and used to predict the distribution of reservoir rocks of the lower Miocene include petrographic analysis documents, petrographic samples, and well geological reports of the study area.

2.2. Methodology

2.2.1. Seismic attribute analysis

Seismic attribute methods have been widely used in oil and gas exploration. Seismic attributes help to detect deep geological structures of mass seismic and effectively calculate reservoir properties such as porosity, permeability, and reservoir saturation. They calculated from seismic and applied it to calculate reservoir parameters (Chopra and Marfurt, 2005; Taner, 2001). In this study, amplitude attributes are

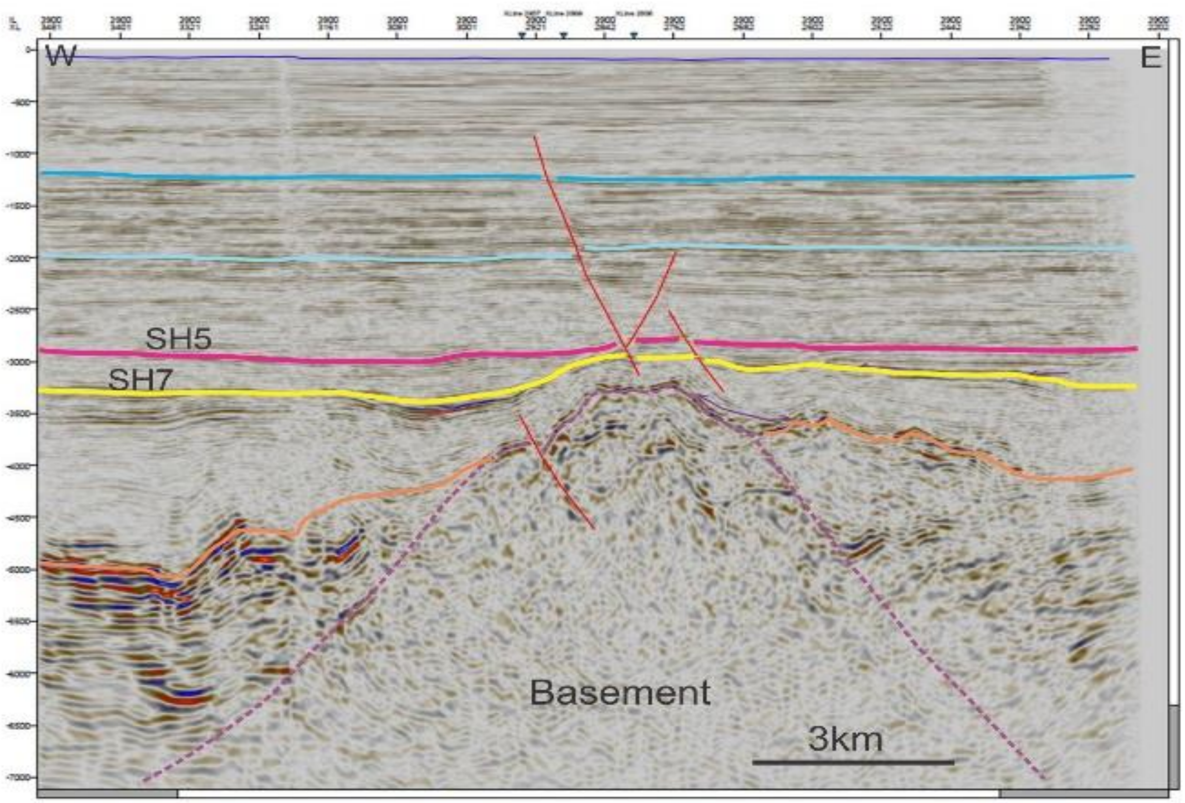


Figure 2. Seismic section through the study area and early Miocene object surrounded by SH5 and SH7.

mainly used to serve reservoir prediction. These attributes reflect quite accurately the changes in lithology and sedimentary facies, thereby providing a clearer view of the distribution of reservoir rocks in the study area. Combining seismic attribute analysis with principal component analysis has been applied to predict the distribution of lower Miocene reservoirs in the Cuu Long basin.

2.2.2 Principal component analysis

Principal Component Analysis (PCA) is a robust method for integrating seismic attributes derived from an initial dataset. The first principal component captures the maximum variability within the data, while each subsequent orthogonal component explains as much of the remaining variability as possible (Guo et al., 2009; Haykin, 2009; Scheevel and Payrazyan, 2001). When applied to seismic attributes from the same initial seismic volume, PCA highlights the attributes that account for the greatest variability, indicating that these combinations are more effective for identifying specific geological features. Essentially, PCA serves as a valuable tool in seismic interpretation by enabling the selection of seismic attribute orientations that yield more meaningful and accurate interpretation outcomes.

In this study, the PCA method is applied to reduce data dimensionality and extract the most meaningful principal components. The process involves calculating the covariance matrix of the seismic attributes, followed by determining the eigenvectors and eigenvalues of that matrix. Principal components are then selected based on the percentage of variability they explain. Here, PCA is used to isolate the principal components, which are subsequently utilized for training an unsupervised network in combination with well data to predict porosity distribution (Figure 3).

2.2.3. Artificial intelligence networks

Artificial Neural Networks (ANN) are computational models inspired by the human brain's structure and functionality, designed to identify patterns in data. These networks consist of multiple layers of interconnected nodes, or "neurons," which process input data and transmit information through activation functions. ANN

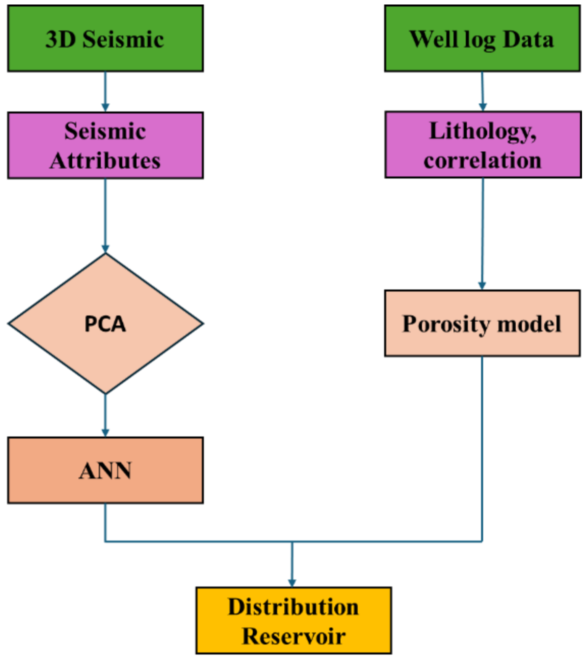


Figure 3. The workflow for prediction of distribution reservoir.

can be trained using two primary methods: supervised and unsupervised learning. In supervised learning, the network is trained with labeled data, where both inputs and the desired outputs are provided, allowing the model to learn by minimizing errors in predictions. On the other hand, unsupervised learning doesn't require labeled data and focuses on organizing the input data or discovering hidden patterns without predefined outcomes (Coleou et al., 2003).

This adaptability makes ANN highly versatile for applications such as seismic data analysis and reservoir characterization (Vu and Nguyen, 2021; Tran et al, 2020; Ta et al, 2019). ANN can be implemented using either supervised or unsupervised training techniques. Supervised training requires labeled training samples and specific desired outputs, while unsupervised training automatically categorizes the input data into layers based on user-defined parameters.

In this study, an unsupervised training method was employed in combination with Principal Component Analysis (PCA) to enhance the accuracy of reservoir distribution prediction. Unlike traditional ANN methods that rely on labeled data, the unsupervised approach here integrates PCA for dimensionality reduction, improving the performance of the model. The combination of PCA with Unsupervised Neural Networks (UNN) has shown significant effectiveness in processing complex seismic data, as it reduces noise and clarifies seismic facies detection. Compared to conventional seismic and geophysical analysis techniques, the UNN + PCA method achieved superior accuracy, exceeding 90%, and delivered more focused results for reservoir characterization.

3. Results and discussion

3.1. Well log interpretation

Three wells (X1, X2, and X3) were used for this study to interpret the geological properties of the subsurface and validate the seismic and PCA-based predictions. The well log analysis provides critical information on lithology, porosity, and facies distribution, which serves as ground truth data for evaluating the results from seismic attributes and PCA.

The analysis of the well logs revealed alternating layers of sand and clay. Specifically, in wells X1 and X2, thick sand layers were observed in the upper part of the section, with thick clay layers in the lower part. This sand-clay sequence suggests a transition from a riverine or deltaic environment to deeper, more lacustrine conditions. The gamma-ray (GR) logs of wells X1 and X2 displayed a cylindrical shape, reflecting a point bar environment, indicative of a fluvial depositional system (Figure 4).

In contrast, well X3 showed thicker sand layers in the lower section and thinner sands near the surface, which corresponds to a shifting depositional environment with more coarse-grained sediments at depth and finer sediments on top. The GR log in this well was bell-shaped, indicating a lacustrine environment (Figure 4). These well log interpretations were used to cross-check and validate the results of the seismic attribute analysis and PCA-derived facies classification.

The well log data not only provided insights into the lithological changes but also helped in calculating the porosity and other reservoir properties that were used as input for predicting the distribution of reservoir rocks.

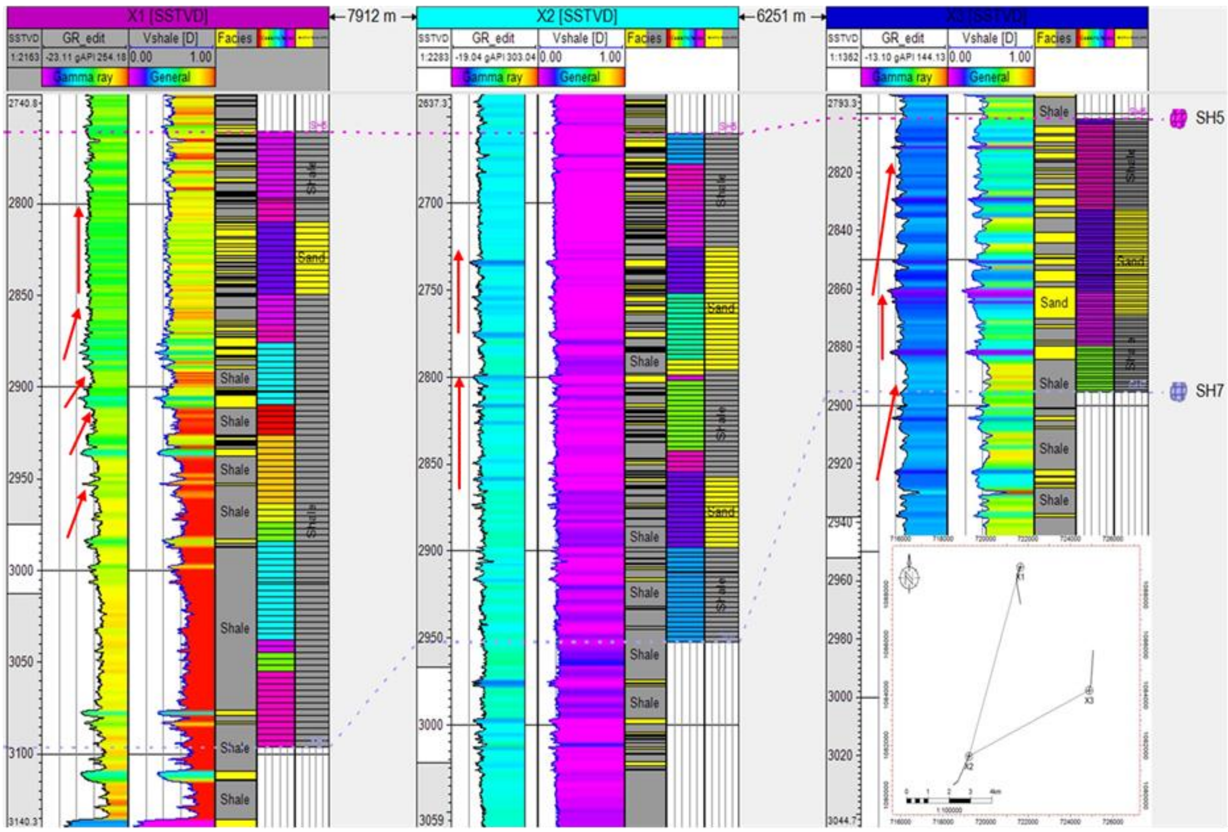


Figure 4. Well correlation and lithofacies classification from well logs.

This interpretation is essential for validating the facies identified from seismic data and ensuring the reliability of the reservoir distribution model.

3.2. Seismic attribute analysis

The seismic attribute analysis is a vital component in this study, as it allows for the identification of key geological features from the seismic data, which are crucial for reservoir prediction. The seismic attributes used in this study include seismicity, RMS amplitude, instantaneous frequency, coverage, RAI, instantaneous phase, sweetness, and t* attenuation (Figure 5).

Among these attributes, RMS amplitude, RAI, and sweetness were particularly effective in distinguishing seismic facies, as they reflect changes in lithology and sedimentary facies. RMS amplitude highlights variations in seismic energy that correlate with lithological boundaries, while RAI (Relative Acoustic Impedance) is useful for identifying variations in lithology, especially in terms of porosity and permeability. Sweetness, a measure of the amplitude spectrum of seismic data, also provided a clear indication of reservoir potential, particularly in areas where porous and permeable rocks were expected.

The frequency attributes, on the other hand, showed weaker results in distinguishing seismic facies. This could be attributed to the influence of noise in the data or the complexity of the geological features in the study area. Nonetheless, the overall seismic attribute analysis provided an essential foundation for identifying potential reservoir rocks by highlighting the zones with higher lithological contrasts and anomalies that are indicative of reservoir presence.

These seismic attributes were further processed through Principal Component Analysis (PCA), where the most significant components were selected based on their ability to reduce data dimensionality while preserving essential geological information. This analysis laid the groundwork for more advanced seismic facies classification and reservoir prediction.

3.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was applied to seismic attribute data to reduce dimensionality and identify the most significant components that account for the variance within the seismic dataset. By extracting key features, PCA enhances seismic interpretation by emphasizing the most critical geological characteristics, making it a highly effective tool for identifying and understanding underlying geological features.

The PCA process involved calculating the covariance matrix of the seismic attributes, followed by the determination of eigenvectors and eigenvalues, which represent the directions

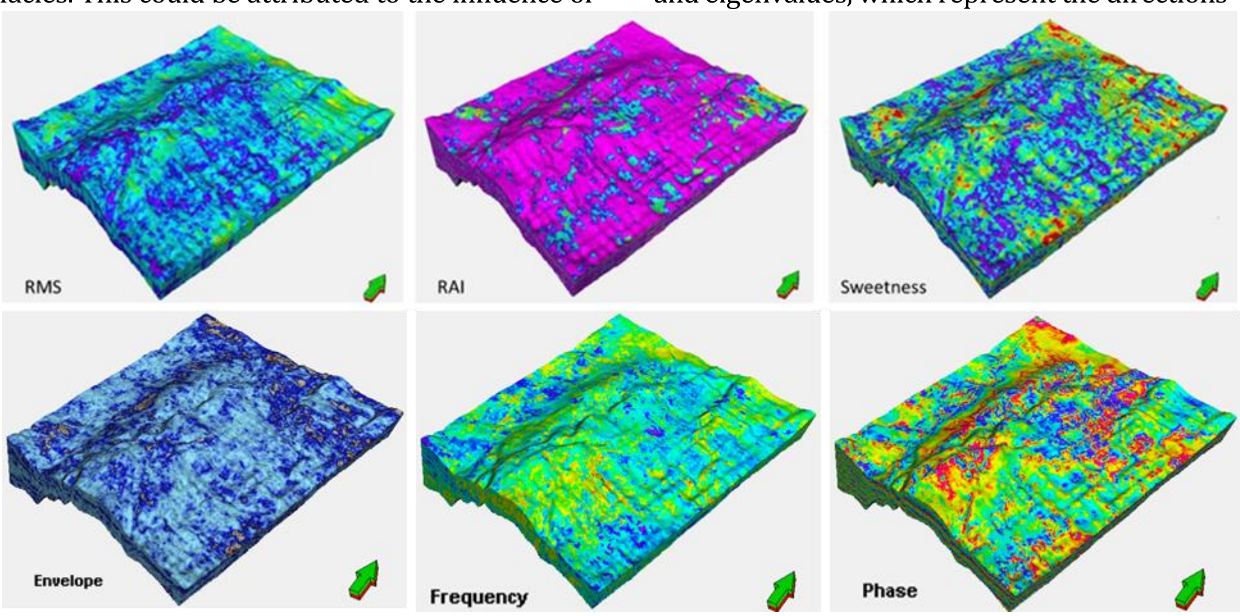


Figure 5. Seismic attribute Maps of Early Miocene sedimentary.

and magnitudes of the variability in the data. The first principal component (PC0) accounted for the highest variance, and each successive component (PC1, PC2, etc.) explained progressively less variance but still retained critical geological information.

In this study, the four principal components, PC0, PC1, PC2, and PC3-collectively accounted for 85.44% of the variance in the dataset (Figure 6). These components were chosen as they captured the most significant geological features while effectively minimizing noise in the seismic data. By concentrating on these four components, the analysis offered a clearer representation of subsurface properties, facilitating more precise seismic facies classification.

The selection of these components was based on both their statistical significance (eigenvalue greater than 1) and their geological relevance, ensuring that the selected components were consistent with the geological understanding of the study area. The PCA results played a crucial role in enhancing the quality of the seismic facies classification, which in turn improved the accuracy of the reservoir distribution predictions (Figure 6).

3.4. Unsupervised Neural Network (UNN)

The UNN model was trained to classify seismic data into distinct facies using the input principal components derived from PCA. This classification process produced a more precise and focused delineation of seismic facies compared to traditional methods. The resulting facies zones were distinctly separated, enabling more accurate predictions of potential reservoir areas.

To further enhance the reservoir distribution model, the UNN output was integrated with well log data, including porosity and facies information. This integration significantly improved the accuracy of predictions, particularly in regions with sparse well data. The performance of the UNN model was evaluated using metrics such as accuracy, sensitivity, and specificity, comparing the model's predictions with actual well data. The results demonstrated that the UNN model achieved an accuracy exceeding 90%, far surpassing the performance of conventional seismic and geophysical analysis methods.

The predicted reservoir distribution map, based on the UNN and PCA method, highlighted key zones with high potential for oil and gas reserves, particularly in the central, eastern, northeastern, and western parts of the study area. These areas are consistent with known geological features and depositional environments, such as river and deltaic systems. The results not only demonstrate the effectiveness of the UNN and PCA method in predicting reservoir distribution but also highlight its ability to reduce noise and improve the overall clarity of seismic facies interpretation. In this study, the input data includes 4 principal components (PC0-PC3) from PCA, and the output is the seismic facies classification. To improve the results, from 4 to 10 seismic facies levels were tested. The results are

Correlation Coefficients	PC0	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Seismic (default)	0.0214	0.4848	0.1525	0.8442	0.0564	0.0009	0.1593	0.0000
RMS	0.8768	-0.0023	-0.1527	0.0107	-0.1242	0.4382	0.0158	0.0000
RAI	0.0041	0.9497	0.0662	-0.0030	-0.0029	0.0216	-0.3054	0.0000
Sweetness	0.9662	0.0289	-0.1507	0.0130	-0.0218	-0.2053	-0.0047	0.0000
Envelope	0.9662	0.0289	-0.1507	0.0130	-0.0218	-0.2053	-0.0047	0.0000
Frequency	0.1935	-0.1078	0.8117	-0.0340	-0.5384	-0.0317	-0.0065	0.0000
Phase	-0.0305	0.8133	-0.0061	-0.5155	-0.0581	-0.0110	0.2616	0.0000
t*Attenuation	0.4070	-0.0554	0.6496	-0.1516	0.6199	0.0448	0.0058	0.0000
Eigenvalue	2.8405	1.8147	1.1773	1.0030	0.6970	0.2800	0.1874	0.0000
Contribution (%)	35.51	22.68	14.72	12.54	8.71	3.50	2.34	0.00
Cumulative Contribution (%)	35.51	58.19	72.91	85.44	94.16	97.66	100.00	100.00

Figure 6. Correlation Coefficients of PCA.

shown in Figure 7. The 10 seismic facies map shows the highest level of facies detail and was selected to be combined and reduced to 2 facies for sand and clay facies classification.

Combining the analysis results of wells X1, X2, X3, and the histogram of seismic facies (Figure 8) to build a model to predict the distribution of the Lower Miocene reservoir. Based on the distribution of the facies and well log analysis, the facies were combined into two groups. The tests were repeated, and the facies were coded. Each facies corresponds to a code numbered from 0÷9. In this study, facies 2, 4, and 7 were combined into a clay-containing facies and coded 0. Facies 1, 3, 5, 6, 8, 9, and 10 were grouped into a sand-containing group and coded 1. Tests were performed and the number of facies was reduced to 2 types: the corresponding attribute group inclined to clay-group 0 and group 1 inclined to sand, to aim at the goal of dividing the clay-sand reservoir (Figure 9).

Figure 9 shows the predicted distribution of sand-clay facies. For potential sand facies such as zones A, B, C, and D (yellow). In which they are mainly concentrated around the central part of

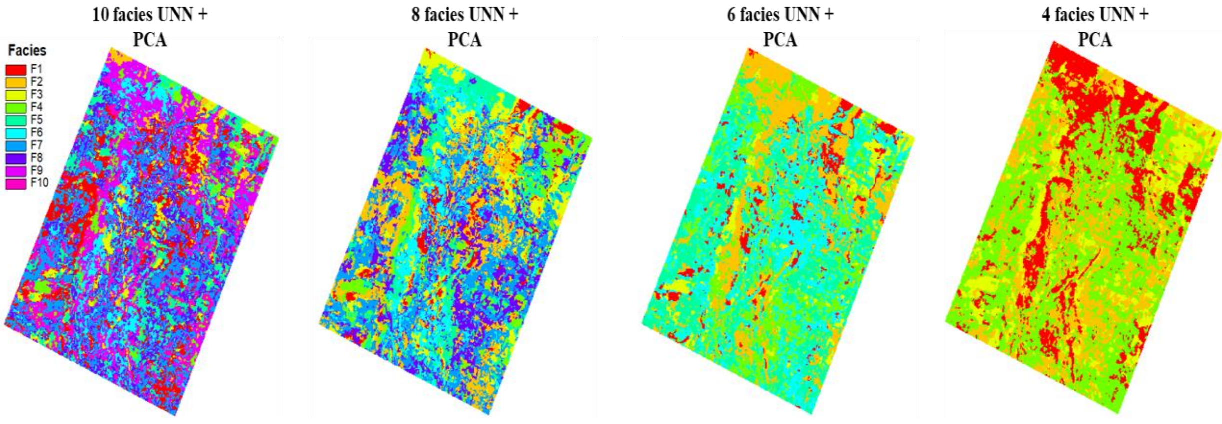


Figure 7. Result of seismic facies classification by UNN method combined with PCA.

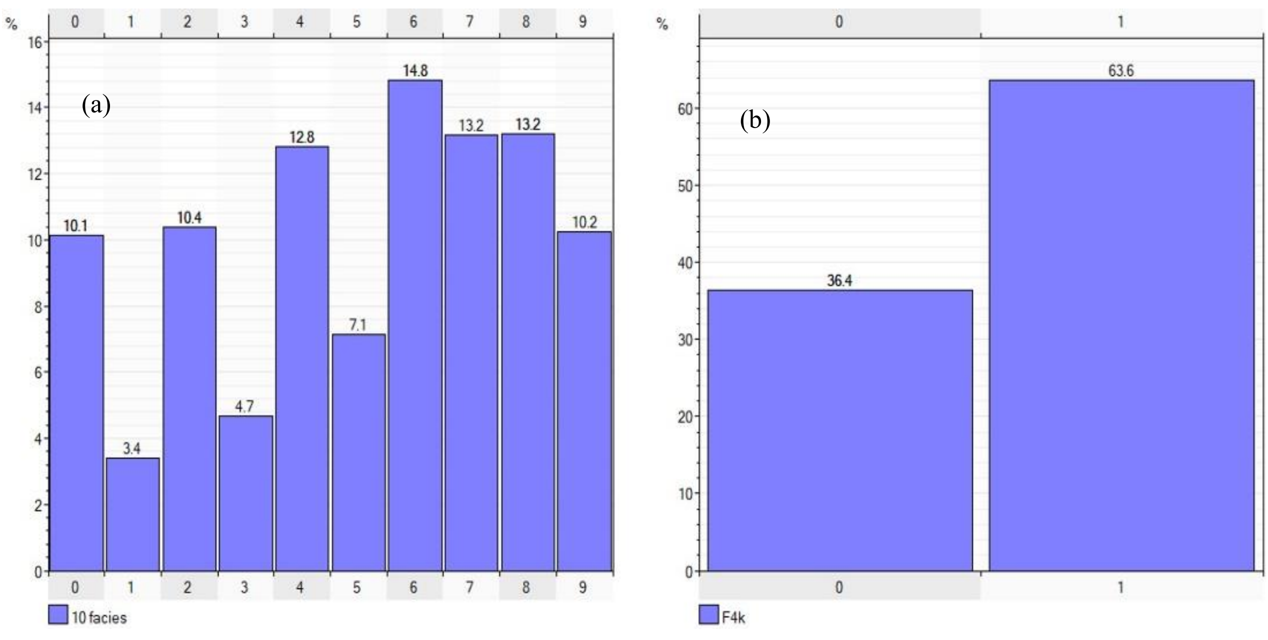


Figure 8. Distribution histogram of 10 seismic facies by UNN method combined with PCA: a) 10 seismic facies classes; b) 2 seismic facies classes (reduced from 10 classes).

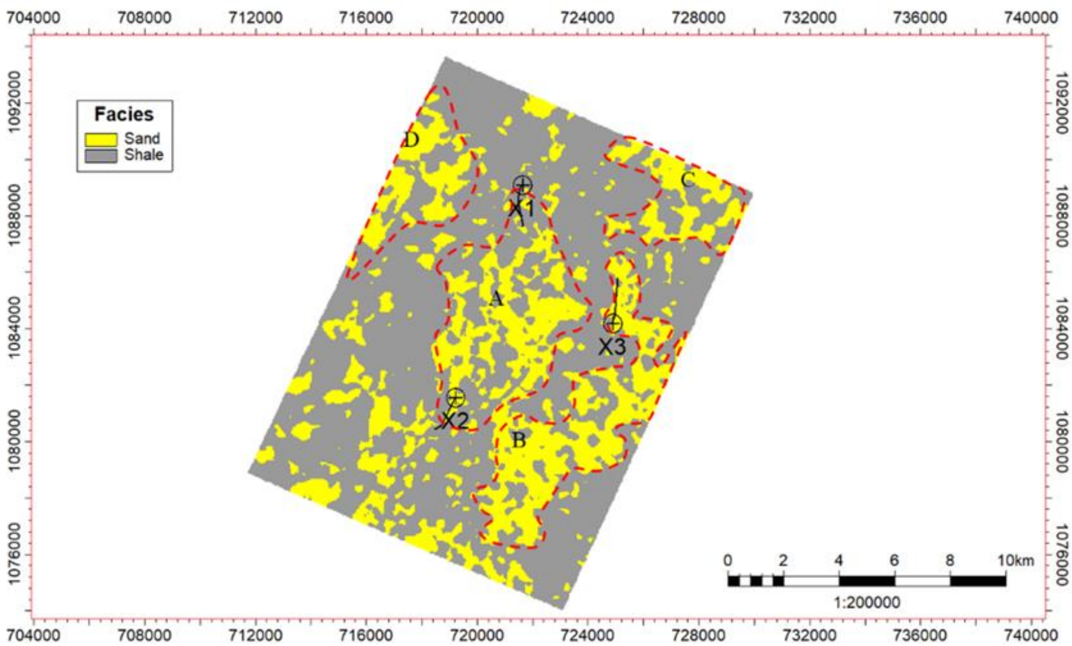


Figure 9. Illustration of sand and clay facies of the Early Miocene in the study area.

the Bach Ho uplift (zone A), the eastern part, and the northeastern part. Oligocene and early Miocene sediments of the Bach Ho field area are lacustrine and deltaic environments (Nguyen, 2009; Nguyen et al., 2023). The association of sedimentary environments with the study area shows that the sand facies were deposited in fluvial and deltaic environments. The results of the analysis of seismic attributes, facies models, and sedimentary conditions show reasonable consistency and reliability.

4. Conclusion

This study presents a significant advancement in the prediction of reservoir distribution in the Northeast Bach Ho field, located in the Cuu Long basin. The primary contribution of this research is the successful integration of seismic attribute analysis with Principal Component Analysis (PCA) and the Unsupervised Neural Network (UNN) method to predict the distribution of early Miocene reservoir rocks. The use of seismic attributes such as RMS amplitude, seismicity, and sweetness,… combined with PCA, allowed for the identification of four key principal components (PC0, PC1, PC2, and PC3) that explained 85.44% of the variance in the seismic data. These components significantly reduced data noise, enhancing the clarity of seismic facies identification.

The key finding of this study is the identification of potential reservoir zones, particularly the early Miocene sandstones, which are located in the central, eastern, northeastern, and western parts of the study area. These areas, which are associated with river and deltaic environments, show the highest potential for oil and gas reserves.

The effectiveness of the PCA and UNN method was further demonstrated by its ability to work with limited well data, a common challenge in exploration areas like the Northeast Bach Ho field. This combination of techniques not only improved the accuracy of the predictions but also provided a more reliable approach for reservoir characterization in regions with sparse data. The results of this study offer valuable insights for optimizing oil and gas exploration strategies in the Cuu Long basin and have the potential to be applied to other sedimentary basins with similar geological conditions.

In conclusion, the integration of PCA with the UNN model offers a powerful and data-driven approach to enhance reservoir prediction accuracy. This study demonstrates that combining seismic attribute analysis, PCA, and machine learning techniques delivers an effective solution to the challenges of reservoir characterization in complex geological settings. The results provide valuable insights that can guide future exploration efforts and improve the efficiency of oil and gas exploitation in the region.

Contributions of authors

Muoi Duy Nguyen - analyzed PCA and ANN and wrote the article; Anh Ngoc Le - analysed seismic attributes; Hong Minh Thi Nguyen - analyzed well logs; Ngan Thi Bui, Hang Thu Thi Nguyen - methodology and database.

Acknowledgments

This work was supported by Hanoi University of Mining and Geology under the project code T24-14.

References

Coleou, T., Poupon, M., & Azbel, K. (2003). Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge*, 22(10), 942-953.

Guo, H., Marfurt, K. J., & Liu, J. (2009). Principal component spectral analysis. *Geophysics*, 74(4), p35-43.

Chopra, S., & Marfurt, K. J. (2005). Seismic attributes-A historical perspective. *Geophysics*, 70(5), 3SO-28SO.

Haykin S., (2009). Neural network and learning machines, 3rd ed.: *Pearson*. 966 p.

Scheevel, J. R., & Payrazyan, K. (2001). Principal component analysis applied to 3D seismic data for reservoir property estimation. *SPE Reservoir Evaluation & Engineering*, 4(01), 64-72.

Taner, M. T. (2001). Seismic attributes: CSEG Recorder. *Rock Solid Images,* Houston, USA, 26(7), 49-56.

Nguyen, H. (2009). The petroleum Geology and Resources of Vietnam. Vietnam oil and Gas Group.

Ta, Q. D., Le, T. H., Pham, D. K. (2019). Using artifical neural network to predict porosity", *Petrovietnam Journal*, 2019, vol 7, pag. 18 - 27.

Tran, O. T., Pham, K. D., Van Hoang, Q., Nguyen, M. D., Bui, N. T., Nguyen, H. H. T., ... & Le Quoc, H. (2020). Application of artificial intelligence network to predict the distribution of volcanic material in sequence D, field X, Cuu Long basin. *Journal of Mining and Earth Sciences*, 61(5), 104-113.

Vu, D. H., & Nguyen, H. T. (2021). Estimation of shale volume from well logging data using Artificial Neural Network. *Journal of Mining and Earth Sciences*, p46-52.

Nguyen, M. D., Nguyen, H. M., & Bui, N. T. (2023). Application of artificial neural network and seismic attributes to predict the distribution of Late Oligocene sandstones in the Cuu Long basin. *Journal of Mining and Earth Sciences*, 64(3), 24-31.