scientific reports

Check for updates

OPEN

Unsupervised neural clustering and seismic attribute analysis for gas channel delineation in the Scarab field

Tarek Khalifa[1], Ahmed Eleslambouly[2], Amin Esmail Khalil[1] & Amir Ismail[1]✉

The western offshore Nile Delta of Egypt represents a significant hydrocarbon province, characterized by complex stratigraphy and challenging reservoir connectivity. This study investigates two vertically stacked, unconnected channel systems, Channels 1 and 2, which serve as the primary reservoirs in the Scarab Field. Conventional seismic interpretation techniques often fall short in delineating subtle stratigraphic features and reservoir architecture. To overcome these limitations, we employed an integrated machine learning workflow combining Principal Component Analysis (PCA) and Self-Organizing Maps (SOMs) to enhance reservoir characterization. After data conditioning and the calculation of seismic attributes, PCA was applied to reduce dimensionality and extract the most informative seismic attributes. The resulting principal components were input into SOMs, an unsupervised neural network approach used to classify seismic meta-attributes into meaningful lithological facies. The SOM successfully identified six distinct facies classes: siltstone, clayey sand, clayey limestone, gas-bearing sandstone, water-bearing sandstone, and interbedded siltstone, validated through well log calibration. In particular, the technique delineated gas-water contacts and subtle stratigraphic trends that are critical for both exploration and development planning. The results and quantitative analysis through classification accuracy and GWC misfit assessments using different data types demonstrate that unsupervised machine learning, supported by multi-attribute seismic analysis, offers important improvements over traditional interpretation methods in detecting Direct Hydrocarbon Indicators (DHIs), refining stratigraphic resolution, and reducing exploration risk.

Keywords Self-Organizing Map (SOM), Machine Learning (ML), Reservoir characterization, Nile delta, Lithofacies classification, Scarab field, Direct Hydrocarbon indicators (DHIs).

The Nile Delta Basin, situated offshore northern Egypt, represents one of the most prolific gas provinces in the Eastern Mediterranean and has been the site of several giant hydrocarbon discoveries (Fig. 1a)[1]. The West Delta Deep Marine (WDDM) concession is located 50–100 km offshore Egypt within the Nile Delta Basin and covers ~ 6150 km² along the northwestern margin of the Nile Delta (Fig. 1b), accounting for nearly two-thirds of Egypt's gas production[2,3]. Gas in the offshore Nile Delta is predominantly produced from clastic reservoirs associated with canyon-fill, turbidite, submarine fan, and slope-channel deposits, accumulated during successive depositional episodes from the Oligocene through the Pleistocene[4–6].

The Scarab Field (Fig. 1c) is located within the WDDM concession, approximately 50–100 km offshore the northwestern Nile Delta in water depths ranging from 250 to 850 m (Fig. 1c)[2,4]. Proven gas accumulations in the field occur within Pliocene channelized sandstones of the El Wastani Formation, which were deposited as slope-channel canyon fills derived from the Nile River's sediment supply into the deepwater deltaic system[7,8]. Research on the Scarab Field has largely concentrated on reservoir-scale characterization, including architectural and reservoir quality analysis of channelized sand bodies[9,10], static reservoir modeling integrating seismic and well-log data[11,12], and regional-to-field scale seismic attributes and interpretation of depositional and structural frameworks[2,3,13–16]. While previous studies of the Scarab Field have largely relied on interpreter-driven seismic analysis, a systematic application of machine learning to integrate seismic attributes with well-log data for unbiased facies classification and gas-channel delineation remains absent. Addressing this gap constitutes the main objective of the present study.

[1]Department of Geology, Faculty of Science, Capital University, Cairo, Egypt. [2]Department of Earth Sciences, Khalifa University, Abu Dhabi, United Arab Emirates. ✉email: amir_ismail@science.capu.edu.eg

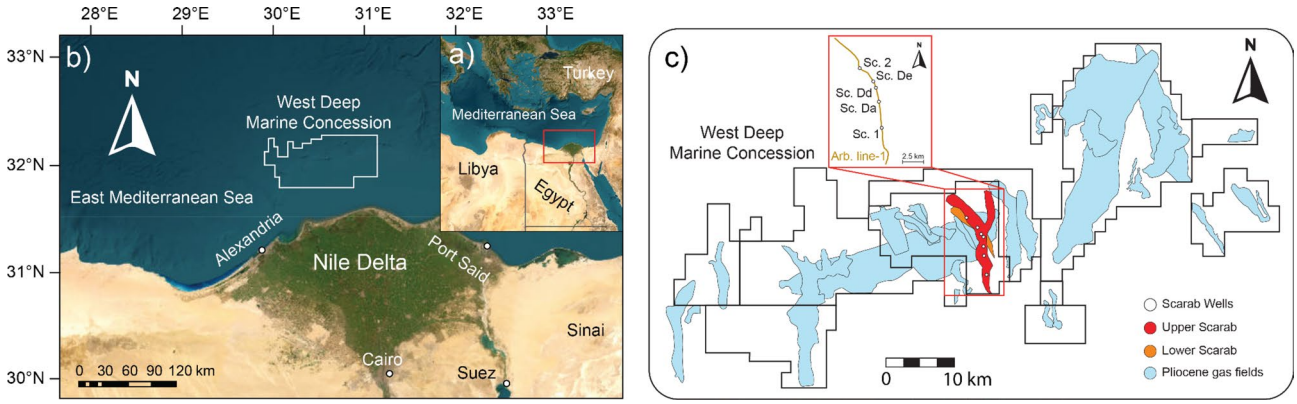


Fig. 1. Satellite and Schematic maps show key geographic domains. (**a**) Regional map showing the location of the Offshore Nile Delta within the Eastern Mediterranean. (**b**) Satellite view highlighting the position of the West Delta Deep Marine (WDDM) concession offshore northern Egypt. (**c**) Schematic map of the WDDM concession, indicating the distribution of Pliocene gas fields, Scarab wells, and the two Scarab channel systems (Upper Channel in red and Lower Channel in orange). The inset highlights the location of the 2D arbitrary seismic line (Line-1) used for subsequent analysis. The figures were created in ArcGIS Pro (V3.6, https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview), a commercial geographic information system from ESRI.

Seismic facies classification is a crucial step in understanding the depositional history and reservoir properties, making it extremely useful in hydrocarbon exploration and development. Traditionally, skilled interpreters delineate seismic facies on seismic lines by visually scanning waveforms for frequency, amplitude, phase, and geometric relationships with neighboring sections to produce a 2D seismic facies map. Seismic attributes provide quantitative estimates of these properties. The number of seismic attributes has also increased dramatically, providing increasingly accurate measurements of reflector morphology while greatly expanding the amount of data to be analyzed. However, conventional seismic interpretation often struggles to resolve subtle stratigraphic features within channelized deepwater reservoirs, particularly when high-dimensional attribute datasets overwhelm interpreter-driven analysis[17]. Unsupervised machine learning offers a powerful alternative by extracting latent patterns from seismic data, thereby eliminating subjective bias[17,18]. PCA reduces dimensionality and isolates the most geologically significant attributes, while SOM enables robust classification of lithofacies and reservoir facies in multi-attribute[19,20]. Although PCA-SOM workflows have been applied in different global basins and settings[17,19], few studies have applied PCA-SOM in the Nile Delta[17], while none have systematically evaluated their utility for direct hydrocarbon indicator (DHI) detection and gas-channel delineation in the Scarab field.

This study establishes an integrated PCA-SOM workflow to advance seismic stratigraphic interpretation in the Scarab field, offshore Nile Delta. Specifically, we (i) apply PCA to identify and select the most informative seismic attributes, (ii) employ SOM for lithofacies classification and high-resolution mapping of channelized reservoirs, and (iii) validate the results against well-log data to demonstrate enhanced detection of gas-bearing intervals and associated DHIs. By minimizing interpreter bias and enhancing facies resolution, this study presents a systematic application of PCA-SOM to the Nile Delta, providing a transferable methodology that reduces exploration risk, refines facies mapping, and improves reservoir characterization.

Geological setting of Scarab field

The West Delta Deep Marine (WDDM) concession occupies the northwestern margin of the Nile Delta cone and represents one of the most productive segments of the basin (Fig. 1b and c). The stratigraphic sedimentary succession of the area spans the Miocene to Holocene and includes key reservoir-bearing formations such as the Qantara, Sidi Salim, Qawasim, Abu Madi, Kafr El Sheikh, and El Wastani formations (Fig. 2)[21–23]. The Pliocene El Wastani Formation hosts the primary reservoirs of the Scarab Field and consists of channelized sandstones deposited as slope-channel canyon fills and turbidites sourced from the Nile River sediment discharge[7,14]. These deposits exhibit rapid lateral facies transitions and vertical stacking, resulting in strong heterogeneity at the reservoir scale[10]. Moreover, these units consist primarily of interbedded sandstones and shales, deposited within a slope-to-basin floor fan system.

The region's sedimentary fill reflects multiple depositional cycles, influenced by regional transgressive-regressive events and tectonic reactivations. Structurally, the WDDM block is bounded by three principal fault systems: (1) the Rosetta Fault (SW-NE trending), the largest in the region and associated with strike-slip to extensional movement; (2) rotated fault blocks along the northwestern margin; and (3) E-W trending faults in the northeast[26,27,28]. These structures play a key role in compartmentalizing the reservoir and influencing fluid flow pathways. The broader tectonic evolution includes compressional deformation during the Syrian Arc Orogeny, Miocene uplift-related faulting, and Messinian regression processes, which contributed to stratigraphic erosion and slope instability[29,30]. Source rocks within the region are primarily associated with Early Pliocene black shales and early Miocene marine sediments, exhibiting high Total Organic Content (TOC) and favorable thermal maturity[31,32].

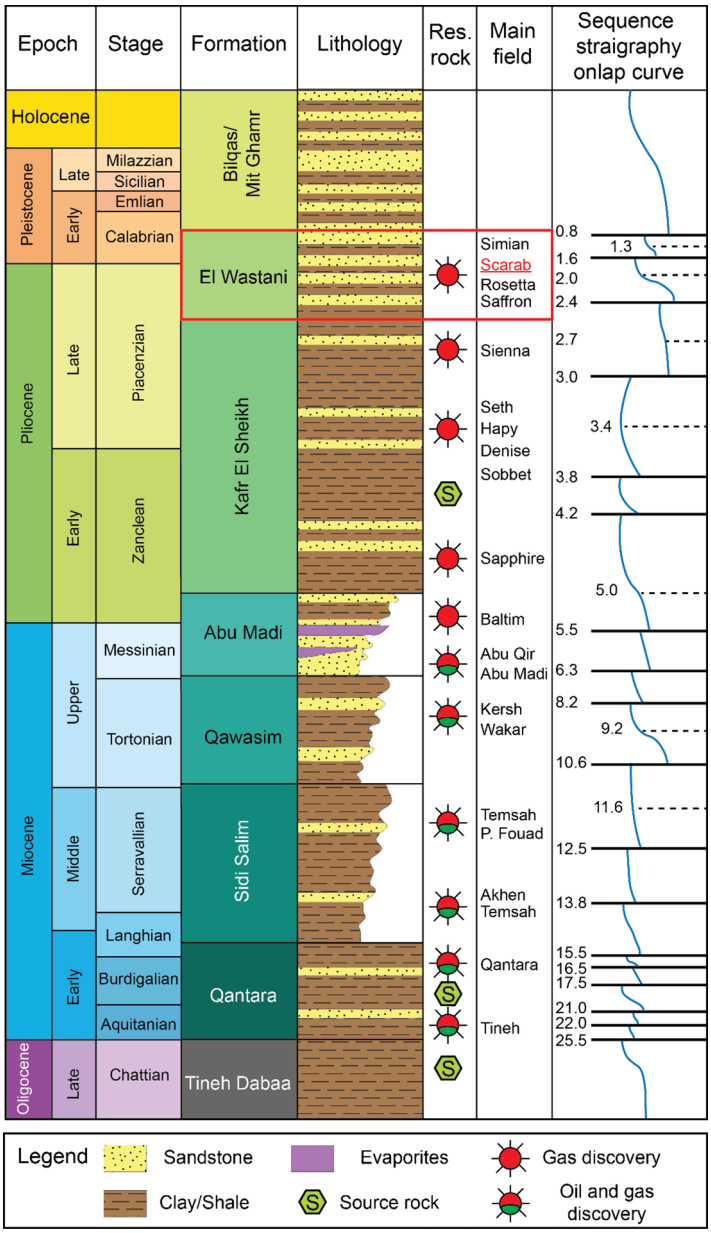


Fig. 2. Generalized stratigraphic column of the Nile Delta Basin spanning the Oligocene to Holocene, highlighting key lithostratigraphic units, reservoir intervals, and main hydrocarbon discoveries. The sand-rich channel and fan deposits serve as the primary gas producers. Interbedded shales act as regional seals, while source rocks are mainly associated with the Qantara, Sidi Salim, and Tineh formations. Modified after Kamel, et al.[24] and Younes[25].

The Scarab Field was discovered in 1998 via the Scarab-1 and Scarab-2 exploration wells (Fig. 1c). The field is part of a broader Pliocene-age submarine channel complex[3,4,9,27], and is currently developed through multiple wells, including Scarab-Da, Scarab-Dd, and Scarab-De. Peak gas production exceeded 800 million standard cubic feet per day (MMSCFD), although current rates have since declined[14]. Geologically, the Scarab Field comprises two intersecting channel systems, the Upper Channel and Lower Channel, which trend in different directions but share comparable depositional environments. These channels exhibit significant lateral and vertical heterogeneity, indicative of a dynamic turbidite-dominated depositional regime. According to El-Mowafy, et al.[33]. The channel complex stratigraphy is characterized by abrupt facies transitions, reflecting rapid sedimentation and varying energy conditions. Petrophysical analysis of the reservoir indicates a gross thickness of approximately 170 m, with net pay intervals reaching up to 50 m. Average porosity and water saturation are both around 28%, highlighting the reservoir's moderate quality and potential for gas accumulation[10].

Methods and materials

Dataset

This study employs 2D PSTM seismic data acquired by the Egyptian General Petroleum Corporation (EGPC) across a 19-km profile within the WDDM concession, complemented by data from two exploration and three production wells (Scarab-1, Scarab-2, Scarab-Da, Scarab-Dd, and Scarab-De). These wells are equipped with a full suite of conventional well-log datasets and Vertical Seismic Profile checkshots. The seismic data were acquired with a bin size of 9.375 × 12.5 m, a sampling interval of 4 ms, and a dominant frequency of ~ 25 Hz. The processed dataset contains 380,231 samples and a spectral bandwidth spanning 10–80 Hz, providing sufficient vertical resolution to interrogate channelized reservoirs.

The Scarab Field lies within a highly prospective Neogene petroleum system where hydrocarbon accumulations are controlled by a combination of structural traps, stratigraphic pinch-outs, and channel-lobe complexes (Figs. 2 and 3). Reservoir quality distribution and performance are strongly influenced by the interplay of fault architecture, facies heterogeneity, and stratigraphic stacking patterns[7,14]. The channel complex system shows a major bright spot across the field but exhibits flat spots at the gas-water contact (GWC). This variability underscores the limitations of conventional seismic interpretation and highlights the need for high-resolution, data-driven approaches.

Workflow

The present study employs an integrated machine learning workflow (Fig. 4) to enhance lithofacies classification and reservoir characterization through a machine learning approach that combines multiple 2D seismic attributes and well-log data in a heterogeneous reservoir setting. The approach combines statistical dimensionality reduction and unsupervised neural clustering to extract hidden patterns and stratigraphic features not visible in conventional amplitude interpretation. The workflow begins with conventional seismic interpretation to define the target reservoir intervals, followed by the generation of 28 seismic attributes, encompassing both instantaneous and geometric properties. PCA is then applied to reduce attribute dimensionality, retaining only the most significant components that preserve geological variability.

Attribute combinations that reveal stratigraphic or anomalous features are selected and fed into a SOM, configured with an 8 × 8 topology. This unsupervised neural network clusters data based on latent-feature similarities, yielding enhanced facies discrimination. Simultaneously, well log data undergo petrophysical analysis and the generation of lithofacies logs. The SOM output is then calibrated by cross-checking against lithofacies logs to ensure the neural clusters are geologically meaningful. The classified SOM facies data is further used for gas sand detection, GWC delineation, and mapping depositional variability across the field.

Seismic-to-well tie

To establish an accurate time-depth relationship across the field, synthetic seismograms were generated for all the utilized wells using a deterministic zero-phased wavelet extracted from the 2D post-stack seismic data (Fig. 5). The wavelet was then applied to reflectivity series derived from acoustic impedance logs to generate synthetic seismograms. The check-shot surveys were used to establish an initial time-depth relationship, which was later refined by manually matching synthetic peaks and troughs to the seismic data. The time-depth relationships derived from the well-seismic ties were used to identify the time window of interest (1700–2300 ms). The computed cross-correlation coefficients indicate good alignment between synthetic and seismic traces (0.78 for Scarab-1 and 0.72 for Scarab-2), with all remaining wells exceeding 0.70, ensuring a robust depth-to-time

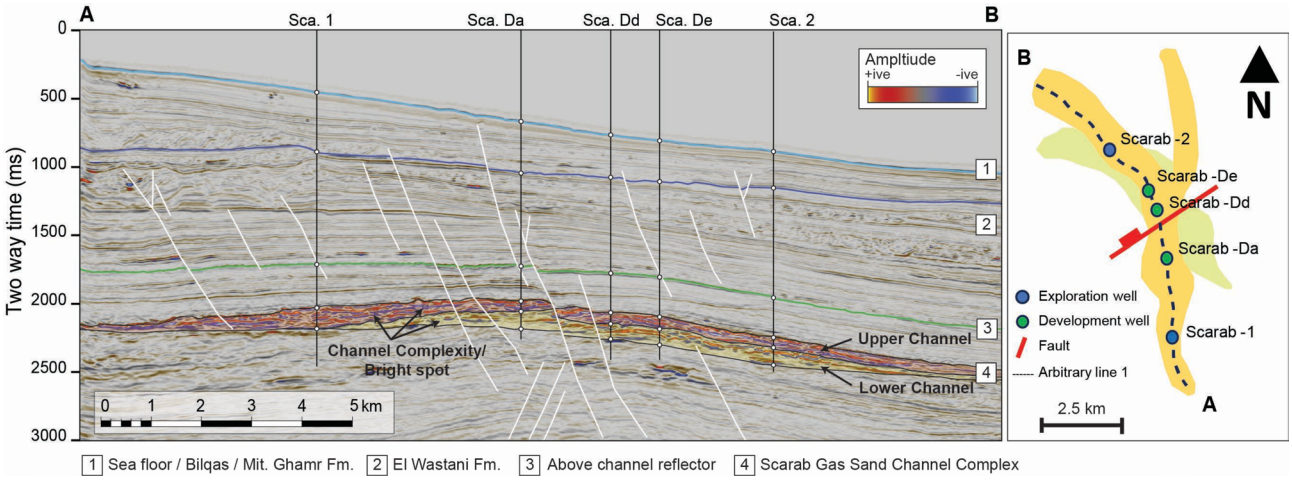


Fig. 3. Arbitrary 2D seismic section across the Scarab Field showing key stratigraphic horizons and channel architecture. The section ties five wells (Scarab-1, Scarab-Da, Scarab-Dd, Scarab-De, and Scarab-2) and highlights the vertically stacked Scarab gas sand channel complexes. The upper and lower channels of the El Wastani Formation are outlined by strong amplitude anomalies. The schematic map shows well locations and the section's position relative to mapped channels.

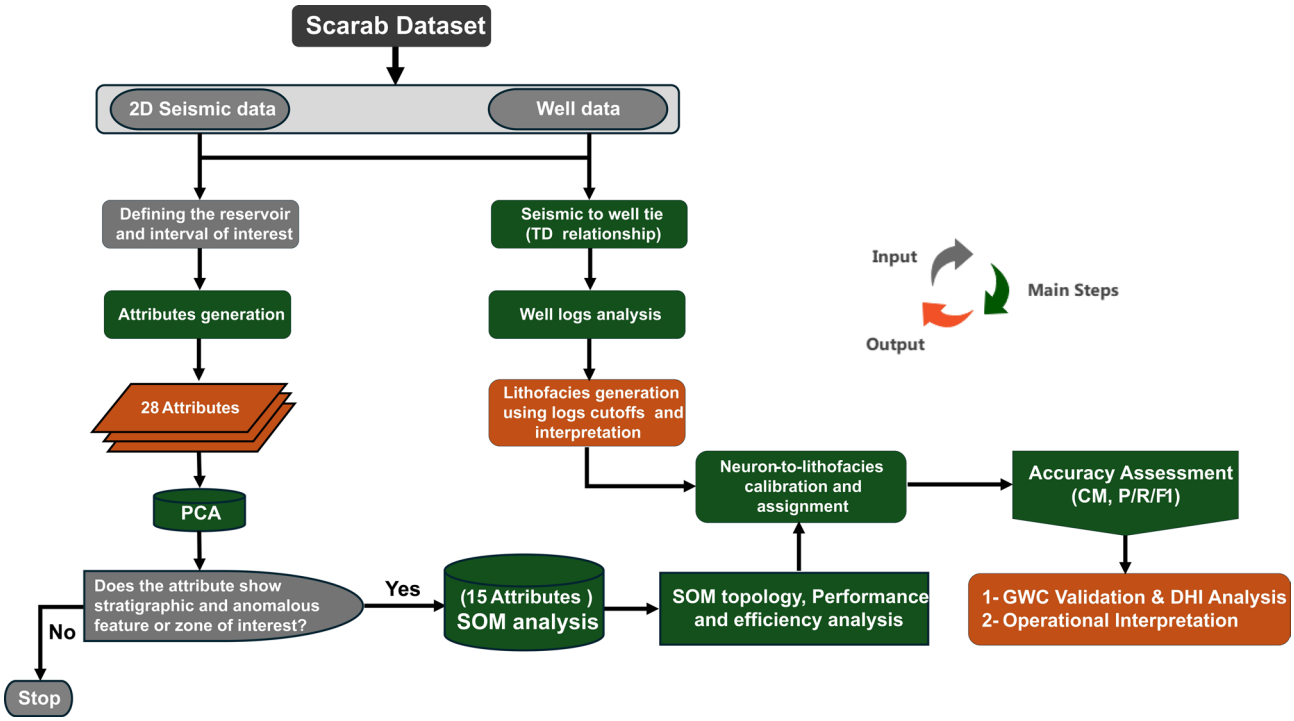


Fig. 4. Workflow for lithofacies classification and gas-channel detection in the Scarab Field. The process integrates 2D seismic and well-log data to delineate lithofacies, detect gas-sand, and identify GWC using the SOM approach.

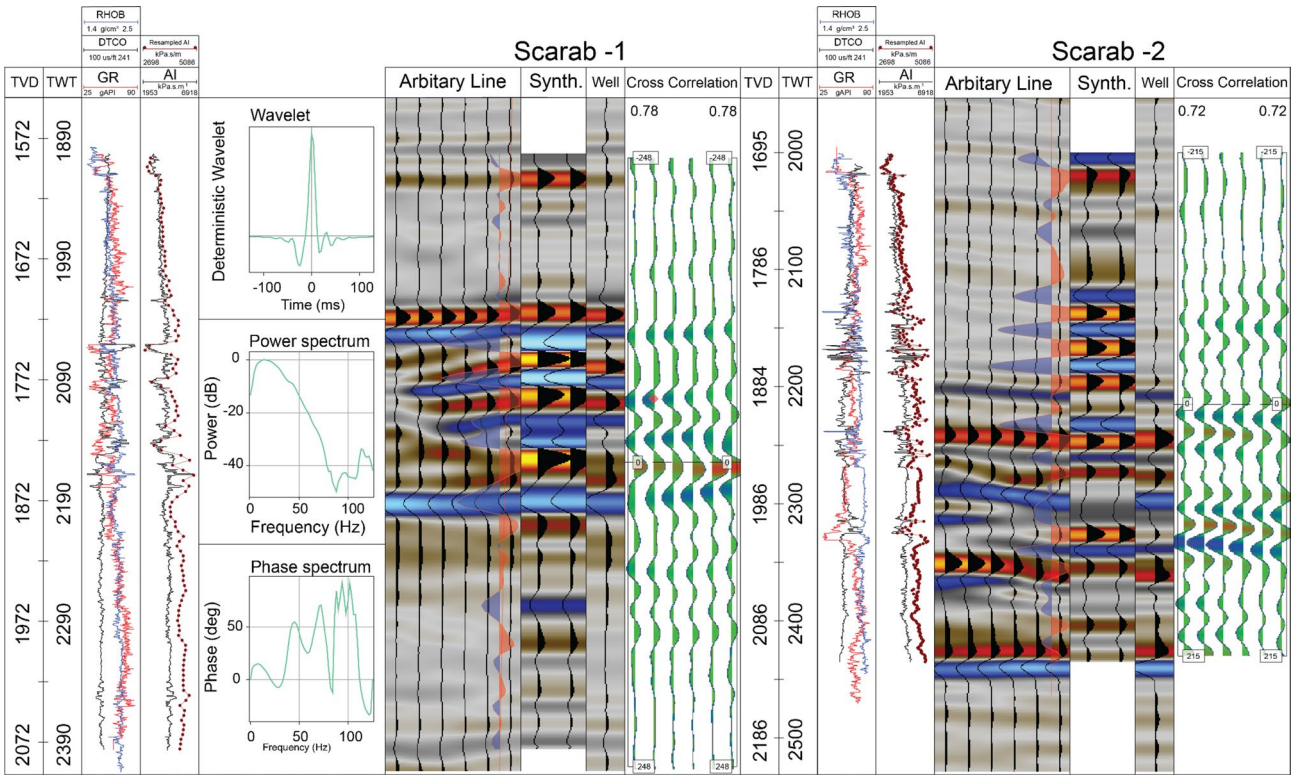


Fig. 5. Well-to-seismic tie for Scarab-1 and Scarab-2 using a single deterministic wavelet extracted from the 2D seismic line. The cross-correlation coefficients are 0.78 for Scarab-1 and 0.72 for Scarab-2.

conversion and validating the seismic polarity. An accurate well tie step is crucial for identifying the reservoir interval, aligning the seismic with the well response, extracting attributes, and subsequently classifying facies[34].

Petrophysical analysis and lithofacies classification

Identification of gas channels is carried out by analyzing the composite logs of four wells with a full suite of well logs for lithofacies classification (Scarab-1, Scarab-Da, Scarab-Dd, and Scarab-De). Figure 6 shows an example of three wells used to generate lithofacies logs and identify channel-complexity zones. Shale volume (V_{shale}) helps to discriminate between the reservoir and non-reservoir rock. The V_{shale} was determined using Larionov's equation for younger rocks (Eq. 1)[35,36], with the gamma-ray index (I_{GR}) calculated using (Eq. 2).

$$V_{shale} = 0.083 \times \left(2^{3.7 I_{GR}} - 1\right) \tag{1}$$

$$I_{GR} = \frac{GR_{\log} - GR_{\min}}{GR_{\max} - GR_{\min}} \tag{2}$$

where I_{GR} is the gamma-ray index, which can be related to and corrected for the rock's shaliness. $GR_{\log}$ is the gamma ray reading of the logging tool, $GR_{\min}$ is the minimum gamma ray, and $GR_{\max}$ = maximum gamma ray. Neutron-density logs are used to calculate the total porosity using an equation adopted from Schlumberger[37] and Tiab and Donaldson[38], whereas the effective porosity (*PHIE*) is calculated and corrected for shale volume. *PHIE* is then derived using an equation from Schlumberger[37]:

$$PHIE = \Phi_{\text{total}} \times (1 - \mathrm{V}_{shale}) \tag{3}$$

Electrofacies were then generated using the neural network, utilizing four logs as input predictors, including V_{shale}, *PHIE*, *GR*, and deep resistivity. These logs were selected because they provide critical sensitivity to lithological variation, capturing net-to-gross ratios, pore fluid effects, and textural heterogeneity, which are essential for data-driven, reliable, and geologically consistent rock classification[39,40]. Meanwhile, the available composite and interpretive logs of the four wells (Scarab-1, Scarab-Da, Scarab-Dd, and Scarab-De) were utilized to generate an electrofacies log based on petrophysical cutoffs[39,41]. Lithology logs were derived using a cutoff-based (Table S1) classification of key petrophysical parameters across the target interval.

The neural network was trained and used with well-log intervals of known lithology from interpreted log cutoffs, formation tops and operator reports to establish a nonlinear mapping between the input log responses and electrofacies classes. Once trained, the model was applied across the full depth interval of each well to produce a continuous electrofacies log (Fig. 6). The drived facies are Sandstone 1 (SS-1; gas saturated sands), Sandstone 2 (SS-2; wet and partially gas saturated), clayey sand (C-S), clayey limestone (C-LS), claystone (CS), and siltstone (SiS).

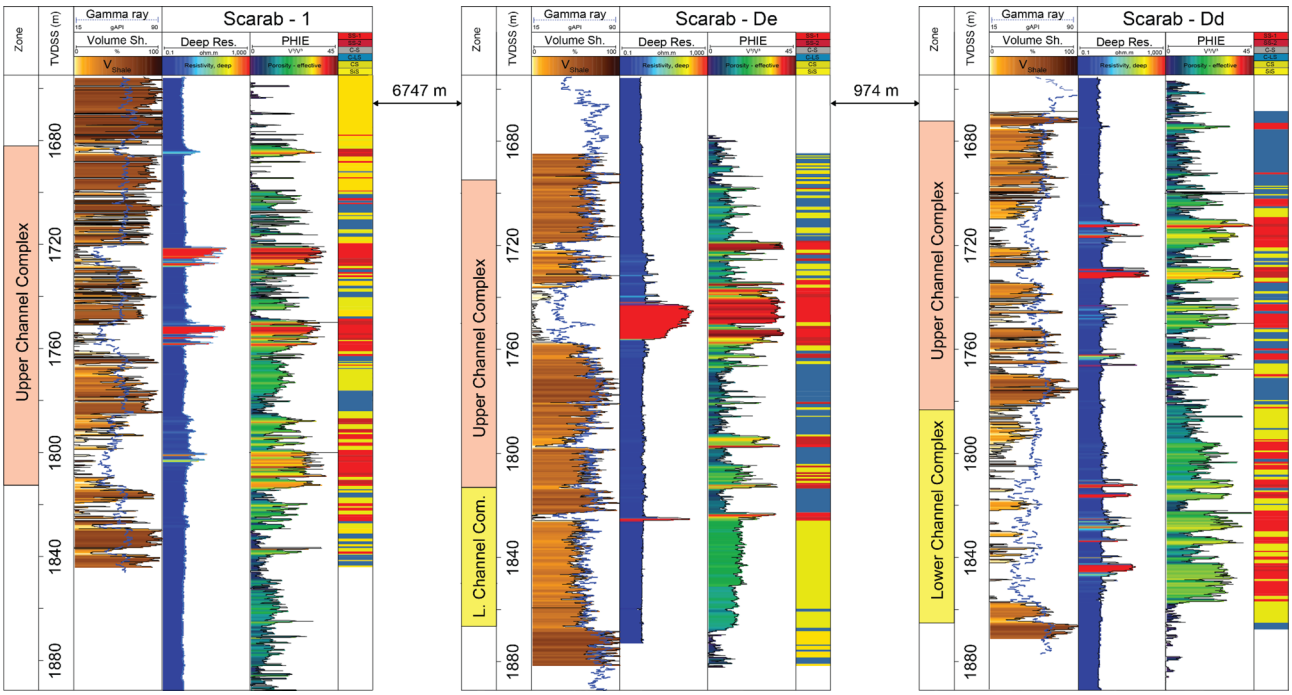


Fig. 6. Petrophysical log view panels for Scarab-1, Scarab-De, and Scarab-Dd showing zonation, gamma ray, shale volume (V_{shale}), deep resistivity (*LLD*), effective porosity (*PHIE*), and electrofacies. The wells highlight variations in channel architecture and reservoir heterogeneity across the Scarab field.

Seismic attribute generation

Seismic attributes are quantitative measures derived from seismic data through mathematical and statistical transformations[42,43]. They are widely used to reveal subtle features that are not easily visible on conventional seismic sections[44]. The extraction of attributes enhances the detection of hidden information such as stratigraphic features, depositional systems, gas-bearing zones, and unconformities, thus supporting more robust geological and geophysical interpretations[45]. Seismic data inherently contains rich information in terms of amplitude, frequency, phase, and texture[46]. By applying different classes of attributes, it becomes possible to achieve both qualitative and quantitative characterization of subsurface geology.

In the present study, we used three main seismic attribute groups: Instantaneous, geometrical, and spectral decomposition attributes. Instantaneous attributes (e.g., sweetness and envelope amplitude) provide insights into lithology, fluid content, and wave-propagation effects, and are commonly applied in reservoir characterization. Geometrical attributes, such as curvature amplitude and curvature structure, help in detecting structural features, including edges and boundaries. Spectral decomposition captures higher-order signal characteristics that are useful for recognizing seismic facies and interpreting stratigraphy. While these attributes often act as direct indicators of hydrocarbon-bearing zones, they must be calibrated with well data to avoid misinterpretation. Spectral decomposition (SD) is particularly effective for enhancing channel detection, mitigating tuning effects, and resolving thin beds that are otherwise below seismic resolution[47,48]. Therefore, we applied spectral decomposition using the Continuous Wavelet Transform (CWT) and both maximum and mean entropy to highlight stratigraphic and depositional-related features.

A comprehensive suite of 28 seismic attributes (Table 1) was extracted from the 2-D PSTM seismic dataset to capture a broad spectrum of geological and reservoir-related features. These types of attributes are known to be highly sensitive to lithologic variations, reservoir property changes, and thickness variations, making them especially valuable in complex channel settings, such as the Scarab Field[49,50]. Once calibrated with well control, these attributes form the basis for facies classification by linking seismic responses to physical rock properties, thereby capturing reservoir heterogeneity and enabling more accurate mapping of lithological and stratigraphic features across the Scarab field.

PCA-based seismic attribute selection

PCA is a powerful statistical method used to reduce dimensionality and highlight the most significant patterns in multi-attribute seismic datasets[51,52]. In multi-attribute analysis, dozens of seismic attributes (e.g., amplitude, coherence, curvature, spectral decomposition) can be computed; however, many are correlated and redundant. PCA transforms these correlated attributes into a smaller set of orthogonal principal components, ranked by the variance they explain in the data[44]. The PCA process was applied to the attributes listed in Table 1 to identify the most significant attributes of the region, enabling the detection of lithological contrasts at the depth of interest. This process enables the interpreter to determine appropriate seismic attributes for the SOM evaluation. In this case, the criterion for selecting the attributes was those whose maximum percentage contribution to the principal component was greater than or equal to 70%. The first ten principal components were reviewed in the Eigen spectrum using this selection technique. Fifteen attributes of the first ten principal components (Eigenvectors) were selected following the criterion applied. Table 2 lists the 15 selected attributes that exhibit significant variance within the total of 28 attributes used in this PCA analysis and are important for use in a SOM analysis.

The contribution of the seismic attributes in the eigenvector represents the relative importance of each attribute to the overall data variability. A higher weight indicates that the attribute is more important in

No. of Attributes selected	Attribute Name	Attribute Group
5	CWT spec voice [from 5 Hz to 85 Hz in 20 Hz increments].	CWT Spectral Voice
1	Curvature Amplitude	Curvature
1	Curvature Structure	Curvature
1	Envelope Amplitude	Instantaneous
5	Spectral ME Mag 3D [from 5 Hz to 85 Hz with 20 Hz increasing step].	Spectral Max Entropy Magnitude
1	Spectral decomposition CWT Peak Frequency	CWT Spectral Peak Frequency
5	CWT Spec Mag [from 5 Hz to 85 Hz with 20 Hz increasing step].	CWT Spectral Magnitude
1	Spectral decomposition Mean Entropy Peak Frequency	Spectral ME Peak Frequency
1	Instantaneous Frequency	Instantaneous
1	Instantaneous Phase	Instantaneous
1	Instantaneous Q	Instantaneous
1	Relative Acoustic Impedance	Acoustic Impedance
1	Sweetness	Sweetness
1	Similarity Coherent Energy	Geometric attributes
1	Hilbert	Instantaneous
1	Thin Bed	Thin Bed Indicator

Table 1. provides a comprehensive summary of the 28 seismic attributes selected for analysis, grouped by their physical meaning and application.

No.	Attribute Name	Attribute Group	Eigenvector	Max %
1	Envelope Amplitude	Instantaneous	1	90.51
2	CWT Spectral Mag 25 Hz	CWT Spectral Magnitude	1	96.4
3	CWT Spectral Mag 45 Hz	CWT Spectral Magnitude	1	94.15
4	CWT Spectral Mag 65 Hz	CWT Spectral Magnitude	1	92.76
5	CWT Spectral Mag 85 Hz	CWT Spectral Magnitude	1	92.73
6	CWT Spec Voice 25 Hz	CWT Spectral Voice	2	95.82
7	CWT Spec Voice 45 Hz	CWT Spectral Voice	2	91.81
8	CWT Spec Voice 65 Hz	CWT Spectral Voice	2	90.95
9	CWT Spec Voice 85 Hz	CWT Spectral Voice	2	90.96
10	Instantaneous Frequency	Instantaneous	3	100
11	Thin Bed	Thin Bed Indicator	3	97.56
12	Curvature Structure	Curvature	8	75.73
13	Instantaneous Q	Instantaneous	9	93.46
14	Hilbert	Instantaneous	10	74.8
15	Sweetness	Instantaneous	1	86.96

Table 2. PCA results showing the top-ranked seismic attributes associated with the selected eigenvectors for facies discrimination eigenvectors for facies discrimination.

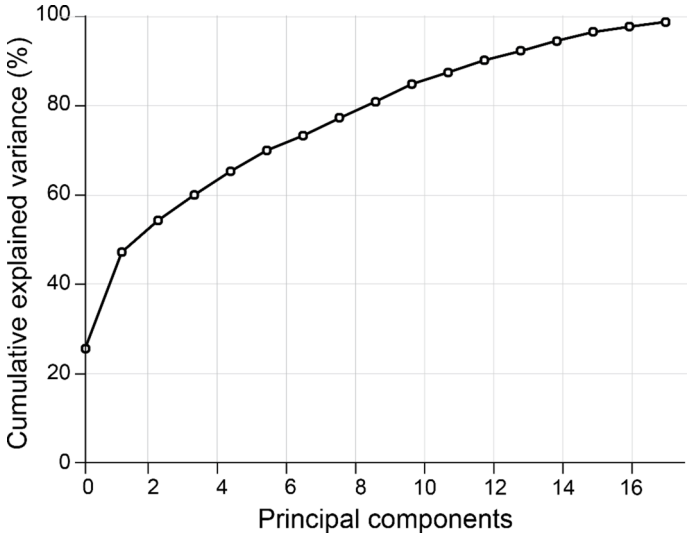


Fig. 7. Scree plot showing cumulative explained variance as a function of the number of principal components. The curve exhibits clear flattening after approximately 10 components, indicating that additional components contribute only marginal gains in explained variance. Consequently, the first 10 principal components, capturing ~ 85% of the total variance, were retained for subsequent SOM analysis.

distinguishing between different well-calibrated geological features. Figure 7 shows the Scree plot of the PCA, which optimizes the dimensionality of seismic attribute data before applying clustering algorithms, such as SOM. This plot illustrates the cumulative explained variance as a function of the number of principal components used. The primary importance of this plot lies in its ability to determine the optimal number of components required to represent the data without overfitting. For this study, we used a cutoff to determine the number of principal components, setting it at 85% of the variance explained, corresponding to approximately the first 10 components.

SOM for seismic attribute analysis and clustering

The SOM, first introduced by Teuvo Kohonen in the early 1980s, is a powerful unsupervised neural network technique for dimensionality reduction and pattern recognition in high-dimensional datasets[53,54]. Unlike conventional wavelet-based classification approaches, SOM treats each seismic sample as an input vector, enabling more granular geological classification of subsurface heterogeneity[55]. Through an iterative learning process consisting of cooperative and competitive training epochs, the neurons in the SOM adaptively organize themselves, capturing the natural clustering structure of the input seismic attributes[44]. The SOM framework is typically implemented as a two-dimensional grid of neurons (N × N), with each neuron representing a unique

data cluster in attribute space. The resolution of clustering depends directly on the SOM topology: coarser grids (e.g., 4 × 4, 16 neurons) emphasize large-scale structural and stratigraphic patterns, while finer grids (e.g., 12 × 12, 144 neurons) capture more detailed variations in lithology, porosity, and reservoir properties[55,56,57]. Although the number of neurons can theoretically be increased indefinitely, practical applications have demonstrated diminishing returns, as excessively fine topologies may introduce overfitting and reduce interpretive clarity[19,58]. To evaluate the influence of neuron grid resolution on SOM performance, multiple grid configurations (4 × 4, 6 × 6, 8 × 8, and 10 × 10) were assessed using identical attribute inputs and training parameters. First, the quantization error (QE) was monitored throughout training to evaluate the representation quality of each grid size. As shown in Fig. S1, all SOM configurations exhibit decreasing QE with increasing epochs, indicating progressive improvement in the mapping of the high-dimensional attribute space. However, the smaller 4 × 4 grid shows persistently higher QE, reflecting insufficient representation capacity, while the 10 × 10 grid achieves lower QE but introduces increased fragmentation and noise sensitivity due to over-segmentation. Topographic error (TE) in Fig. S2 quantifies the extent to which the first- and second-best-matching units are not adjacent, serving as an indicator of topology preservation in the SOM. Lower, more stable TE values across training epochs indicate improved neighborhood consistency and a more reliable structural mapping of the input attribute space.

To provide a more systematic quantitative comparison, several derived efficiency metrics normalized between 0 and 1 were calculated from the SOM training statistics, including Convergence Efficiency (C), Compactness Efficiency ($C_{(comp)}$), Utilization Efficiency (U), Stability Efficiency (S), Information Efficiency (I), and Training Economy (T). Convergence efficiency quantifies the reduction in feature-space disorder during training, while compactness efficiency measures cluster tightness based on the average distance between samples and their best-matching units. Utilization efficiency evaluates how effectively the neuron grid is used, and stability efficiency assesses the robustness and continuity of the SOM topology. These metrics are fully described and presented in Tables S2 and S3. Information efficiency measures the structural signal captured by the attribute set, whereas training economy reflects learning efficiency relative to training effort and time.

The comparison across grid sizes shows systematic trends: smaller grids produce higher compactness but insufficient representational resolution, while larger grids increase information capacity but reduce compactness and training efficiency due to increased dispersion in feature space. The 8 × 8 SOM provides the most balanced performance, maintaining strong convergence, high stability and neuron utilization, while preserving compact clusters and efficient learning behavior. These results, summarized in Fig. S3, demonstrate that the 8 × 8 topology (64 neurons) achieves a balanced trade-off between model complexity and computational efficiency, while reflecting sufficient resolution to capture stratigraphic heterogeneity, lithofacies variability, and channel boundary delineation.

Neuron-to-lithofacies calibration and assignment

The SOM output represents unsupervised clustering in multi-attribute seismic space. Geological lithofacies classification is not inherent to the SOM itself but is achieved through a subsequent calibration step, in which SOM neurons are correlated with lithofacies defined from well-log data. This post-training neuron-lithofacies mapping enables transformation of attribute-based clusters into geologically meaningful facies classes, which are then subjected to quantitative accuracy assessment against independent well data. To quantitatively link SOM neurons to geological lithofacies and to assess classification performance, a two-stage validation strategy was adopted. First, SOM neurons were assigned to lithofacies classes using a probabilistic calibration based on the overlap of probability density functions (PDFs) between neuron-associated petrophysical responses and well-log defined lithofacies[59,60,61]. This process and the associated parameters are defined in Table S5. For each neuron intersected by the calibration wells, PDFs of key petrophysical properties (e.g., *GR*, V_{shale}, *PHIE*, and resistivity) were computed and compared against the corresponding PDFs derived from true lithofacies intervals in the wells. Each neuron was assigned to the lithofacies with the greatest statistical overlap, ensuring an objective, reproducible mapping from unsupervised clusters to geological classes.

Following neuron-to-lithofacies assignment, quantitative accuracy metrics were computed to evaluate how well the SOM-derived facies reproduce the well-log interpretations. Confusion matrices were constructed by comparing SOM-predicted lithofacies against the reference electrofacies at the well locations. From these matrices, standard performance measures-including precision, recall, and F1-score-were calculated for each lithofacies. Because reservoir performance and fluid-contact prediction are most sensitive to sand-prone units, this quantitative assessment focused specifically on sandstone intervals, providing a targeted and operationally relevant measure of classification accuracy.

After selecting the 8 × 8 grid size based on quantization error (QE), topographic error (TE), and overall performance metrics, a classification accuracy assessment was conducted using well-log-defined lithofacies as reference data. The resulting confusion matrix and corresponding precision, recall, and F1-score metrics quantify the agreement between SOM-derived facies assignments and well-log interpretations (Figs. S4 and S5). The normalized confusion matrix shows strong diagonal dominance, particularly for Sandstone-1 (0.85), indicating reliable class identification. The associated precision, recall, and F1-scores further confirm consistent classification performance across sandstone facies. These results demonstrate that the selected SOM configuration yields robust and reproducible lithofacies assignments consistent with independent well control.

Results and interpretation

Seismic attribute and SOM-based facies classification

The selected attributes were derived from various categories, including instantaneous attributes, spectral decomposition (magnitude and voice), thin-bed indicators, and curvature, as identified through PCA in Table 2. The diverse set of attributes captures distinct features of seismic responses, enabling more reliable unsupervised facies classification. Figure 8 illustrates four representative seismic attributes used as SOM inputs,

each emphasizing different aspects of the subsurface while consistently revealing a bright spot feature linked to gas-bearing sand channels. Cross-section A (Envelope Amplitude; Instantaneous group, 90.5% contribution to Eigenvector 1) highlights the total reflection energy, clearly outlining gross stratigraphic geometries and channel trends, though it does not resolve subtle lithological variations. Cross section B (Hilbert attribute; Instantaneous group, Eigenvector 10) enhances phase-related reflectivity and improves continuity, particularly along thin beds that border stratigraphic traps, making the channel geometry more continuous. Cross-section C (Sweetness attribute; ratio of envelope to the square root of instantaneous frequency) distinctly emphasizes low-frequency, high-amplitude anomalies, shown here as the bright spot response diagnostic of gas-charged sands within the channel system. Cross-section D (CWT Spectral Voice at 45 Hz; Spectral group, Eigenvector 2) provides frequency tuning that accentuates thin-bed interference patterns and lithological boundaries, highlighting the lateral variability of the channel fill. The selected seismic attributes (Fig. 8) consistently reveal a persistent bright spot anomaly coinciding with the gas-bearing channel complex. Although each attribute emphasizes different aspects of the seismic response, such as reflector continuity, frequency-dependent tuning, or amplitude variations, they collectively validate the same stratigraphic feature, providing robust evidence for its interpretation as a hydrocarbon-bearing sandstone body.

Figure 9a presents the post-stack seismic amplitude section along Arbitrary Line 1 with interpreted stratigraphy and well markers. This section displays the main channel complex boundaries and the gross architecture of the WDDM system in the Scarab Field; however, as is typical for amplitude data alone, it provides limited lithologic discrimination within the channel fill. To improve lithofacies resolution, this amplitude section was transformed into a multi-attribute input and processed through a SOM using an 8×8 neuron topology. The result, shown in Fig. 9b, converts the original amplitude image into a classified facies section where each neuron cluster corresponds to a calibrated lithology. This multi-attribute SOM classification reveals much finer internal structure and fluid-related anomalies that are not apparent in the amplitude data. On Fig. 9b, siltstone is mapped primarily by neuron 15, claystone by neurons 7, 8, and 16, clayey limestone by neuron 24, clayey sand by neuron 31, and sandstone facies by neurons 56, 63, and 64 (Sandstone-1) and neurons 62 and 55 (Sandstone-2). This mapping demonstrates a correlation between seismic attribute space and lithofacies distribution, highlighting stratigraphic heterogeneity within the sandstone bodies, including transitions between clayey sands and cleaner sandstone intervals. Finally, Fig. 9c shows the filtered SOM output after applying a 10% probability threshold to suppress noise-related neurons and incoherent features. Removing these non-reservoir signals yields a much cleaner and geologically consistent facies map, enhancing interpretability and reducing the risk of misclassification compared to traditional single-attribute analysis.

Figure 10 illustrates SOM facies classifications to emphasize reservoir-related neurons and suppress noise. In section (a), SOM classification using an 8×8 topology highlights gas-prone sand facies (neurons 55, 56, 62, 63, 64) alongside claystone and transitional lithologies. Section (b) isolates and shows only the most confident reservoir neurons, effectively removing incoherent responses and leaving a cleaner representation of gas-bearing sand intervals across the channel complexes.

To demonstrate the stability of the SOM classification, we analyzed the evolution of winning neuron shifts across training epochs. Although metrics such as quantization error and topographic error are commonly used in SOM analysis, this study emphasizes winning neuron shift as a complementary convergence indicator, as it directly reflects the stability of neuron assignments that support subsequent lithofacies interpretation. Figure 11 illustrates the evolution of winning neuron shifts (%) across training epochs during the SOM training process. The winning neuron shift metric quantifies the percentage of input vectors that change their best-matching neuron between consecutive epochs, thus serving as a convergence indicator for the SOM training. Initially, the shift rate remains high above 85%, indicating active competition among neurons and significant reorganization of the attribute space. Between epochs 10 and 20, the shift rate exhibits a gradual decline with intermittent

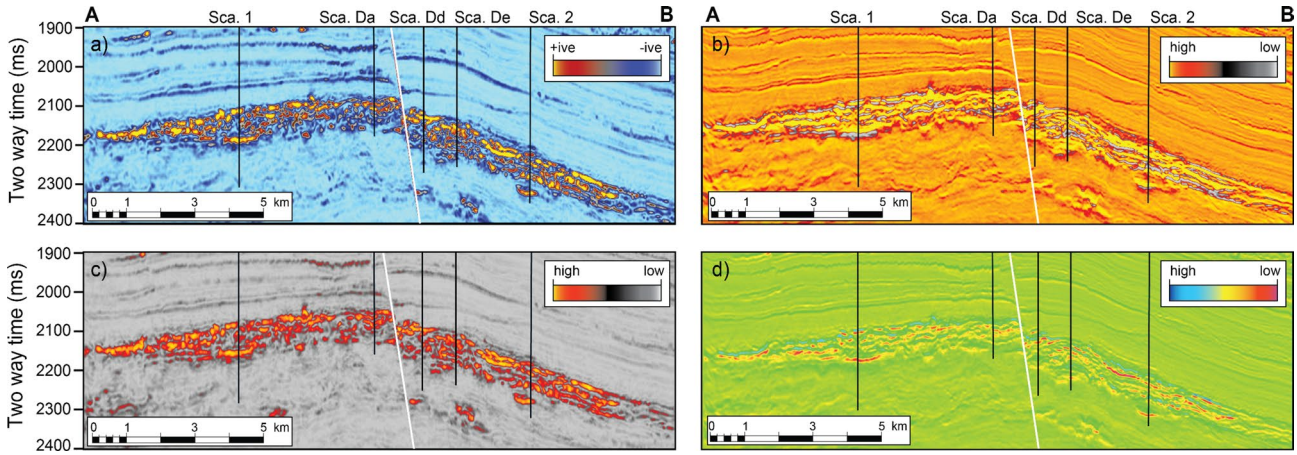


Fig. 8. Selected seismic attributes after PCA analysis that were used as inputs for the Scarab SOM classification. (**A**) Envelope Amplitude (**B**) Hilbert attribute (**C**) Sweetness (**D**) CWT Spectral Voice (45 Hz). These attributes help to delineate the bright spots associated with gas-bearing sand channels.

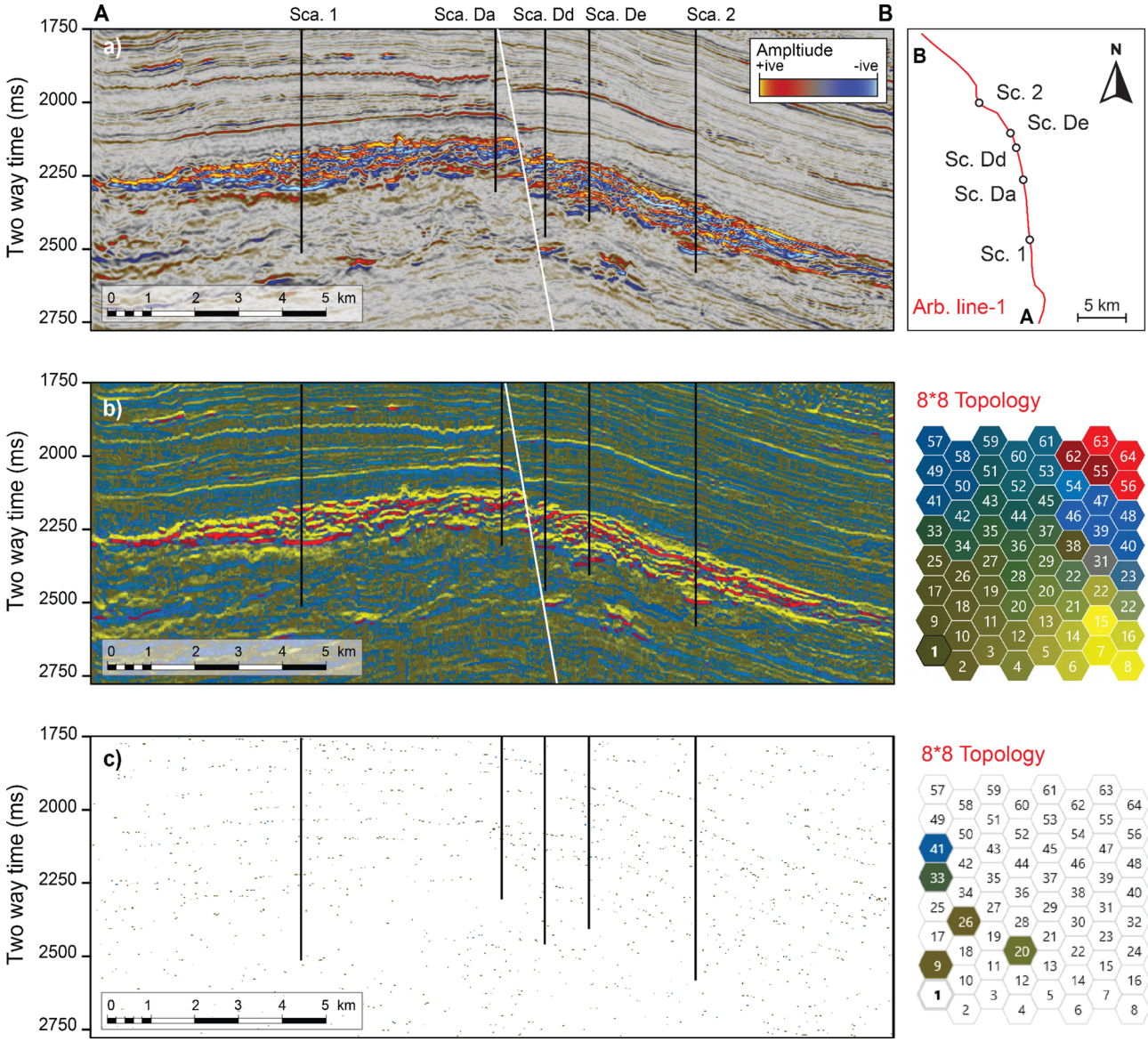


Fig. 9. Vertical seismic section along Arbitrary Line 1 showing: (**A**) Post-stack seismic amplitude section with interpreted stratigraphy and well markers; (**B**) SOM classification result using 8 × 8 neuron topology, highlighting all classified seismic facies; (**C**) Filtered SOM result showing only neurons representing noise and incoherent features, using a probability threshold of 10%. The well positions are annotated for reference, and neuron color maps are provided for both full and filtered SOM results.

fluctuations, reflecting the transition from a global to a more localized topological ordering. A sharper decrease is observed after epoch 20, indicating that SOM has entered the fine-tuning phase, where only minor adjustments are made to the neuron map. After approximately 40 epochs, the shift rate drops below 5%, suggesting that the network has stabilized and reached a near-converged state. The final epoch shows minimal change, confirming that the training has sufficiently captured the data structure and that further training would yield diminishing returns. This convergence behavior validates the robustness of the SOM classification, ensuring that the neuron-to-facies relationships established are stable and representative of the underlying seismic attribute distribution.

SOM lithofacies to well correlation

The SOM-based facies classification was validated through integration with well log and electrofacies data from four key wells intersecting the seismic survey: Scarab-1, Scarab-Da, Scarab-De, and Scarab-Dd. This cross-validation enabled the SOM-derived neurons to be correlated with lithological and petrophysical indicators, resulting in a geologically consistent and spatially continuous facies model. Figure 12 presents four representative well correlations (labeled A through D), each correlating the original post-stack seismic amplitude, SOM-derived facies section, and corresponding well logs. These subpanels illustrate the consistency between SOM neuron classes and the petrophysical responses observed in *GR*, V_{shale}, deep resistivity, and *PHIE* logs. In subpanel A

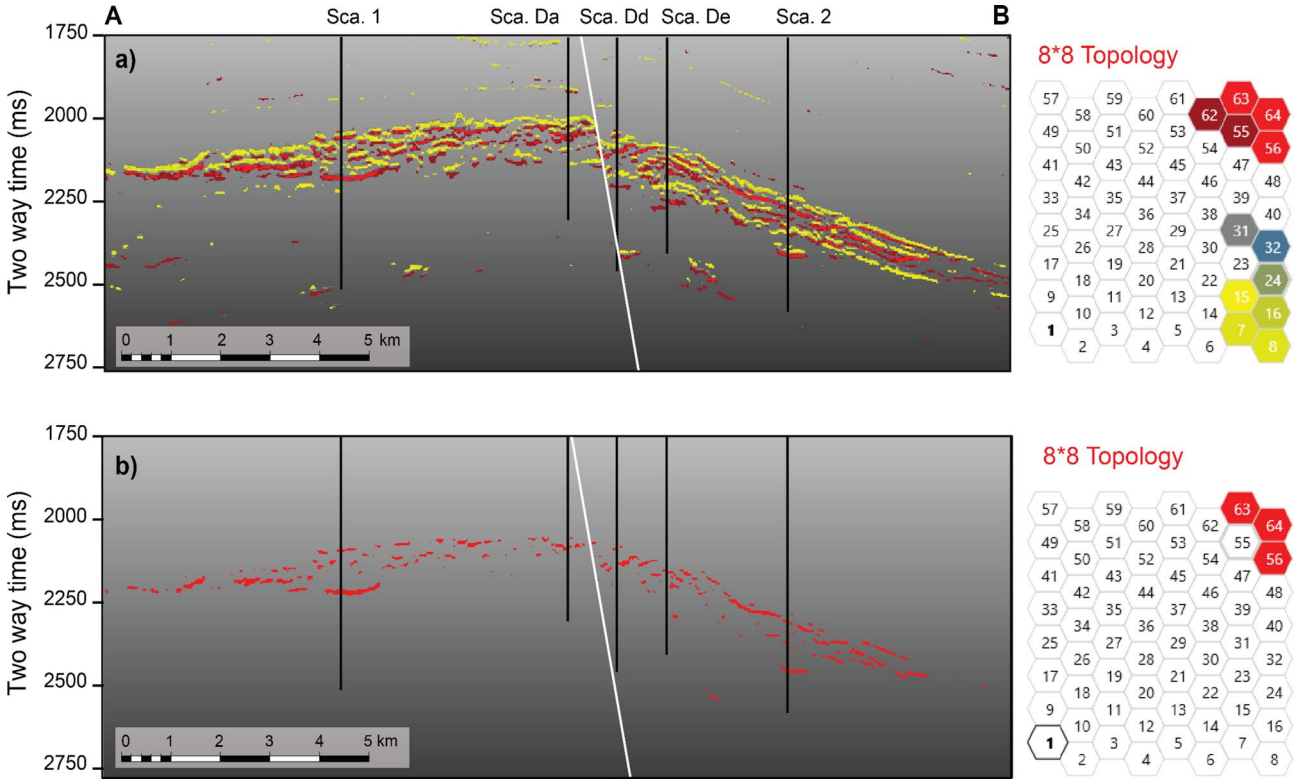


Fig. 10. SOM-based facies classification along Arbitrary Line 1 using an 8 × 8 neuron topology shows neurons of interest. (**a**) SOM classification section showing reservoir neuron classes in the lower and upper channels, including both sand and non-sand facies. (**b**) SOM classification displaying only the high-quality Sandstone-1 gas-saturated reservoir neurons (56, 62, 64), providing a clearer delineation of this interval.

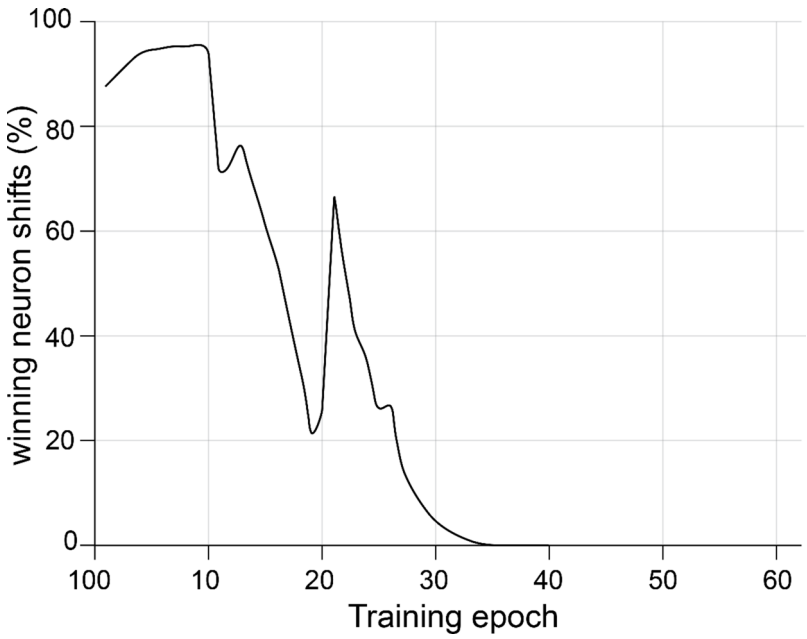


Fig. 11. Percentage of winning neuron shifts during SOM training, illustrating the three main learning phases. High shift rates during the initial ordering phase (epochs ~ 1–10) reflect the rapid global organization of the map. The transition phase (epochs ~ 10–25) shows decreasing and oscillatory shifts as local structure is refined. After approximately 25–30 epochs, the winning neuron shifts approach zero, indicating convergence and stable SOM clustering. Training was terminated automatically once the convergence criterion was met.

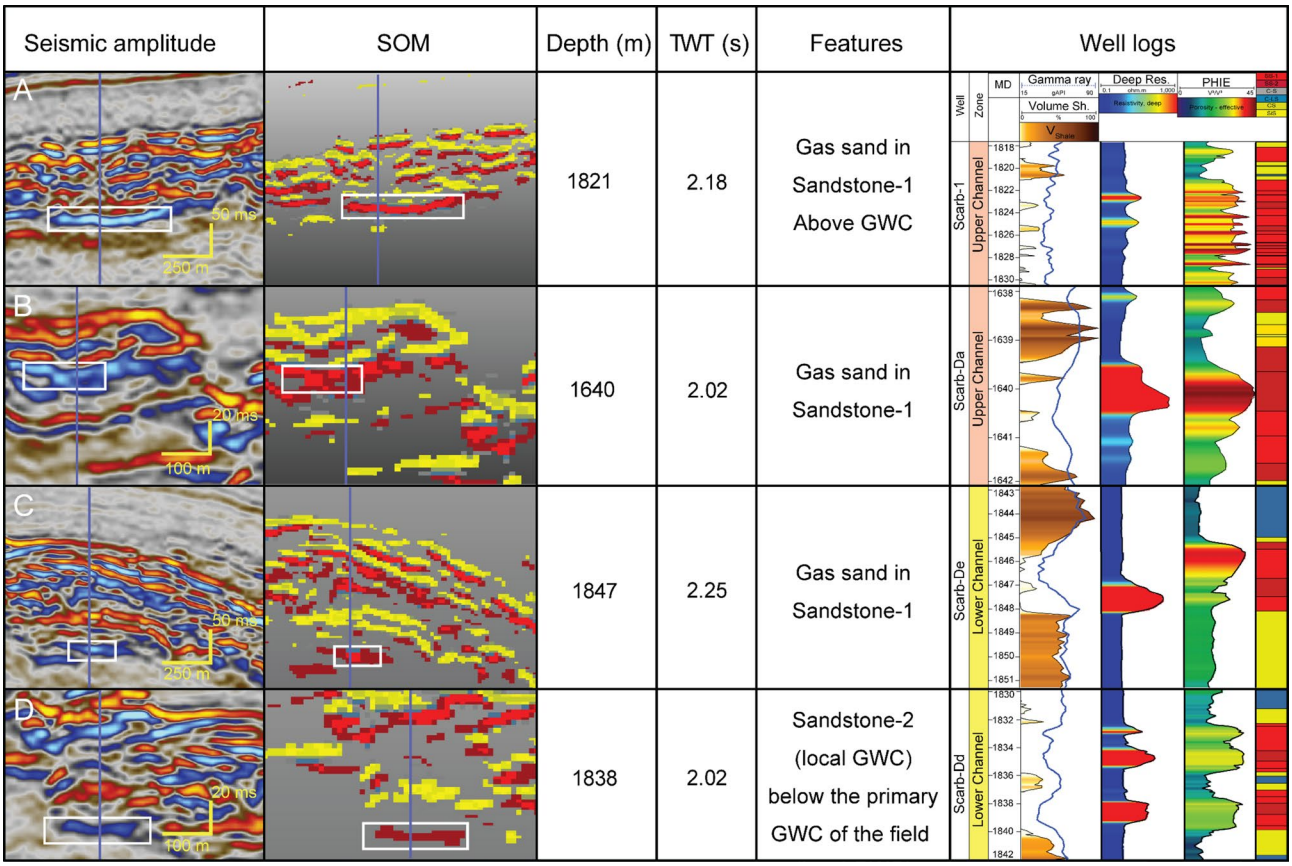


Fig. 12. Lithofacies classification of Scarab channel units and their validation using well log data. (**A**-**D**) Seismic amplitude sections (left) and corresponding SOM-based lithofacies classification (right) for selected regions in Scarab Upper and Lower channels. Lithofacies classes are color-coded as described in Fig. 5, with yellow representing the Upper channel and orange representing the lower channel, calibrated using well data. Dashed boxes highlight zones of interest characterized by potential reservoir-quality sand bodies. The right-side panel shows well log responses (*GR*, V_{shale}, Deep resistivity, and *PHIE*) from four wells (Scarab-1, Scarab-Da, Scarab-De, and Scarab-Dd), confirming the lithofacies classification and supporting the channel-scale facies interpretations. Dark blue lines highlight the wells trajectories.

(Scarab-1), a gas-charged sandstone interval (Sandstone-1) at a depth of 1821 m (2.18 s) is clearly delineated above the GWC. SOM neurons 56, 63, and 64 dominate this zone and correlate with low *GR* and high *PHIE*, reinforcing the interpretation of a clean, thick-bedded reservoir unit. Subpanel B (Scarab-Da) captures another gas-bearing Sandstone-1 unit at 1640 m (2.02 s). The SOM classification is dominated by the same reservoir neurons, highlighting continuity in the stratigraphic unit. The petrophysical logs confirm this interpretation, showing a strong porosity development and minimal shale content. Subpanel C (Scarab-De) shows a similar Sandstone-1 interval within the lower channel at 1847 m (2.25 s). Here, the SOM sharply separates the reservoir interval from overlying and underlying muddy facies, which are represented by claystone neurons (7, 8, 16). This improved vertical resolution is particularly valuable for stratigraphic interpretation in complex fluvial-deltaic environments. For Scarab-Dd, seismic and well-log integration indicate that the primary GWC occurs around 1850 m measured depth MD (1827.75 m total vertical depth subsea TVDSS), representing the primary contact across the field. However, an additional contact is observed at 1872 m MD (1849.69 m TVDSS), approximately 22 m deeper than the principal GWC. This deeper interface is interpreted as a perched-water zone rather than a regional contact, likely resulting from stratigraphic barriers and compartmentalization within the channel system. Such localized hydraulic isolation is consistent with the structural configuration of the Scarab Field, where stratigraphic pinch-outs, dip closures, and fault-related segmentation (Rosetta trend) create discrete reservoir compartments[11]. This provides a mechanism for hydraulically isolated compartments, supporting the interpretation that this deeper GWC may not represent the regional contact but rather the effect of a perched water body. Moreover, subpanel D (Scarab-Dd) reveals a Sandstone-2 interval at 1838 m (2.02 s), interpreted as a local gas pay zone intercalated by perched water positioned below the primary field-wide GWC. Neurons 62 and 55 dominate this zone, corresponding to thinner, gas-saturated sands affected by perched water trapped by the deeper clayey interval. This configuration supports the interpretation of a hydraulically isolated perched-water body rather than continuous field-wide contact.

Discussion

SOM configuration and performance analysis (8 × 8 grid)

The SOM was trained on the Scarab 2D seismic dataset using a hexagonal 8 × 8 grid (64 neurons) and a multi-attribute feature space composed of amplitude, instantaneous, spectral CWT components, sweetness, and thin-bed attributes. Parameters of the 2D seismic dataset and the SOM configuration, including network architecture, training controls, learning parameters, and convergence metrics used in this study are included in Table S4. The training process employed an initial learning rate of 0.3 with smooth exponential decay ($\tau = 10$) and a large initial neighborhood radius ($\sigma_0 = 7$), promoting strong global ordering during early iterations. Training converged efficiently after 76 epochs (maximum allowed: 100), with early stopping activated at a 0.01 threshold, indicating stable and rapid organization of the feature space (Fig. S6). Quality statistics show a reduction in standard deviation from 0.9317 to 0.3382, yielding a convergence efficiency of 0.637, which reflects substantial reduction of initial feature-space disorder. The average Euclidean distance per sample resulted in a compactness efficiency of 0.577, indicating relatively tight clustering and coherent pattern separation. No dead neurons or unclassified samples were observed, producing utilization and stability efficiencies of 1.0, confirming robust topology preservation and appropriate grid sizing. The information efficiency derived from the Akaike Information Criterion (AIC) reached ≈ 0.70, suggesting that the attribute set captures meaningful structural variation without excessive redundancy. Hence, the 8 × 8 configuration provides a balanced trade-off between clustering compactness, learning stability, and information content, supporting its suitability for representing the multi-attribute seismic feature space.

Reservoir heterogeneity, fluid contacts, and perched water in the Scarab field

Scarab-1 intersects a thick, gas-prone sandstone interval within the Upper Channel Complex, whereas Scarab-De exhibits a laterally continuous and comparatively cleaner reservoir section with limited internal facies heterogeneity[4,9]. In contrast, Scarab-Dd penetrates both the Upper and Lower Channel complexes, providing a clear example of vertically stacked channelized reservoirs and pronounced lateral and vertical facies variability across the field[4,9,15] (Fig. 6). Such stratigraphically complex deep-water systems are highly sensitive to uncertainty in fluid-contact interpretation, making accurate identification of GWCs critical for reserve estimation and completion design.

Quantitative comparison between log-defined and seismic-derived contacts demonstrates the improved predictive capability of the SOM-based interpretation. In Scarab-1, the SOM-interpreted GWC occurs at ~ 1824 m MD, closely matching the operator-defined log GWC at 1821 m MD, with an absolute misfit of ~ 3 m. In Scarab-Dd, the SOM predicts the regional GWC at 1850 m MD, in exact agreement with the log-based reference depth (Fig. S7). By comparison, amplitude-only interpretation yields misfits on the order of ~ 20 m. Overall, the SOM-based workflow reduces GWC depth error by approximately 85–92.5%, depending on the benchmark used, providing a robust quantitative basis for the improved performance observed in this study (Fig. S7).

The SOM results further reveal facies-consistent transitions at both the seismic GWC and a deeper log-defined water zone, but with important differences in their attribute expression. At the seismic GWC, SOM cluster changes are primarily driven by fluid effects, expressed through attributes sensitive to lithology and pore fluids while preserving sand-dominated lithological signatures. In contrast, the deeper contact associated with the regional aquifer corresponds to a combined change in fluid state and depositional facies. Importantly, the interval interpreted as perched water in Scarab-Dd is represented by distinct SOM clusters that retain sandstone-like petrophysical characteristics, rather than collapsing into shale- or silt-dominated facies. These clusters occupy a transitional region in attribute space, reflecting partial water saturation within reservoir-quality sands rather than a fundamental change in rock type. This behavior supports interpretation of perched water as a stratigraphically trapped fluid body within a sand-dominated depositional unit, controlled by internal heterogeneity and reduced vertical connectivity, rather than a misidentified lithological boundary.

From an operational perspective, the integration of SOM clustering with seismic attributes and petrophysical data in this manner can inform completion decisions in the Scarab Field. In Scarab-Dd, the well was strategically plugged back at 1850 m MD-coincident with the SOM-predicted onset of non-reservoir facies, thereby avoiding both the perched-water interval and the regional GWC. In contrast, Scarab-De is characterized by laterally continuous gas-bearing facies with no indication of perched water, consistent with uninterrupted completion and the SOM-derived facies architecture. These outcomes demonstrate that SOM-guided interpretation provides a quantitatively validated and operationally effective framework for resolving fluid contacts, identifying localized stratigraphic compartmentalization, and mitigating water-production risk in stratigraphically complex deep-water reservoirs. The ability of the SOM to independently resolve these critical facies and fluid-contact relationships using pre-drill seismic information demonstrates its predictive capability. These results indicate that the same workflow can be applied prior to drilling to identify fluid contacts, recognize stratigraphic compartmentalization, and reduce water-production risk in stratigraphically complex deep-water reservoirs, thus supporting more informed well placement and completion planning.

Log response and interpretation of the perched-water interval

In Scarab-Dd, well logs identify an additional water-bearing interval at approximately 1872 m MD, ~ 22 m deeper than the regional GWC. This deeper contact has been interpreted by the operator as a perched-water zone within the Lower Channel complex. The petrophysical response across this interval is systematic and internally consistent, distinguishing it from both the overlying gas-bearing sands and the deeper regionally water-saturated interval. *PHIE* remains relatively high and comparable to the overlying reservoir sands (typically ~ 0.15–0.25), indicating preservation of reservoir-quality pore structure rather than lithological degradation. In contrast, deep resistivity (Rt) shows a marked decrease relative to the gas zone but remains higher than in the deeper regional aquifer, consistent with partial water saturation rather than complete water filling. V_{shale} remains low

to moderate and does not exhibit a sharp increase across the interval, indicating that the resistivity drop is not driven by increased shale content.

This log response supports the interpretation of the interval as a locally water-filled sand body hydraulically isolated from the regional aquifer, rather than **a** simple downward transition into the main GWC. Thin shale and silt interbeds inferred from seismic facies analysis and depositional interpretation are not individually resolved at the log scale but are expressed indirectly through subtle increases in V_{shale} variability and small-scale resistivity fluctuations. These features are therefore interpreted as sub-log-scale heterogeneities that reduce vertical permeability and enable localized fluid trapping, rather than thick, laterally continuous shale barriers.

Alternative scenarios and justification of perched water

Several alternative interpretations were considered, including (1) a gradational gas-water transition zone, (2) capillary pressure-controlled saturation effects, and (3) lithological deterioration due to increasing shale content. However, these scenarios are inconsistent with the observed log responses. A capillary transition zone would be expected to show progressive porosity reduction and stronger shale influence[62], which are not observed. Similarly, a lithology-driven interpretation would require a systematic increase in V_{shale} coincident with the resistivity decrease, which is not supported by the data. Instead, the persistence of good porosity, low shale volume, and an abrupt but limited decrease in resistivity favor a perched-water interpretation controlled by stratigraphic compartmentalization and reduced vertical connectivity.

This interpretation is consistent with previous studies in the Scarab Field and analogous Nile Delta deep-water channel systems, where thin-bedded heterogeneity, channel abandonment surfaces, and internal stratigraphic barriers have been shown to generate localized perched-water accumulations above the regional GWC. Comparable patterns have been reported in analogous geological settings (e.g., Abreu, et al.[63] and Hulea[64]). Moreover, such behavior, in which high-frequency depositional layering limits vertical fluid equilibration despite an overall sand-dominated lithology, is expected in slope-channel and channel-levee systems.

Facies-guided interpretation of DHIs and channel edge variability

When comparing Scarab-De and Scarab-Dd in Fig. 12, Scarab-Dd clearly exhibits greater stratigraphic heterogeneity and fluid-contact complexity, including a primary regional GWC and a deeper perched-water interface. The SOM-derived lithofacies section in Fig. 12d highlights this complexity by resolving compartmentalized Sandstone-2 intervals and delineating localized fluid segregation below the field-wide GWC. In this case, SOM enhances interpretation by identifying subtle stratigraphic barriers and vertically stacked channel elements that are not fully apparent in amplitude alone. In contrast, Scarab-De represents a comparatively more laterally continuous and cleaner Sandstone-1 development within the lower channel system. Although geologically simpler than Scarab-Dd, SOM plays a critical role in Scarab-De by sharply separating the gas-bearing sandstone from adjacent surrounding facies and clearly delineating the bright-spot interval at 1847 m (2.25 s). The dominant reservoir neurons reliably align with low GR, elevated PHIE, and increased deep resistivity responses in the well logs (Fig. 12c), confirming the gas-prone nature of the interval. Additionally, SOM enhances vertical resolution by capturing the continuity of both Sandstone-1 and Sandstone-2, separating reservoir sands from muddy facies and highlighting fluid contacts, such as the GWC.

Figure 13 illustrates the integration of SOM-classified facies with seismic amplitude anomalies intersecting a the key producing well Scarab-2 to extract high-confidence indicators of reservoir architecture and hydrocarbon presence. The SOM results clearly delineate key facies, particularly Sandstone-1 (neurons 56, 63, 64) and Sandstone-2 (neurons 62, 55), which correlate with prominent amplitude anomalies such as bright spots and flat spots. A distinct bright spot at ~ 2.42 s depth, directly above a flat spot, coincides with gas-filled sandstone layers underlain by water-bearing facies. This feature is interpreted as the regional GWC, supported by both facies boundaries and amplitude terminations. A second, more localized flat spot in the eastern part of the channel system suggests a compartmentalized GWC, possibly influenced by stratigraphic or structural controls. The facies classification also reveals characteristic changes at the channel edges, where transitions from sand-dominated to clay-rich facies are sharply defined, highlighting potential flow barriers. These lateral facies changes are critical in identifying sweet spots and heterogeneities that may not be obvious from seismic amplitude alone.

Spatial representativeness and limitations of the 2D seismic dataset

The analysis is based on a single 2D PSTM seismic profile integrated with five calibration wells, where an 8×8 SOM was applied to a PCA-screened multi-attribute dataset. Along the studied line, SOM clusters show consistent correspondence with well-constrained lithofacies, particularly in separating sand-dominated reservoir intervals from clay-rich non-reservoir facies, with balanced classification accuracy of ~ 70% for the 8×8 grid and higher precision and recall for sandstones. SOM-based prediction of the GWC closely matches log-defined contacts, with absolute misfits of 0–3 m and an error reduction of ~ 85–92.5% relative to amplitude-only interpretation. Because the analysis is restricted to a single 2D profile, the results describe facies variability and fluid-contact behavior along the line, rather than full three-dimensional reservoir geometry; channel width, sinuosity, and lateral connectivity cannot be fully resolved, and apparent facies continuity may be influenced by line orientation relative to the depositional trend. Nevertheless, the consistency of SOM-well correlations across all wells and the quantitative improvements in GWC prediction demonstrate that the workflow is robust at the profile scale, while also being directly transferable to 3D seismic volumes, where volumetric SOM mapping would enable improved characterization of channel architecture, connectivity, and stratigraphic compartmentalization at the field scale.

Conclusions

This study demonstrates the effectiveness of an integrated PCA-SOM workflow for improving seismic stratigraphic interpretation and DHI analysis in the Scarab Field, offshore Nile Delta. The fluvial-deltaic

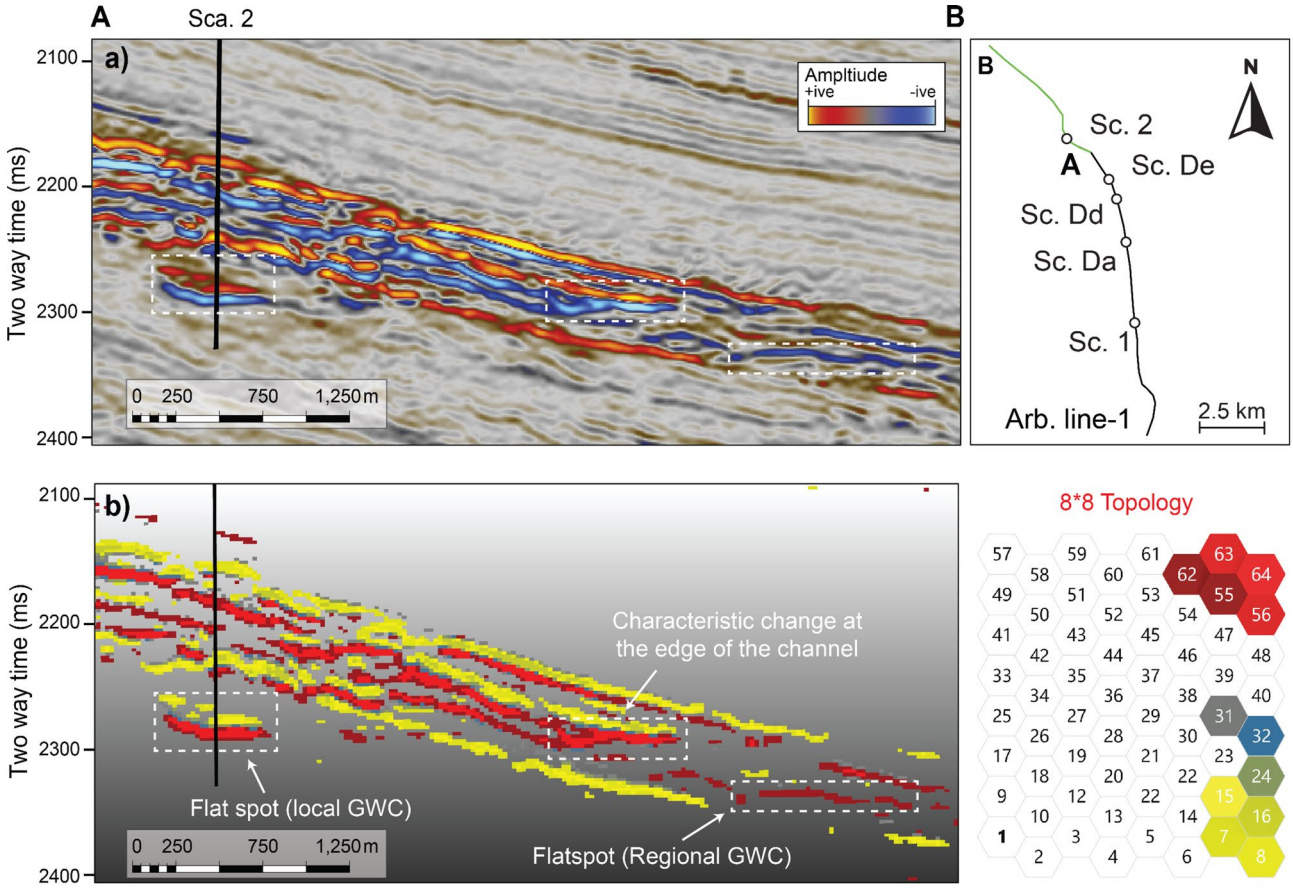


Fig. 13. This figure summarizes the integration of SOM-based lithofacies classification with seismic amplitude to enhance the interpretation of DHIs and channel architecture. Sandstone-1 and Sandstone-2 facies (neurons 56, 63, 64, and 55, 62) correlate with observed bright and flat spots, including a prominent GWC around 2.42 s. Lateral facies changes at channel edges reveal stratigraphic barriers, while compartmentalized GWCs suggest heterogeneity.

architecture of the field, characterized by vertically stacked channel complexes and pronounced lithological heterogeneity, poses significant challenges for conventional amplitude-based interpretation, particularly for reliable GWC identification. The proposed machine-learning-driven approach addresses these limitations by combining multi-attribute dimensionality reduction with unsupervised clustering, yielding a more objective, higher-resolution facies framework.

The key outcomes of this study are as follows.

- PCA reduced the original set of 28 seismic attributes to 15 attributes, preserving more than 85% of the cumulative variance while retaining the most geologically meaningful information. This step minimized attribute redundancy and optimized the input space for unsupervised SOM analysis.
- The SOM analysis implemented using an 8 × 8 neuron topology clustered seismic samples into six data-driven facies groups that were subsequently interpreted as gas-prone Sandstone-1, lower-quality or wet Sandstone-2, clayey sand, clayey limestone, siltstone, and claystone.
- Calibration against well logs confirmed consistent relationships between SOM clusters, petrophysical and fluid-related seismic responses, establishing a robust link between seismic attribute space and subsurface fluids and lithofacies variations.

Data availability

The data that support the findings of this study are available from the Egyptian General Petroleum Corporation (EGPC), but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of the EGPC.

Received: 20 December 2025; Accepted: 6 October 2026
Published online: 16 November 2026

References

1. Ghassal, B. I., Atfy, E., Sachse, H., Littke, R. & V. & Source rock potential of the Middle Jurassic to middle Pliocene, onshore Nile Delta basin, Egypt. *Arab. J. Geosci.* **9**, 744 (2016).
2. Ismail, A., Ewida, H. F., Al-Ibiary, M. G., Gammaldi, S. & Zollo, A. Identification of gas zones and chimneys using seismic attributes analysis at the Scarab field, offshore, Nile Delta, Egypt. *Pet. Res.* **5**, 59–69 (2020).
3. Ghoneimi, A., Farag, A., Bakry, A. & Nabih, M. A new deeper channel system predicted using seismic attributes in scarab gas field, west delta deep marine concession, Egypt. *J. Afr. Earth Sc.* **177**, 104155 (2021).
4. Samuel, A., Kneller, B., Raslan, S., Sharp, A. & Parsons, C. Prolific deep-marine slope channels of the Nile Delta, Egypt. *AAPG Bull.* **87**, 541–560 (2003).
5. Abuamarah, B. A., Nabawy, B. S., Shehata, A. M., Kassem, O. M. & Ghrefat, H. Integrated geological and petrophysical characterization of Oligocene deep marine unconventional poor to tight sandstone gas reservoir. *Mar. Pet. Geol.* **109**, 868–885 (2019).
6. Turner, R. et al. Structural and stratigraphic controls on reservoir architecture: A case study from the lower oligocene Vicksburg formation, Brooks county, Texas. *Mar. Pet. Geol.* **160**, 106627 (2024).
7. Cross, N. E. et al. Three-dimensional seismic geomorphology of a deep-water slope-channel system: The Sequoia field, offshore west Nile Delta, Egypt. *AAPG Bull.* **93**, 1063–1086 (2009).
8. Mohamed, I. A., El-Mowafy, H. Z. & Fathy, M. Prediction of elastic properties using seismic prestack inversion and neural network analysis. *Interpretation* **3**, T57–T68 (2015).
9. Mokhtar, M., Saad, M. & Selim, S. Reservoir architecture of deep marine slope channel, Scarab field, offshore Nile Delta, Egypt: application of reservoir characterization. *Egypt. J. petroleum*. **25**, 495–508 (2016).
10. Farouk, S. et al. Petrographical, petrophysical, and geomechanical characterization of the Pliocene sandstone reservoirs of the Scarab Gas Field, deepwater Nile Delta, Egypt–Inferences on reservoir development. *J. Afr. Earth Sc.* **214**, 105259 (2024).
11. Abd El-Gawad, E. A. et al. Static reservoir modeling of El Wastani formation, for justifying development plans, using 2D seismic and well log data in Scarab field, offshore Nile Delta, Egypt. *J. Afr. Earth Sc.* **158**, 103546 (2019).
12. Nabih, M. et al. Rock physics analysis from predicted Poisson's ratio using RVFL based on Wild Geese Algorithm in Scarab gas field in WDDM concession, Egypt. *Mar. Pet. Geol.* **147**, 105949 (2023).
13. Mohamed, I. A. in *LIPE & AAPG European Region Northern Arabia Geoscience Conference & Exhibition in Lebanon on May 27–29, 2014. Neural Network Analysis versus Pre-Stack Seismic Inversion for Reservoir Characterization, Scarab Field, Offshore Nile Delta, Egypt*.
14. Mohamed, I. A., El-Mowafy, H. Z., Kamel, D. & Heikal, M. Prestack seismic inversion versus neural-network analysis: A case study in the Scarab field offshore Nile Delta, Egypt. *Lead. Edge* **33**, 498–506 (2014).
15. Ismail, A., Ewida, H. F., Al-Ibiary, M. G. & Zollo, A. Application of AVO attributes for gas channels identification, West offshore Nile Delta, Egypt. *Pet. Res.* **5**, 112–123 (2020).
16. Othman, A., Fathy, M. & Mohamed, I. A. Application of artificial neural network in seismic reservoir characterization: A case study from offshore Nile Delta. *Earth Sci. Inform.* **14**, 669–676 (2021).
17. El-Dabaa, S. A., Metwalli, F. I., Maher, A. & Ismail, A. Unsupervised machine learning-based multi-attributes analysis for enhancing gas channel detection and facies classification in the serpent field, offshore Nile Delta, Egypt. *Geomech. Geophys. Geo-Energy Geo-Resources*. **10**, 185 (2024).
18. Song, C. et al. Unsupervised seismic facies analysis with spatial constraints using regularized fuzzy c-means. *J. Geophys. Eng.* **14**, 1535–1543 (2017).
19. Roden, R., Smith, T. & Sacrey, D. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation* **3**, SAE59–SAE83. https://doi.org/10.1190/INT-2015-0037.1 (2015).
20. Lubo-Robles, D., Bedle, H., Marfurt, K. J. & Pranter, M. J. Evaluation of principal component analysis for seismic attribute selection and self-organizing maps for seismic facies discrimination in the presence of gas hydrates. *Mar. Pet. Geol.* **150**, 106097 (2023).
21. Aly Ismail, A. & Abdel Kader Boukhary, M. Ibrahim AbdelNaby, A. Subsurface stratigraphy and micropaleontology of the Neogene rocks, Nile Delta, Egypt. *Geologia Croatica*. **63**, 1–26 (2010).
22. El-Kahawy, R. M., Aboul-Ela, N., El-Barkooky, A. N. & Kassab, W. G. Bio-sequence stratigraphy of the Neogene: an example from El-Wastani gas field, onshore Nile Delta, Egypt. *J. Micropalaeontology*. **42**, 147–169 (2023).
23. Makled, W. A., Mandur, M. M. & Langer, M. R. Neogene sequence stratigraphic architecture of the Nile Delta, Egypt: A micropaleontological perspective. *Mar. Pet. Geol.* **85**, 117–135 (2017).
24. Kamel, H., Eita, T. & Sarhan, M. in *Proceedings of the 14th Egyptian General Petroleum Corporation Exploration and Production Conference, Cairo, Egypt. Nile Delta hydrocarbon potentiality* 485–503.
25. Younes, M. A. A. Natural Gas Geochemistry in the Offshore Nile Delta, Egypt. *Advances Petrochemicals*, 27–40 (2015).
26. Farouk, S., El-Din, A. A., Shennawy, E., Al-Kahtany, T., Hu, Y. & K. & Integration of seismic and well log data for structural evaluation of the West Delta Deep Marine, Egypt: implications for petrophysical analysis of the Messinian Abu Madi Reservoir. *Mar. Geophys. Res.* **46**, 16 (2025).
27. Aal, A. A. et al. Tectonic evolution of the Eastern Mediterranean Basin and its significance for hydrocarbon prospectivity in the ultradeepwater of the Nile Delta. *Lead. Edge* **19**, 1086–1102 (2000).
28. Abd-Allah, A. M., Aal, M. H. A. & Ghandour, A. Structural characteristics and tectonic evolution of the northwestern margin of the Nile Delta, Egypt. *J. Afr. Earth Sc.* **68**, 82–95 (2012).
29. Barber, P. M. Messinian subaerial erosion of the proto-Nile Delta. *Mar. Geol.* **44**, 253–272 (1981).
30. Pigott, J. D. & Abdel-Fattah, M. I. Seismic stratigraphy of the Messinian Nile Delta coastal plain: Recognition of the fluvial regressive systems tract and its potential for hydrocarbon exploration. *J. Afr. Earth Sci.* **95**, 9–21 (2014).
31. Khairy, A. et al. Geochemical assessment and hydrocarbon potential of Oligocene–Pliocene source rocks from northeast onshore Nile Delta, Egypt. *Int. J. Coal Geol.* **274**, 104283 (2023).
32. Vandré, C., Cramer, B., Gerling, P. & Winsemann, J. Natural gas formation in the western Nile delta (Eastern Mediterranean): Thermogenic versus microbial. *Org. Geochem.* **38**, 523–539 (2007).
33. El-Mowafy, H. Z., Ibrahim, M. & Dunlap, D. B. Unlocking gas reserves in bypassed stratigraphic traps in a deepwater brownfield using prestack seismic inversion: A case study from offshore Nile Delta, Egypt. *Lead. Edge* **37**, 502–509 (2018).
34. Eleslambouly, A., Zeynalli, M., Moncada, A., Fathy, A. & Rouxel, S. in *Carbon Capture, Utilization, and Storage Conference and Exhibition. Advancing carbon storage in offshore Louisiana: Evaluation and modeling potential of two major depleted reservoirs in Vermilion_014 Field* 198–217.
35. Szabó, N. P. & Dobróka, M. Extending the application of a shale volume estimation formula derived from factor analysis of wireline logging data. *Math. Geosci.* **45**, 837–850 (2013).
36. Larionov, V. *Radiometry of boreholes 127* (Nedra, 1969).
37. Schlumberger. *Log interpretation principles/applications* (Schlumberger Educational Services, 1991).
38. Tiab, D. & Donaldson, E. C. *Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties* (Elsevier, 2024).
39. Eleslambouly, A., Khalifa, T., Aldhanhani, O., Zeynalli, M. & Abdelmaksoud, A. Integrated field characterization and static hydrocarbon reserve estimations of the Penobscot Field, Nova Scotia, Canada. *Energy Geoscience*. **100448** https://doi.org/10.1016/j.engeos.2025.100448 (2025).

40. Yahia, L., Toni, M., Metwalli, F. I., Mansour, M. & Ismail, A. Integration of 3D Structural Modelling and Well Logging Analysis in the Evaluation Of October Oil field, Gulf of Suez, Egypt. *Trends Adv. Sci. Technol.* **1**, 5 (2024).
41. Eleslambouly, A. K., Zeynalli, M., Al-Shalabi, E. W. & Alsuwaidi, M. Probabilistic approach for static carbon storage capacity estimation: A case study on the VR014 depleted gas field in offshore Louisiana, USA. *Geoenergy Sci. Eng.* **260**, 214397. https://doi.org/10.1016/j.geoen.2026.214397 (2026).
42. Chopra, S., Marfurt, K. J. & 3SO-28SO. Seismic attributes—A historical perspective. *Geophysics* **70** https://doi.org/10.1190/1.2098670 (2005).
43. Eleslambouly, A., Aldhanhani, O., Fathy, A., Zeynalli, M. & Alsuwaidi, M. in *Abu Dhabi International Petroleum Exhibition and Conference. Geothermal Plays Feasibility and Potential in a Transform Basin: A Case Study from Southern Vienna Basin* D031S100R003 (SPE).
44. Chopra, S. & Marfurt, K. J. *Seismic attributes for prospect identification and reservoir characterization* (Society of Exploration Geophysicists and European Association of..., 2007).
45. Chopra, S. & Marfurt, K. J. Spectral decomposition and spectral balancing of seismic data. *Lead. Edge.* **35**, 176–179 (2016).
46. West, B. P., May, S. R., Eastwood, J. E. & Rossen, C. Interactive seismic facies classification using textural attributes and neural networks. *Lead. Edge* **21**, 1042–1049 (2002).
47. Ismail, A., Ewida, H. F., Al-Ibiary, M. G. & Zollo, A. Integrated prediction of deep-water gas channels using seismic coloured inversion and spectral decomposition attribute, west offshore, Nile Delta, Egypt. *NRIAG J. Astron. Geophys.* **9**, 459–470 (2020).
48. Torrado, L., Mann, P. & Bhattacharya, J. Application of seismic attributes and spectral decomposition for reservoir characterization of a complex fluvial system: Case study of the Carbonera Formation, Llanos foreland basin, Colombia. *Geophysics* **79**, B221–B230 (2014).
49. Chopra, S. & Marfurt, K. Seismic Attributes–a promising aid for geologic prediction. *CSEG recorder*. **31**, 110–120 (2006).
50. Oumarou, S. et al. Seismic attributes in reservoir characterization: an overview. *Arab. J. Geosci.* **14**, 1–15. https://doi.org/10.1007/s12517-021-06626-1 (2021).
51. Xu, Z. et al. Software defect prediction based on kernel PCA and weighted extreme learning machine. *Inf. Softw. Technol.* **106**, 182–200. https://doi.org/10.1016/j.infsof.2018.10.004 (2019).
52. Zhang, Y. et al. Unsupervised seismic facies analysis technology based on SOM and PSO. *Chin. J. Geophys.* **58**, 3412–3423. https://doi.org/10.6038/cjg20150933 (2015).
53. Kohonen, T. *Self-organizing maps* Vol. 30 (Springer Science & Business Media, 2012).
54. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **37**, 52–65 (2013).
55. Coléou, T., Poupon, M. & Azbel, K. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *Lead. Edge*. **22**, 942–953. https://doi.org/10.1190/1.1623635 (2003).
56. de Matos, M. C., Osorio, P. L. & Johann, P. R. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics* **72**, P9–P21 (2007).
57. Zhao, T., Jayaram, V., Roy, A. & Marfurt, K. J. A comparison of classification techniques for seismic facies recognition. *Interpretation* **3**, SAE29–SAE58. https://doi.org/10.1190/INT-2015-0044.1 (2015).
58. Roden, R. & Chen, C. W. Interpretation of DHI characteristics with machine learning. *First Break* **35**(5). https://doi.org/10.3997/1365-2397.35.5.88069 (2017).
59. Assenza, A., Valle, M. & Verleysen, M. A comparative study of various probability density estimation methods for data analysis. *Int. J. Comput. Intell. Syst.* **1**, 188–201. https://doi.org/10.2991/ijcis.2008.1.2.9 (2008).
60. Davis, J. Justification of probability density function for resolution in statistical models of overlap. *Chromatographia* **44**, 81–90. https://doi.org/10.1007/BF02466520 (1997).
61. Montanari, A. & Calò, D. G. Model-based clustering of probability density functions. *Adv. Data Anal. Classif.* **7**, 301–319. https://doi.org/10.1007/s11634-013-0140-8 (2013).
62. Masalmeh, S. K., Shiekah, I. A. & Jing, X. Improved characterization and modeling of capillary transition zones in carbonate reservoirs. *SPE Reservoir Eval. Eng.* **10**, 191–204 (2007).
63. Abreu, V., Sullivan, M., Pirmez, C. & Mohrig, D. Lateral accretion packages (LAPs): an important reservoir element in deep water sinuous channels. *Mar. Pet. Geol.* **20**, 631–648. https://doi.org/10.1016/j.marpetgeo.2003.08.003 (2003).
64. Hulea, I. N. Perched water contacts: Understanding fundamental controls. *Petrophysics* **60**, 438–449 (2019).

Acknowledgements

We are grateful to the Egyptian General Petroleum Corporation (EGPC) for granting us access to the data set required for our investigation. The Department of Geology & Geophysics, Texas A&M University, and the Department of Geophysics, Faculty of Science, Capital University, are gratefully acknowledged for providing the research environment and computational resources to support this work. The authors are grateful to Geophysical Insight for providing access to the Paradise software for the SOM analysis. Additionally, we express our gratitude to the Fulbright Egypt Commission for awarding Tarek Khalifa a Master's scholarship to Texas A&M University-College Station.

Author contributions

T.K.: Methodology, Data Analysis, Investigation, Writing-Original Draft, Software, Visualization, and Review. A.E.: Writing-Original Draft, Software, Visualization, Writing - Review & Editing. A.E.K.: Supervision, Writing-Review & Editing. A.I. : Methodology, Data Analysis, Investigation, Data Curation, Validation, Supervision, Writing-Review & Editing.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/1

0.1038/s41598-026-75298-2.

Correspondence and requests for materials should be addressed to A.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026