Physics-guided attribute selection for improved direct hydrocarbon indication

Mohamed Ahmed Gad[1*] and John Castagna[1]

University of Houston, Houston, Texas 77204, USA
mgad@uh.edu*

The non-uniqueness of seismic responses to rock and fluid properties remains a major challenge in hydrocarbon exploration because similar seismic anomalies may arise from different geological factors. Consequently, identifying seismic attributes that are specifically sensitive to pore-fluid effects is critical for reducing the exploration risk associated with direct hydrocarbon indicators. In this study, we introduce a Physics-Guided Local Variability attribute-selection methodology that integrates synthetic sensitivity analysis with local seismic variability measurements to identify physically meaningful seismic attributes for hydrocarbon detection. The methodology is demonstrated in the Hutton Sandstone reservoir of the Eromanga Basin, Australia which is often targeted on structural highs where dry holes may have seismic responses much like nearby commercial reservoirs. Due to strong lateral stratigraphic variability, conventional direct hydrocarbon indicators sometimes fail. In this case study, statistical attribute-selection approaches, including principal component analysis followed by self-organizing map classification, also failed to distinguish commercial reservoirs from nearby dry holes. Synthetic perturbation analysis at a single calibration well was used to quantify the relative influence of fluid, porosity, and thickness variations on seismic attributes, whereas local variability analysis applied to real seismic data was used to evaluate attribute behavior around wells with known fluid saturations. Spectral components in the 30–35 Hz range were identified as the most diagnostic attributes for pore-fluid detection, with the strongest responses observed in the far-angle data. These attributes improved self-organizing map classification, seismic inversion, and amplitude-versus-offset interpretation, resulting in improved discrimination between hydrocarbon-bearing and water-saturated intervals. Blind validation at an independent well further confirmed the robustness of the selected attributes.

Introduction:

Seismic attributes are widely used in reservoir characterization to identify lithology and pore-fluid content [1, 2]. Traditional interpretation tools such as bright spot methodology, seismic inversion, and amplitude variation with offset (AVO) analysis have been successfully used in many geological settings, but they sometimes fail when lateral geological facies variability is strong due to non-uniqueness of the seismic response. In such cases, anomalous seismic amplitudes may reflect lithologic or stratigraphic variability (such as "shaling out" or thickness changes) rather than pore-fluid effects, thereby leading to dry holes.

Some studies have shown that spectral analysis of seismic data can improve pore-fluid discrimination from seismic data [3, 4]. These findings highlight that reservoir can be most anomalous at specific frequency components. The challenge is: with so many seismic attributes available, and in a particular locality, how can we select those attributes that are truly diagnostic of pore-fluid effects rather than lithologic or stratigraphic changes?

A traditional approach to attribute selection is principal component analysis (PCA), which combines attributes based on global variance [5]. Although PCA is widely used, it does not distinguish whether variance arises from lithological or fluid effects. Further, it is data driven, weighting by the variance of the entire sample analyzed while commercial reservoirs may contribute only a small amount to that variance. As a result, attributes affected by lithology may be preferentially weighted and, hence, dominate machine learning applications. The most important attributes such as spectral components most strongly related to hydrocarbon presence may be overlooked [6]. These limitations highlight the need for an attribute-selection strategy that evaluates seismic responses locally and distinguishes pore-fluid effects from other sources of variability.

To overcome the PCA limitation, we have developed a physics-guided alternative approach: the physics-guided local variability (PGLV) method. This approach quantitatively estimates the local attribute variance around wells with known fluid saturations and synthetic seismograms involving perturbations to rock and fluid properties in those wells. By implementing the condition, for example, that an attribute's variance decreases when hydrocarbons are absent, PGLV emphasizes attributes that are most sensitive to pore-fluid changes. These selected attributes are then used as input to a self-organizing map (SOM), resulting in seismic facies classifications that are consistent with the pore fluids at the wells. In addition, the same attributes enhance the reliability of conventional physics-based interpretation tools such as seismic inversion and AVO analysis.

The physics-guided local variability (PGLV) workflow introduced in this study is shown in Figure 1. We first extract a set of seismic attributes, including complex trace attributes [7], AVO attributes [8, 9], and spectral decomposition attributes [10]. All attributes are initially evaluated at a single calibration well using sensitivity analysis to quantify the relative influence of pore-fluid, porosity, and thickness variations on attribute responses. The attributes are subsequently validated using the seismic data through local variability measurements to assess whether hydrocarbon-bearing regions exhibit greater variability than wet regions. Attributes are retained when their rankings are consistent between the calibration-well sensitivity analysis and the seismic local-variability evaluation. The resulting attribute set is then subjected to blind validation at an independent well that was not used during attribute selection or workflow development to assess the transferability and predictive capability of the methodology.

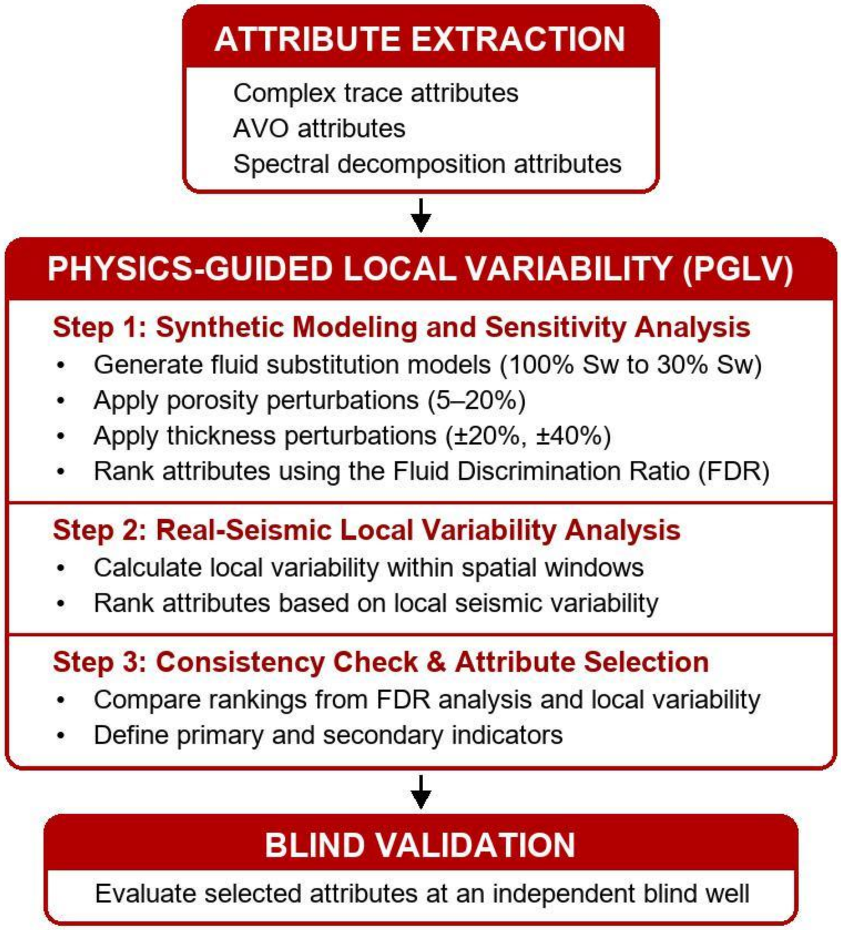

Figure 1: The physics-guided local variability (PGLV) workflow.

Geological and Dataset Background:

The study area is in the Queensland Field, Eromanga Basin, Onshore Australia (Figure 2) where a target reservoir is the Hutton Sandstone Formation. The regional structural framework of the Eromanga Basin is represented by four-way dip closed anticlinal trends in a regional sag basin [11]. Previous studies of the Hutton Sandstone within the Eromanga Basin describe the reservoir as a fluvial sandstone system composed of amalgamated channel sands and sheet-like sand bodies interbedded with mudstones [12, 13]. These sandstones were deposited in a high-energy braided fluvial environment and exhibit fining-upward cycles, with limited shale interbeds and strong vertical and lateral facies variability. Local scour and incision by the overlying Birkhead Formation further introduce stratigraphic compartmentalization within the upper Hutton interval [13, 14].

This heterogeneity contributes to seismic non-uniqueness and motivates the need for localized attribute analysis.

The available dataset includes 3D seismic data over the study area. The Togar 3D seismic survey was acquired using a vibroseis source with a 7–110 Hz sweep, 2 ms sample rate, and a recording template of 10 receiver lines × 96 live channels, giving 960 active channels per shot. The receiver line spacing was 240 m, and the receiver station interval was 60 m. The source line spacing was also 240 m, with a source station interval of 42.4 m. The survey was acquired with receiver lines oriented at approximately 45°, and source lines oriented at 0° and 90°, giving broad/full-azimuth coverage. The nominal fold was 60, with recorded offsets ranging from 0 m to 2400 m. The processed 3D volume was loaded with inline and crossline increments of 1 and a subsurface bin spacing of 15 m × 15 m. The inline axis has an azimuth of 315°, and the crossline axis has an azimuth of 45°.

The seismic data are provided as 1) a full-stack volume, 2) a near-stack volume corresponding to 0°–15° incidence angles and 0–600 m offsets, 3) a mid-stack volume corresponding to 15°–30° incidence angles and 600–1200 m offsets, 4) a far-stack volume corresponding to 30°–45° incidence angles and 1200–1800 m offsets, and 5) an ultra-far-stack volume corresponding to 45°–60° incidence angles and 1800–2400 m offsets. The calculated P-wave critical angle at the Hutton interface is approximately 73°, which is greater than the maximum incidence angle represented in the available seismic data. AVO compliant seismic data conditioning was applied to the angle gathers prior to interpretation, including Radon filtering, time-variant spectral balancing, trim statics, phase correction, and band-pass filtering. These steps were intended to improve data quality and facilitate AVO analysis.

Three suites of conventional well-log curves from three wells (Andromedae South, Pyxis-1 and Pyxis-2) in the study area include two wells (Andromedae South and Pyxis 1) that have hydrocarbons production in the Hutton formation and one dry hole (Pyxis-2). We also used Andromedae-1, a well located outside the calibration area, as a blind validation well. Andromedae-1 has hydrocarbons with 30% water saturation. All wells have caliper log, resistivity log, SP log, gamma-ray log, sonic log, density log and neutron log. The seismic to well tie at the Andromedae South well for the far stack (30°–45° angle stack) provides confidence in identifying the Hutton formation (Figure 3).

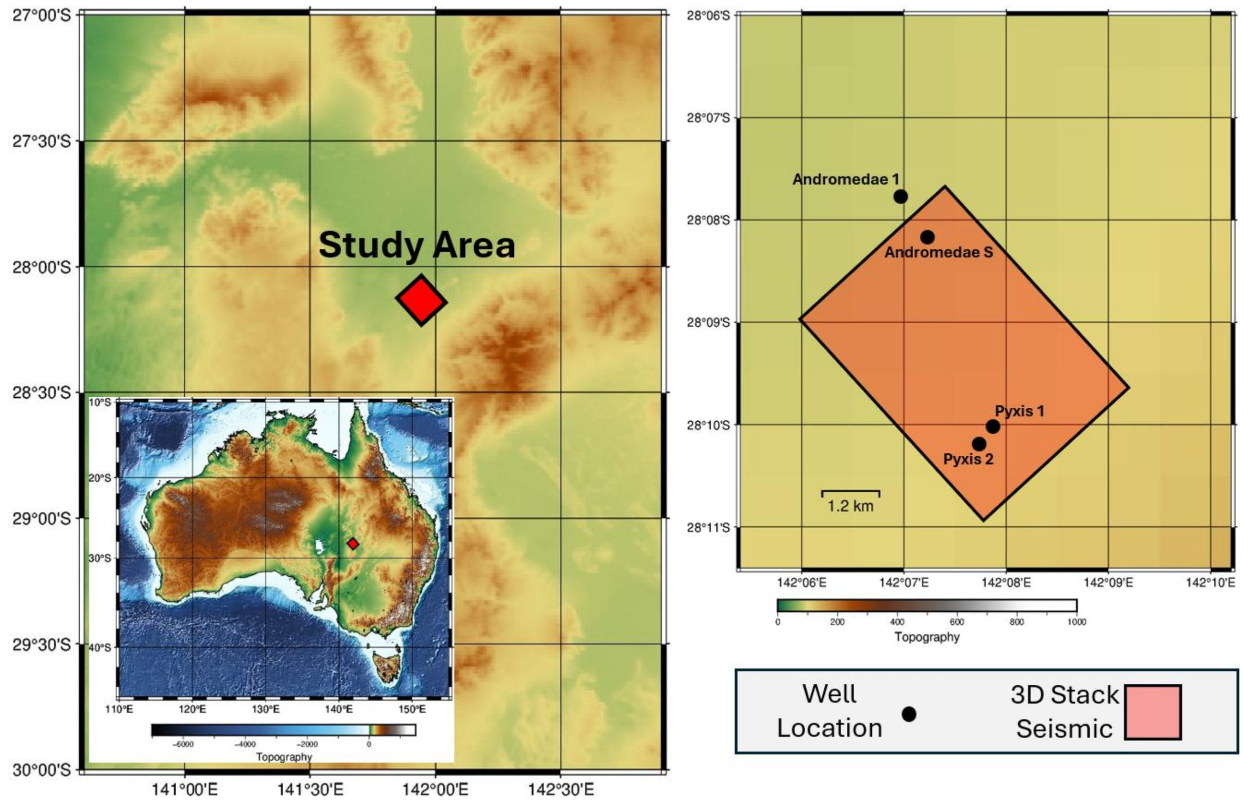

Figure 2: Topographic maps show the location of the study area onshore Australia.

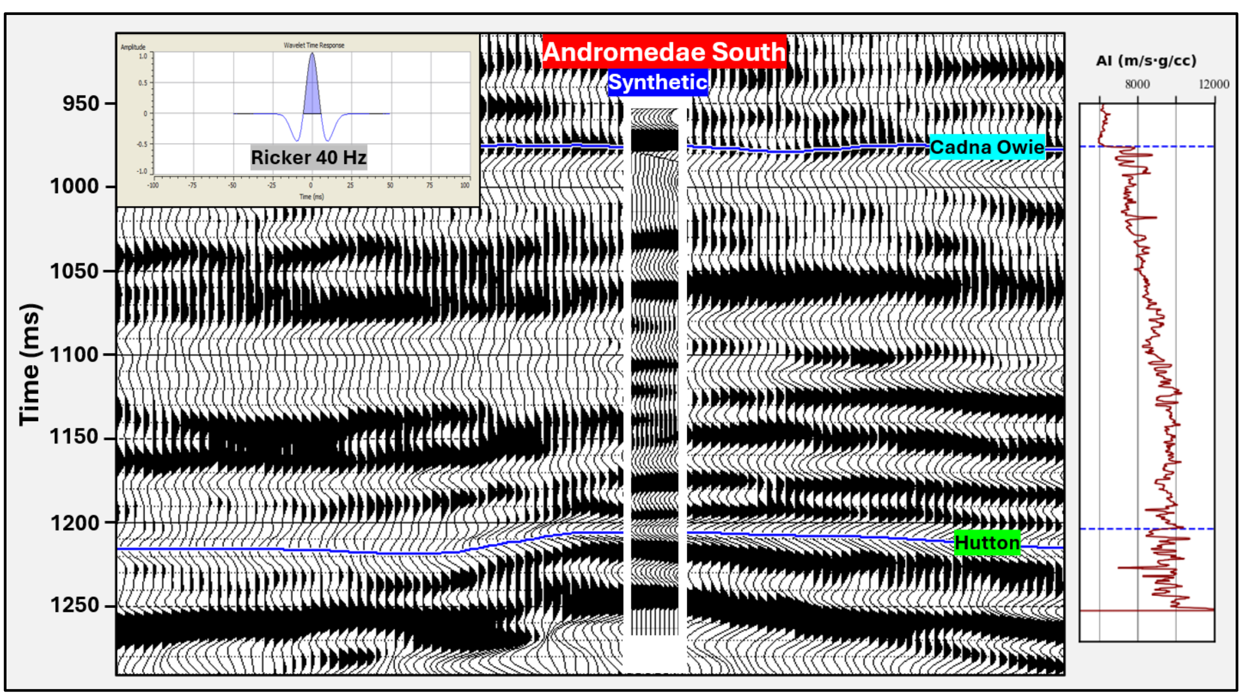

Figure 3: Seismic to well tie at Andromedae South well. Synthetic and real seismic data represent the far stack. Synthetic seismograms were generated using a zero-phase 40 Hz Ricker wavelet with a bandwidth of approximately 20–60 Hz. The Cadna–Owie formation produces a regional seismic marker at about 1200 m depth in the study area and the Hutton formation at approximately 1600 m depth is the main reservoir target. AI is the acoustic impedance log. The two blue lines on the acoustic impedance track indicate the tops of the Cadna–Owie and Hutton formations at the well location.

An amplitude map along the Hutton/Birkhead interface is shown in Figure 4 with structural contours showing four-way plunging anticlinal closures which are prospective traps for hydrocarbons. The key challenge in this case study is the strong lateral variability, which makes it difficult to predict hydrocarbon distribution in the Hutton Sandstone. For example, the two wells (Pyxis-1 and Pyxis-2) drilled on the same anticlinal closure gave contrasting results: the up-dip well, Pyxis-2, was dry, while the mid-dip possibly stratigraphically or fault separated well, Pyxis-1, produced hydrocarbons from the same stratigraphic interval. What makes it even harder, although the Hutton here is low impedance, the amplitude map at the top reservoir shows anomalous amplitude response at the water-saturated zone and weak amplitude at the hydrocarbon-bearing zone. Looking at the target reservoir in the near, mid and far-angle stacks in Figure 5, the amplitude changes with offset at the Pyxis-2 dry hole but there is no clear AVO anomaly at Pyxis-1 which has hydrocarbons.

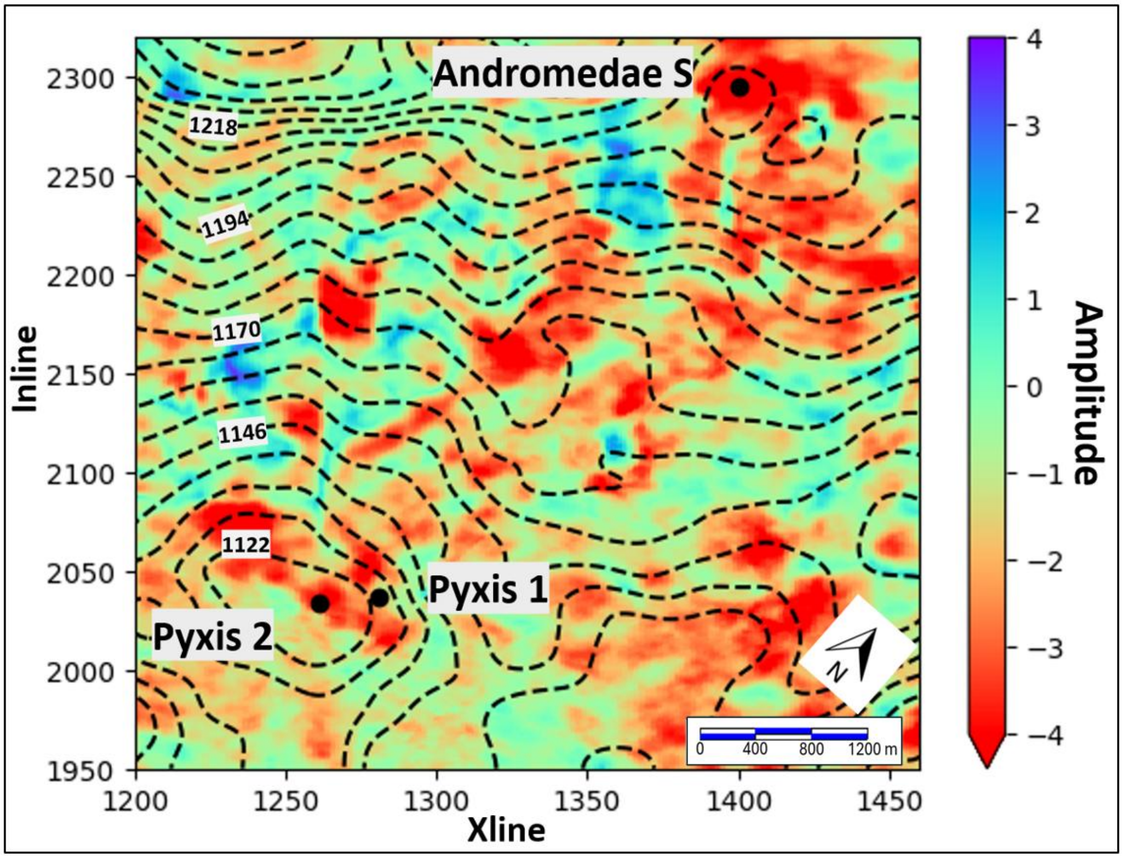

Figure 4: Amplitude map at the Hutton–Birkhead interface with structural contours (24 ms interval) showing four-way plunging anticlinal closures. Despite being located within the same structural closure, the Pyxis-1 well (hydrocarbon-bearing) is associated with weak amplitude response, whereas the up-dip Pyxis-2 well (dry) exhibits strong amplitudes.

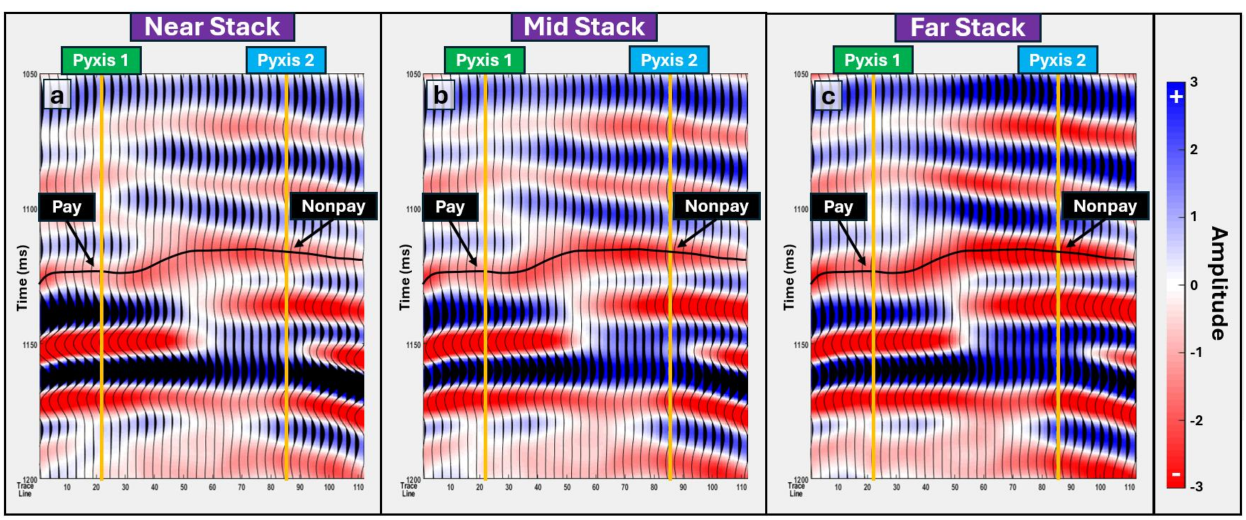

Figure 5: A vertical display of seismic data along Pyxis-1 and Pyxis-2 wells. a) Near stack. b) Mid stack. c) Far stack. The dry Pyxis-2 well shows amplitude variation with offset at the target interval, whereas the hydrocarbon-bearing Pyxis-1 well exhibits no clear AVO anomaly.

The petrophysical analysis of the well logs shows that the water saturations at Andromedae South, Pyxis-1 and Pyxis-2 are 30%, 60% and 100%, respectively, estimated using the Archie Equation [15]. Despite an average water saturation of approximately 60% in Pyxis-1, drill stem test (DST) results and subsequent production indicate the presence of commercial hydrocarbons in this well. Calculated water saturations above 50% sometimes commercially produce hydrocarbons for a variety of reasons (e.g., pay zones thinner than log resolution or presence of conductive minerals in the reservoir).

Considering plausible ranges of Archie parameters and log uncertainties, the estimated water saturation uncertainty is approximately ±15.7–17.4%. The water saturation at Andromedae South looks consistent with the seismic amplitude response. However, the water saturations at Pyxis-1 and Pyxis-2 are inconsistent with the expectation in a low impedance reservoir rock that hydrocarbons will exhibit brighter amplitudes. The stratigraphic lateral variation is evident from mineral volumetrics logs at the three wells (Figure 6). The ratio of quartz to clay volume fractions for the interval decreases toward Pyxis-2. This suggests that the strong amplitude at Pyxis-2 is related to stratigraphic variation rather than a bright spot due to hydrocarbons.

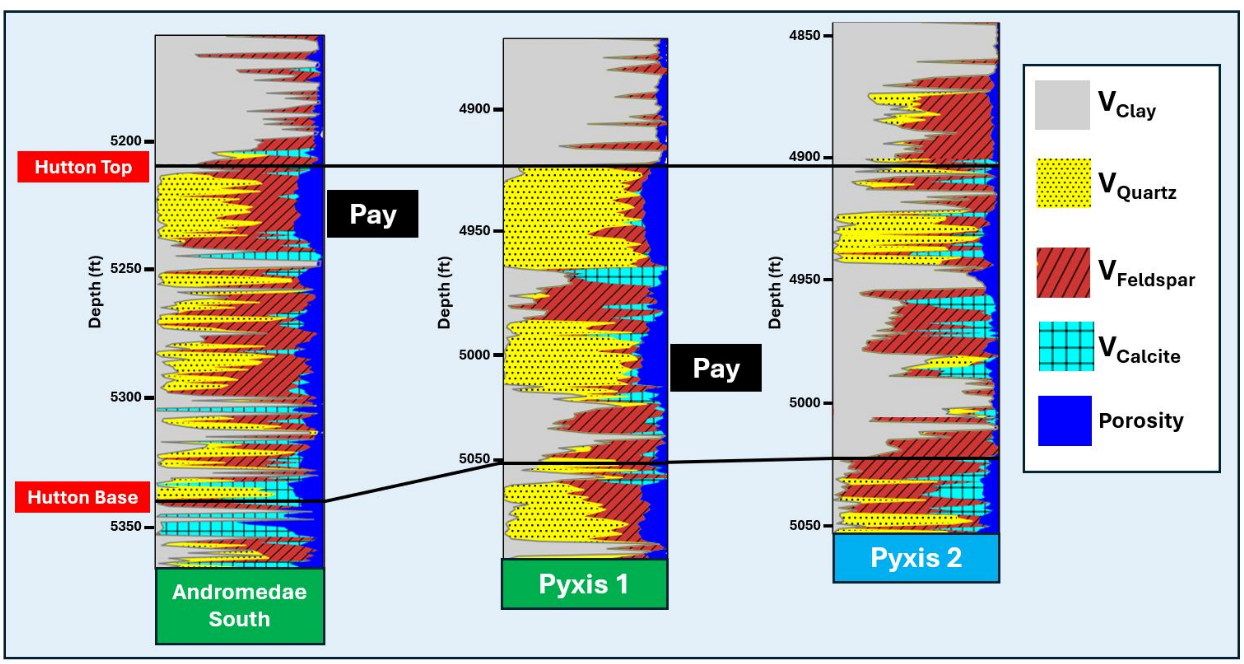

Figure 6: Mineral volumetrics at Andromedae South, Pyxis 1 and Pyxis 2 at the Hutton Sandstone Formation. V's are volume fractions of the subscripted constituents. The wells are displayed using the Hutton Top as a datum reference to facilitate stratigraphic correlation between wells.

Methods:

Physics-Guided Local Variability (PGLV):

The Physics-Guided Local Variability (PGLV) methodology combines sensitivity evaluation at one or more calibration wells with local seismic variability measurements to identify seismic attributes that are the most sensitive to pore-fluids. Prior to the analysis, normalization or standardization is applied to the seismic attributes to ensure that all attributes are treated equally during variance estimation. The variance per attribute is estimated using the following equation:

$$s_{j,Z}^{2} = \frac{1}{n_Z - 1} \sum_{(i,x)\in Z} \left(a_j(i,x) - \overline{a}_{j,Z} \right)^2 \tag{1}$$

where $S_{j,Z}^2$ is the variance of attribute a_j within an analysis zone Z, n_Z is the number of samples within Z, i and x denote the inline and crossline indices, respectively, and $\bar{a}_{j,Z}$ is the mean value of attribute j within the same zone. For 1D synthetic models, Z represents a fixed time window at the target interval along the synthetic trace. For real seismic data, Z represents a local spatial window around each well sufficiently large to extend below any possible water contact given expected column heights in the area and defined by ranges of inlines and crosslines within a predefined time window.

The attribute selection process proceeds through three main steps. All attributes are first evaluated at the calibration well using perturbations of pore fluid saturation, porosity, and layer thickness to quantify the degree to which attribute variability is controlled by pore-fluid effects relative to porosity and thickness variations. Secondly, the attributes are validated in the real seismic data using local-variability analysis to determine whether they exhibited the predicted high variability in hydrocarbon-bearing local areas relative to wet regions. Finally, the PGLV-selected attributes are identified based on the consistency between the calibration-well and real-seismic variability rankings.

To establish the expected physical behavior of each attribute, synthetic seismic responses are generated under different fluid-saturation, porosity, and thickness scenarios while holding the remaining parameters constant. Attribute variability is then estimated from these synthetic responses across a range of perturbation scenarios, capturing how each attribute responds to changes in pore-fluid content, porosity, and layer thickness independently. In this case study, a single calibration well was used to construct the perturbation models. A schematic illustrating this sensitivity workflow is shown in Figure 7.

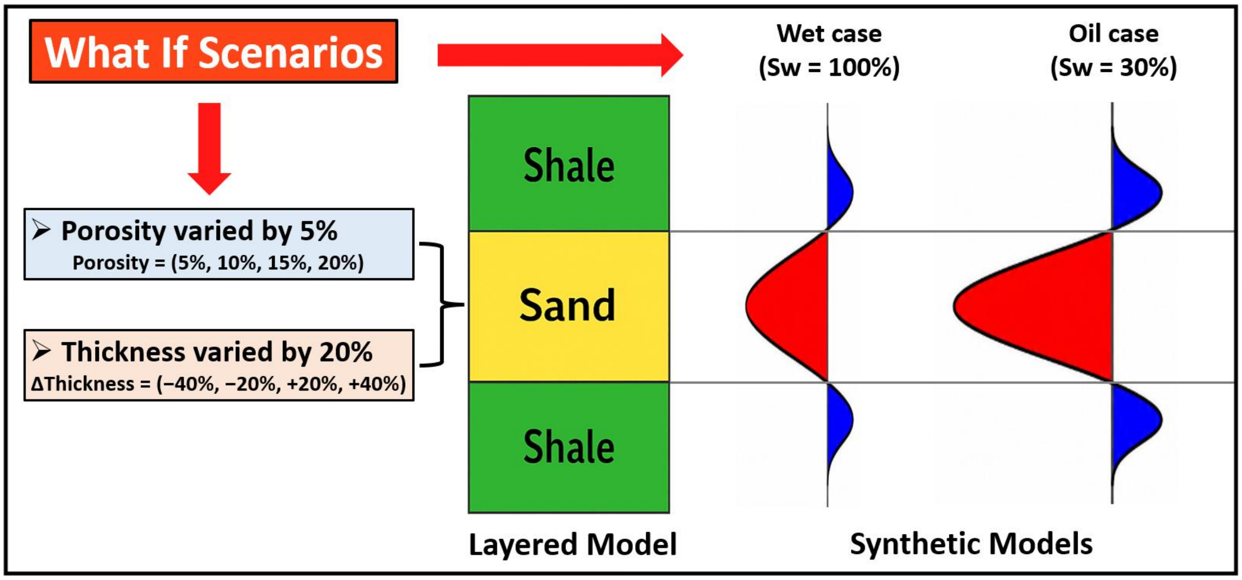

Figure 7. Conceptual schematic illustrating the sensitivity-analysis framework. Porosity and thickness perturbations are applied to a reference one-dimensional layered sand–shale model. The modeled reservoir interval occurs near tuning-thickness conditions, where interference between top and base reflections influences the resulting seismic response. Synthetic seismic responses for the selected ranges of porosities and thicknesses are compared for two water-saturation cases: a water-saturated case (Sw = 100%) and an oil case (Sw = 30%).

The Fluid Sensitivity Index (FSI) is defined as the variance of the difference between the hydrocarbon-bearing and water-saturated cases:

$$\mathrm{FSI}(A) = \mathrm{Var}(A^{\mathrm{HC}} - A^{\mathrm{wet}}) \tag{2}$$

where A^{HC} and A^{wet} are the attribute responses for the hydrocarbon-bearing and fully water-saturated models, respectively. Synthetic fluid-saturation scenarios are generated using Gassmann fluid substitution [16] to estimate the corresponding elastic and seismic responses. When shear-wave velocity logs were not available, the Greenberg and Castagna method [17] was used to predict shear-wave velocity from the available P-wave velocity logs. Brine and hydrocarbon elastic properties were computed using the Batzle and Wang fluid-property relationships [18].

The Porosity Sensitivity Index (PSI) is defined as the average variance of the difference between porosity-perturbed models and the water-saturated baseline:

$$\mathrm{PSI}(A) = \frac{1}{N_\phi} \sum_{i=1}^{N_\phi} \mathrm{Var}\left(A^{\phi_i} - A^{\mathrm{wet}}\right) \tag{3}$$

where A^{ϕ_i} represents the attribute response for the i-th porosity-perturbed model and N_ϕ is the number of porosity scenarios. Porosity perturbations were modeled using the Raymer–Hunt–Gardner velocity–porosity relationship [19] to estimate the effect of porosity variations on elastic properties and the corresponding seismic response.

The Thickness Sensitivity Index (TSI) is defined as the average variance of the difference between thickness-perturbed models and the water-saturated baseline:

$$\mathrm{TSI}(A) = \frac{1}{N_h} \sum_{j=1}^{N_h} \mathrm{Var}\left(A^{h_j} - A^{\mathrm{wet}}\right) \tag{4}$$

where A^{h_j} represents the attribute response for the j-th thickness-perturbed model and N_h is the number of thickness scenarios.

The Background Variability Index (BVI) represents non-fluid-driven variability and is defined as the combined contribution of porosity and thickness effects:

$$\mathrm{BVI}(A) = \text{Porosity Sensitivity Index} + \text{Thickness Sensitivity Index} \tag{5}$$

The Fluid-to-Porosity Ratio (FPR) compares fluid-related variability to variability caused by porosity changes, while the Fluid-to-Thickness Ratio (FTR) compares fluid-related variability to variability caused by thickness changes. The Fluid Discrimination Ratio (FDR) compares fluid sensitivity to the total background variability resulting from non-fluid effects:

$$\mathrm{FPR} = \frac{\text{Fluid Sensitivity Index}}{\text{Porosity Sensitivity Index}},$$

$$\mathrm{FTR} = \frac{\text{Fluid Sensitivity Index}}{\text{Thickness Sensitivity Index}},$$

$$\mathrm{FDR} = \frac{\text{Fluid Sensitivity Index}}{\text{Background Variability Index}} \tag{6}$$

Values greater than 1 indicate that fluid-related variability dominates over non-fluid contributions, implying that the attribute is more effective for fluid discrimination. The attributes are subsequently ranked according to their FDR values, with higher FDR values indicating stronger fluid sensitivity relative to background geological variability.

In parallel with the calibration-well sensitivity estimates, the set of the attributes is evaluated in the real seismic data using local variance measurements around hydrocarbon-bearing and water-saturated wells. The objective is to test that the identified attributes from synthetics consistently exhibit high local variability in hydrocarbon-bearing local areas relative to water-saturated local regions while also demonstrating strong fluid sensitivity at the calibration well.

Attributes are retained only if the hydrocarbon-related variability exceeds the water-saturated-case variability:

$$S_{j\,HC}^{2} > S_{j\,wet}^{2} \quad (7)$$

where HC and wet denote the hydrocarbon-bearing and fully water-saturated cases, respectively. More specifically in this case study:

$$S_{j\,30\%}^{2} > S_{j\,60\%}^{2} > S_{j\,100\%}^{2} \quad (8)$$

where the suffix represents the interpreted water saturation used to order the calibration zones. The numerical saturation values are not used directly in calculating the attribute ranks. Large uncertainties in water saturation estimates or subtle fluid effects may affect this criterion. In such cases, the selection criterion can be simplified to a pay versus non-pay comparison.

For the attributes retained from the real-seismic local-variability analysis, diagnostic ratios are computed:

$$VR_j^{30/100} = \frac{S_{j\,30\%}^{2}}{S_{j\,30\%}^{2}+S_{j\,100\%}^{2}}, VR_j^{60/100} = \frac{S_{j\,60\,\%}^{2}}{S_{j\,60\%}^{2}+S_{j\,100\%}^{2}}, VR_j^{30/60} = \frac{S_{j\,30\%}^{2}}{S_{j\,30\%}^{2}+S_{j\,60\%}^{2}} \quad (9)$$

where VR is the variance ratio between the superscripted water saturation cases, $S_{j\,30\%}^{2}$ $S_{j\,60\%}^{2}$, $S_{j\,100\%}^{2}$ are the variance of attribute a_j in zones with 30%, 60%, and 100% water saturation, respectively. The attributes are then ranked using a composite local variability score (S_j):

$$S_j = VR_j^{30/100} + VR_j^{60/100} + VR_j^{30/60} \tag{10}$$

Final PGLV-selected attributes are identified based on the consistency between the calibration-well fluid-sensitivity ranking and the real-seismic local-variability ranking. Attributes exhibiting strong agreement between both rankings are interpreted as robust fluid-sensitive attributes, whereas attributes exhibiting high ranking in only one analysis are treated as secondary indicators.

Principal Component Analysis (PCA):

PCA is an unsupervised method commonly used to reduce dimensionality in seismic attribute analysis [5, 6, 20, 21, 22] and can be used for attribute selection based on global variance. The first principal component (1^{st} PCA) represents the maximum variation in the data while the second principal component (2^{nd} PCA) captures the next level of variation (Figure 8), and so forth, with the number of dimensions being the number of input attributes.

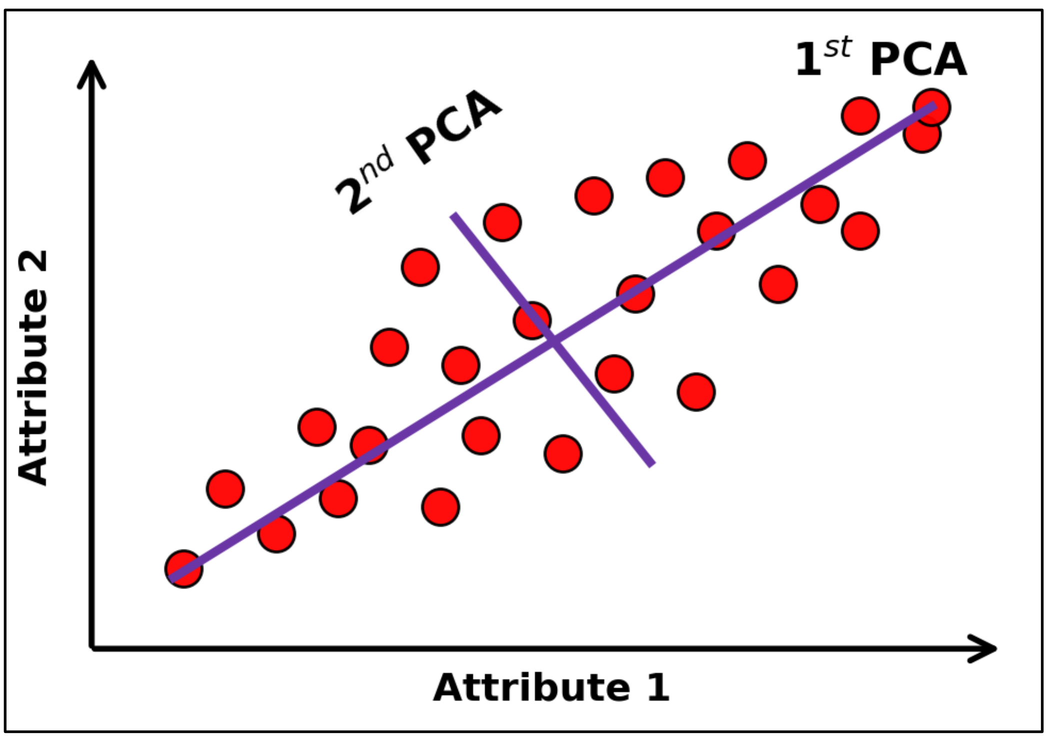

Figure 8: The first and the second principal components for a two-dimensional space.

The covariance matrix is:

$$C = \frac{1}{N-1} X^{\top} X \tag{11}$$

where X is the attribute data matrix with N observations. The eigenvectors of C represent the principal directions of variation, and the corresponding eigenvalues estimate the variance explained by each component. PCA finds the directions (eigenvectors) that maximize the global variance, then projects attributes on them. Because PCA ranks attributes based on global variance, it does not distinguish between variance caused by lithologic, thickness, or pore-fluid effects, which can limit its effectiveness for fluid discrimination in geologically heterogeneous settings. As hydrocarbon effects are present in only a small fraction of the data, applying PCA to the entire dataset may be inappropriate for hydrocarbon detection, as those effects may be masked in the principal components accounting for most of the variance.

Self-Organizing Map (SOM)

A self-organizing map (SOM) is an unsupervised machine learning algorithm that projects high-dimensional input data (seismic attributes) onto a two-dimensional lattice of neurons [20, 21, 22, 23, 24, 25, 26, 27, 28]. It has been used in clustering seismic attributes into classes of different facies. SOM is not the primary focus of this study but is used as an unsupervised classification to evaluate how different attribute-selection methods influence facies discrimination.

During the training process, neurons update their weights so that similar inputs are mapped close to each other. This allows the SOM to group seismic data into clusters representing facies, without the need for labeled data. Figure 9 shows the structure of the self-organizing map neural network [29].

The following equation [26, 27, 28] shows the weight vector update of the neuron *i* in the SOM:

$$w_i(t+1) = w_i(t) + \alpha(t)\, h_{bi}(t)\, [x(t) - w_i(t)] \tag{12}$$

where $\alpha(t)$ is the learning rate and $h_{bi}(t)$ is the neighborhood function that estimates the influence of Best Matching Unit (Neuron b) on a neuron *i* at iteration t.

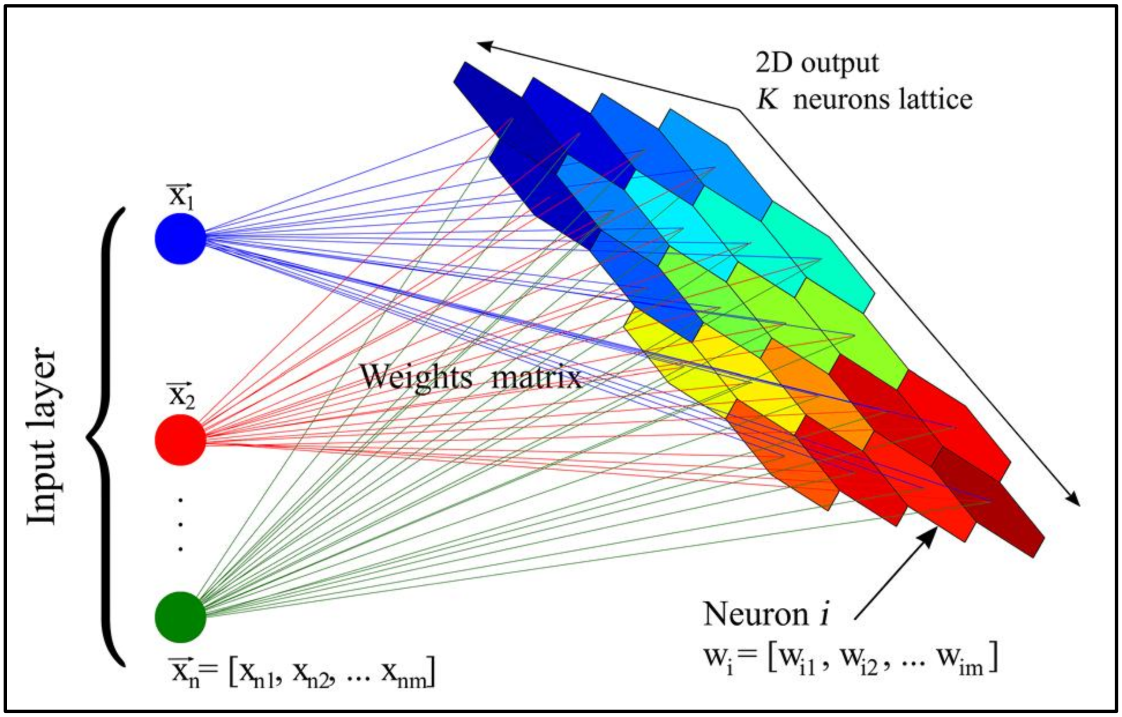

Figure 9: The Self-Organizing Map Neural Network structure [29]. Each input vector (x_i) contains the normalized seismic-attribute values at one seismic location and is connected to the neurons in the two-dimensional output lattice through their weight vectors.

Conventional Seismic Analysis

Seismic inversion was conducted using a wavelet-independent, multiscale reflectivity inversion approach [30], which does not require wavelet estimation, horizon constraints, well-log calibration, or user-defined regularization parameters. This method produces a band-limited impedance suitable for relative comparison and interpretation. In addition, angle-stack seismic data were analyzed using conventional AVO-based interpretation. The near, mid, and far stacks correspond approximately to incidence-angle ranges of 0°–15°, 16°–30°, and 31°–45°, respectively. Only these angle stacks were used for the AVO analysis conducted in this study. Frequency-dependent AVO responses were further evaluated through spectral decomposition applied to the angle stacks. Seismic inversion and AVO analysis are not the primary focus of this study but were used as interpretation tools to evaluate how different attribute-selection methods influence the resulting seismic responses and reservoir characterization.

Results:

Conventional attribute analyses, including complex trace attributes [7], AVO attributes [8, 9], spectral decomposition using the continuous wavelet transform (CWT) [31], and Constrained Least-Squares Spectral Analysis (CLSSA) [32] were extracted for reservoir-scale interpretation. These attributes formed the basis of an initial seismic facies classification workflow. Some examples of the extracted seismic attributes are shown in Figure 10. We tested two methods for seismic attribute selection including principal component analysis (PCA) and the physics-guided local variability (PGLV) approach developed in this study.

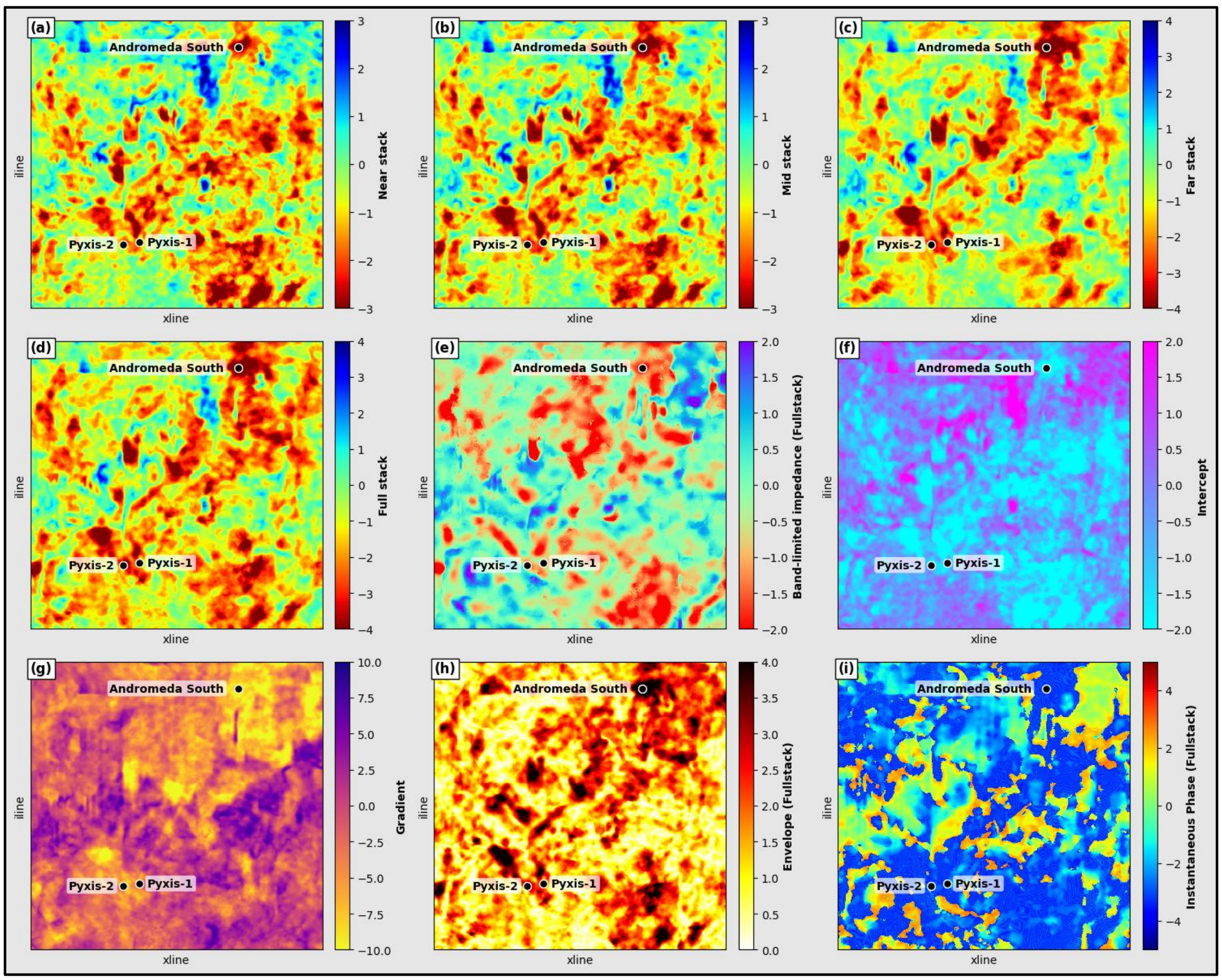

Figure 10: Examples of input seismic attributes extracted along the top Hutton Sandstone horizon slice: (a) Near stack, (b) Mid stack, (c) Far stack, (d) Full stack, (e) Band-limited seismic inversion of the full stack, (f) AVO intercept, (g) AVO gradient, (h) Envelope of the full stack, and (i) Instantaneous phase of the full stack.

PCA Analysis:

We first conducted PCA and examined the first and second principal components shown in Figure 11. Each attribute's contribution is represented by a variance value. Attributes with the highest variance in the first and second principal components were selected. The seismic attributes selected by PCA include the mid- and far-offset stacks, the full stack and the AVO gradient. Because PCA selects attributes that explain the largest amount of global variance, the resulting attribute set reflects the dominant variability present across the entire survey without distinguishing whether that variability is related to pore-fluid effects, lithologic variations, reservoir thickness, or other geological and geophysical factors. These PCA-selected attributes were subsequently used for comparison with the attributes selected by the proposed physics-guided local variability (PGLV) approach.

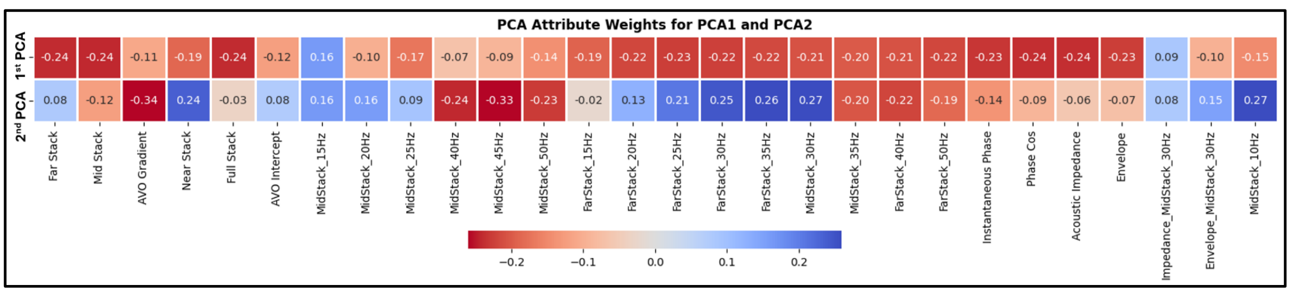

Figure 11: The attribute variances for the first and second principal components.

PGLV Attribute Selection:

To determine the expected attribute sequence using the PGLV method, synthetic modeling was first performed at a single calibration well (Pyxis-2). Sensitivity analysis started with fluid substitution from the in-situ case with 100% water saturation to an oil case with 30% water saturation. Iso-frequency gathers were then generated to examine the frequency response in the far stack, as shown in Figure 12. A pronounced trough anomaly at the 30–35 Hz components corresponds closely to the reservoir top.

The frequency-dependent spectral analysis of the near, mid, and far stacks shows that the largest spectral-amplitude differences between the oil and water-saturated cases occur within the 30–35 Hz frequency range, as shown in Figure 13. The hydrocarbon-bearing case consistently exhibits stronger spectral amplitudes than the water-saturated case within the 30–35 Hz frequency range. The frequency-dependent AVO analysis in Figure 14 further supports the sensitivity of the 30–35 Hz components to pore-fluid changes. The 30–35 Hz components exhibit the largest negative intercept and gradient differences between the oil-bearing and water-saturated cases, indicating the strongest fluid-related AVO response within this frequency range. In contrast, frequencies outside the 30–35 Hz range exhibit weaker separation and reduced fluid discrimination. In addition, the finite-band far-stack synthetic responses in Supplementary Fig. S1 also preserve a clear hydrocarbon–wet separation within the Hutton interval for the 30–35 Hz band. The results suggest that the pore-fluid effect is concentrated within a limited spectral band rather than across the full broadband response.

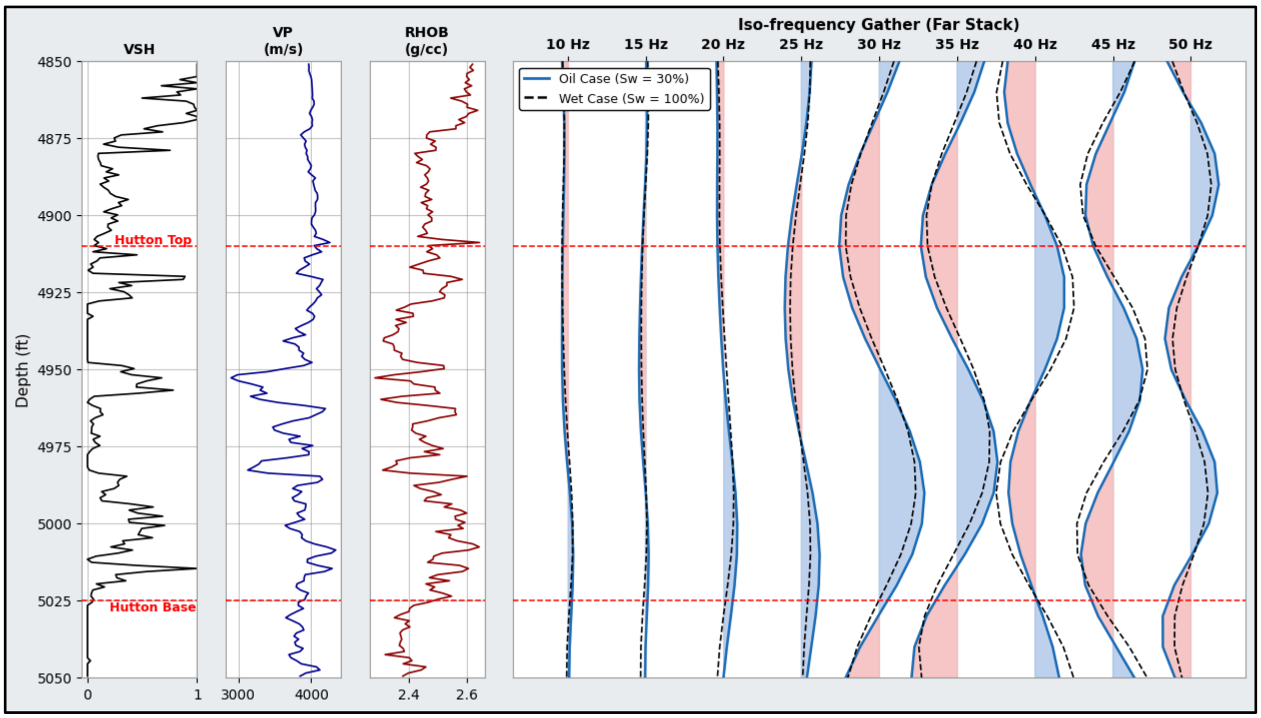

Figure 12: Volume of shale (VSH), P-wave velocity (VP), and bulk density (RHOB) logs together with CLSSA-derived iso-frequency responses of the far-stack synthetic data at the Pyxis-2 well for the oil case (solid blue) and the fluid-substituted water-saturated case (dashed black) within the Hutton reservoir interval. At the reservoir conditions, the oil has an API gravity of approximately 44°, a pore pressure of 19.7 MPa, a temperature of 26.7 °C, a gas gravity of 0.7, a gas-to-oil ratio (GOR) of 200 ft^3 bbl^{-1}, and a formation-water salinity of 3,400 ppm.

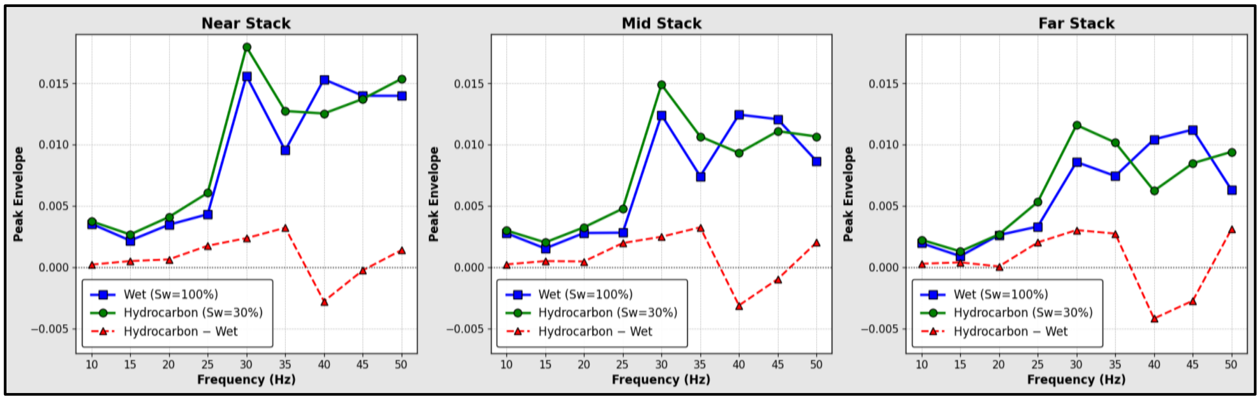

Figure 13: Comparison of CLSSA spectral envelope amplitudes between the water-saturated in-situ case (100% Sw) and the fluid-substituted oil case (30% Sw) for the near, mid, and far stacks at Pyxis-2. The left, middle, and right panels correspond to the near, mid, and far stacks, respectively. The red dashed curves represent the difference between the oil and wet cases.

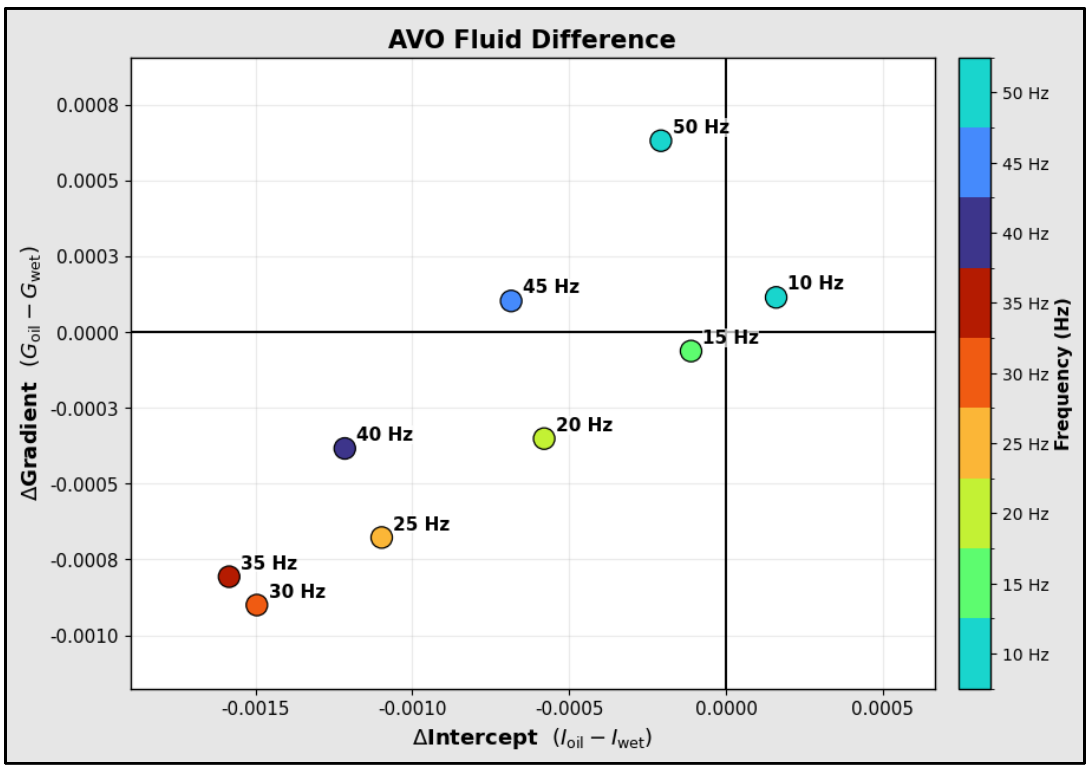

Figure 14. Synthetic frequency-dependent AVO fluid-difference crossplot at the Pyxis-2 well, comparing the oil case (30% Sw) and water-saturated case (100% Sw) within a window at the Hutton Top. The crossplot displays the differences in AVO intercept (ΔIntercept = $I_{oil} - I_{wet}$) and gradient (ΔGradient = $G_{oil} - G_{wet}$) for each CLSSA-derived frequency component. The 30 Hz and 35 Hz attributes exhibit the largest differences.

Sensitivity analysis was further performed to evaluate the contribution of porosity-related variability. Synthetic models at the Pyxis-2 calibration well were generated by systematically varying porosity over realistic ranges (5–20%) under fully water-saturated conditions to isolate non-fluid effects. The resulting porosity perturbation responses produced measurable waveform variations across multiple frequencies and angle stacks (Figure 15). The porosity perturbation responses exhibit a systematic decrease in variability with increasing offset angle. The near-stack responses show the largest waveform variations, followed by the mid stack, whereas the far stack displays the weakest response. This behavior indicates that the far stack is less sensitive to porosity variations.

Sensitivity analysis was also applied to sand thickness in Figure 16. Thickness was systematically perturbed by 20% and 40% of the original sand thickness under fully water-saturated conditions. The perturbed sand interval was embedded within a larger reservoir analysis window to account for variations in the sand/shale ratio. This approach allows the modeled responses to include the influence of surrounding shale layers and capture thickness-related seismic variability associated with changes in sand/shale proportions. The resulting thickness perturbation responses also produced measurable waveform variations across multiple frequencies and angle stacks (Figure 17). The response exhibits frequency-dependent behavior. At lower frequencies, such as 25 Hz, larger thickness values generally produce stronger amplitude responses, whereas at higher frequencies, such as 40 Hz, stronger amplitudes are associated with smaller thickness values. This behavior indicates that thickness effects vary across the spectral domain and are likely influenced by tuning-related effects. In contrast, the far-stack 35 Hz component exhibits relatively weak sensitivity to thickness changes, implying a reduced contribution from thickness-related variability.

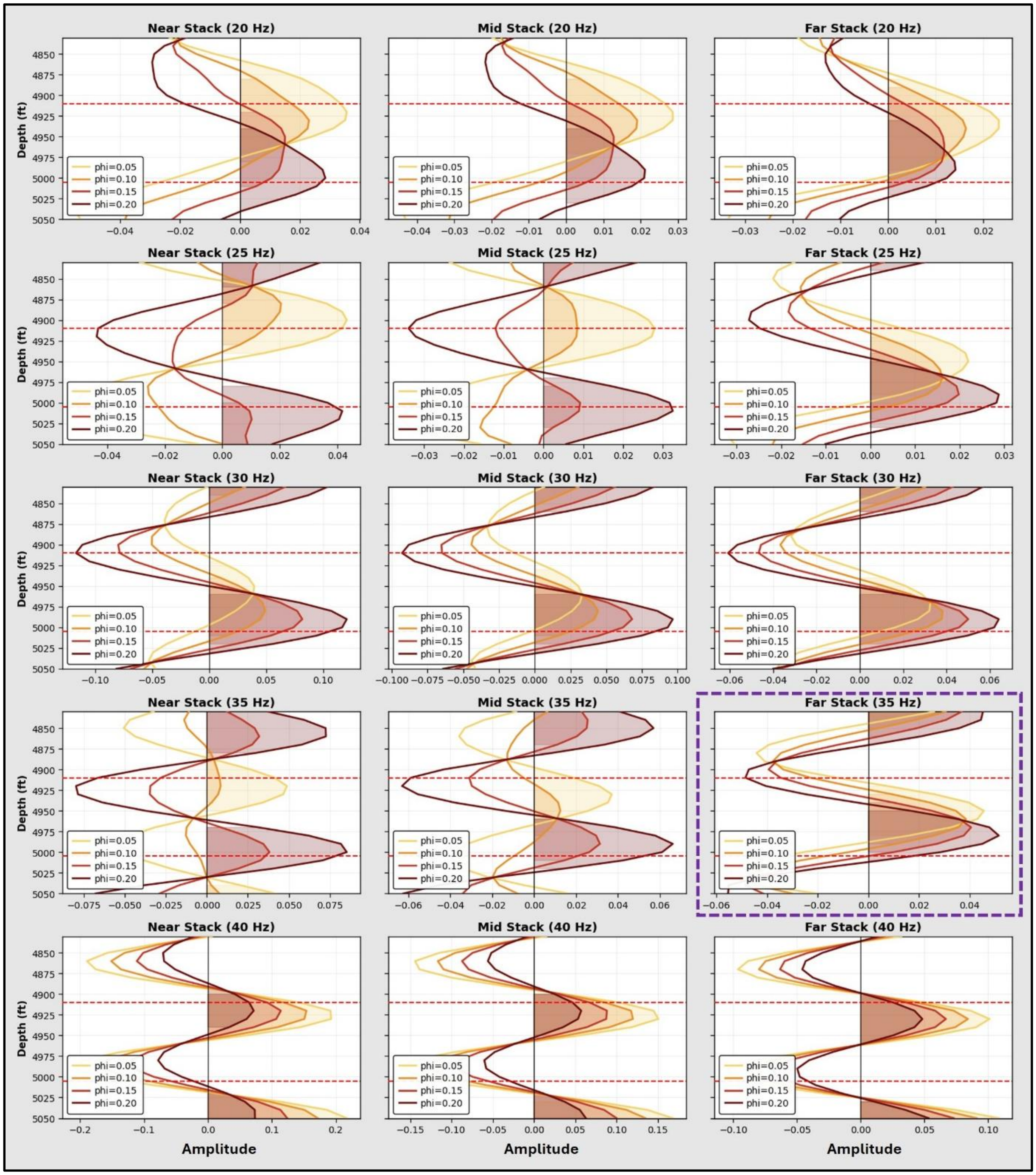

Figure 15: Porosity-perturbation responses for near-, mid-, and far-stack CLSSA-derived iso-frequency components generated from synthetic models at the Pyxis-2 calibration well. Porosity was systematically varied from 5% to 20% under fully water-saturated conditions to isolate non-fluid effects. The resulting waveform responses are displayed for frequencies from 20–40 Hz across different angle stacks. Red dashed lines indicate the boundaries of the perturbed reservoir interval. The far-stack 35 Hz component exhibits comparatively limited waveform variation across the tested porosities.

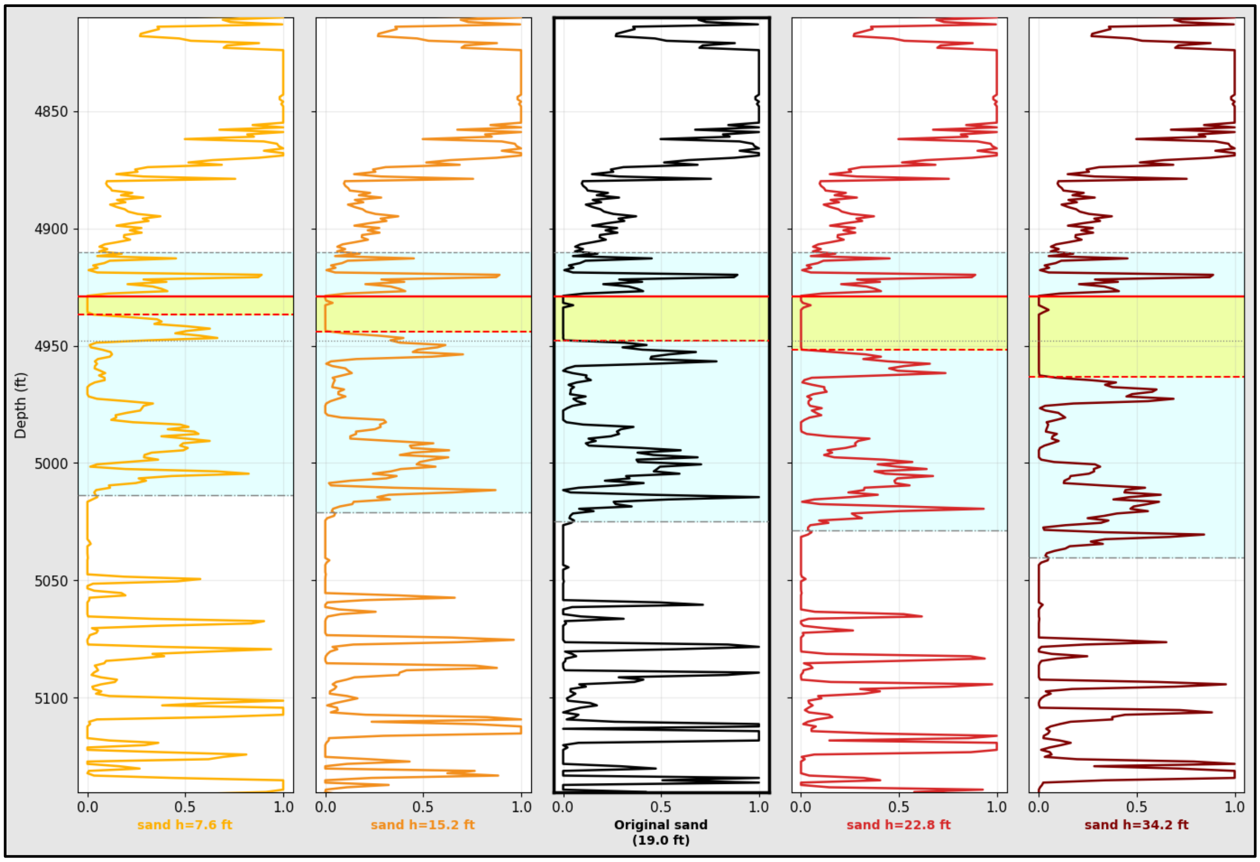

Figure 16: Log-based sand-thickness perturbation models used in the sensitivity analysis at the Pyxis-2 calibration well. The original sand interval (center panel) was modified to generate thinner and thicker sand bodies (7.6 ft, 15.2 ft, 22.8 ft, and 34.2 ft) under fully water-saturated conditions while preserving the surrounding stratigraphic framework and original log characteristics. The perturbed sand intervals are highlighted in yellow, whereas the blue shaded area represents the overall reservoir window used for evaluation.

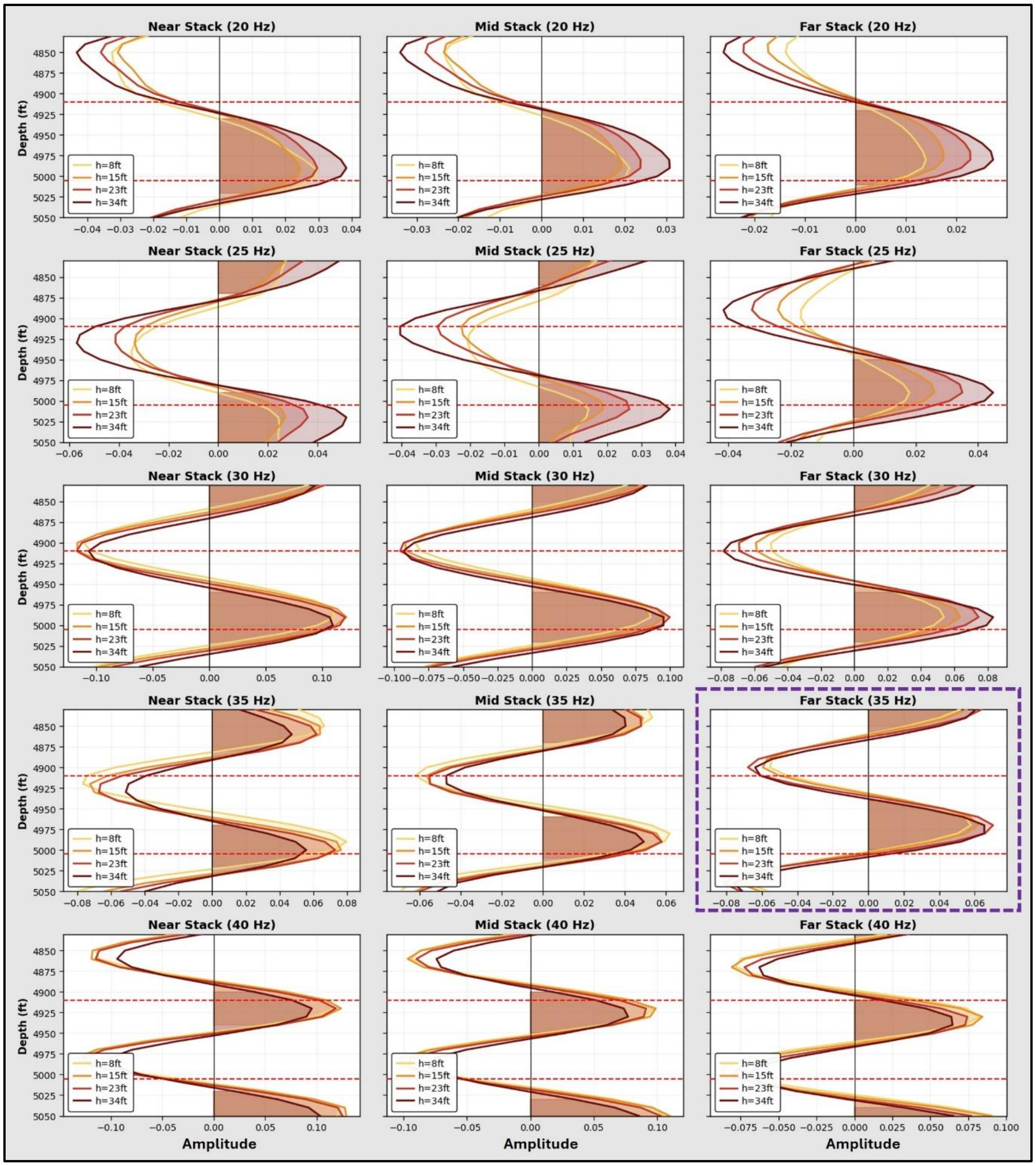

Figure 17: Thickness-perturbation responses for near-, mid-, and far-stack CLSSA-derived iso-frequency components generated from synthetic models at the Pyxis-2 calibration well. Sand thickness was systematically varied (8 ft, 15 ft, 23 ft, and 34 ft) under fully water-saturated conditions to isolate non-fluid effects while preserving the surrounding geological framework. The resulting waveform responses are displayed for frequencies from 20–40 Hz across different angle stacks. The far-stack 35 Hz component exhibits comparatively limited waveform variation across the tested thicknesses.

All attributes generated from the synthetic modeling at the Pyxis-2 calibration well were subsequently ranked using the Fluid Discrimination Ratio (FDR), defined as the ratio between fluid-related variability and competing non-fluid variability. The FDR ranking was used to evaluate the extent to which the seismic response was controlled by pore-fluid effects relative to porosity- and thickness-related variations. As shown in Figure 18, the 30–35 Hz spectral components consistently achieved the highest rankings. The far-stack 35 Hz component exhibits the strongest discrimination capability (FDR = 6.56) and is followed by the far-stack 30 Hz component (FDR = 2.01).

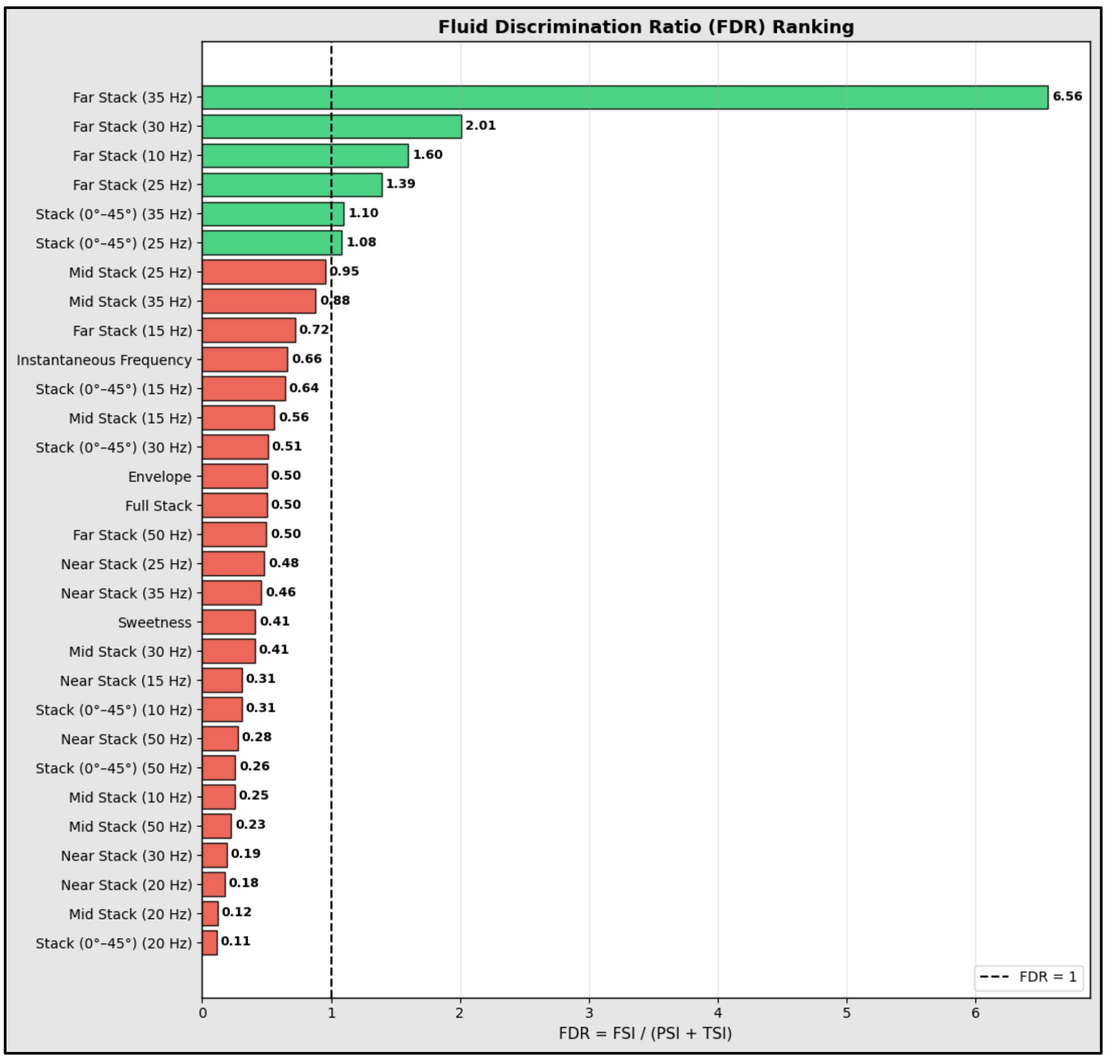

Figure 18: Fluid Discrimination Ratio (FDR) ranking of 30 seismic attributes generated from synthetic modeling at the Pyxis-2 calibration well. values greater than 1 indicate that fluid-related variability dominates over competing porosity- and thickness-related effects.

PGLV was then applied within a horizon-guided analysis window extending 5 ms below the Hutton horizon. Within this interval, local seismic variability was evaluated over spatial windows of 21 inlines × 21 crosslines (approximately 315 m × 315 m). The same spatial window size was used for all wells, while window locations were selected based on the structural and stratigraphic continuity of the target reservoir to reduce mixing with surrounding zones. The selected lateral spatial window was sufficiently large to capture seismic variability associated with the expected lateral extent of potential fluid-transition zones and possible water-contact positions. This approach helps ensure that fluid-related seismic variability is adequately represented within the variance calculation around each well. The seismic attributes were selected only if they exhibited high local variability within the analysis windows surrounding the hydrocarbon-bearing wells, Andromeda South and Pyxis-1, but low local variability within the corresponding window surrounding the water-saturated well, Pyxis-2. The selected attributes were then ranked according to their composite scores.

To evaluate the stability of the spatial window selection using real seismic data, a sensitivity analysis was conducted by varying the window size to 19 × 19 and 17 × 17 traces while keeping all other parameters fixed. The top-ranked attributes remained identical across all tested window sizes. These results indicate that the PGLV attribute-selection outcome is stable and not sensitive to moderate variations in spatial window dimensions providing possible water contact levels are captured in the window. The attribute ranking obtained only from real seismic data shows dominant 30–35 Hz components extracted from mid- and far-angle stacks. The eight highest-ranked attributes and their variances estimated at each well are shown in Table 1.

Table 1: Variance and variance-ratio metrics of candidate seismic attributes computed from real seismic data within spatial windows around the calibration well (Pyxis-2) and two additional wells (Andromedae South and Pyxis-1). The table shows the eight highest-ranked attributes based on the composite PGLV ranking score.

Rank	Selected Attributes	Var_30%	Var_60%	Var_100%	VR_30/100	VR_60/100	VR_30/60	Composite Ranking Score
1	Amplitude_FarStack_30Hz	0.000826838	0.00041227	0.00015608	0.8412105	0.72538662	0.667282999	2.233880122
2	Envelope_MidStack_30Hz	0.000240812	0.00017266	5.54E-05	0.8130565	0.7571838	0.582414459	2.152654763
3	Impedance_MidStack_30Hz	2.33E-05	1.73E-05	8.92E-06	0.72327746	0.6601137	0.57370326	1.957094422
4	Amplitude_FarStack_35Hz	0.000768563	0.00073696	0.00031542	0.70901977	0.70028138	0.510495949	1.919797101
5	AVO_Intercept	0.550706165	0.51954606	0.23826357	0.6980067	0.68558915	0.514557363	1.898153214
6	Amplitude_MidStack_35Hz	0.000193496	0.00016224	9.95E-05	0.66042324	0.61987342	0.543928923	1.824225583
7	Amplitude_MidStack	0.655610406	0.64643647	0.45458968	0.59053356	0.58712182	0.50352289	1.681178265
8	Envelope_FullStack	0.741423295	0.72715456	0.68756656	0.51884434	0.51399145	0.504858013	1.537693803

The 30–35 Hz spectral components extracted from the far-stack were identified as primary indicators because they consistently ranked highly in both the synthetic fluid-discrimination analysis and the local seismic-variability assessment. In contrast, the 30–35 Hz spectral components extracted from the mid-stack were considered secondary indicators. Although they produced stable and geologically consistent responses within the calibration area and ranked among the top attributes in the local seismic variability analysis, they exhibited lower synthetic fluid discrimination and therefore weaker fluid sensitivity than the far-stack spectral components.

Impact of Attribute Selection

Once attributes are selected using PGLV methodology, the selected attributes extracted from real seismic data are used as input for SOM to be compared against attributes selected by PCA. This comparison was performed to evaluate how different attribute-selection strategies and input attributes influence facies discrimination performance.

Self-organizing map (SOM) classification was conducted on standardized attribute vectors using a 2 × 3 grid and trained for 50 epochs with a random seed of 30 to ensure reproducibility. The SOM topology was constrained because the objective of this study focuses on hydrocarbon-bearing versus water-saturated discrimination rather than detailed multi-facies stratigraphic characterization. The resulting SOM nodes were subsequently grouped into three geological facies. To evaluate repeatability and sensitivity to training parameters, the SOM was retrained multiple times using training epochs from (10–200) and by using different random initialization seeds (30, 35, 40, 45, and 50). The resulting classifications remained consistent. To assess the internal structure and stability of the SOM, a U-matrix was computed. It revealed clear node-to-node distance contrasts and internally coherent regions, supporting a well-organized and repeatable SOM topology for this case study.

The facies classification was performed using a fully unsupervised SOM trained solely on seismic data. The wells were used only after the clustering step for facies identification and validation and were not included in SOM training. Two separate SOM classifications were generated. The SOM architecture and training parameters were kept identical, and only the input attributes were changed. The PCA-selected attribute set consisted of the mid- and far-offset stacks, the full stack, and the AVO gradient. The resulting facies map (Figure 19a) shows that the red facies cluster occurs not only at the hydrocarbon-bearing Andromedae South well but also at the Pyxis-2 dry hole, indicating poor discrimination between hydrocarbon-related and water-saturated seismic responses. In contrast, the PGLV-selected attribute set consisted of the 30 Hz and 35 Hz far-stack spectral amplitudes, as well as the 30 Hz mid-stack spectral amplitude, spectral envelope, and impedance attributes. The corresponding SOM classification (Figure 19b) produced a red cluster aligned only with the hydrocarbon wells, no longer appearing at the dry well, indicating improved discrimination of hydrocarbon-related seismic variability.

The three wells (Andromedae South, Pyxis-1, and Pyxis-2) were subsequently used as a post-classification assessment, summarized by the confusion matrices (Figures 19c and 19d). The confusion matrices show that the SOM classification derived from the PGLV-selected attributes is more consistent with the known well outcomes than the classification derived from the PCA-selected attributes. While the small number of wells does not permit a rigorous assessment of classification accuracy, the results indicate that the PGLV-selected attributes provide improved separation between hydrocarbon-bearing and water-saturated facies in this study area. The PGLV-SOM result identifies an undrilled follow-up lead within the structural closure marked by the black star in Figure 19b. This location is classified as red hydrocarbon facies in the PGLV-SOM map but is not similarly highlighted in the PCA-SOM result.

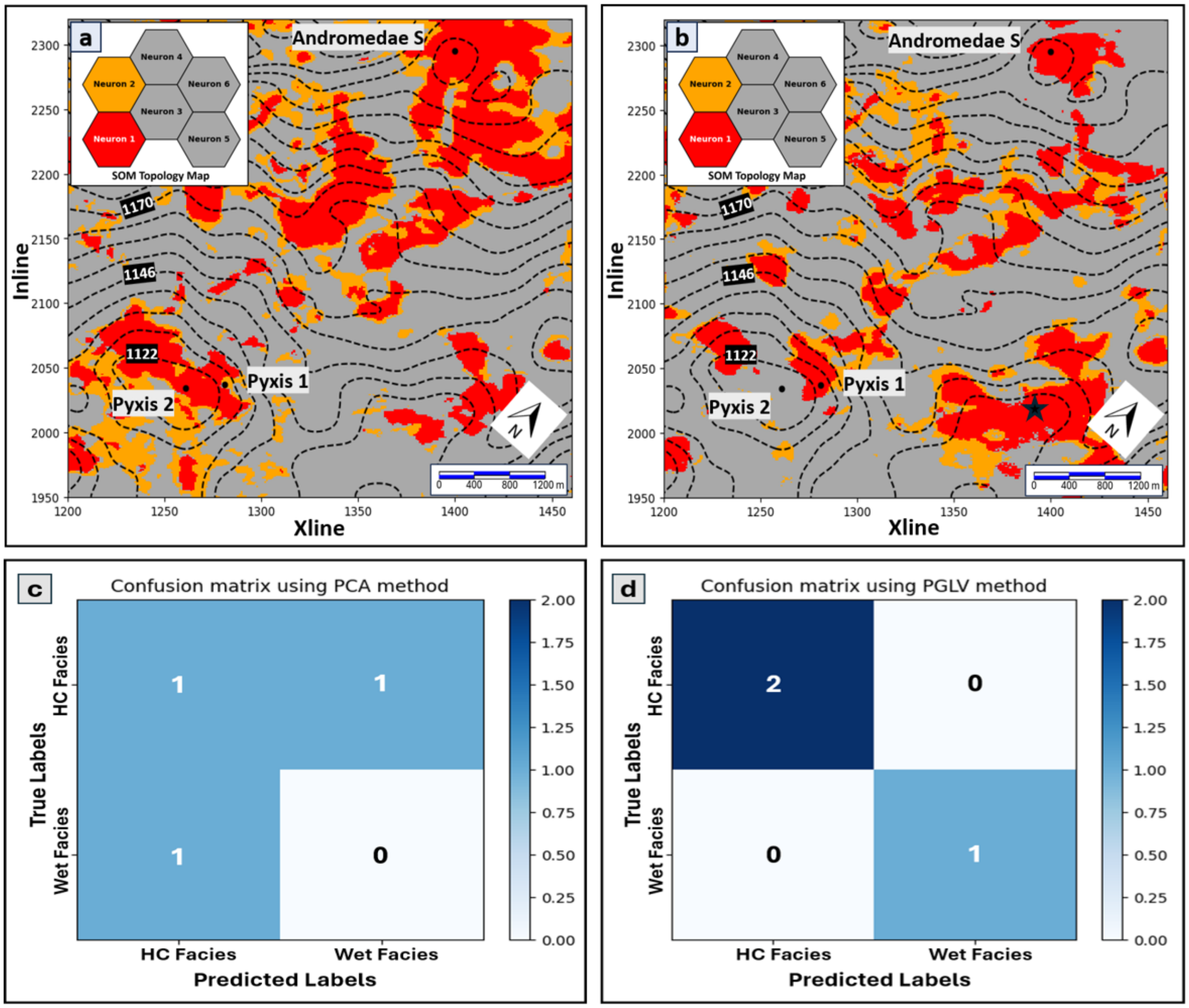

Figure 19: Well-based validation of SOM facies using different attribute-selection approaches. (a) Seismic SOM facies map generated using PCA-selected attributes and (b) seismic SOM facies map generated using PGLV-selected attributes. The structural contour interval is 24 ms. The inset in each panel shows the six neurons of the 2 × 3 SOM topology. Neuron 1, shown in red, represents the hydrocarbon-bearing Facies 1. Neuron 2, shown in orange, represents Facies 2, whereas Neurons 3–6, shown in gray, are grouped as Facies 3. Facies 2 and Facies 3 correspond to non-productive or water-saturated seismic responses. The black star in panel (b) marks an undrilled follow-up lead identified by the PGLV-SOM hydrocarbon facies within a structural closure; this location is not similarly highlighted by the PCA-SOM result in panel (a). (c) Confusion matrix for SOM classification using PCA-selected attributes and (d) confusion matrix for SOM classification using PGLV-selected attributes. Rows represent the true well outcomes (hydrocarbon-bearing and water-saturated facies), whereas columns represent the predicted classes from the SOM classification.

To quantify the relationship between the SOM hydrocarbon-bearing facies (red cluster) and the input attributes, a Jaccard-overlap analysis was performed using a 70th percentile attribute-anomaly threshold (Table 2). The analysis evaluates how much of each individual attribute anomaly is captured by the SOM red facies. This provides a direct measure of whether the SOM result reproduces a single attribute anomaly or represents a more integrated multi-attribute response.

For the PCA-SOM result, the far-stack anomaly has the highest Jaccard overlap of 0.630, with 63.16% of the far-stack anomaly captured by the PCA-SOM red facies. The full-stack attribute also shows strong agreement, with a Jaccard overlap of 0.574 and 59.57% of the full-stack anomaly captured by the SOM red facies. For the PGLV-SOM result, the mid-stack 30 Hz component has the highest Jaccard overlap of 0.481, with 49.74% of the attribute anomaly captured by the SOM red facies. The far-stack 35 Hz and far-stack 30 Hz components show similar levels of agreement, with Jaccard overlaps of 0.471 and 0.445 and anomaly-capture values of 49.00% and 47.11%, respectively.

Table 2. Quantitative overlap between individual attribute anomalies and the SOM hydrocarbon facies using a 70th percentile threshold. SOM HC-facies capture (%) represents the percentage of each attribute-threshold anomaly captured by the SOM hydrocarbon facies.

Method	**Input Attribute**	**Jaccard overlap**	**SOM HC-facies capture (%)**
PCA-SOM	Far Stack	0.630	63.16
	Full Stack	0.574	59.57
	Mid Stack	0.418	48.19
	AVO Gradient	0.523	56.11
PGLV-SOM	Far Stack (35 Hz)	0.471	49.00
	Mid Stack (30 Hz)	0.481	49.74
	Far Stack (30 Hz)	0.445	47.11
	Envelope-Mid Stack (30 Hz)	0.436	46.46
	Impedance-Mid Stack (30 Hz)	0.443	46.96

These observations demonstrate that the SOM classification is strongly influenced by the input attribute space, but the resulting SOM hydrocarbon facies is not simply a thresholded of any single attribute. Instead, the SOM combines the information contained in the input attributes and identifies a multi-attribute facies pattern. As shown in Figure 20, the SOM hydrocarbon facies is not simply a thresholded representation of the input attribute (the far-stack 30 Hz) because the overlap is incomplete and the map contains both SOM-only and attribute-only areas. The SOM-only zones indicate that the SOM integrates information from multiple PGLV-selected attributes and can identify hydrocarbon-facies responses that are not defined by the far-stack 30 Hz anomaly alone. However, both hydrocarbon-bearing wells occur within the overlap areas, and the overlap covers a large proportion of the SOM hydrocarbon facies, demonstrating the strong influence of the PGLV-selected attributes on the classification. Therefore, the primary added value originates from the PGLV attribute selection, whereas the SOM provides additional value by integrating these attributes and refining their combined response.

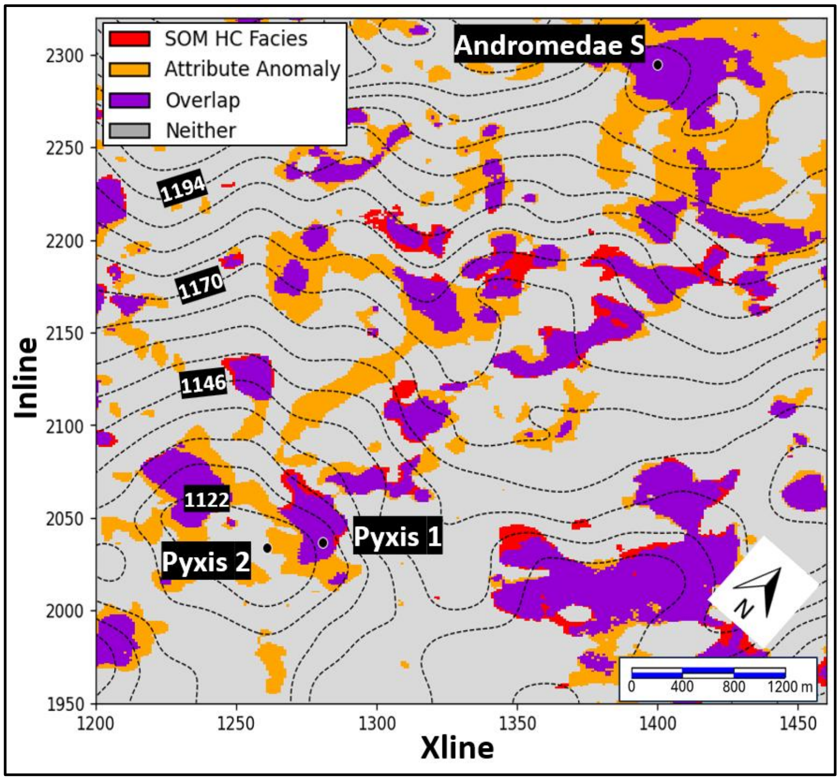

Figure 20. Spatial overlap between the PGLV-SOM hydrocarbon facies and the far-stack 30 Hz amplitude anomaly using a 70th-percentile threshold. Purple areas indicate agreement between the SOM hydrocarbon facies and the attribute anomaly. Red areas represent the SOM hydrocarbon facies only. Orange areas represent the far-stack 30 Hz attribute anomaly only. Gray areas indicate neither response.

Blind Validation

To further assess the robustness of the PGLV method, an additional independent validation was performed using Andromedae-1, which was not used during attribute selection, validation, or workflow development. The 30–35 Hz far-stack spectral components, identified as primary indicators during the PGLV attribute-selection stage, were evaluated at the location of this blind validation well and compared with the full-stack and original far-stack attributes selected by PCA.

As shown in Figure 21, the PGLV-selected attributes provide improved fluid discrimination relative to the PCA-selected attributes. For both the full stack and original far stack, the trough-ratio values are less than 1 (0.681 and 0.811, respectively), indicating that the hydrocarbon-bearing response at Andromedae-1 is weaker than the wet-reference response at Pyxis-2. In contrast, the far-stack 30 Hz and 35 Hz spectral components yield trough-ratio values greater than 1 (1.622 and 1.199, respectively), indicating a stronger hydrocarbon response relative to the wet reference. These results are consistent with the PGLV attribute-selection analysis, which identified the 30–35 Hz far-stack spectral components as being more sensitive to fluid-related variability and less affected by competing geological factors than the conventional stack attributes. Figure 22 further shows that Andromedae-1 is located along the periphery of a distinct amplitude anomaly identified in the far-stack 35 Hz spectral component. This observation is consistent with the estimated 30% water saturation at the well and provides additional support for the hydrocarbon sensitivity of the PGLV-selected attribute.

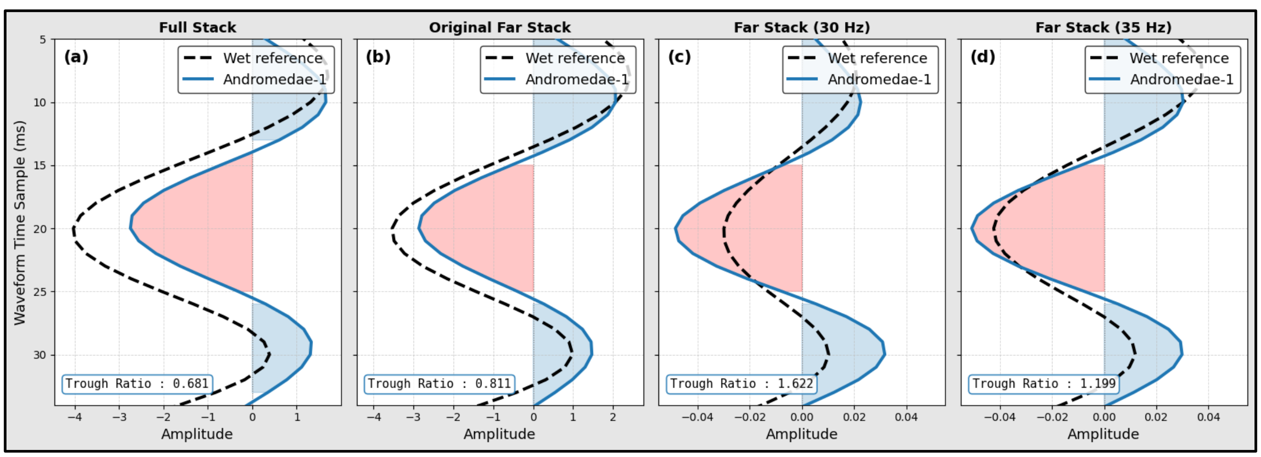

Figure 21: Comparison between the wet-reference waveform extracted at the water-saturated Pyxis-2 well (black dashed line) and the waveform extracted at the hydrocarbon-bearing Andromedae-1 blind validation well (blue solid line) for (a) full stack, (b) original far stack, (c) far stack at 30 Hz, and (d) far stack at 35 Hz. Trough-ratio values correspond to the ratio between the hydrocarbon trough amplitude at Andromedae-1 and the wet-reference trough amplitude at Pyxis-2. Values greater than 1 indicate a stronger hydrocarbon response relative to the wet reference, whereas values less than 1 indicate a weaker response.

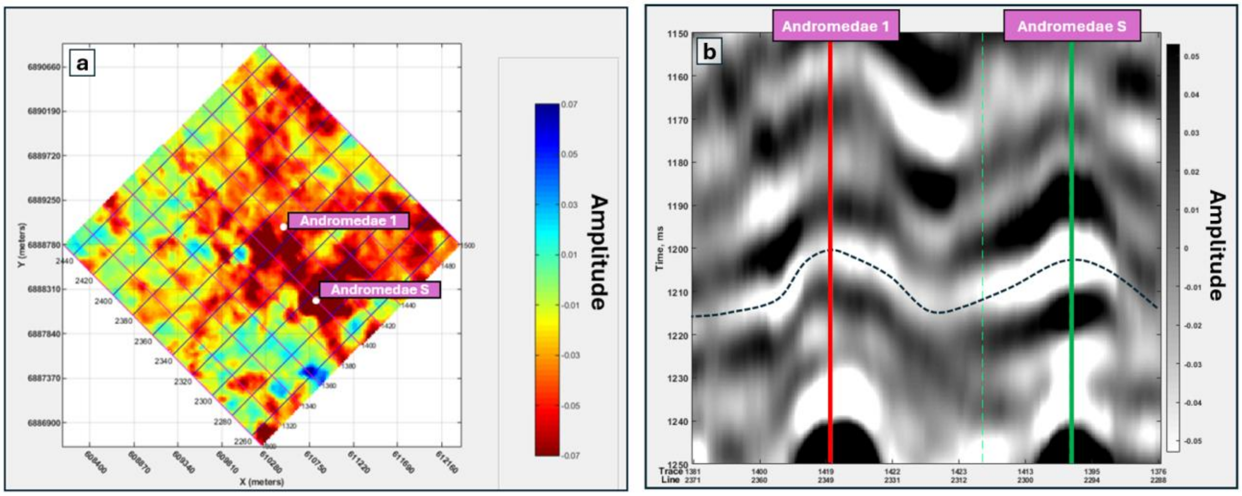

Figure 22: a) Amplitude map around validation well Andromedae-1 extracted from far stack at the spectral component of 35 Hz along the reservoir top. b) Corresponding vertical seismic display showing the same spectral component. Andromedae-1 lies along the periphery of a distinct amplitude anomaly, consistent with its estimated 30% water saturation and supporting the hydrocarbon sensitivity of the PGLV-selected attribute.

Discussion:

Conventional interpretation tools such as analysis of the bright spots of stacked amplitudes and AVO analysis of angle gathers failed to discriminate hydrocarbons from non-productive facies in this case study, especially at the anticlinal closure penetrated by the Pyxis wells. Because bright-spot and AVO anomalies were observed at both hydrocarbon-bearing and dry wells, these approaches did not provide a reliable basis for distinguishing productive from non-productive intervals. This limitation motivated the development of the PGLV methodology to identify seismic attributes specifically sensitive to pore-fluid-related variability while reducing the influence of anomalies caused by porosity, thickness, and other non-fluid geological variations.

Two different attribute-selection strategies were tested. First, seismic attributes were selected using the conventional PCA-based approach. The seismic attributes selected by PCA include the mid- and far-offset stacks, the full stack and the AVO gradient. These attributes exhibit relatively high variance at the dry hole, as these responses are influenced by competing geological factors, such as porosity variations, and thickness effects rather than only pore-fluid effects. Because both the oil and the dry wells showed anomalous amplitudes, those anomalies were carried into the SOM classification (Figure 19a). The red SOM cluster appeared not only at the Andromedae South well, which produces oil, but also at the Pyxis-2 dry hole.

On the other hand, the PGLV method selects attributes based on physically meaningful variability associated with pore-fluid effects rather than global variance alone. Spectral components in the 30–35 Hz range are identified as the most diagnostic attributes for these data. This is consistent with previous studies showing that hydrocarbon effects are often most pronounced in certain frequency bands [3, 4]. Re-running the SOM with these attributes, the hydrocarbon wells have the same SOM cluster while the dry well clusters differently (Figure 19b). These results demonstrate that the effectiveness of SOM classification for hydrocarbon detection significantly depends on the selection of input attributes.

The difference between PCA- and PGLV-selected attributes is also evident in the seismic sections. Figure 23a shows the full stack attribute selected by PCA, where anomalous amplitude response is observed at the dry hole and weak amplitude at the pay zone The opposite is shown in the spectral component of 30 Hz obtained with the Constrained Least-Squares Spectral Analysis (CLSSA) extracted from the mid-angle stack (Figure 23b). The amplitude dims at the dry hole well and strengthens at the pay zone.

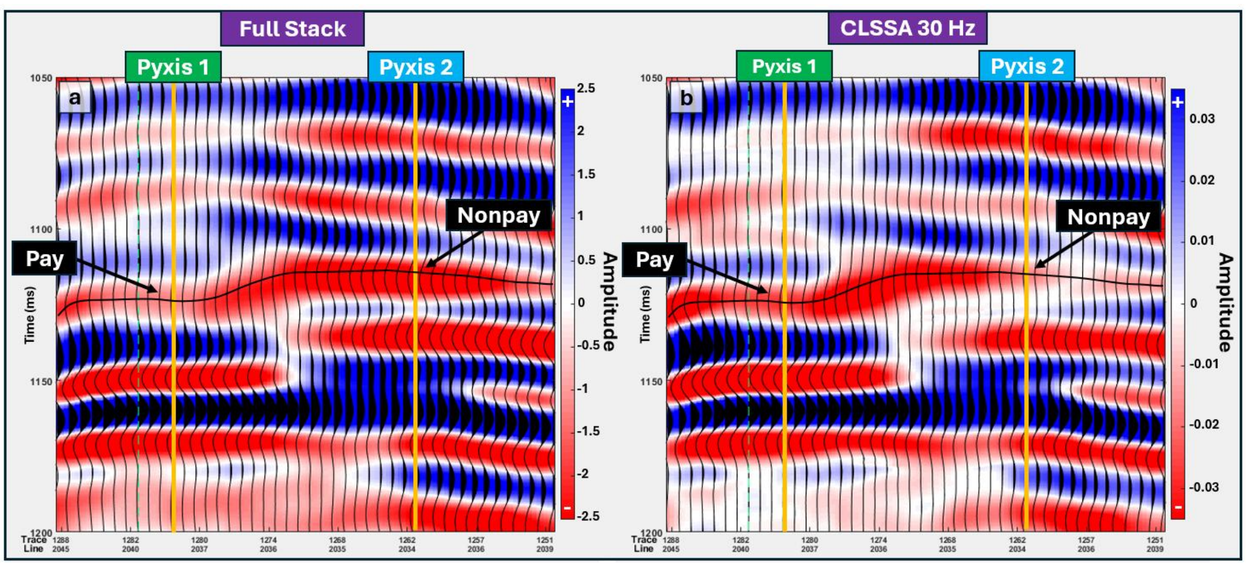

Figure 23: a) A seismic attribute selected by PCA. b) A seismic attribute selected by PGLV.

We applied conventional physics-based seismic inversion to the PGLV-selected attributes and compared the results with inversion derived from the PCA-selected full-stack attribute. The inversion based on the PGLV-selected 30 Hz spectral component exhibits a more geologically consistent distribution of low acoustic impedance than the full-stack inversion (Figure 24). In the full-stack inversion (Figure 24a), the anomalously low acoustic impedance covers the entire closure and appears downdip at the Andromedae South well. In the updated inversion using the 30 Hz mid-stack attribute selected by PGLV (Figure 24b), however, the low impedance is restricted to the top of the closure. At the Pyxis closure, the full-stack inversion shows low impedance within the water-saturated zone and relatively higher impedance within the hydrocarbon-bearing zone, which is inconsistent with the expected fluid-related impedance response. By comparison, the inversion derived from the 30 Hz mid-stack attribute shows low impedance within the oil zone and relatively high impedance within the water-saturated reservoir interval. Similarly, the initial AVO product using original gathers shows anomalous response at the dry hole and weak response at the oil zone as shown in Figure 25. The updated AVO product at 35 Hz, however, shows a dimming at the dry hole and strengthening of the AVO product at the pay zone of 60% water saturation. The response becomes anomalous updip of Pyxis-1 implying decreased water saturation. Together, the updated seismic inversion and AVO results confirm the robustness of the PGLV approach in reducing exploration risk. PGLV is seen to be useful, not only for machine learning methods such as SOM, but also for conventional geophysical quantitative analysis.

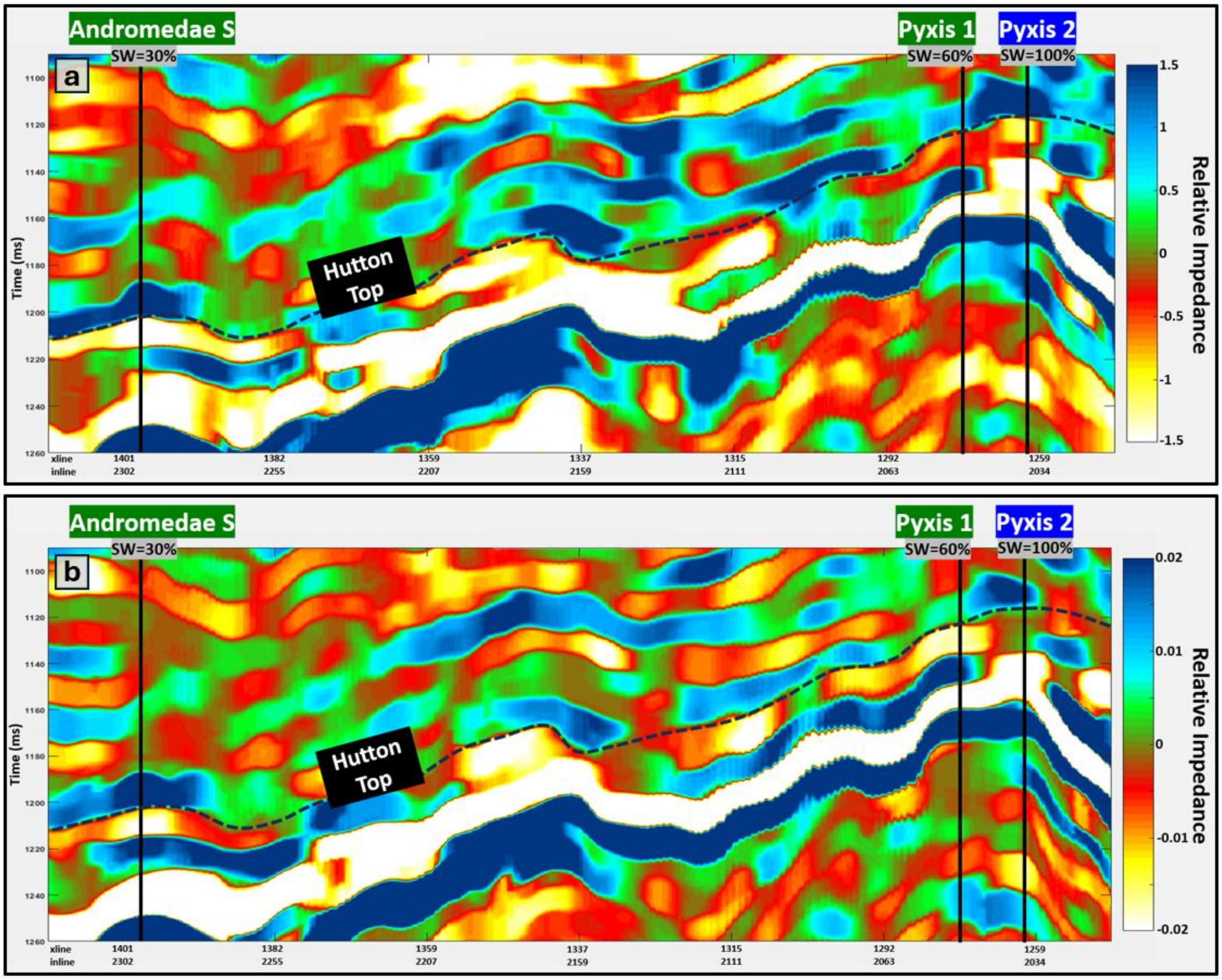

Figure 24: Band-limited seismic inversion of (a) the original full-stack seismic data and (b) the 30 Hz isofrequency component extracted from the mid-angle stack.

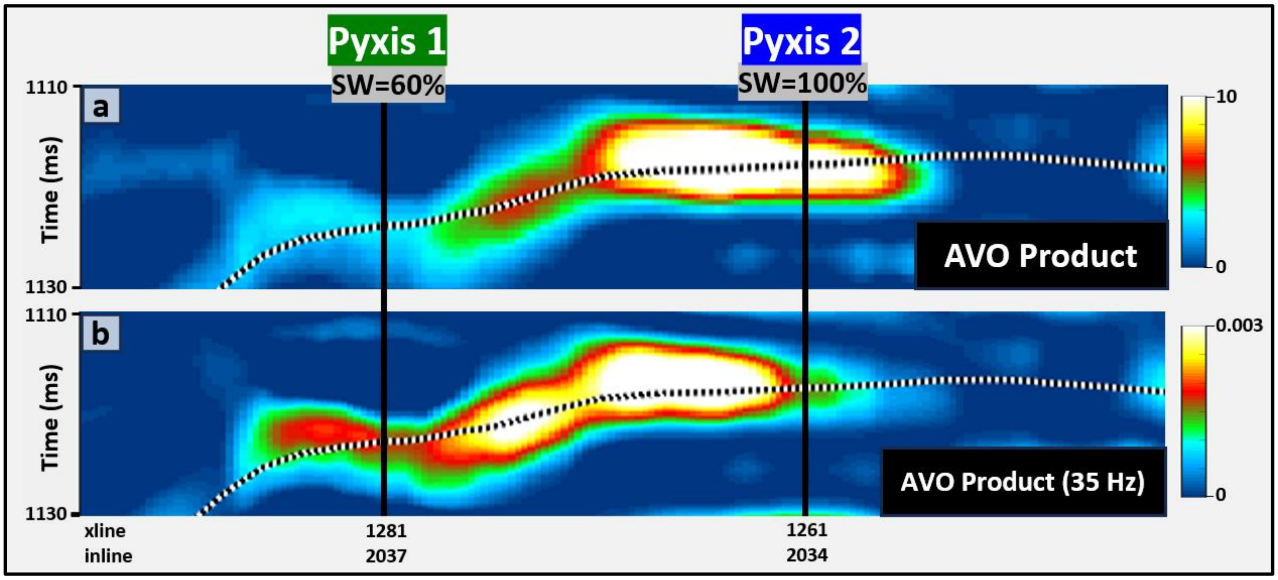

Figure 25: a) AVO product estimated from original angle gathers. b) AVO product estimated from angle gathers at the spectral component of 35 Hz.

The apparent contradiction between the full-band dry-hole AVO product anomaly and the PGLV-selected 30–35 Hz spectral attributes is explained by frequency-dependent tuning and stratigraphic interference. In a thin or near-tuning reservoir, the recorded amplitude is not the response of a single interface, but the composite interference of the top and base reservoir reflections and nearby stratigraphic events. This interference varies with frequency because it depends on bed thickness, wavelet bandwidth, and phase. Therefore, some frequency components may enhance the reservoir response, whereas others may suppress or mask it [33, 34, 35, 36].

The frequency-gather comparison (Figure 26) shows that the 30–35 Hz far-stack response is the most diagnostic frequency range for separating the hydrocarbon-bearing Pyxis-1 well from the water-saturated Pyxis-2 well. At lower frequencies, Pyxis-2 exhibits stronger amplitudes than Pyxis-1, which is opposite to the expected hydrocarbon response. Similarly, at higher frequencies around 40–50 Hz, the water-saturated Pyxis-2 response is also stronger than Pyxis-1. Therefore, these frequency ranges are not interpreted as fluid-sensitive in this case. In contrast, the 30–35 Hz components show enhanced negative spectral amplitude at Pyxis-1 and a comparatively dim response at Pyxis-2. This supports the use of the 30–35 Hz far-stack components as fluid-sensitive attributes. Moreover, the synthetic fluid-substitution results show that the largest oil–wet intercept and gradient differences occur near 30–35 Hz, implying that this band maximizes the modeled hydrocarbon/water-saturated contrast for the Hutton interval.

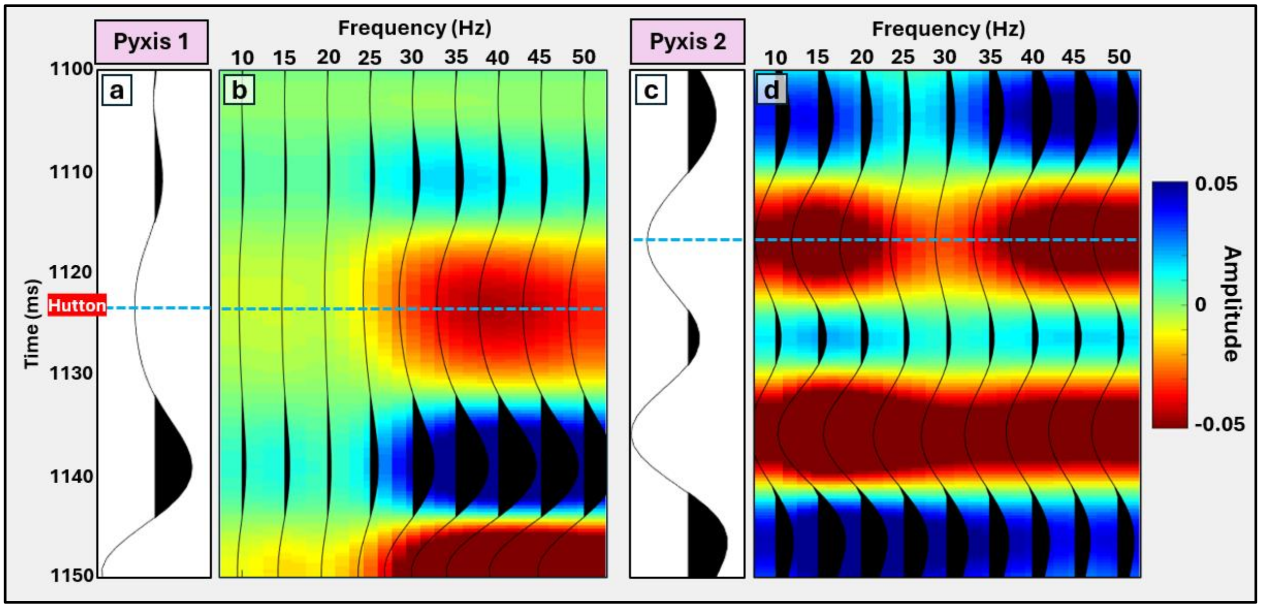

Figure 26: Far-stack seismic traces and corresponding iso-frequency gathers at the locations of the Pyxis wells. (a) Far-stack seismic trace extracted at the hydrocarbon-bearing Pyxis-1 well location and (b) corresponding CLSSA iso-frequency gather; (c) far-stack seismic trace extracted at the water-saturated Pyxis-2 well location and (d) corresponding CLSSA iso-frequency gather. The iso-frequency gathers display spectral amplitude as a function of time and frequency from 10 to 50 Hz, where spectral amplitude = spectral magnitude × cosine of phase. Each black wiggle represents the spectral-amplitude trace of an individual frequency component extracted from the far-stack seismic data at the well location, while the color background displays the corresponding spectral amplitude. At the 30–35 Hz components, Pyxis-1 exhibits enhanced negative spectral amplitude at the Hutton interval, whereas the equivalent response at the water-saturated Pyxis-2 well is comparatively dimmer and less diagnostic. In contrast, at lower and higher frequencies, the water-saturated Pyxis-2 response is stronger than Pyxis-1. The dry-hole anomaly exhibits a broader spectral response, with strong amplitudes occurring both below and above the 30–35 Hz band.

Because variability in seismic attributes can arise from multiple sources, including fluid effects, porosity variations, thickness variations, and noise, high variability alone does not necessarily indicate fluid sensitivity. Therefore, sensitivity indices and Fluid Discrimination Ratios (FDR) were computed to quantitatively evaluate the relative contributions of fluid and competing geological effects. Attributes with Fluid Discrimination Ratios (FDR) greater than 1 are interpreted as being predominantly controlled by fluid effects, whereas values less than 1 indicate dominance of competing non-fluid contributions. Figure 27 shows that the PGLV-selected far-stack spectral components at 30–35 Hz consistently yield FDR values greater than 1, with the 35 Hz component producing the highest value (FDR = 6.56) followed by the 30 Hz component (FDR = 2.01). These results indicate that fluid-related variability dominates over competing geological effects at these attributes. In contrast, the PCA-selected full-stack attribute exhibits an FDR below unity (FDR =

0.50), indicating that its variability is primarily controlled by porosity and thickness effects rather than fluid effects.

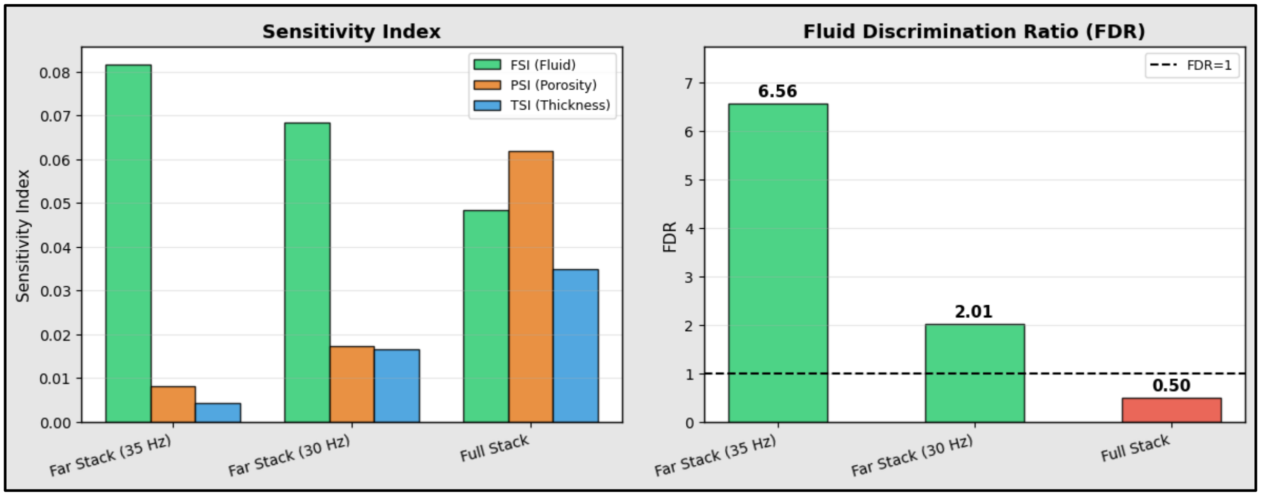

Figure 27: Comparison of fluid-related and competing geological sensitivities for representative seismic attributes. The left panel shows the Fluid Sensitivity Index (FSI), Porosity Sensitivity Index (PSI), and Thickness Sensitivity Index (TSI) computed from synthetic perturbation analysis. The right panel shows the Fluid Discrimination Ratio (FDR), defined as the ratio between fluid-related variability and background variability associated with porosity and thickness effects. The dashed line represents FDR = 1, which separates attributes dominated by fluid effects from those dominated by competing geological effects.

The influence of competing geological factors on the 35 Hz seismic response is illustrated in Figure 28. The black arrows represent the spread of responses caused by the perturbations, where longer arrows indicate larger variability and shorter arrows indicate smaller variability. For porosity perturbations at 35 Hz, the near-stack response exhibits a larger spread near the top of the reservoir, indicating greater sensitivity to porosity variations, whereas the far-stack response shows a narrower spread and is less sensitive to porosity changes. Similarly, for thickness perturbations at 35 Hz, the far-stack response displays a smaller spread compared with the near-stack response, indicating reduced sensitivity to thickness variations. These observations suggest that the far-stack at 35 Hz is less sensitive to porosity and thickness variations. Therefore, fluid-related effects become more dominant within this PGLV-selected attribute.

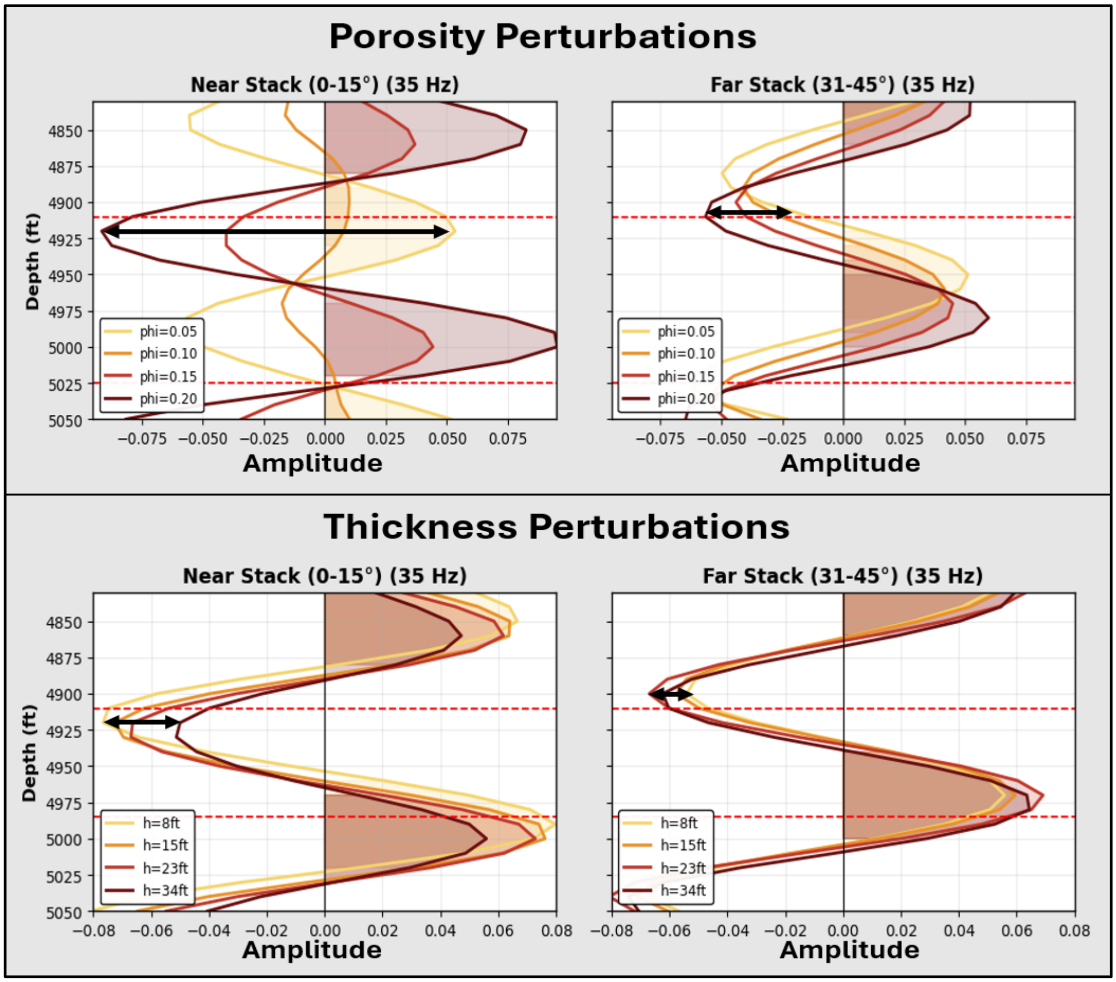

Figure 28: Synthetic seismic responses to porosity and thickness perturbations at 35 Hz for near-stack (0–15°) and far-stack (31–45°) data. Top panels show porosity perturbations for porosity values ranging from 5% to 20%, while bottom panels show thickness perturbations for reservoir thicknesses ranging from 8 ft to 34 ft. Red dashed lines indicate the reservoir interval boundaries. The black arrows represent the spread of responses caused by the perturbations.

Consistency between synthetic modeling and real seismic observations provides additional confidence that the selected attributes are responding to physically meaningful reservoir variations rather than noise-related artifacts or spurious correlation. When the seismic data are too noisy for reliable analysis, inconsistent attribute rankings between synthetic predictions and real data are expected, which increases the likelihood of incorrect predictions at additional wells. Although perfect agreement is not expected, confidence in the selected attributes is strengthened when the same attributes are consistently identified across different wells and in both synthetic and real data. In this study, the repeated appearance of the 30–35 Hz far-stack spectral components in both synthetic modeling and spatial analysis provides this consistency. Thus, the 30–35 Hz spectral components extracted from the far-stack were identified as primary indicators. The corresponding 30–35 Hz mid-stack spectral components were considered secondary indicators because, although they showed stable and geologically consistent behavior and ranked highly in the local seismic variability analysis, they exhibited weaker fluid sensitivity in the synthetic modeling results.

To further evaluate the pore-fluid sensitivity of the PGLV-selected attributes, a waveform-based comparison was performed between the hydrocarbon-bearing Andromedae-1 blind validation well and the wet-reference Pyxis-2 well (Figures 21). The Andromedae-1 well was not used during workflow development, attribute selection, or validation. The trough ratio is defined as the ratio between the trough amplitude extracted at the hydrocarbon-bearing Andromedae-1 well (30% Sw) and the corresponding trough amplitude extracted from the wet-reference waveform at the Pyxis-2 well (100% Sw). Trough-ratio values less than 1 indicate that the hydrocarbon-bearing response is weaker than the wet-reference response. This reflects the presence of false bright-spot behavior at the water-saturated Pyxis-2 well as observed in the broadband full-stack and original far-stack. In contrast, the far-stack spectral components at 30–35 Hz produced higher trough ratios that exceed 1, indicating suppression of the false bright-spot response. These results confirm the effectiveness of the PGLV method in selecting seismic attributes with enhanced pore-fluid sensitivity and improved discrimination between hydrocarbon-bearing and water-saturated intervals.

More generally, synthetic modeling at the wells is essential for defining the expected variance behavior. In this case study, the PGLV rule highlighting that variance increases with decreasing water saturation, is only applicable where synthetic models at the wells show systematic variance increase with presence of commercial hydrocarbons. However, the criteria used in the PGLV approach can be modified for other reservoirs and geological settings.

The PGLV method is designed to be applicable under realistic exploration conditions where well control is limited. The attributes ranking can be established using synthetic modeling at a single calibration well, while additional wells are used for validation and for increasing confidence in the results. Without validation wells, however, there will be greater uncertainty in the predictions. With no prior knowledge of rock and fluid properties, there would be less confidence in predictions, although they could still be interpreted as possible direct hydrocarbon indicators, confidence of which could be increased using standard hydrocarbon exploration principles (such as conformity to structure etc.).

The PGLV method has limitations in geological settings where fluid-related seismic responses are weak relative to lithologic, thickness, or structural effects, and it would be less effective in high-impedance reservoirs. Therefore, the use of variance as an attribute selection criterion in this study is only appropriate when hydrocarbon effects have a stronger influence on seismic response than competing factors. Where rock physics modeling indicates that hydrocarbon effects are smaller than other variable factors, reliable hydrocarbon detection may not be possible. In such cases, the approach can be adapted to target other reservoir properties, such as porosity, by modifying the PGLV selection rule to reflect porosity-driven trends, which may correspond to either increased or decreased local seismic variability. Similarly, the Fluid Discrimination Ratio (FDR) can be reformulated as a porosity discrimination ratio by comparing porosity-related variability against other competing sources of geological variability.

More broadly, the PGLV methodology integrates physics-guided synthetic modeling with local seismic variability analysis to identify attributes sensitive to target reservoir properties and is potentially applicable to a wide range of reservoir-characterization objectives, including hydrocarbon saturation, porosity, mineral volumetrics, net-to-gross ratio, thickness, shale content, pore pressure, fracture density, brittleness and organic richness in unconventional shale reservoirs, and time-lapse reservoir monitoring.

Another potential application of the PGLV methodology is frequency selection for RGB spectral blending. RGB spectral blending is commonly applied through interpreter-driven selection of frequencies that visually enhance geological features such as channels. However, the selected frequencies do not necessarily correspond to hydrocarbon-sensitive responses, as similar channel geometries may also be associated with shale-filled or water-saturated systems. The PGLV methodology provides a physics-guided framework for identifying frequency components that are specifically sensitive to hydrocarbon-related variability. Thereby, the PGLV method can be used for frequency selection for RGB blending and reduces the risk of emphasizing non-fluid-related geological features.

Table 3 summarizes the key differences between the Principal Component Analysis (PCA) and the Physics-Guided Local Variability (PGLV) methods in terms of attribute-selection strategy, use of well control, selection of spectral components, data requirements, applicability conditions, and computational cost and scalability.

Table 3: A Comparison between PCA and PGLV methods.

Criteria	Principal Component Analysis (PCA)	Physics-Guided Local Variability (PGLV)
Basis of selection	Global variance	Local variance
Approach type	Purely statistical	Physics-guided
Use of well control	Not considered in attribute ranking	Attributes evaluated locally around wells with known fluid saturations
Spectral components	Sometimes overlooked [6]	Selected spectral components with high pore-fluid sensitivity
Selected attributes	Does not distinguish between high variance related to fluid or lithology effect	Only selected attributes that have pore-fluid sensitivity
Data requirements	Seismic attribute volumes only	Seismic data, well control, and rock physics modelling
Applicability conditions	Best suited for problems dominated by global variance patterns	Best suited for problems dominated by local variance associated with pore-fluid effects
Computational cost and scalability	Low (One-time global variance calculation; readily scalable to larger datasets)	Moderate (Local variance evaluation and synthetic perturbation tests; cost increases with perturbation scenarios)

Conclusion:

The proposed Physics-Guided Local Variability (PGLV) methodology integrated synthetic sensitivity analysis with local seismic variability measurements to identify attributes most sensitive to pore-fluid effects.

In the Hutton Sandstone case study, spectral components in the 30–35 Hz range were identified as the most diagnostic attributes for pore-fluid detection. The far-stack 30–35 Hz components served as the primary indicators, whereas the mid-stack 30–35 Hz components served as secondary indicators.

The selected PGLV attributes improved not only SOM-based facies classification but also conventional geophysical interpretation workflows, including seismic inversion and AVO analysis. PGLV attributes resulted in improved discrimination between hydrocarbon-bearing and water-saturated intervals.

Blind validation at the independent Andromedae-1 well showed that the PGLV-selected attributes successfully predicted hydrocarbon-related seismic responses at a new step-out prospect location that was not used during workflow development, attribute selection, or validation.

The optimal frequency identified in this study (30–35 Hz) is specific to the reservoir properties, thickness, and geological conditions of the Hutton Sandstone. Other reservoirs may exhibit maximum sensitivity at different frequencies. Therefore, synthetic sensitivity analysis remains a critical step of the PGLV workflow.

Because seismic-attribute variability can arise from multiple factors. Sensitivity analysis is essential for determining the dominant controls on each attribute. Consequently, the PGLV workflow can be adapted to identify attributes sensitive to a wide range of reservoir properties beyond hydrocarbon saturation.

This study confirms the idea “good attributes in, good results out”. Whatever the algorithm used, the outcome depends on the quality of the input data. Therefore, selecting meaningful input attributes that are most relevant to the problem at hand, such as hydrocarbon detection, is a critical step for obtaining reliable interpretation results regardless of the analytical method employed.

References:

1. Chopra, S. & Marfurt, K. J. *Seismic attributes for prospect identification and reservoir characterization.* (Soc. Expl. Geophys., 2007). https://doi.org/10.1190/1.9781560801900
2. *Marfurt, K. J. Seismic attributes as the framework for data integration. (Soc. Expl. Geophys., 2018). https://doi.org/10.1190/1.9781560803522*
3. Deng, J.-X., Han, D.-H., Liu, J. & Yao, Q. Application of spectral decomposition to detect deep-water gas reservoir. *SEG Tech. Program Exp. Abstr.* **26**, 2502–2506 (2007). https://doi.org/10.1190/1.2792766
4. Fahmy, W. A., Matteucci, G., Butters, D., Zhang, J. & Castagna, J. Successful application of spectral decomposition technology toward drilling of a key offshore development well. *SEG Tech. Program Exp. Abstr.* **27**, 344–348 (2008). https://doi.org/10.1190/1.2144316
5. Jolliffe, I. T. *Principal component analysis.* 2nd edn. (Springer, 2002). https://doi.org/10.1007/b98835
6. La Marca, K. & Bedle, H. User vs. machine-based seismic attribute selection for unsupervised machine-learning techniques: Does human insight provide better results than statistically chosen attributes? In *Advances in subsurface data analytics*, 3–30 (Elsevier, 2022). https://doi.org/10.1016/B978-0-12-822295-9.00002-9
7. Taner, M. T., Koehler, F. & Sheriff, R. E. Complex seismic trace analysis. *Geophysics* **44**, 1041–1063 (1979). https://doi.org/10.1190/1.1440994
8. Ostrander, W. J. Plane-wave reflection coefficients for gas sands at non-normal angles of incidence. *Geophysics* **49**, 1637–1648 (1984). https://doi.org/10.1190/1.1441571
9. Aki, K. & Richards, P. G. *Quantitative seismology.* (W. H. Freeman & Co., 1980).
10. Partyka, G., Gridley, J. & Lopez, J. Interpretational applications of spectral decomposition in reservoir characterization. *Lead. Edge* **18**, 353–360 (1999). https://doi.org/10.1190/1.1438295
11. Radke, B. M. Hydrocarbon and geothermal prospectivity of sedimentary basins in central Australia: Warburton, Cooper, Pedirka, Galilee, Simpson and Eromanga basins. *Geosci. Aust. Rec.* 2009/25 (2009). Available at: https://www.ga.gov.au/bigobj/GA15314.pdf (accessed 11 Nov 2025).
12. Dodman, A. P. & Rodrigues, J. T. The Jackson oil field development. *APPEA Journal* **29**, 279–289 (1989).
13. Hamilton, D. S., Holtz, M. H., Ryles, P. G. & Hillyer, M. Approaches to identifying reservoir heterogeneity and reserve growth opportunities in a continental-scale bed-load fluvial system: Hutton Sandstone, Jackson Field, Australia. *AAPG Bulletin* **82**, 2192–2219 (1998).
14. Lanzilli, E. The Birkhead Formation: reservoir characterisation of the Gidgealpa South Dome and sequence stratigraphy of the Eromanga Basin, Australia. PhD thesis, Univ. of Adelaide (1999).
15. Archie, G. E. The electrical resistivity log as an aid in determining some reservoir characteristics. *Trans. AIME* **146**, 54–62 (1942). https://doi.org/10.2118/942054-G

16. Gassmann, F. Elastic waves through a packing of spheres. *Geophysics* **16**, 673–685 (1951).
17. Greenberg, M. L. & Castagna, J. P. Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and application. *Geophys. Prospect.* **40**, 195–209 (1992). https://doi.org/10.1111/j.1365-2478.1992.tb00371.x
18. Batzle, M. & Wang, Z. Seismic properties of pore fluids. *Geophysics* **57**, 1396–1408 (1992). https://doi.org/10.1190/1.1443207
19. Raymer, L.L., Hunt, E.R. & Gardner, J.S. An improved sonic transit time-to-porosity transform. In SPWLA 21st Annual Logging Symposium (1980).
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. Petrol. Geol.* **150**, 106097 (2023). https://doi.org/10.1016/j.marpetgeo.2023.106097
21. Zhao, T., Jayaram, V., Roy, A. & Marfurt, K. J. A comparison of classification techniques for seismic facies recognition. *Interpretation* **3**, SAE71–SAE83 (2015). https://doi.org/10.1190/INT-2015-0044.1
22. Chopra, S. & Marfurt, K. J. Seismic characterization with unsupervised machine learning applications for facies classification. *Lead. Edge* **44**, 934–943 (2025). https://doi.org/10.1190/tle44120934.1
23. Roden, R., Smith, T. & Sacrey, D. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation* **3**, SAE59–SAE83 (2015). https://doi.org/10.1190/INT-2015-0037.1
24. 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-Energ. Geo-Resour.* **10**, 185 (2024). https://doi.org/10.1007/s40948-024-00907-1
25. La Marca, K., Bedle, H., Stright, L. & Marfurt, K. J. Uncertainty assessment in unsupervised machine-learning methods for deepwater channel seismic facies using outcrop-derived 3D models and synthetic seismic data. *Interpretation* **12**, T303–T319 (2024). https://doi.org/10.1190/INT-2023-0094.1
26. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69 (1982). https://doi.org/10.1007/BF00337288
27. Kohonen, T. *Self-organizing maps.* 3rd edn, Springer Ser. Inf. Sci., Vol. 30 (Springer, Berlin, 2001). https://doi.org/10.1007/978-3-642-56927-2
28. Yin, H. The self-organizing maps: Background, theories, extensions and applications. *Stud. Comput. Intell.* **115**, 715–762 (2008). https://doi.org/10.1007/978-3-540-78293-3_17
29. Carrasco Kind, M. & Brunner, R. J. SOM z: Photometric redshift PDFs with self-organizing maps and random atlas. *Mon. Not. R. Astron. Soc.* **438**, 3409–3421 (2014). https://doi.org/10.1093/mnras/stt2456

30. Castagna, J. P., Portniaguine, O., Gil, G., Oyem, A. & Liang, C. Multiscale transform–based seismic reflectivity inversion using a convolutional neural network. *arXiv* **2506.11337** (2025).
31. Sinha, S., Routh, P. S., Anno, P. & Castagna, J. P. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics* **70**, P19–P25 (2005). https://doi.org/10.1190/1.2127113
32. Puryear, C. I., Portniaguine, O. & Cobos, C. M. Constrained least-squares spectral analysis: Application to seismic data. *Geophysics* **77**(5), V143–V167 (2012). https://doi.org/10.1190/geo2011-0210.1
33. Marfurt, K. J. & Kirlin, R. L. Narrow-band spectral analysis and thin-bed tuning. *Geophysics* **66**, 1274–1283 (2001). https://doi.org/10.1190/1.1487075
34. Castagna, J. P., Sun, S. & Siegfried, R. W. Instantaneous spectral analysis: detection of low-frequency shadows associated with hydrocarbons. *Lead. Edge* **22**, 120–127 (2003). https://doi.org/10.1190/1.1559038
35. Backus, M. M., Nepomuceno, F. & Cao, J. The reflection seismogram in a solid layered earth. In *SEG Technical Program Expanded Abstracts 1982*, 416–417 (Society of Exploration Geophysicists, 1982). https://doi.org/10.1190/1.1826899
36. Chopra, S. & Castagna, J. P. *AVO* (Society of Exploration Geophysicists, 2014). https://doi.org/10.1190/1.9781560803201

Contributions:

Author M.A.G. implemented the methods and drafted the manuscript. The second author, J.C., discussed the results with the first author and reviewed the manuscript.

Data Availability:

The seismic data analyzed in this work is a subcube extracted from the publicly available TOGAR 3D 2011 seismic survey. The dataset was originally provided by GeoKinetics (now SEA Exploration) in 2018 and is currently available through the Queensland Government and can be accessed at [TOGAR 3D 2011 - TOGAR 3D 2011 - Seismic - GSQ Open Data Portal].

Acknowledgements:

The authors thank Lumina Geophysical for providing the software license used for extracting seismic attributes in this research.

Funding Declaration:

The authors did not receive external funding for conducting this research.