UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INTEGRATING SEISMIC INVERSION PRODUCTS INTO MULTI-ATTRIBUTE ANALYSIS AND UNSUPERVISED MACHINE LEARNING FOR RESERVOIR CHARACTERIZATION, POHOKURA, TARANAKI BASIN, NEW ZEALAND

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By Yasin Üzüm

Norman, Oklahoma

2026

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/342488

INTEGRATING SEISMIC INVERSION PRODUCTS INTO MULTI-ATTRIBUTE ANALYSIS AND UNSUPERVISED MACHINE LEARNING FOR RESERVOIR CHARACTERIZATION, POHOKURA, TARANAKI BASIN, NEW ZEALAND

A THESIS APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Matthew J. Pranter

Dr. Hao Hu

© Copyright by YASIN UZUM 2026

All Rights Reserved.

To those who work hard, strive, and earn their place on their own merit.

To those who truly deserve.

Acknowledgements

First and foremost, I would like to thank my selfless parents, my mother Aynur and my father Halil İbrahim, who dedicated their lives to their children and supported us under all circumstances, even at the cost of their own dreams. Everything I have achieved has been possible because of you. I am also deeply grateful to my beautiful sisters, Merve and Yasemin, for their unwavering encouragement and constant presence.

I would like to express my sincere gratitude to my advisor, Dr. Heather Bedle, for her mentorship, guidance, and constant understanding, and for everything I have learned from her, as well as her kind and consistently positive attitude.

I am grateful to all members of the AASPI (Attribute-Assisted Seismic Processing & Interpretation) research group, especially Dr. Kurt Marfurt, whose legacy we strive to carry forward, and Dr. David Lubo-Robles. I am particularly grateful to my colleagues, with whom I had the opportunity to work, learn, and grow together: Nour, Hilmi, Aniq, Alex and April, as well as many others. I also thank Hy and Pamela for welcoming me to Norman; I will carry your kindness and generosity forward.

I would like to thank my esteemed committee members, Dr. Hao Hu and Dr. Matthew Pranter, for their valuable feedback.

I am grateful to the OU Geoscience staff, especially Rebecca Fay, for their continuous support and invaluable assistance. I would also like to thank all OU staff whose hard work, often unseen, supports our daily lives.

I am sincerely grateful to the Ministry of National Education of the Republic of Türkiye and the Turkish Petroleum Corporation (TPAO) for fully funding my master's studies under the

Study Abroad Program (Law No. 1416). I also thank my TPAO advisors, Taner Arpacı, Uygar Ceyhan, and Ufuk Durmuş, for their guidance and support.

I gratefully acknowledge the Ministry of Business, Innovation and Employment (MBIE) and New Zealand Petroleum and Minerals for providing open access to the datasets used in this research.

Finally, and most importantly, I would like to thank my friends. I am especially grateful to Melih and Emre for their friendship and support from afar. My heartfelt thanks go to Abby and Afshin, my very first friends in Norman, and to the many wonderful people who came into my life through them. I will truly miss every moment we shared. I am also deeply grateful to Harun, Alperen, Şaban, Beşir and their families, who made me feel at home.

I would also like to honor the memory of my dear friend Tibet, who is no longer with us. Not a single day has gone by without me thinking of you. The art we created together on the theater stage in Taşkışla and the wonderful people I shared it with will always remain a part of me.

Table of Contents

Acknowledgements v

Table of Contents vii

List of Figures viii

List of Tables xi

Abstract xii

1. Introduction 1

2. Study Area and Datasets 7

3. Geologic Background 14

4. Methods and Workflow 22

5. Results 40

6. Discussion 59

7. Conclusion 68

8. References 70

9. Appendix 76

List of Figures

Figure 1.1: Amplitude spectrum comparison of different angle stacks extracted from 3D seismic data over the Pohokura Field. 3

Figure 2.1: Map showing **(a)** the location of Pohokura Field and **(b)** enlarged view of Pohokura 3D Seismic Survey (PR4544) and well locations used in the analysis. Amplitude data of W-E Profile (Inline 595) is shown in Figure 2.2. **Note.** Adapted from Geodata Catalogue (New Zealand Petroleum and Minerals, n.d.)..... 7

Figure 2.2: Seismic amplitude section along the W-E Profile (Inline 595). The green rectangular area shows the cropping boundaries (3200-4800). Figure 2.1 shows the direction of the profile. Inline 595 serves as a representative seismic section capturing the main structural and stratigraphic features of the Pohokura Field, including the anticlinal structure of the reservoir.. 11

Figure 3.1: Structural framework of the Taranaki Basin, modified after New Zealand Petroleum & Minerals (2014), with the study area outlined by a black box and highlighted in red. 14

Figure 3.2: Generalized Cretaceous-Cenozoic stratigraphic cross-section of the Taranaki Basin with the study interval outlined by a black box and highlighted in red (New Zealand Petroleum & Minerals, 2014). 16

Figure 3.3: Schematic stratigraphic cross-section across the Pohokura Field showing the internal stratigraphic architecture of the Mangahewa Formation. The offshore study interval is outlined by a black box and highlighted in red. Measured depth (MD) values shown alongside the figure are approximate and interpreted from well logs. Table 3.1 provides interpretations of the reservoir units 1-5. Modified after Shell Exploration NZ Ltd. (2012b). 18

Figure 4.1: Generalized workflow illustrating the comparative multi-attribute scenarios: multi-attribute analysis based solely on seismic amplitude data and multi-attribute analysis

incorporating seismic inversion products alongside other attributes. Red boxes represent processing steps, whereas blue boxes indicate 3D data volumes used in the workflow. 23

Figure 4.2: Well log correlation panel showing gamma ray (GR) logs and interpreted stratigraphic horizons across the study wells: **(a)** panel flattened on the Turi Top, illustrating the interval from the Turi Formation to the Mangahewa Base **(b)** panel flattened on the Mangahewa Top, providing a detailed view of the Mangahewa reservoir section. 25

Figure 4.3: Cropped seismic section (Inline 595) showing the interpreted upper and lower boundaries define the analysis interval. 26

Figure 4.4: Correlation matrix showing the relationships between the generated seismic attributes. Numerical correlation values are displayed only for the attributes selected for the final attribute set. 31

Figure 4.5: Schematic representation of PCA showing the first two principal components. 33

Figure 5.1: Percentage contributions of individual seismic attributes to the first two principal components (PC1 and PC2) derived from PCA for the evaluated input datasets. 41

Figure 5.2: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 555) result without inversion products **(b)** Depth slice at 4000m without inversion products **(c)** WE Profile (Inline 555) result with inversion products **(d)** Depth slice at 4000m with inversion products. The observed enhancements are highlighted as follows: structural discontinuities (yellow arrows), reservoir boundaries (blue arrows), and reservoir heterogeneity zones (red arrows). 44

Figure 5.3: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 537) result without inversion products **(b)** enlarged view of the highlighted interval without inversion products **(c)** WE Profile (Inline 537) result with inversion products **(d)** enlarged view of the

highlighted interval with inversion products. The gamma ray (GR) log of the Pohokura-2 well is included for calibration. 48

Figure 5.4: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 648) result without inversion products **(b)** enlarged view of the highlighted interval without inversion products **(c)** WE Profile (Inline 648) result with inversion products **(d)** enlarged view of the highlighted interval with inversion products. The gamma ray (GR) log of the Okoki-1 well is included for calibration. 51

Figure 5.5: Crossplot of two K-means clustering axis **(a)** WE Profile (Inline 595) result without inversion products **(b)** Depth slice at 4000m without inversion products **(c)** WE Profile (Inline 595) result with inversion products **(d)** Depth slice at 4000m inversion products. 54

List of Tables

Table 2.1: Summary of well information and log data..... 10

Table 2.2: Acquisition parameters of the component surveys of the Pohokura Merge 3D Survey. 12

Table 3.1: Summary of the interpreted reservoir units within the Mangahewa Formation in the Pohokura Field. The reservoir units (Units 1-5) correspond to the stratigraphic intervals illustrated in Figure 3.3. 19

Table 4.1: Formation tops interpreted from well logs and tied to seismic data. Depths are reported in meters measured depth (MD). 24

Table 4.2: List of seismic attributes generated from the seismic amplitude volume. Attributes selected for multi-attribute analysis are shown in bold. 29

Abstract

Recent advances in machine learning have enhanced the capability of multi-attribute analysis to resolve relationships within high-dimensional data, enabling more sophisticated reservoir characterization. However, the ability of multi-attribute analysis to produce geologically meaningful results remains strongly controlled by the diversity, relevance, and spatial extent of the input data. Seismic amplitude data is inherently band-limited and does not adequately capture low-frequency components, which limits its ability to represent subsurface variability. Inversion products derived from low-frequency models provide a complementary source of information to address this limitation.

This study investigates the effects of incorporating pre-stack simultaneous AVO inversion products into multi-attribute analysis. To evaluate the impact of inversion products, a comparative framework based on two input datasets is established. The first scenario utilizes attributes derived solely from seismic amplitude data, whereas the second incorporates P- and S-impedance volumes into the attribute set. The Pohokura Field in the Taranaki Basin is used as a case study, where multi-attribute analysis is performed for both scenarios using Principal Component Analysis (PCA) and K-means clustering.

The results indicate that the inclusion of inversion products leads to a shift in data representation from a structure primarily governed by seismic amplitude response to one that is more consistent with subsurface geology and controlled by elastic properties. This transition improves the delineation of reservoir and seal units and enhances the definition of reservoir boundaries, lithological trends, heterogeneity zones, structural discontinuities, and facies distributions. Sand-prone intervals that are not clearly resolved in the attribute-only scenario

become detectable when inversion products are included. Furthermore, the inclusion of inversion products results in reduced noise and improved vertical and lateral resolution, which enhances interpretability and enables a more reliable characterization of the complex Mangahewa reservoir.

The achievements gained by incorporating inversion products into multi-attribute analysis highlight the importance of enriching the frequency content of the input data. In this regard, multi-attribute analysis should be regarded not simply as a data integration approach, but as a design-oriented process in which the selection and organization of the input dataset fundamentally control the quality of the results.

1. Introduction

Seismic attributes are powerful tools as mathematical representations designed to highlight geological or physical features that may not be explicitly distinguishable in raw seismic data. Whether using conventional visualization methods (e.g. crossplotting, co-rendering, color blending) or advanced statistical methods (e.g. regression, clustering, predictive modelling), the use of more than one attribute in combination is called multi-attribute analysis. Recent developments in machine learning (ML) have improved the ability to capture both linear and nonlinear relationships among input variables. In particular, advances in dimensionality reduction and clustering techniques have enabled effective analysis of high-dimensional data by reducing data redundancy, revealing the organization of the data in attribute space, and improving the identification of significant patterns. Despite these advances, the ability of multi-attribute analysis to yield geologically meaningful results is strongly conditioned by the diversity, relevance and spatial extent of the input data.

Seismic inversion can be defined as methods that convert seismic amplitude data into geological models describing the physical rock properties of the subsurface. The fundamental idea behind seismic inversion is to recover the acoustic impedance, which is the primary factor causing the observed amplitude variations in seismic reflection data, and to obtain physically interpretable subsurface representations that are more directly related to reservoir properties.

Multi-attribute analysis and seismic inversion are widely used approaches in reservoir characterization and reveal different levels of information about the subsurface. Seismic data and attributes derived from it primarily reflect amplitude-based variations, whereas seismic inversion products provide model-based representations of subsurface physical properties. In this respect,

the integration of inversion products into multi-attribute analysis has the potential to improve input diversity, add geological relevance to the analysis, and provide higher-resolution spatial information.

In this study, the effects of incorporating pre-stack simultaneous AVO (Amplitude Versus Offset) inversion products into multi-attribute analysis are investigated. To this end, two scenarios were constructed based on different input datasets: (1) an attribute-only dataset derived from seismic amplitude data, and (2) a dataset incorporating seismic inversion products alongside amplitude attributes. The same workflow was applied to the input sets of both scenarios without any modification to the analysis steps or parameters, and the resulting outputs were compared to investigate the effects of including inversion products into multi-attribute analysis.

The main motivation of this research arises from the band-limited nature of seismic amplitude data and the complementary low-frequency content that inversion products can provide. Due to acquisition and recording limitations, attenuation, and various processing steps (e.g. low-cut filtering, deconvolution, noise suppression) seismic amplitude volumes contain energy only within a restricted frequency band (typically ~ 5-80 Hz). Figure 1.1 shows the amplitude spectra of different angle stacks from the Pohokura 3D seismic dataset (PR4544). A comparison of the amplitude spectrums indicates that, although minor differences are observed, all the stacks exhibit similar bandwidths. The spectral ranges largely overlap, and the energy distributions for all stacks are concentrated in a similar dominant frequency range. Consequently, both very low and very high frequencies are not resolved, and attributes derived from such data can only capture variations occurring within this limited spectral range. In contrast, seismic inversion products are generated using a low-frequency model (LFM) constructed by calculating

acoustic impedances from well logs and propagating it along interpreted horizons; as a result, they contain low-frequency information that is not present in seismic data. Including inversion products in multi-attribute analysis is expected to enrich the dataset with low-frequency content, improving the ability to characterize lithological variations and facies distribution.

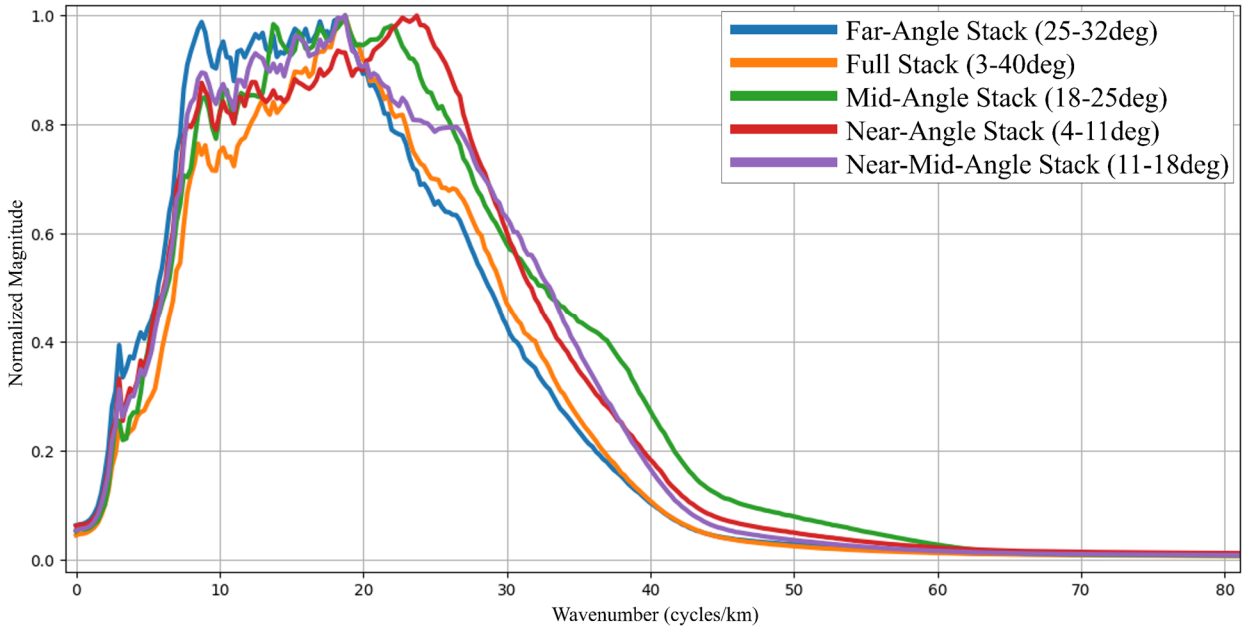


Figure 1.1: Amplitude spectrum comparison of different angle stacks extracted from 3D seismic data over the Pohokura Field.

Beyond this spectral complementarity introduced by inversion products, several additional hypotheses can be formulated regarding their contribution to multi-attribute analysis. First, the inclusion of inversion volumes is expected to reduce the overall noise content of the input set, since attribute calculations often amplify noise present in the original seismic amplitudes, whereas the processing workflow used to generate inversion products generally suppresses random noise. Second, their inclusion may help reduce the influence of shared noise across multiple attributes derived from the same seismic volume, which can otherwise appear as spurious correlations and lead to misleading geological interpretations. Third, by combining the

high-resolution variability captured by seismic attributes with the smoother and more geologically consistent behavior of inversion products, multi-attribute analysis can integrate information across different spectral scales, stabilize spatial variability, and produce more coherent spatial patterns. Finally, inversion products represent quantitative elastic parameters directly related to subsurface rock and fluid properties, their inclusion is expected to enhance the discrimination of lithological variations and facies distribution. Based on these theoretical expectations, it is hypothesized that incorporating seismic inversion products into multi-attribute analysis will provide significant improvements.

While the potential advantages of seismic inversion products are substantial, their inherent limitations and possible impacts on analysis outcomes must also be considered. A key limitation is that model-based filtering and regularization applied during the inversion process may suppress high-frequency geological content by treating such variations as noise. In addition, inversion products are strongly dependent on the initial model and are inherently dominated by low-frequency components. The inclusion of these products within multi-attribute analysis may therefore reduce the relative contribution of high-frequency information provided by seismic attributes. Furthermore, the reliability and availability of inversion products are critical factors controlling their integration into the analysis. In early-stage exploration, where well control is limited or absent, inversion products cannot be reliably generated and may not be feasible at all. In such cases, reservoir characterization must rely solely on seismic amplitude data and attributes. Considering these limitations, it is essential to rigorously evaluate the contribution of seismic inversion products to multi-attribute analysis, as well as the reliability of attribute-only approaches.

Numerous studies have integrated seismic inversion products and seismic attributes within a variety of methodological frameworks. These studies can broadly be classified under three main categories: interpretation workflows, comparative approaches, and integrated analytical models. The majority of these studies aim to improve reservoir characterization by incorporating them within interpretation workflows. Mahdy et al. (2026) improved the identification of thin gas-sand channels by using elastic parameters obtained from pre-stack simultaneous inversion together with seismic attributes in a physics-based interpretation workflow. Lee et al. (2017) combined amplitude and inversion attributes in a K-means multi-attribute analysis and quantitatively delineated the distribution of gas hydrate reservoirs. Some studies have evaluated the two approaches comparatively. Leal et al. (2019) compared the results obtained by separately applying traditional seismic attributes and inversion-derived elastic parameters to multi-attribute classification using a Self-Organizing Map (SOM) and showed that both approaches yielded spatially consistent results, with the traditional attribute-based SOM providing greater lithological detail. There are also studies that jointly use inversion and attribute data within the same analytical model. Calderon and Castagna (2007) estimated the spatial distribution of porosity and lithology using acoustic impedance derived from model-based inversion together with seismic attributes. Their approach, which combines multi-attribute transforms with neural network inversion, leads to improved prediction accuracy. Verma et al. (2025) developed a machine learning approach using a Probabilistic Neural Network (PNN), in which Principal Component Analysis (PCA) was applied to reduce the dimensionality of seismic attributes prior to integration with impedance volumes and accurately estimated the porosity distribution between wells. Harryandi and Tura (2019) modeled reservoir heterogeneity by using elastic parameters obtained from pre-stack seismic inversion together with properties estimated

using multi-attribute transformation in Bayesian facies modeling, resulting in the identification of sweet-spot regions.

These studies demonstrate that integrating seismic inversion products with seismic attributes can enhance both interpretation and analytical workflows. However, existing work rarely isolates and systematically evaluates the role of inversion products by performing a controlled comparison between attribute-only and inversion-augmented datasets under identical analytical conditions. As a result, the specific contributions and potential limitations of incorporating inversion products into multi-attribute analysis remain insufficiently constrained. Critically, no study to date provides a unified comparative framework that explicitly assesses the respective contributions and limitations of seismic attributes and inversion products within machine learning-assisted reservoir characterization workflows.

The impact of incorporating pre-stack simultaneous AVO inversion products into multi-attribute analysis is systematically investigated through a controlled comparative workflow applied to the Pohokura 3D dataset. The proposed approach integrates PCA and K-means clustering as complementary unsupervised techniques to evaluate how inversion products modify the structure of the attribute space and influence reservoir characterization outcomes, while assessing the capabilities and limitations of attribute-only and inversion-included input sets. The results demonstrate that integrating seismic inversion products into multi-attribute analysis leads to a shift from amplitude-driven variability toward a more geologically consistent, elastically controlled representation. This transformation improves the delineation of reservoir architecture and internal heterogeneity, enhances facies discrimination, and reduces noise, resulting in more interpretable and spatially coherent outcomes.

2. Study Area and Datasets

The Pohokura hydrocarbon field, located in the Taranaki Basin on the west coast of New Zealand's North Island, approximately 10 km offshore to the north of Waitara (Figure 2.1), serves as the study area of this research. The field provides an excellent case study given its structural and stratigraphic complexity, the availability of a comprehensive geophysical dataset including high-resolution 3D seismic data and pre-stack inversion products, and extensive well control across the reservoir.

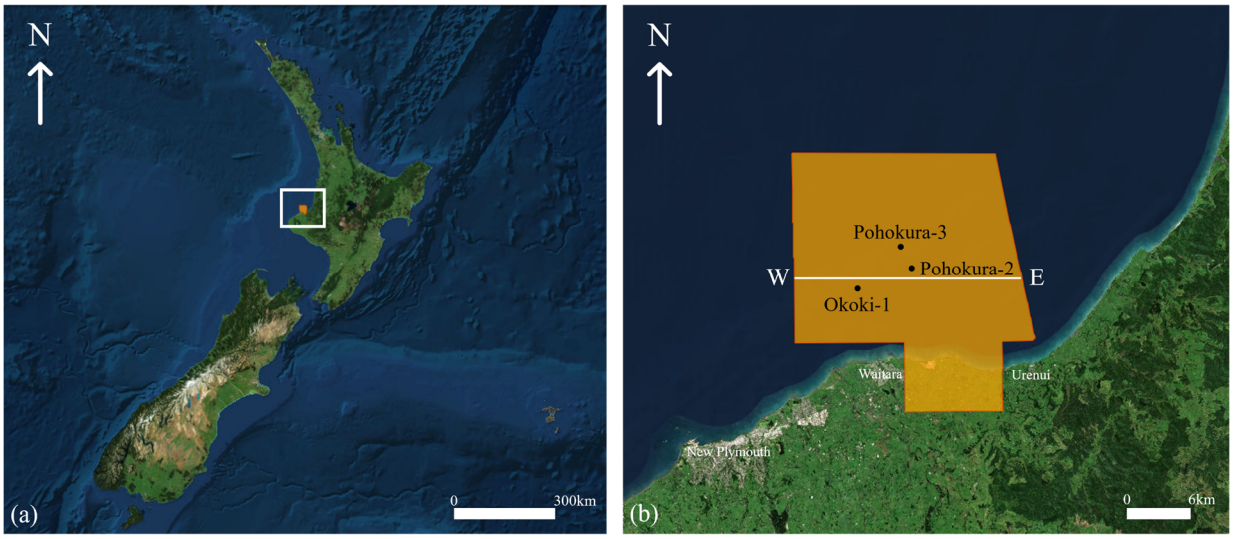


Figure 2.1: Map showing **(a)** the location of Pohokura Field and **(b)** enlarged view of Pohokura 3D Seismic Survey (PR4544) and well locations used in the analysis. Amplitude data of W-E Profile (Inline 595) is shown in Figure 2.2. **Note.** Adapted from Geodata Catalogue (New Zealand Petroleum and Minerals, n.d.)

Exploration activities in the Taranaki Basin, which hosts the majority of New Zealand's hydrocarbon production, accelerated following the discovery of the Kapuni gas field in 1959 and the Maui gas field in 1969, and the Pohokura field was subsequently discovered in February 2000 (New Zealand Petroleum & Minerals, 2014). The first discovery well, Pohokura-1,

intersected a gas column approximately 130 m thick within the Mangahewa Formation at a depth of 3480 m, recorded significant gas flow rates, and revealed substantial gas and condensate accumulations within sandstone reservoirs developed in an inverted anticlinal structure (New Zealand Petroleum & Minerals, 2014). Drilled to determine the structural position of the reservoir and the distribution of sandstones, to test reservoir continuity in the northern part of the anticline, and to obtain hydrocarbon samples, the Pohokura-2 well encountered a gas column of approximately 115 m within the Maui sandstones of the Kapuni Group and exhibited high gas and condensate flow rates, thereby confirming the presence of a commercially viable hydrocarbon accumulation in the field (Murray, 2000; New Zealand Petroleum & Minerals, 2014). Following these discoveries, numerous exploration and appraisal wells were drilled in the area, reservoir lateral continuity and commercial production potential were confirmed, and ultimately the first commercial gas production commenced in 2006 (New Zealand Petroleum & Minerals, 2014). Today, the Pohokura field continues to play a critical role in New Zealand's hydrocarbon production and energy supply. According to the Ministry of Business, Innovation and Employment (2025, Figure 15), the Pohokura field contains estimated remaining reserves of 18.7 petajoules of condensate and 180.63 petajoules of natural gas, making it the second largest hydrocarbon accumulation in New Zealand.

A key strength of the Pohokura field for this research lies in the completeness and quality of the available geophysical datasets, which provide a strong foundation for the multi-attribute analysis. High-resolution 3D seismic data in both time and depth domains, together with multiple angle stacks and advanced post-processing volumes, constitute the seismic dataset required for attribute generation. In addition, pre-stack simultaneous AVO inversion products offer a broad range of inputs for comparative analysis.

The availability of pre-stack inversion products is particularly important, as pre-stack inversion was selected over post-stack inversion due to its several advantages. Once seismic data is stacked, reflections from different angles are averaged to represent normal incidence conditions. Under normal incidence, no mode conversion occurs and seismic wave propagation is governed primarily by compressional behavior (Chopra & Marfurt, 2024). Consequently, stacked seismic data exhibits a predominantly acoustic response, and post-stack inversion is limited to estimating only P-wave impedance. In addition, stacking removes offset-dependent AVO (Amplitude Versus Offset) behavior and tuning effects associated with thin layers, and this information cannot be recovered during post-stack inversion. In contrast, pre-stack seismic gathers preserve amplitude variations with offset, allowing analysis of angle-dependent reflectivity and separation of components related to shear elastic properties. This enables estimation of S-wave velocity and, consequently, S-impedance, as well as improved characterization of thin layers and recovery of a broader range of elastic parameters. For these reasons, pre-stack inversion products were preferred, and the Pohokura field was chosen as the case study partly due to the availability of pre-stack inversion products.

In addition to the geophysical data, numerous exploration and production wells across the reservoir provide well control. Well-log data from selected wells are used for calibration and validation of the analysis results and to support the interpretations. The well locations used in the analysis are illustrated in Figure 2.1, and their key properties are summarized in Table 2.1.

Most importantly, all of these datasets are openly accessible via the Geodata Catalogue provided by New Zealand Petroleum and Minerals under the Crown Minerals Act 1991 (New Zealand Petroleum & Minerals, n.d.).

Table 2.1: Summary of well information and log data.

Well Name	Location (Lat-Long)	Elevation (m AMSL)	Total Depth (m)	Logs Used
Okoki-1	38°55'38.9"S 174°12'17.7"E	31.4	4257.5	SP, GR, RHOB, NPHI
Pohokura-2	38°54'35.616"S 174°15'42.973"E	30.2	3817	SP, GR, DENS, NEUT
Pohokura-3	38°53'28.1663"S 174°15'00.7551"E	23.2	3815	SP, GR, DENS, NEUT

Beyond the advantages of the available datasets, the Pohokura field is particularly well suited from a geological perspective, as it contains well-defined structural features for investigation, including a fault-controlled anticlinal structure (Figure 2.2) that forms the primary hydrocarbon trap. The reservoir is characterized by pronounced multi-scale heterogeneity and structural segmentation, which pose challenges for reservoir characterization and provide additional layers of insight to the study.

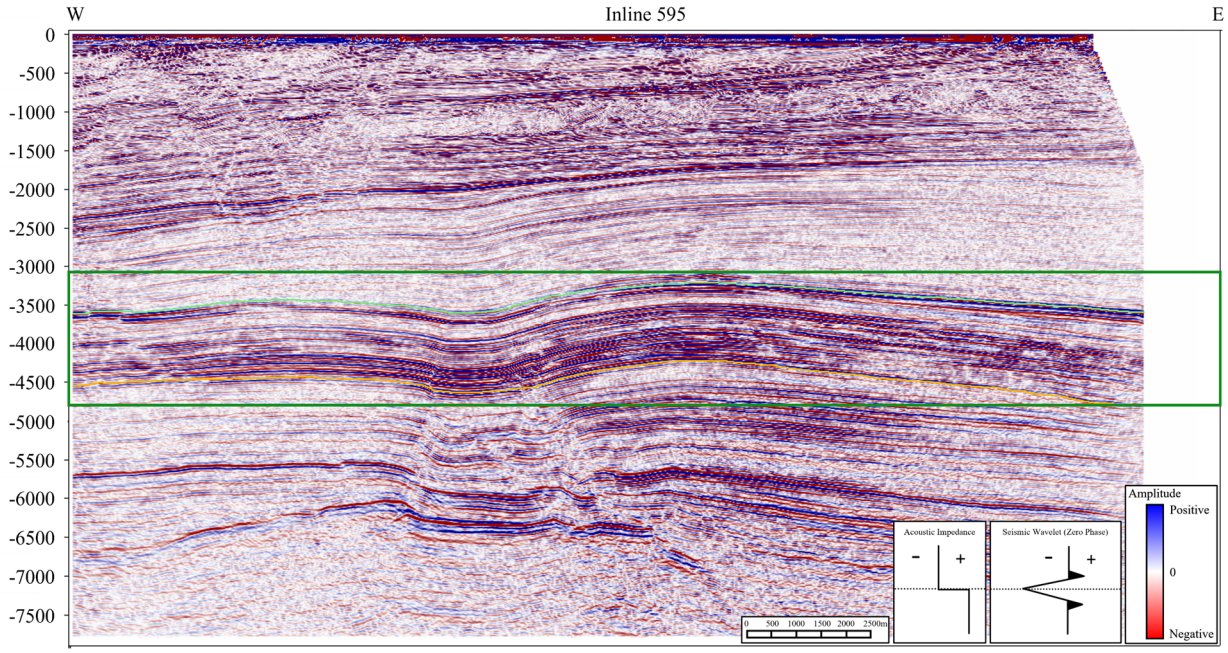


Figure 2.2: Seismic amplitude section along the W-E Profile (Inline 595). The green rectangular area shows the cropping boundaries (3200-4800). Figure 2.1 shows the direction of the profile. Inline 595 serves as a representative seismic section capturing the main structural and stratigraphic features of the Pohokura Field, including the anticlinal structure of the reservoir.

The seismic datasets are derived from the Pohokura Merge 3D Survey (PR4544), which is a combination of marine, transition zone, and land seismic data (Table 2.2) from the Pohokura and Okaki fields. The merged dataset integrates approximately 324 km² of marine data, 69 km² of transition zone data, and 177 km² of onshore seismic data. Due to overlap between the individual surveys, the effective coverage of the merged dataset is approximately 358 km². The seismic data were reprocessed, as reported by WesternGeco, to enhance hydrocarbon exploration by addressing limitations of the previous processing, including imaging challenges, noise contamination, and acquisition footprint (Shell Exploration NZ Ltd., 2012a).

Table 2.2: Acquisition parameters of the component surveys of the Pohokura Merge 3D Survey.

Survey	Sample Increment (ms)	Record Length (s)	Bin Size (m x m)
Marine 3D	2	6	12.5 x 25
TZ 3D	2	6	12.5 x 25
Land 3D	2	5.2	30 x 30

The inversion products were generated using simultaneous AVO inversion applied to pre-stack depth-migrated seismic data, as part of a study carried out by WesternGeco Australia for Shell Exploration NZ Ltd (Shell Exploration NZ Ltd., 2012b). In the initial stage, elastic parameters were calculated from compressional wave velocity (Vp) and density (ρ) derived from well logs. Shear wave velocity (Vs) measurements were used directly where available; in wells lacking shear measurements, Vs was estimated empirically from Vp using Greenberg-Castagna relationships (Greenberg & Castagna, 1992) calibrated to local geological conditions. Subsequently, the elastic parameters derived from the wells were used to construct a low-frequency model (LFM), which served as the initial model for inversion. The angle-dependent amplitude variations observed in the pre-stack seismic data were then modeled to determine the elastic property contrasts responsible for these variations, and this relationship was inverted to estimate the spatial distribution of elastic parameters (Shell Exploration NZ Ltd., 2012b). Simultaneous AVO inversion is an iterative inverse modeling process that seeks the distribution of elastic parameters that minimizes the mismatch between the observed seismic amplitudes and the seismic response predicted from the elastic model. In this process, the seismic response computed from the initial elastic model is compared with the observed data, and the model

parameters are progressively updated at each iteration to reduce the misfit. Through this iterative optimization, inversion volumes representing the three-dimensional distribution of elastic properties across the study area were obtained. These products include primary attributes (P-impedance, S-impedance, density), and as well as secondary attributes (Vp/Vs ratio, lambda-rho, mu-rho, poisson's ratio). In addition, reservoir characterization volumes, including net-to-gross, were generated using rock property predictions derived from the elastic parameters.

3. Geologic Background

The Taranaki Basin represents New Zealand's principal hydrocarbon province and hosts the majority of the country's commercial oil and gas accumulations. Located along the western margin of New Zealand's North Island, it developed as a Late Cretaceous intra-continental rift basin associated with Tasman Sea spreading during the breakup of Gondwana and was later modified by Miocene compressional deformation and structural inversion (King & Thrasher, 1996). Structurally, the basin is subdivided into the relatively stable Western Platform and the more tectonically active Eastern Mobile Belt, where the Pohokura Field is located (Figure 3.1).

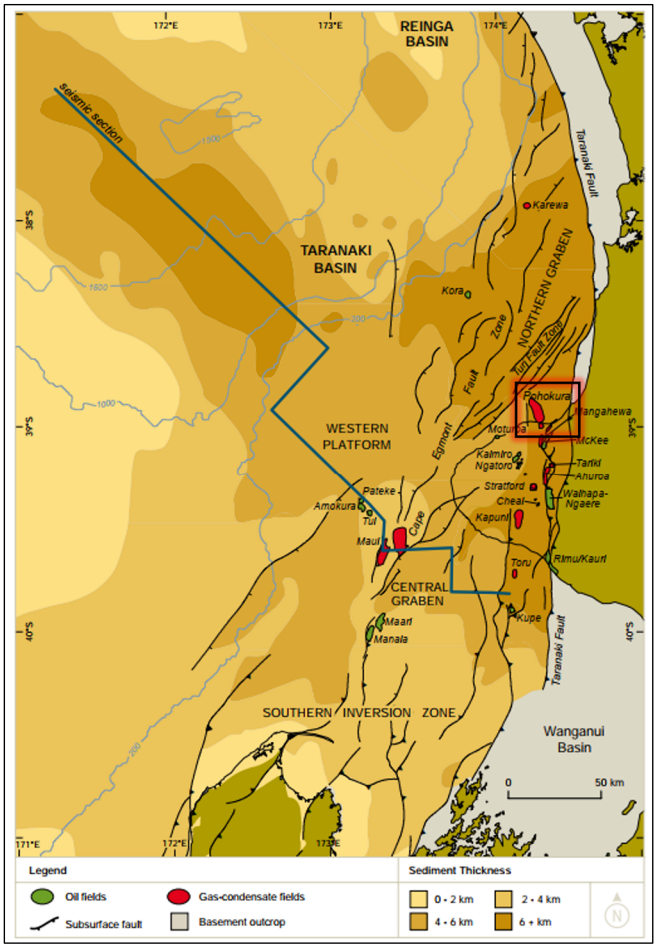


Figure 3.1: Structural framework of the Taranaki Basin, modified after New Zealand Petroleum & Minerals (2014), with the study area outlined by a black box and highlighted in red.

This structural framework reflects the basin's early extensional architecture established during the Late Cretaceous. The Late Cretaceous extensional tectonic regime generated fault-controlled depocenters that accommodated thick accumulations of terrestrial and nearshore transitional sediments; these successions progressively evolved into increasingly marine shelf environments during the Paleogene (King & Thrasher, 1996; New Zealand Petroleum & Minerals, 2014). Continued thermal subsidence during the Eocene to Early Oligocene passive-margin phase led to a regional marine transgression. This resulted in the development of a broad shelf-to-slope system, within which terrestrial to marginal-marine successions of the Kapuni Group, including the Mangahewa Formation sandstones, and the overlying marine mudstone successions of the Turi Formation were deposited both vertically and laterally (King & Thrasher, 1996). These depositional systems formed the fundamental elements of the basin's petroleum system. In particular, organic-rich coals and carbonaceous mudstones within Late Cretaceous to Eocene terrestrial and paralic successions provided predominantly gas-prone source rock potential, while associated sandstone bodies served as reservoirs (King & Thrasher, 1996). Regionally extensive transgressive marine mudstones of the Turi Formation acted as effective seals (Robinson & King, 1988).

Beginning in the Miocene, changes in the interaction between the Pacific and Australian plates during oblique convergence introduced compressional deformation and localized structural inversion, reactivating earlier rift-related normal faults and generating fold- and fault-related structural traps (King & Thrasher, 1996). This tectonic phase led to the uplift of structural highs, inversion of earlier extensional structures, and the development of fault-bounded anticlines that enhanced hydrocarbon entrapment. The combined effects of early extensional basin architecture, long-term subsidence, diverse marine depositional systems, and subsequent fault reactivation

established the Taranaki Basin as New Zealand's most prolific hydrocarbon province (King & Thrasher, 1996; Robinson & King, 1988).

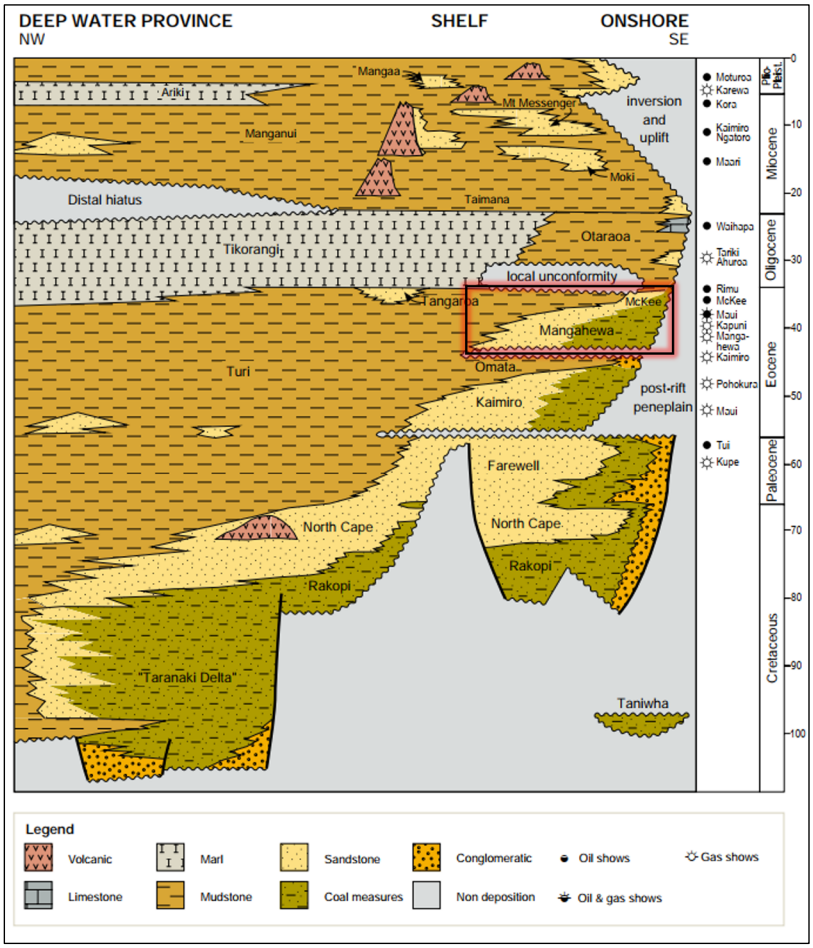


Figure 3.2: Generalized Cretaceous-Cenozoic stratigraphic cross-section of the Taranaki Basin with the study interval outlined by a black box and highlighted in red (New Zealand Petroleum & Minerals, 2014).

Among the hydrocarbon accumulations discovered in the Taranaki Basin, the Pohokura Field, located in the northern offshore sector of the basin (Figure 3.1), is one of the largest gas-condensate accumulations in New Zealand. Its structural framework reflects the broader tectonic evolution of the basin. Normal faults developed during the Late Cretaceous rift phase. During the Miocene, oblique convergence between the Pacific and Australian plates reactivated these faults

under a compressional regime, leading to the development of a north-south trending anticline (Figure 2.2) approximately 16 km long and 5 km wide (King & Thrasher, 1996; Kalid & Hamzah, 2014). This reactivation transformed earlier extensional structures into fault-controlled closures, establishing the structural configuration necessary for hydrocarbon accumulation within the field. Furthermore, the fault-controlled anticlinal geometry contributes to structural segmentation within the reservoir, promoting the development of structural heterogeneity.

Stratigraphically, Pohokura represents a structurally uplifted and Miocene-deformed version of the Late Cretaceous to Neogene succession of the Taranaki Basin (Figure 3.2). This deformation resulted in folding and faulting of pre-existing units, leading to local thickness variations and the development of structurally controlled trap geometries. The Paleocene-Eocene units, deposited prior to this deformation, formed under passive-margin conditions during a phase of thermal subsidence, which resulted in the accumulation of thick sedimentary sequences (King & Thrasher, 1996).

Across the Taranaki Basin, the principal source of hydrocarbon charge is provided by Late Cretaceous-Eocene coaly and organic-rich paralic to marginal-marine deposits primarily within the Pakawau and Kapuni groups. These units are characterized by gas-prone facies deposited in coastal swamps, delta plain, and transitional environments, particularly during the Eocene (King & Thrasher, 1996; Robinson & King, 1988). Continued basin subsidence and progressive burial facilitated thermal maturation of these organic-rich intervals, enabling significant gas generation (King & Thrasher, 1996). The stratigraphic proximity between source and reservoir units results in a relatively short migration pathway. Miocene structural reactivation not only contributed to the development of trap geometries but also influenced potential migration pathways within the field.

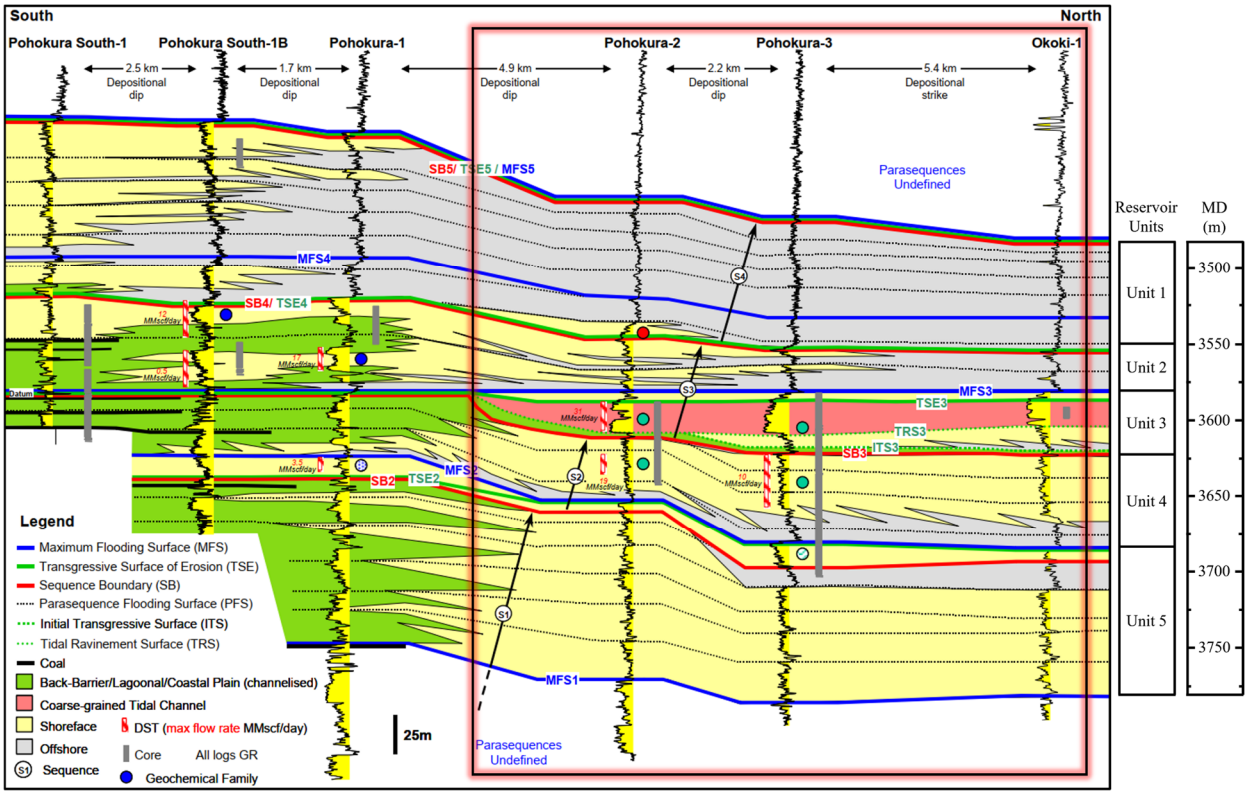


Figure 3.3: Schematic stratigraphic cross-section across the Pohokura Field showing the internal stratigraphic architecture of the Mangahewa Formation. The offshore study interval is outlined by a black box and highlighted in red. Measured depth (MD) values shown alongside the figure are approximate and interpreted from well logs. Table 3.1 provides interpretations of the reservoir units 1-5. Modified after Shell Exploration NZ Ltd. (2012b).

The Paleocene-Eocene Mangahewa Formation, part of the Kapuni Group, constitutes the main reservoir unit in the Pohokura Field and is one of the most economically significant producing intervals in the basin. Lithologically, it comprises interbedded quartz-rich sandstones, siltstones, mudstones, and subordinate coal horizons, deposited in marginal-marine to shallow-marine environments (Higgs, 2001). The reservoir sandstones do not form a laterally homogeneous body. Instead, they were deposited within a marginal-marine system influenced by relative sea-level fluctuations, sediment supply, and local structural controls during the Paleogene (King & Thrasher, 1996). Repeated transgressive-regressive cycles produced

vertically stacked reservoir units bounded by flooding surfaces and sequence boundaries (Figure 3.3), reflecting changes in accommodation space and shoreline migration. Consequently, the reservoir architecture exhibits laterally variable sand-body geometries and internal stratigraphic complexity. Regressive phases typically generated thicker shoreface and tidally influenced sand bodies with comparatively better reservoir properties, whereas transgressive fine-grained intervals contributed to vertical segmentation within the reservoir. Within this stratigraphic framework, the Mangahewa reservoir interval can be subdivided into five units (Units 1-5) reflecting variations in depositional facies and reservoir quality, summarized in Table 3.1.

Table 3.1: Summary of the interpreted reservoir units within the Mangahewa Formation in the Pohokura Field. The reservoir units (Units 1-5) correspond to the stratigraphic intervals illustrated in Figure 3.3.

Reservoir Units	Stratigraphic Position	Dominant Facies	Depositional Environment	Reservoir Quality
Unit 1	Lower Turi	Marine heterolithic deposits (mudstone-siltstone with minor very fine sandstone)	Marine flooding to offshore transition	Poor reservoir / baffle
Unit 2	Upper Mangahewa	Shoreface sandstone bodies	Prograding shoreface system	Moderate reservoir
Unit 3	Central Mangahewa	Coarse-grained, channelized sandstone complexes	High-energy tidal channel-bar system	High-quality reservoir
Unit 4	Central Mangahewa	Lagoonal-back-barrier heterolithic deposits (interbedded sandstone, mudstone and coal)	Back-barrier lagoonal transition	Heterogeneous reservoir / internal baffle
Unit 5	Lower Mangahewa	Heterolithic sandstones with local coal-bearing intervals	Transgressive offshore-shoreface transition	Heterogeneous reservoir / internal baffle

The latest Eocene to earliest Oligocene Turi Formation, which conformably overlies the Mangahewa Formation, consists predominantly of marine mudstones and siltstones and functions as a regional seal across the basin (Robinson & King, 1988; King & Thrasher, 1996). The fine-grained and low-permeability character of the Turi Formation provides effective sealing capacity for hydrocarbons accumulated within the underlying Mangahewa reservoir.

From a petrophysical perspective, the sandstones of the Mangahewa Formation exhibit low-to-moderate porosity characteristics, largely influenced by deep burial and diagenetic processes. Integrated petrophysical and rock-physics analyses conducted by Hossain and Rahman (2025), based on well-log data from depths of 3200-4000 m, indicate that total porosity ranges between 17-21%, while effective porosity varies between 8-19%. Permeability values are strongly controlled by grain-size distribution, clay content, and diagenetic processes. Consistently, the integrated reservoir characterization study carried out by Radwan et al. (2023) reports an average effective porosity of approximately 11.4% for the Mangahewa Formation and highlights significant variability in net pay thickness between wells. These petrophysical characteristics indicate that the Mangahewa reservoir exhibits moderate storage capacity but significant spatial variability, conditions that are expected to influence the distribution of elastic properties and associated seismic responses within the reservoir interval.

Reservoir heterogeneity within the Pohokura Field is multi-scale in nature. At the structural scale, inversion-related faulting produces structural segmentation and localized reservoir compartmentalization through reactivation of earlier rift-related faults (King & Thrasher, 1996). At the stratigraphic scale, cyclic stacking of shoreface, estuarine, and marginal-marine sand bodies generate vertical heterogeneity between reservoir units, reflecting variations in accommodation and depositional processes during the Paleogene (Higgs, 2001). At the facies

scale, lateral transitions from channelized coarse-grained tidal deposits to more distal mud-rich intervals create significant permeability contrasts (Radwan et al., 2023). At the diagenetic scale, spatially variable cementation and dissolution further modify pore structure and elastic properties, as reflected in rock-physics analyses and seismic responses (Hossain & Rahman, 2025).

In summary, the Pohokura Field represents a Paleocene-Eocene marginal-marine sandstone reservoir system that was structurally inverted and segmented during Miocene compressional deformation. The stratigraphic succession comprises gas-productive Mangahewa sandstones at the base, overlain by the regionally sealing Turi Formation, and capped by more open-marine carbonate and mudstone units, clearly illustrating the reservoir-seal-overburden relationship of the petroleum system. As a result of this tectono-stratigraphic evolution, the field constitutes a structurally controlled gas-condensate reservoir characterized by significant heterogeneity.

4. Methods and Workflow

A comparative analysis workflow (Figure 4.1) with two scenarios has been designed to investigate the effects of including inversion products into multi-attribute analysis. In the first scenario, multi-attribute analysis was performed using only attributes derived from seismic amplitude data, without incorporating inversion products. In the second scenario, all parameters and the input attribute set were kept constant, and inversion products were added as additional inputs.

The workflow begins with preliminary processing steps, including data preparation, seismic interpretation, and cropping the datasets to the study interval. In the attribute-only case, spectral balancing was applied to the full-stack seismic data, followed by seismic attribute generation, and the final input dataset was constructed based on correlation matrix analysis. For the inversion-included case, the optimal dataset was determined through a series of tests conducted on candidate input combinations. After the preparation of the input datasets for both scenarios, multi-attribute analysis was performed using PCA and K-means clustering. In the final stage, the results of the two scenarios were systematically compared and evaluated.

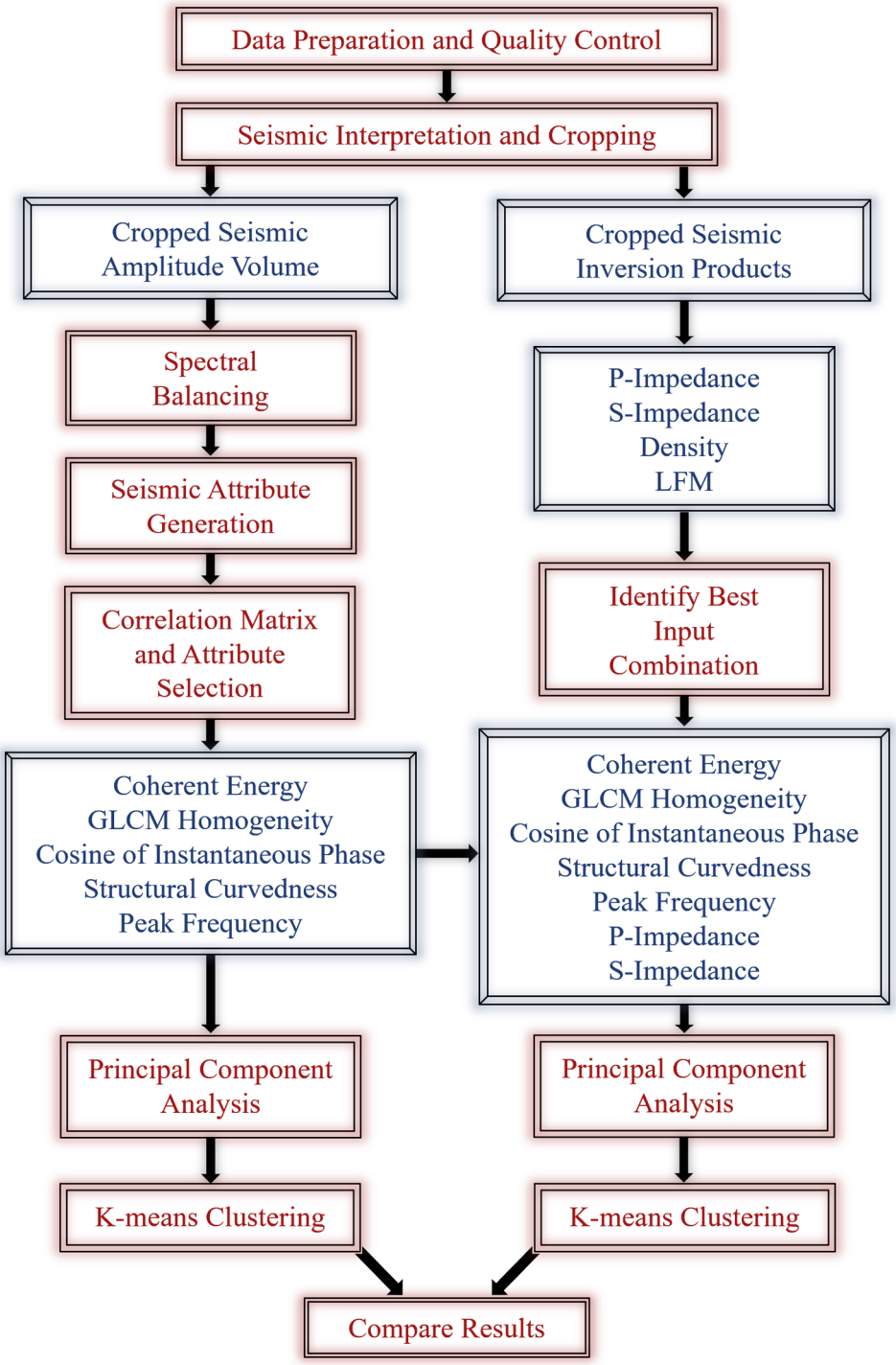


Figure 4.1: Generalized workflow illustrating the comparative multi-attribute scenarios: multi-attribute analysis based solely on seismic amplitude data and multi-attribute analysis incorporating seismic inversion products alongside other attributes. Red boxes represent processing steps, whereas blue boxes indicate 3D data volumes used in the workflow.

In the initial preparation phase, a data inventory was compiled, including seismic amplitude volumes, inversion products, and well logs. All datasets subsequently underwent systematic quality control. Taking the seismic amplitude volumes as a reference, inversion volumes were aligned in terms of survey geometry, bin grid, and trace headers. Well logs were calibrated to the full-stack seismic volume using check shot data and deviation surveys, and formation tops (Table 4.1) were identified and correlated (Figure 4.2).

Table 4.1: Formation tops interpreted from well logs and tied to seismic data. Depths are reported in meters measured depth (MD).

Formation	Pohokura-2 (m)	Pohokura-3 (m)	Okoki-1 (m)
Turi Top	3254.9	3225.9	3568.9
Mangahewa Top	3532.6	3515	3851.8
Tidal Channel Top	3580.8	3550.8	3878.2
Tidal Channel Base	3603.1	3574.9	3915.2
Basal Mangahewa Top	3648.4	3644.2	3980
Mangahewa Base	3811.8	3807.5	4153.5

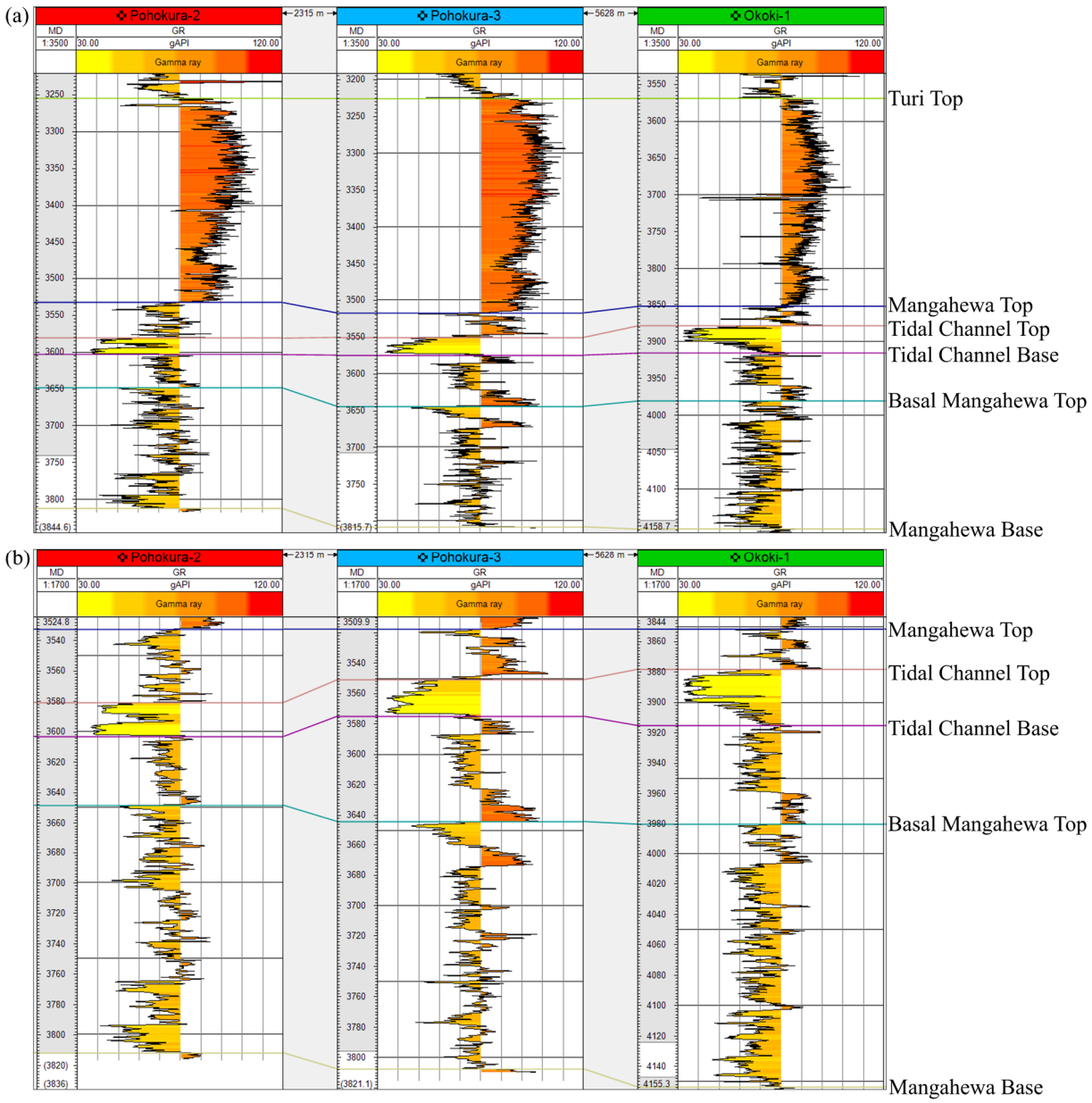


Figure 4.2: Well log correlation panel showing gamma ray (GR) logs and interpreted stratigraphic horizons across the study wells: **(a)** panel flattened on the Turi Top, illustrating the interval from the Turi Formation to the Mangahewa Base **(b)** panel flattened on the Mangahewa Top, providing a detailed view of the Mangahewa reservoir section.

Following data preparation, seismic interpretation was conducted to establish the stratigraphic framework of the study area. Reflector continuity and lateral correlations were evaluated to delineate the upper and lower boundaries of the seismic reflection package (Figure 4.3) containing the reservoir. Using these interpreted horizons, structural maps were generated, and all volumes were cropped to the 3200-4800 m interval to include the reservoir reflection package together with the surrounding stratigraphic units. This broader interval was intentionally selected to enable a comparative evaluation of the responses of the two scenarios across different geological settings. Rather than restricting the analysis to the reservoir area, an interval encompassing the entire anticlinal structure was defined to incorporate all relevant geological and structural elements into the comparative analysis.

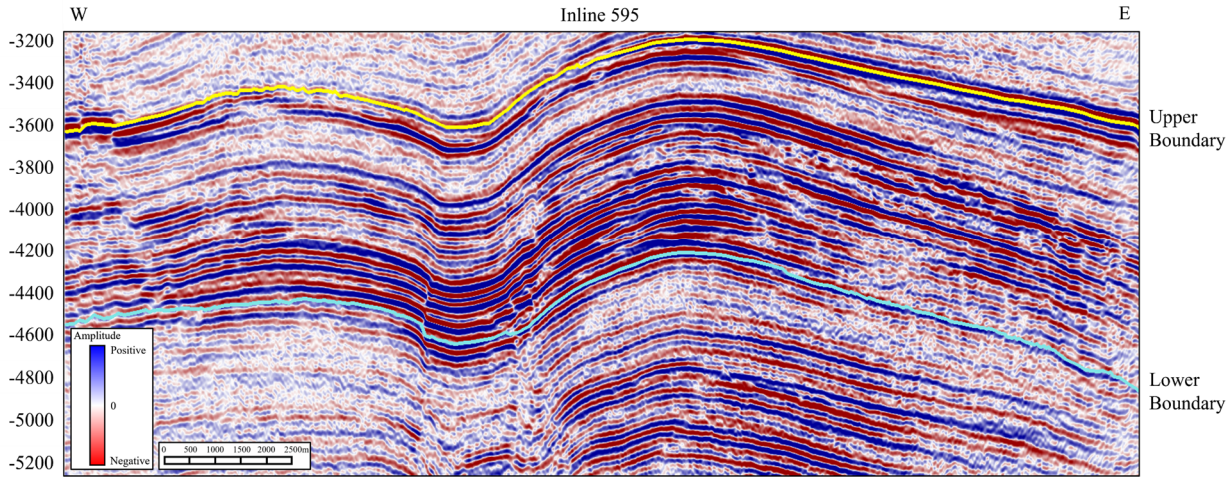


Figure 4.3: Cropped seismic section (Inline 595) showing the interpreted upper and lower boundaries define the analysis interval.

To generate seismic attributes and perform multi-attribute analysis with machine-learning algorithms, the cropped seismic amplitude volumes, inversion products and the interpreted top and base horizons were converted to AASPI format and imported into the workstation.

Prior to attribute generation, spectral balancing was applied to the full-stack seismic volume to eliminate imbalances in the frequency spectrum and ensure that the seismic signal had a homogeneous frequency content. As a result, high-frequency components are enhanced, leading to improvements in reflector continuity and vertical resolution. The effects of this process are illustrated in both seismic section and amplitude spectrum in Appendix A.

Volumetric dip components (inline and crossline dip) represent a fundamental input for the calculation of structurally oriented attributes and for multi-attribute analysis workflows. The accurate and reliable estimation of these components is therefore critical for the quality of the attribute volumes and for the overall performance of multi-attribute analysis. Dip components are typically calculated from the seismic volume used for attribute generation. However, in this workflow, the potential of impedance volumes to improve dip estimation was considered. The impedance volume exhibits a broader effective bandwidth, improved phase stability, and reduced noise compared to the full-stack amplitude data. For this reason, it was expected to provide more reliable dip estimates. Accordingly, dip was computed from both the full-stack amplitude volume and the P-impedance volume to evaluate the influence of the input volume on dip estimation. A comparison of the results indicated that the dip calculated from the impedance volume produced structurally smoother and more geologically consistent trends. Consequently, the impedance-derived dip was selected and used for subsequent attribute calculations and multi-attribute analyses. Although the dip components were derived from the P-impedance volume in both scenarios, this choice was made deliberately to ensure that differences in the results arise solely from the inclusion of inversion products in the input dataset, rather than from differences arising from the use of different dip components. A comparison of the dip estimation results is provided in Appendix B.

Using the full-stack amplitude volume together with the dip components derived from the P-impedance volume, a range of seismic attributes were generated. All the attributes calculated are summarized in Table 4.2. The generated attributes represent different characteristics of the seismic signal. Instantaneous attributes describe local signal behavior derived from the analytic signal, providing information on amplitude strength, phase continuity, and frequency variations. Spectral attributes characterize variations in the dominant frequency content of the seismic signal. Similarity attributes quantify reflector continuity between neighboring traces and assist in the identification of potential discontinuities. Structural curvature attributes describe the local geometric deformation of reflector surfaces and are particularly sensitive to subtle structural features. Texture attributes based on the Gray Level Co-occurrence Matrix (GLCM) quantify spatial variations in seismic amplitude patterns. Together, these attributes provide a complementary description of the seismic response and form a diverse attribute dataset for multi-attribute analysis.

Table 4.2: List of seismic attributes generated from the seismic amplitude volume. Attributes selected for multi-attribute analysis are shown in bold.

Attribute Families	Attributes
Instantaneous Attributes	
	Hilbert transform
	Instantaneous phase (and its cosine)
	Instantaneous envelope
	Instantaneous frequency
	Average frequency
	Average bandwidth
	Sweetness
Spectral Attributes	
	Peak magnitude
	Peak frequency
Similarity Attributes	
	Total energy
	Coherent energy
	Energy ratio similarity
	Outer product similarity
	Sobel filter similarity
Structural Curvature Attributes	
	Most positive (K1) curvature
	Most negative (K2) curvature
	Curvedness
GLCM Attributes	
	GLCM Homogeneity
	GLCM Energy
	GLCM Contrast
	GLCM Dissimilarity
	GLCM Entropy
	GLCM Mean
	GLCM Variance
	GLCM Correlation

Multi-attribute analysis benefits from a wide variety of input data to produce results that can collectively represent different characteristics of the subsurface. At the same time, to minimize redundancy and improve interpretability, each attribute selected for the analysis should ideally represent a single physical property or geological feature (Chopra & Marfurt, 2005). Using attributes of similar type may cause a single attribute to dominate the analysis, while other components may primarily reflect computational artifacts rather than meaningful geological information (Chopra & Marfurt, 2014). To this end, one attribute from each class (geometric, texture, shape, spectral, and instantaneous) was selected, ensuring a rich dataset while avoiding redundancy among attributes of the same type.

To determine which attribute from each class should be selected, the Pearson correlation coefficient method was used to evaluate the linear relationships among the generated attributes. The correlation coefficient is calculated by normalizing the covariance between two data series by the product of their standard deviations (Pearson, 1896). For this purpose, each generated attribute volume was first converted into a one-dimensional data vector. To reduce computational cost and establish a common sampling set for all attributes, randomly selected samples from the data volume were used. Based on the length of the shortest data vector, common indices were determined for all attributes, and correlation coefficients were calculated for all attribute pairs using a randomly sampled subset representing approximately 10% of the dataset. These coefficients were then used to construct the correlation matrix (Figure 4.4). The correlation coefficient is a statistical measure that quantifies the strength and direction of the linear relationship between two attributes and ranges between -1 (perfect negative linear relationship) and +1 (perfect positive linear relationship). Since the strength of the linear relationship is more important than its direction for attribute selection, the absolute values of the correlation

coefficients were used. Higher absolute values indicate stronger correlations between attributes. Among the extracted attributes, coherent energy, GLCM homogeneity, cosine of instantaneous phase, curvedness, and peak frequency were chosen as the main attribute set used in both multi-attribute analysis scenarios and are provided in Appendix C.

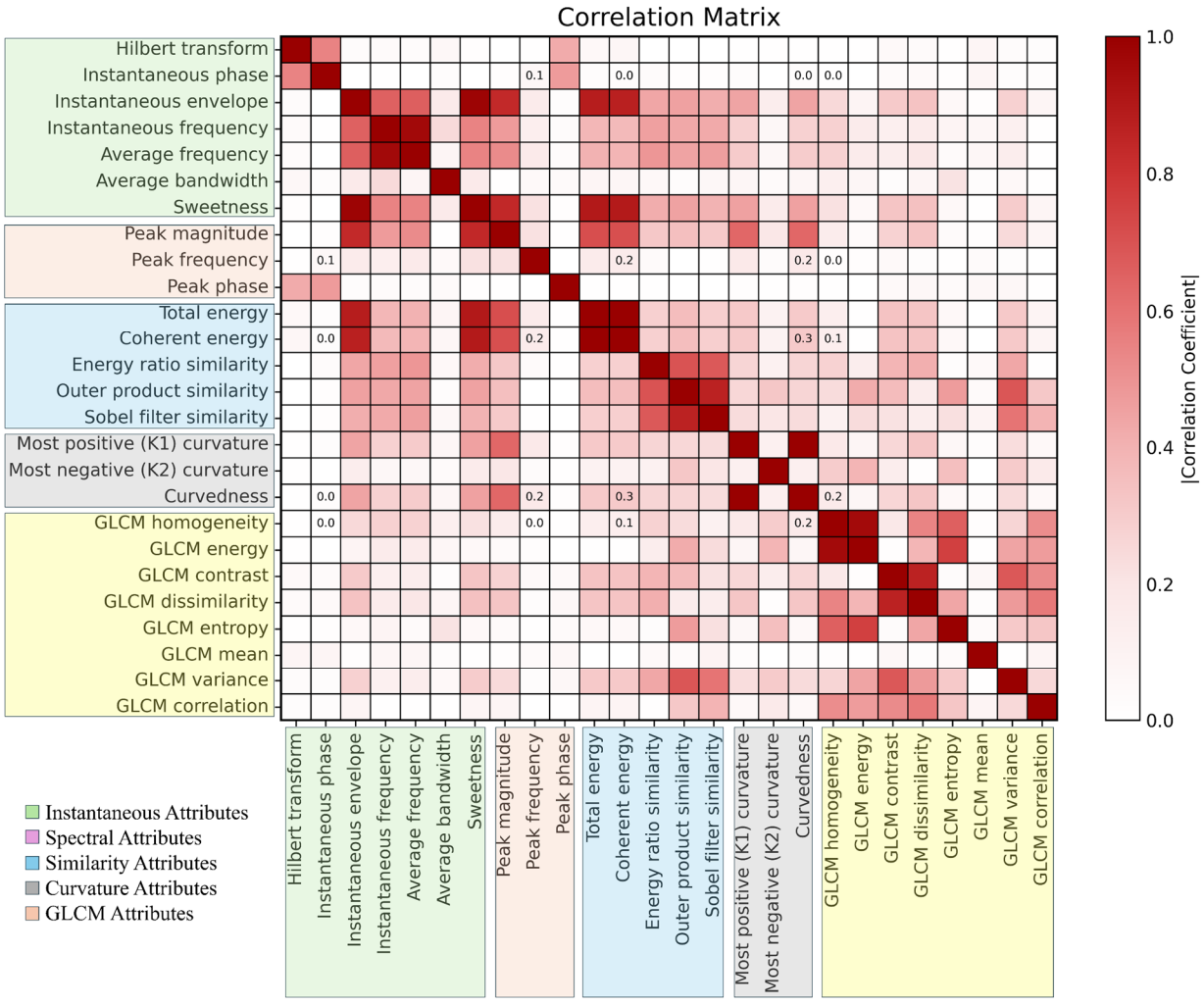


Figure 4.4: Correlation matrix showing the relationships between the generated seismic attributes. Numerical correlation values are displayed only for the attributes selected for the final attribute set.

Pre-stack inversion products can be grouped into two categories: primary products (P-impedance, S-impedance, density) and secondary products (Vp/Vs ratio, Poisson's ratio, Lambda-Rho, Mu-Rho, Net-to-Gross). Primary products represent the fundamental elastic properties estimated directly from the inversion process, whereas secondary products are derived from these primary volumes and therefore represent combinations or transformations of the same underlying physical parameters. Because of this mathematical dependency, including both types simultaneously in the analysis may introduce redundancy in the input dataset. Accordingly, only primary inversion products were considered as candidates in the analysis (P-impedance, S-impedance, and density). The V_P/V_S low-frequency model (LFM) used during inversion was also incorporated as an additional input. Because the LFM is derived solely from well-log impedances and structural interpolation, it does not contain seismic amplitude information. Including it in the inversion-based scenario allows the contribution of the low-frequency trend to be evaluated directly within the multi-attribute analysis.

The inversion products were provided in both absolute and band-limited forms. Absolute inversion volumes incorporate the low-frequency trends derived from well data through the low-frequency model, whereas band-limited volumes represent only the frequency content resolved by the seismic data. For this reason, absolute inversion products were employed.

To determine the most appropriate combination among these candidates (Appendix D), a series of PCA experiments was conducted. In these tests, different input combinations were generated using the inversion products, and the resulting PCA outputs were compared to identify the most optimal configuration. The evaluation was based on the ability of the PCA results to represent key geological features, including structural discontinuities, reservoir boundaries, and heterogeneity zones. The results obtained from a total of 11 different combinations indicate that

the combination of P-impedance and S-impedance provides the most effective and sufficient input dataset, outperforming the alternative combinations tested. Representative examples of the PCA test results are provided in Appendix E.

Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction in multivariate datasets by identifying orthogonal axes that explain the maximum variance in the data. The method was first introduced by Pearson (1901) as a geometric projection problem aimed at determining the line that best fits a multidimensional system of points. In this formulation, Pearson defined a line that minimizes the mean squared distance of multidimensional data points from the line, thereby revealing the dominant directions of variability within the dataset. Hotelling (1933) later reformulated the method within a statistical framework based on eigenvalue-eigenvector decomposition of the covariance matrix, establishing the modern formulation of PCA. Hotelling's work demonstrated that the total variance in a dataset can be maximized along mutually orthogonal directions, which can be identified through the eigenvectors of the covariance matrix.

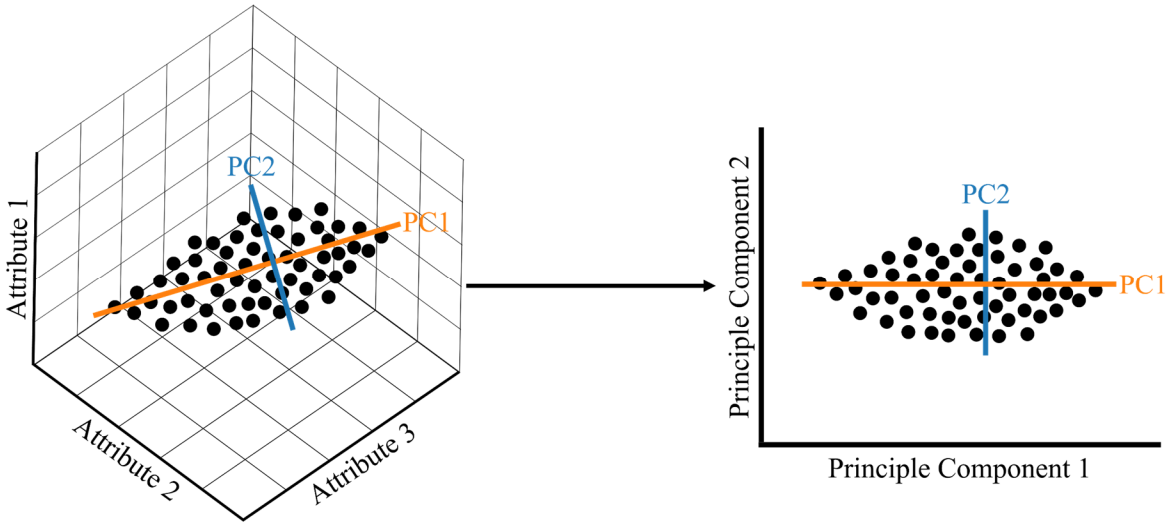


Figure 4.5: Schematic representation of PCA showing the first two principal components.

PCA transforms a dataset composed of correlated attributes into a new coordinate system defined by mutually orthogonal axes. These new axes are referred to as principal components, and each component can be expressed as a linear combination of the original variables. As schematically illustrated in Figure 4.5, the eigenvector corresponding to the largest eigenvalue represents the first principal component (PC1) and defines the direction of maximum variance in the dataset. The second principal component (PC2) is defined to be orthogonal to PC1 and represents the second direction that explains the largest portion of the remaining variance (Hotelling, 1933). In this way, attributes that initially contain complex and often implicit relationships can be represented through independent components that capture the shared variance structure of the dataset, allowing the dominant patterns of variation in the data space to be expressed in a clearer and more interpretable form.

Multi-attribute analysis is inherently prone to high dimensionality and strong correlation problems. For this reason, it is necessary to transform attribute sets into a lower-dimensional and better-organized representation domain, which can be robustly achieved using PCA. In a comparative study of unsupervised classification methods used in seismic facies analysis, Coléou et al. (2003) demonstrated that PCA provides an effective data reduction approach by transforming high-dimensional seismic datasets into a lower-dimensional representation space while preserving the fundamental patterns of variation within the data. Another common challenge in multi-attribute analysis arises from the fact that different attributes may be sensitive to the same geological features, resulting in strong correlations and redundancy within the dataset. The role of PCA in multi-attribute analysis is not limited to dimensionality reduction; the method can also be used to evaluate attribute datasets and identify suitable combinations. Roden et al. (2015) reported that PCA reduces redundancy within the attribute data space and produces

a more organized feature space for the SOM algorithm, thereby improving the reliability of geological pattern recognition. Similarly, Kim et al. (2019) showed that PCA can assist in determining more meaningful attribute combinations in seismic facies classification studies. La Marca (2020) demonstrated that PCA applied to a dataset containing 28 attributes from deep-marine environments reduced the dimensionality of the dataset and enabled the selection of nine representative attributes, while also revealing the importance of certain attributes that had previously been overlooked in user-driven analyses.

To observe the effects of including inversion products in the comparison scenarios, it is necessary to examine how the selected inputs influence the results. This problem essentially involves identifying the fundamental relationships among the input variables and addressing a dimensionality reduction task, for which PCA is a well-established method. When new variables are incorporated into the attribute set, PCA makes it possible to observe the additional variance introduced by these variables within the data structure, allowing the effects of different attribute combinations to be evaluated directly. In this respect, PCA provides an appropriate multi-attribute analysis framework that satisfies the methodological requirements.

Accordingly, the comparative multi-attribute analysis was carried out using PCA. In the first stage, values sampled from the selected attribute volumes were combined to construct a multidimensional feature matrix in which rows represent data points and columns represent the different attributes. Subsequently, z-score normalization was applied to the dataset in order to make attributes defined at different scales and units comparable. In this procedure, the mean value was subtracted from each attribute, and the result was divided by the standard deviation to standardize the data; otherwise, the covariance matrix may be influenced by unit differences among attributes (Chopra & Marfurt, 2014). Following normalization, the covariance matrix was

calculated, and eigenvalue-eigenvector decomposition was performed to determine the principal component axes of the data space. The eigenvectors represent orthogonal directions of variation in the data space, while the corresponding eigenvalues indicate the amount of variance explained by each component. The first two principal components (PC1 and PC2) were used, as the initial components of PCA are known to represent the dominant signal variations in the dataset, whereas subsequent components may often correspond to uncorrelated noise components (Zhao et al., 2015). In addition, the use of the first two principal components allows the dominant variation structure of the data space to be directly visualized in a two-dimensional plane and enables the distribution of the data to be represented through PC1-PC2 crossplots. The attribute dataset was projected onto these two orthogonal axes to obtain a two-dimensional feature space for the analysis. In the following step, the PCA model was applied to the entire dataset, and PC1 and PC2 values were calculated for each data point; these values were then generated as new volumes. In the final stage, the resulting PC1 and PC2 volumes were crossplotted. Two-dimensional histogram approach was used to represent the distribution of the data points, and the data space was divided into 64 × 64 color bins along the PC1 and PC2 axes. The resulting distribution was visualized using 4096 colors, allowing the density distribution of the data points in the principal component plane to be clearly represented. The same analysis workflow was applied separately to the input datasets constructed for the two scenarios.

K-means clustering is an unsupervised machine learning method that divides multivariate datasets into a predefined number of clusters (k) based on similarity in the feature space. In this approach, the data space is divided into regions, each represented by a cluster centroid (i.e., the mean of the data points assigned to the cluster), and each data point is assigned to the nearest centroid. To define within-cluster variance, each cluster centroid is expressed as the mean of the

data points assigned to it, and the clustering process is based on minimizing the sum of squared distances between data points and their respective centroids. This optimization is achieved through a least-squares-based iterative procedure in which cluster assignments and centroids are updated sequentially to minimize the total squared error (Lloyd, 1982), following the original formulation introduced by MacQueen (1967). This process results in the dataset being partitioned into internally consistent clusters that represent the dominant similarity structure of the data.

Multi-attribute analysis involves high-dimensional attributes that are often sensitive to similar geological features, resulting in strong correlations and implicit relationships within the dataset. This complexity makes the direct interpretation of the data structure challenging. Unsupervised clustering methods such as K-means reveal similarity patterns in the attribute space by grouping data points with similar attribute combinations into the same clusters. In this context, the resulting clusters can be interpreted as discrete segments representing different seismic facies, allowing the spatial distribution of variability within the data space to be more clearly identified.

K-means is not only a clustering method that reveals similar structures in the data space, but also a suitable analytical tool for evaluating the effects of different attribute combinations on data organization. Troccoli et al. (2022) reported that K-means can effectively partition multi-attribute seismic datasets into discrete facies groups and, when used in combination with PCA, facilitates the organization of the data within a more structured representational space. This integration is particularly important because the principal components obtained from PCA are mathematical linear combinations that do not always have a direct geological correspondence (Chopra & Marfurt, 2014; La Marca, 2020) and therefore require clustering-based segmentation to be translated into interpretable data patterns.

K-means clustering algorithm was employed as a complementary method to the PCA analysis. While PCA reveals the contributions of different attribute combinations to the data structure in terms of variance, K-means transforms this variability into discrete data segments, enabling a direct evaluation of the effects of inversion products on facies distribution within the dataset. After comparing the contributions of different attribute and inversion product combinations at the variance level through PCA, the K-means algorithm was applied separately to the input datasets constructed for each scenario to investigate the impact of inversion products on data segmentation. This approach allows for a direct comparison of facies distribution across different scenarios.

Accordingly, K-means clustering analysis was performed using the input datasets constructed for each scenario. In the initial stage, to assess the sensitivity of the clustering results to the number of clusters, the analysis was performed for cluster numbers ranging from 2 to 12. In addition, a fixed random seed (1000) was used to control the initial positions of the cluster centroids and ensure reproducibility of the results. During this parameter analysis, the sum of squared distances to the cluster centroids was calculated for different cluster numbers, and the results were evaluated using distance curves. Elbow method plots showing the variation of within-cluster sum of squares (WCSS) as a function of the number of clusters, used to determine the optimal cluster number. The resulting curves (Appendix F) showed that the rate of error reduction slowed beyond a certain point as the number of clusters increased, and this inflection point was interpreted as the optimal number of clusters. For both input sets, the optimal cluster number representing the data structure most effectively was determined as $k = 7$. Following the parameter analysis, the final K-means model was constructed using the selected cluster number and applied to the entire dataset. As a result, each data point was assigned to a corresponding

cluster, and the data space was segmented into discrete regions. The clustering results were visualized using the same crossplot approach applied in the PCA analysis, allowing for a direct comparison of the results.

5. Results

Two different input datasets are compared to evaluate the impact of incorporating inversion products into multi-attribute analysis. The first dataset consists of attributes derived solely from seismic amplitude data, whereas the second dataset includes these attributes together with inversion-derived volumes, specifically P-impedance and S-impedance. To ensure that any observed differences arise exclusively from the input data composition, all processing steps and parameters within the workflow were kept constant. In addition, a consistent crossplot-based visualization framework was applied to both scenarios to enable a direct and unbiased comparison of the results.

The comparison focuses on evaluating how the results from the two scenarios differ in their ability to represent overall geological trends, resolve intra-formational heterogeneity zones, delineate structural discontinuities, track reservoir boundaries, and distinguish between facies. The effects of these differences on the interpretability and geological relevance of the results are analyzed, and the findings are comprehensively evaluated in terms of their contributions to reservoir characterization.

The percentage distribution of the first two principal components for both scenarios is shown in Figure 5.1, illustrating the relative contribution of each attribute to the variance defining these components and how different attribute combinations organize the data space. In addition, the total variance explained by each principal component is also presented.

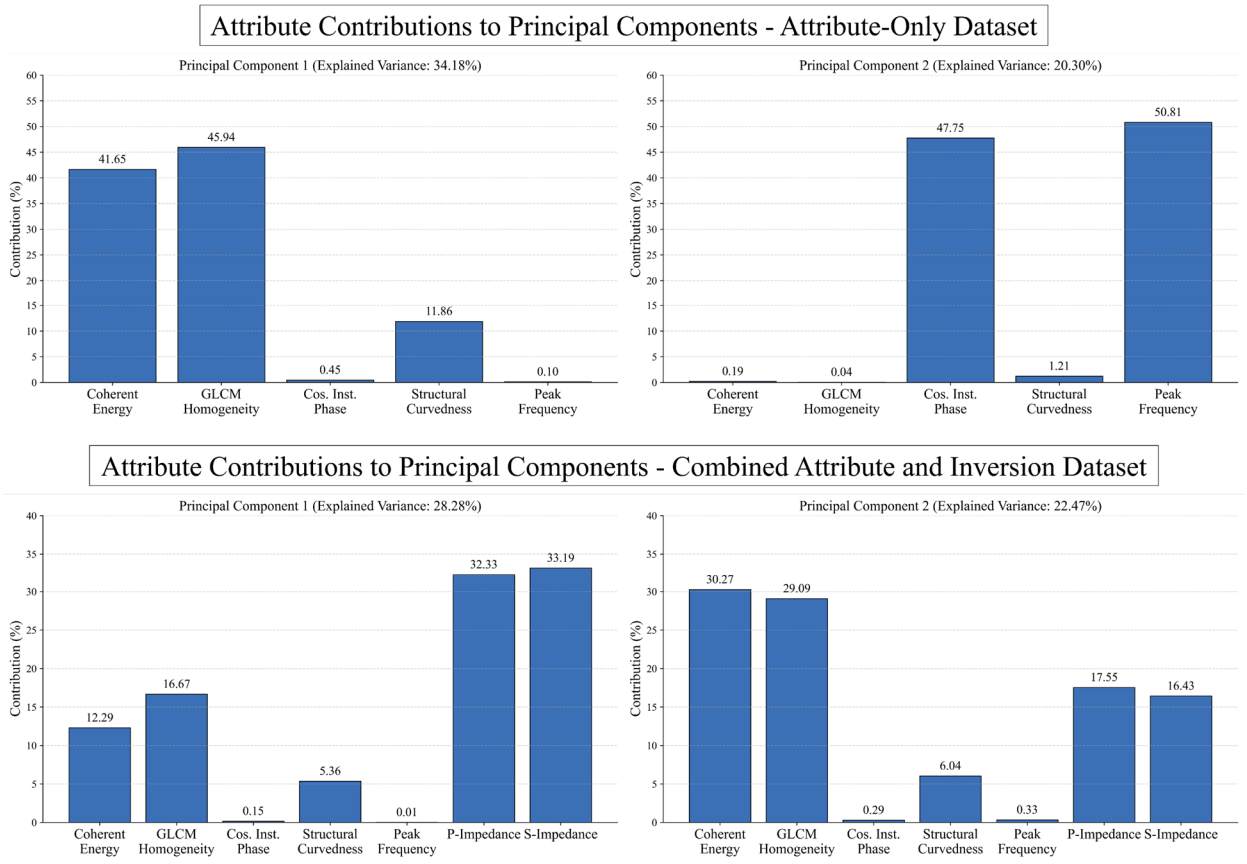


Figure 5.1: Percentage contributions of individual seismic attributes to the first two principal components (PC1 and PC2) derived from PCA for the evaluated input datasets.

The cumulative variance explained by the first two principal components is 54.48% for the attribute-only dataset and 50.75% for the combined dataset. These values indicate the proportion of total variance that can be represented within a reduced-dimensional subspace. The similar levels of explained variance observed in both scenarios suggest that the data structure can be meaningfully summarized using two components; however, this representation is achieved through different distributions of input attributes.

In the attribute-only dataset, the first principal component (PC1) is primarily defined by GLCM Homogeneity (45.94%) and Coherent Energy (41.65%), with Structural Curvature contributing to a lesser extent (11.86%). Considering the dominant contributing attributes, PC1

in this scenario represents textural homogeneity, reflector continuity, and reflector geometry. On the other hand, the second principal component (PC2) is largely controlled by Peak Frequency (50.81%) and Cosine of Instantaneous Phase (47.75%). Based on the dominant contributions, PC2 represents frequency- and phase-related variations that are largely independent of amplitude effects. For this scenario, the derived principal components provide representation across all seismic attributes within the input set.

With the inclusion of inversion products, a clear redistribution of input contributions is observed. In this scenario, the first principal component (PC1) is dominated by P-impedance (32.33%) and S-impedance (33.19%). The strong shared variance between these impedance inputs results in the introduction of a new dominant trend within the dataset, leading to their control over PC1. Although the contributions of Coherent Energy (12.29%), GLCM Homogeneity (16.67%), and Structural Curvature (5.36%) are reduced, their relative ranking remains consistent with the previous scenario. This indicates that these seismic attributes continue to define the structural framework of the data. However, elastic properties introduced by impedance inputs take a primary role in defining the main direction of variance in this component. The information introduced through impedance does not replace the existing structure but instead contributes as a dominant additional variance component that is consistent with and reinforces the existing variance structure. In contrast, the second principal component (PC2) is primarily controlled by Coherent Energy (30.27%) and GLCM Homogeneity (29.09%), with additional contributions from P-impedance (17.55%) and S-impedance (16.43%). This structure indicates a more balanced representation between seismic continuity, textural attributes, and elastic properties. Although the principal components in this scenario adequately represent the overall variance of the dataset, it is observed that Peak Frequency and Cosine of

Instantaneous Phase do not contribute to the first two principal components. This behavior is associated with the expansion of the feature space and the increase in total variance resulting from the inclusion of P- and S-impedance volumes. As the number of attributes and the information content increase, the variance is distributed across a larger number of components, limiting the representation of lower-contribution attributes such as those related to frequency and phase within the first two principal components.

The redistribution of attribute contributions observed between PC1 and PC2 is consistent with the orthogonal decomposition nature of PCA, where PC2 captures the variance not explained by PC1. In this context, attributes that show high contributions within the same component exhibit similar variance behavior and are therefore likely to be correlated, although PCA does not provide direct information on the magnitude or direction of this relationship. This variance structure can be examined spatially. For this purpose, the PC1 and PC2 component volumes derived for each scenario were crossplotted, where the distribution of data points in PC1-PC2 space was represented using a two-dimensional histogram, and this distribution was subsequently projected back into the three-dimensional domain as a color-encoded volume.

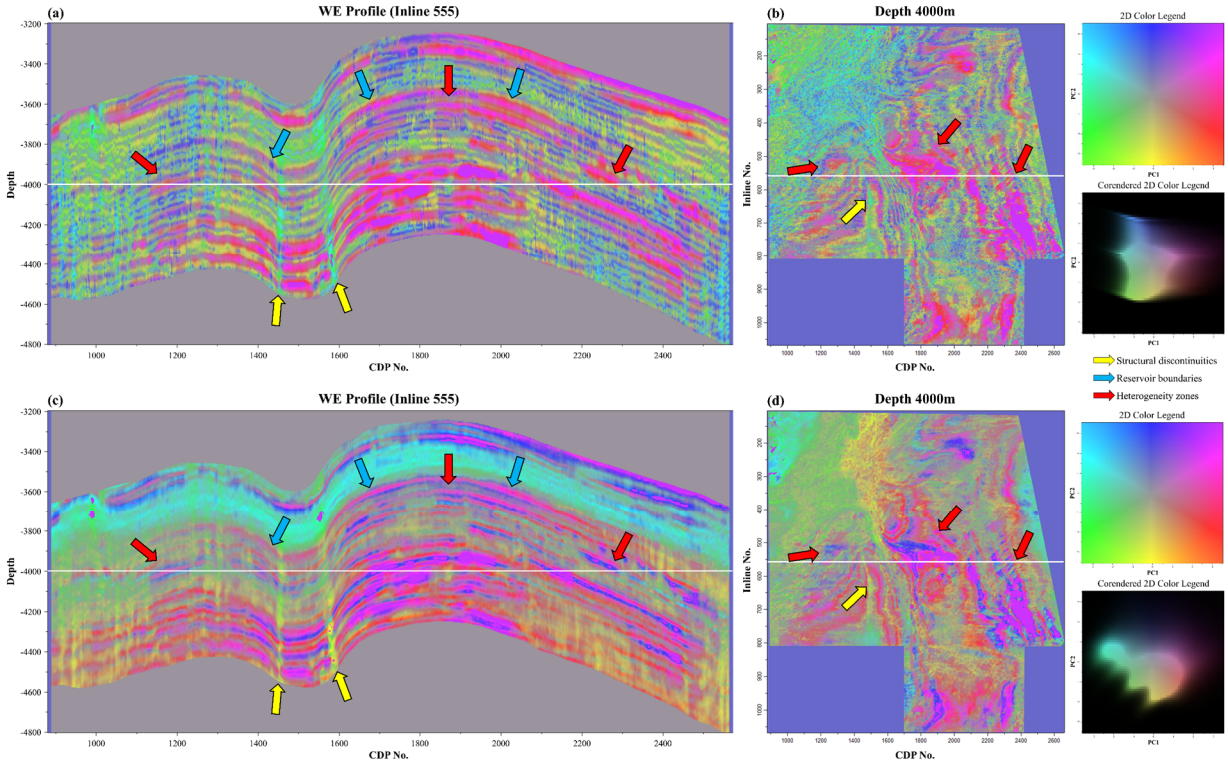


Figure 5.2: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 555) result without inversion products **(b)** Depth slice at 4000m without inversion products **(c)** WE Profile (Inline 555) result with inversion products **(d)** Depth slice at 4000m with inversion products. The observed enhancements are highlighted as follows: structural discontinuities (yellow arrows), reservoir boundaries (blue arrows), and reservoir heterogeneity zones (red arrows).

Figure 5.2 presents the PCA results along the WE profile (Inline 555) and the depth slice at 4000 m, selected for the comparison of the two scenarios. At first glance, the most noticeable difference is in the overall trends and the representation of the color distribution. In the attribute-only scenario, the resulting color distribution exhibits a more heterogeneous character with locally sharp transitions. In contrast, when inversion products are incorporated, color transitions appear smoother and the overall distribution becomes more homogeneous. These findings indicate that results obtained using only seismic attributes largely reflect the seismic amplitude response, whereas the inclusion of inversion products leads to results that are more geologically

meaningful and interpretable. This primary improvement is accompanied by several notable advantages, as well as certain limitations.

First, structural discontinuities are well resolved in both scenarios, although minor differences are evident. The attribute-only PCA results are capable of detecting these structural features at a level comparable to the inversion-integrated case. However, additional vertical linear features that are not directly related to faults or fractures are also observed. This behavior is primarily associated with frequency- and phase-sensitive attributes that retain representation in the attribute-only scenario. While these attributes enhance the visibility of structural discontinuities, they also introduce a higher level of noise in the results.

In contrast, in the inversion-integrated scenario, the suppression of Peak Frequency and Cosine of Instantaneous Phase leads to smoother and less noisy results, although it may reduce the detectability of some subtle structural discontinuities. Despite these differences, both scenarios are able to resolve structural discontinuity zones to a considerable extent, and linear discontinuities appear clearly defined in both the profile and depth slice views. However, in the attribute-only case, distinguishing discontinuities from noise remains more challenging, whereas the inclusion of inversion products improves the clarity and interpretability of these features.

Second, the inclusion of inversion products leads to significant improvements in the representation of geological boundaries, stratigraphic layering, and reservoir boundaries. In particular, the contrast between the marine mudstones of the Lower Turi Formation, which cap the reservoir, and the sandstones of the Mangahewa Formation become markedly clearer when inversion products are included. The mudstone interval above the reservoir is distinctly separated and represented by a characteristic light cyan color in the inversion-integrated results. In contrast, this transition between the two formations is not clearly resolved in the attribute-only

scenario. This improvement reflects the direct impact of incorporating low-frequency, large-scale trends into the analysis and represents a critical enhancement for reservoir characterization. In addition, vertical resolution is noticeably improved with the inclusion of inversion products. Boundaries that appear as chaotic color clusters in the attribute-only results become more clearly defined, exhibiting higher detail and more coherent, continuous contours in the combined dataset. In particular, color transitions along reflector continuity become more organized, and reservoir boundaries are more distinctly delineated.

Third, beyond the identification of inter-formational zones, the inclusion of inversion products enables higher-resolution detection of both vertical and lateral variations within formations. Features associated with the multi-scale heterogeneity of the Pohokura field become more clearly identifiable in the inversion-integrated scenario. The additional information introduced by impedance volumes reduces the apparent uniformity in color distribution associated with lithological variations, leading to improved differentiation and more accurate identification of reservoir heterogeneity zones. Furthermore, the high-frequency, scattered color variations observed in the attribute-only scenario are transformed into a more organized and spatially consistent pattern in the combined dataset. As a result, structural segmentation and localized reservoir compartmentalization associated with faulting can be identified with higher resolution, leading to a more accurate characterization of reservoir heterogeneity.

The two-dimensional color palettes associated with the crossplot diagrams provide further insight into the organization of the data in attribute space. In the attribute-only scenario, the color distribution is concentrated in localized clusters representing attribute groups largely governed by seismic character. The distribution observed along the PC2 axis, which is sensitive to frequency and phase variations of the seismic amplitude data, indicates that the attribute-only

scenario is strongly controlled by seismic response characteristics. In contrast, when inversion products are included, the color distribution spreads over a broader region across both principal components and exhibits a more organized, trend-aligned structure. This indicates that the distribution in crossplot space reflects not only amplitude-driven variations, but also more systematic trends associated with elastic properties. This reorganization of the color palette suggests that the incorporation of inversion products results in a more structured and better-separated attribute space.

Overall, these results indicate that the inclusion of inversion products leads to clearer reservoir boundaries and a more accurate representation of heterogeneity, thereby improving the overall quality of reservoir characterization. However, it should also be noted that the analysis based solely on seismic amplitude-derived attributes still provides a considerable level of accuracy.

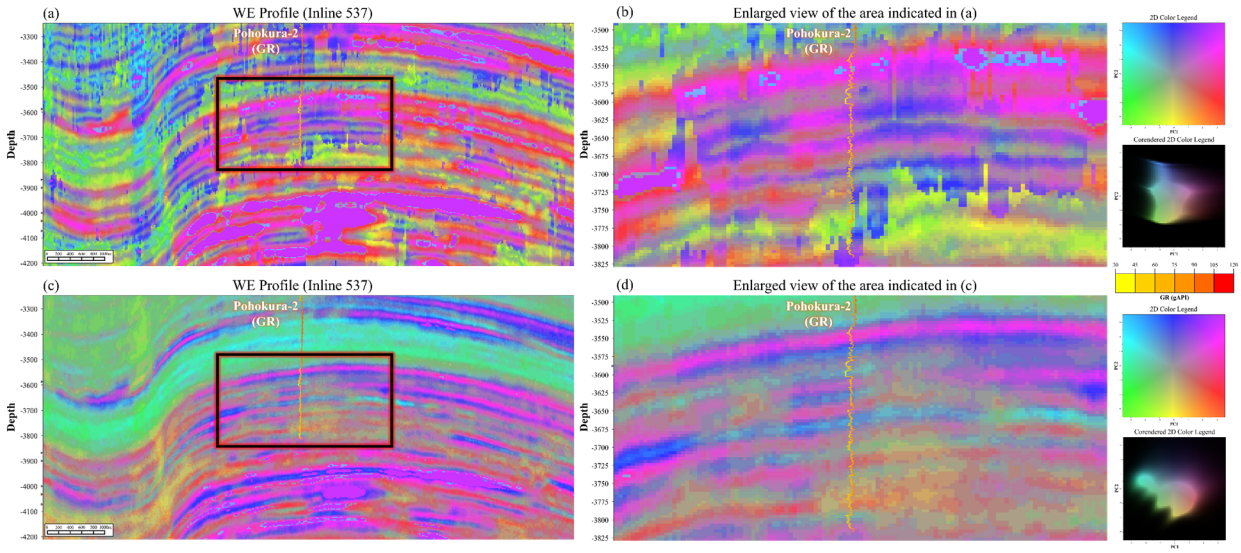


Figure 5.3: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 537) result without inversion products **(b)** enlarged view of the highlighted interval without inversion products **(c)** WE Profile (Inline 537) result with inversion products **(d)** enlarged view of the highlighted interval with inversion products. The gamma ray (GR) log of the Pohokura-2 well is included for calibration.

To validate the analysis results and comparatively assess their impact on heterogeneity zones within the reservoir interval, the PCA results were integrated with the seismic-well tie. For both scenarios, the results were tied to the GR log of the Pohokura-2 well along and evaluated using the zoomed view focused on the reservoir interval (Figure 5.3).

A comparison of the PCA results reveals that the most noticeable difference is the higher level of noise present in the attribute-based scenario. This noise negatively affects reflector continuity, lithological discrimination, and the identification of layer continuity, particularly within the target reservoir interval. Part of this noise is interpreted to originate from anomalies and irregularities associated with gas effects, while another component is associated with scattering behavior introduced by the mathematical transformations applied during seismic attribute computation. Although such effects may contain information related to gas-bearing

zones, they complicate the identification of stratigraphic variations and reduce overall interpretability. When inversion products are incorporated, these effects are significantly reduced, resulting in a more regular reflector geometry and improved interpretability of intra-reservoir boundaries.

The contrast between cap rock and reservoir, which is well represented by the gamma ray log, is identifiable in both cases; however, this contrast becomes sharper with the inclusion of inversion products. The associated gain in resolution enables better detection of vertical variations. The transition from the mudstones of the Turi Formation to the Mangahewa reservoir is more difficult to interpret in the attribute-based results due to the influence of noise. In contrast, this interface can be tracked more reliably when inversion-derived information is included. Additionally, the more homogeneous color distribution provides a representation that is more consistent with geological units and enables higher-resolution discrimination between formations.

The identification of different facies within the Mangahewa Formation is also enhanced with the incorporation of inversion products. The tidal channel, characterized by a low GR response, can be clearly distinguished in the inversion-included scenario; however, this interval cannot be identified in the attribute-only scenario. Within heterolithic deposits, productive intervals can be differentiated from interbedded mudstones exhibiting higher GR responses, with significantly improved vertical resolution following the use of inversion products. In addition, the lateral continuity of reservoir facies can be tracked more effectively. In contrast, the distinction between the intervals defining the reservoir is not sufficiently clear in the attribute-only scenario to allow reliable characterization. These improvements indicate that the multi-

attribute workflow incorporating inversion products enables the detection of heterogeneity not only at the stratigraphic scale but also at the facies scale.

Due to the ability of inversion-based data to provide a representation more consistent with lithological variations, their inclusion facilitates a transition from a reflectivity- and seismic signal-driven representation toward a more geology-driven model based on layer properties. This transition contributes to a clearer definition of reservoir trends and enhances the visibility of heterogeneity zones.

Alongside these improvements, smoother transitions are observed along some reflection surfaces due to the suppression of high-frequency variations. The inherently lower-frequency content of inversion-derived volumes leads to partial smoothing of fine-scale lateral variations that may be captured by certain texture-based attributes. Nevertheless, when evaluated with well-log correlations, the reflectors that can be reliably tied to the well indicate that inversion-derived volumes provide a more geologically consistent and higher-resolution representation.

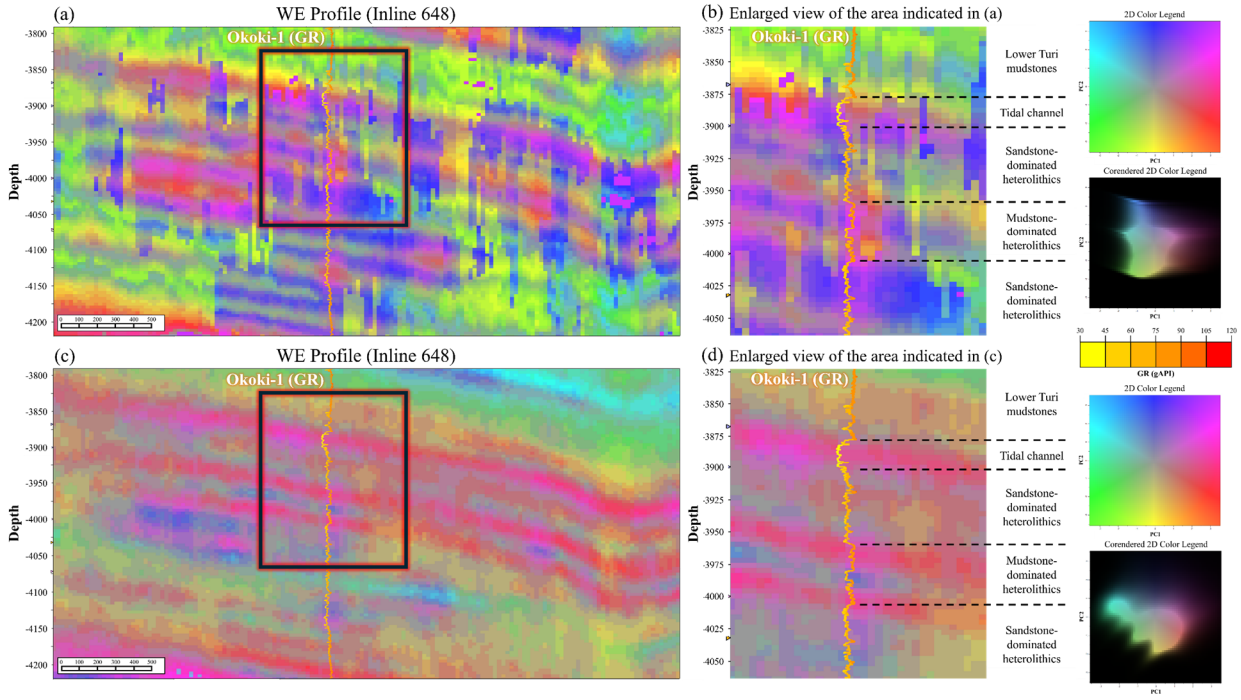


Figure 5.4: Crossplot of two principal components (PC1 vs. PC2) **(a)** WE Profile (Inline 648) result without inversion products **(b)** enlarged view of the highlighted interval without inversion products **(c)** WE Profile (Inline 648) result with inversion products **(d)** enlarged view of the highlighted interval with inversion products. The gamma ray (GR) log of the Okoki-1 well is included for calibration.

The results obtained along Inline 648, visualized to compare the representation of reservoir-scale heterogeneity zones, reveal clear differences in the level at which facies boundaries are resolved (Figure 5.4).

In both scenarios, the transition between the Lower Turi mudstones, represented by high GR values, and the tidal channel facies characterized by a lower GR trend can be successfully identified. However, due to the higher noise level and limitations in vertical resolution, this transition is represented in a more chaotic manner in the attribute-only scenario. In contrast, in the inversion-included scenario, reservoir boundaries and their lateral continuity can be tracked more reliably.

With the inclusion of inversion products, not only the overall geological trends but also the boundaries defining internal variations within each geological interval can be delineated more sharply and represented with a more consistent color separation. The distinction between sandstone-prone and mudstone-prone intervals, as well as the internal interbedding within each, becomes clearly resolved. In particular, the interbedded sandstone interval observed between approximately 3975 and 3990 m within the mudstone-dominated heterolithics can only be detected in the inversion-included scenario, where its lateral continuity can also be consistently tracked. This stands out as one of the clearest examples of the contribution of inversion products to reservoir characterization within the applied multi-attribute workflow and provides direct evidence of resolution enhancement. Sand-prone intervals with potentially good reservoir quality, which are not detected in the attribute-only results, become visible in the PCA outputs when inversion products are incorporated.

Significant differences are also observed in the color-palette representation of facies distributions. In the scenario based only on amplitude-derived attributes, reservoir zones are predominantly represented by dark blue colors, and this concentration corresponds to areas where PC2 takes higher values. In this case, PC2 is defined mainly by Peak Frequency and Cosine of Instantaneous Phase. This component primarily highlights variations in waveform geometry, phase behavior, and spectral content rather than reflector continuity. The dominance of this spectral- and phase-sensitive component may weaken facies clustering by emphasizing certain reflection packages and waveform variations, thereby creating overlap in color representation between reservoir intervals and other reflectors. As a result, local inconsistencies in the color representation of facies are observed in the attribute-only scenario, and the

representation of reservoir zones is partially obscured by the influence of other reflection surfaces.

In the scenario incorporating inversion products, the dominance of impedance inputs leads to the development of a color palette structure that enables the clustering of facies representations, with distinct color clusters observable along both axes. Consequently, the color representations in the inversion-integrated scenario become more consistent with the facies variations observed in the GR log.

Moreover, the local inconsistencies observed in the color representation in both scenarios indicate that the results are still influenced by seismic response behavior. Although this effect is reduced with the inclusion of inversion products, inconsistencies in color representation are still observed. In particular, while the tidal channel is predominantly represented by magenta tones, other sand-prone intervals are associated with different color signatures, and certain mudstone intervals are also represented by similar magenta tones. This overlap reduces the uniqueness of the color-facies relationship and indicates that the color distribution is still partially controlled by seismic response rather than purely geological variability. While the abrupt variations observed in the GR log are more effectively resolved in the inversion-included scenario, their color representation still reflects a response behavior driven by high interlayer contrast.

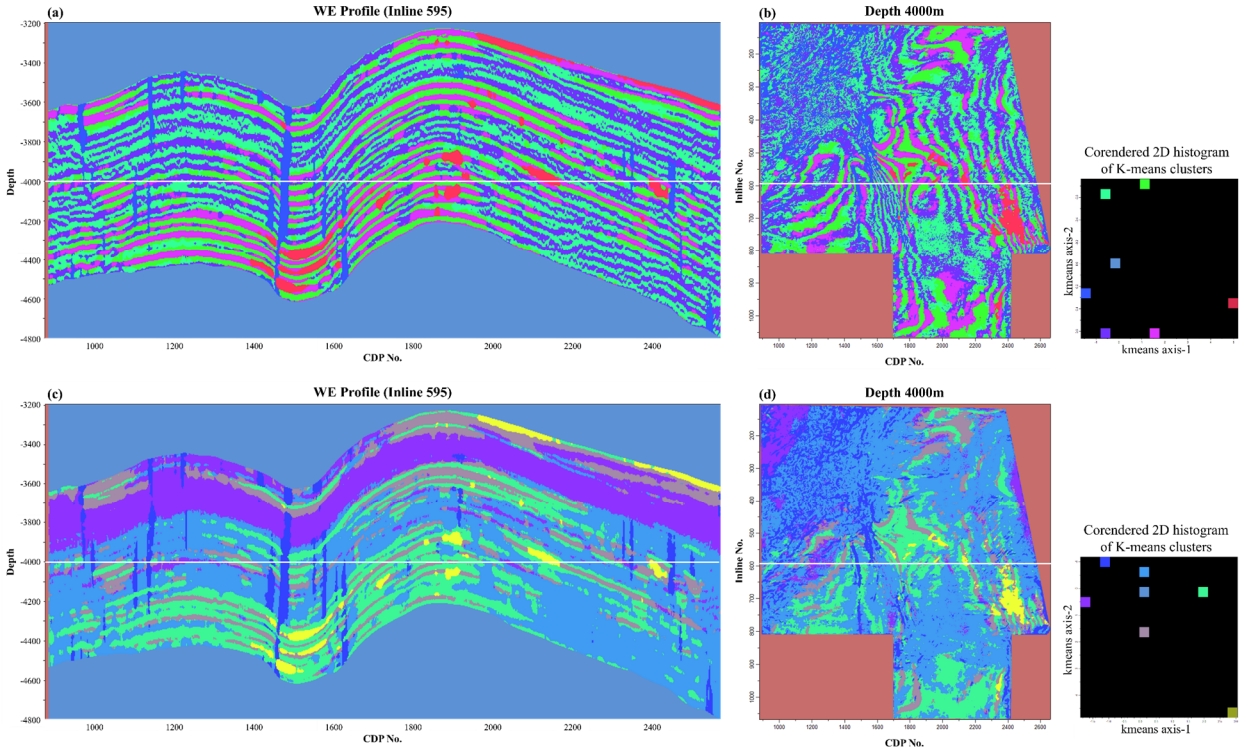


Figure 5.5: Crossplot of two K-means clustering axis **(a)** WE Profile (Inline 595) result without inversion products **(b)** Depth slice at 4000m without inversion products **(c)** WE Profile (Inline 595) result with inversion products **(d)** Depth slice at 4000m inversion products.

The K-means clustering results were visualized in clustering space using a crossplot-based representation (Figure 5.5). The distribution of samples along the K-means axes was expressed as a two-dimensional histogram and subsequently mapped back into the three-dimensional domain as a color-coded volume.

A comparison of the K-means clustering results reveals, similar to the PCA outcomes, that the primary distinction lies in the resolution of general geological trends, particularly the separation between seal and reservoir intervals. In the inversion-integrated scenario, this contrast resolved with higher confidence, with the mudstones of the Turi Formation being represented by a distinct dark-purple cluster, effectively separating them from the sandstone intervals of the Mangahewa Formation. In this case, clusters that are more meaningfully separated in terms of

lithological characteristics are arranged in a manner consistent with the overall geological framework. In contrast, the clusters obtained from the attribute-based scenario display less pronounced contrasts within the same zone. The clusters representing the mudstone interval in this scenario also partially overlap with the reservoir interval, making the boundary between seal and reservoir less distinct and more difficult to interpret.

In addition, the K-means results derived solely from amplitude-based attributes exhibit vertical streaking patterns along horizons. This effect weakens the representation of stratigraphic continuity. Such behavior is consistent with the strong noise influence observed in the PCA results, indicating that the impact of noise persists in the clustering outcomes as well. When inversion products are incorporated, this vertical smearing effect becomes smoother, and the boundaries between layers are represented with improved lateral continuity, allowing geological trends to be more clearly differentiated. Furthermore, horizons appear more parallel and consistent with the geological structure, resulting in a stronger representation of stratigraphic continuity. This contrast suggests that the attribute-based scenario is primarily controlled by waveform characteristics, noise, and amplitude variations, whereas the inversion-integrated scenario moves beyond seismic response toward a representation closer to elastic facies. As a result, overlap in crossplot space is reduced, lateral continuity and stratigraphic consistency are enhanced, and the separation between reservoir and non-reservoir intervals becomes clearer. Clusters also show improved correspondence with intra-formational heterogeneity zones and reservoir quality trends, indicating a significant contribution to reservoir characterization.

Consistent with the PCA results, the inclusion of inversion products in the K-means analysis leads to more geologically consistent representations of heterogeneity zones. Both inter-formational contrasts and intra-formational variations emerge as interpretable clusters in the

inversion-integrated scenario. Heterogeneity intervals within the Mangahewa reservoir, associated with variations in depositional environment, are represented by clusters that can be related to one another within a coherent geological framework. This behavior indicates that the inclusion of inversion products enriches the input space in terms of petrophysical content. Although the attribute-based scenario is also capable of detecting these heterogeneity zones, interpreting the relationships between clusters remains more challenging compared to the inversion-integrated case.

The color bars representing the distribution of clusters along the two K-means axes and their corresponding color assignments also provide important insights into the clustering results. The proximity of clusters within this two-dimensional space indicates similarity in their characteristics. In other words, colors that are close to each other on the colormap are expected to represent geologically related facies.

In the attribute-based scenario, clusters generally follow the overall geological trend; however, some clusters that are located close to each other in this space do not correspond to similar geological features. A key implication of this behavior is that attribute-based clustering, due to its sensitivity to amplitude variations, tends to emphasize reflective surfaces rather than true facies distinctions. As a result, both inter-formational and intra-formational contrasts are not fully captured.

In contrast, the inversion-integrated scenario produces a much more consistent distribution of clusters, both across lithological boundaries and within heterogeneous zones of the same geological unit, significantly improving reservoir characterization. With the inclusion of inversion products, cluster volumes shift from being primarily controlled by seismic

amplitude responses toward representing probable facies behavior, enabling clearer discrimination of reservoir-related features.

Structural discontinuities are represented by a distinct dark-blue cluster in both scenarios. Examination of these representations indicates that both approaches are capable of resolving structural discontinuities effectively. In the attribute-based PCA results, noise-induced scatters made the identification of faults and fractured structures more challenging. However, in the K-means results, this effect is reduced, allowing fault-related clusters to be more distinctly separated and interpreted. While the scatter observed in PCA does not significantly affect the cluster representing structural discontinuities, its influence is still evident in the representation of other clusters.

All these findings demonstrate that incorporating P- and S-impedance into multi-attribute analysis shifts the data representation from a structure governed by seismic response behavior toward a more geologically meaningful one. With the inclusion of inversion products, overall geological trends, intra-formational heterogeneity zones, facies distributions, and structural discontinuities can be visualized more clearly and consistently. In addition, the improvement in both lateral and vertical resolution, together with the reduction in noise levels, significantly enhances the interpretability of the results and enables more reliable inferences for reservoir characterization. As a direct consequence of this improvement, sand-prone intervals with potentially good reservoir quality that are not detectable in the attribute-only scenario can be identified, and clustering patterns that better represent reservoir facies are obtained.

On the other hand, this comparative analysis also provides an opportunity to evaluate the performance of the attribute-only scenario. While this scenario yields acceptable results in representing overall geological trends and structural discontinuities, it exhibits more limited

performance in delineating reservoir boundaries, resolving heterogeneity zones, and distinguishing between facies. Nevertheless, analyses based solely on seismic amplitude attributes still provide valuable insights for reservoir characterization.

6. Discussion

In this section, the results obtained from the two scenarios are evaluated within the framework of the theoretical expectations, and the effects of incorporating inversion products into multi-attribute analysis are discussed.

The findings not only highlight the contributions of inversion products to multi-attribute analysis but also provide a basis for broader insights into the methodology through the analytical framework established by the comparative two-scenario approach. Within this context, the capabilities and limitations of the applied workflow are critically assessed, and the results are used to outline potential directions for future research and methodological improvements.

PCA and K-means clustering results demonstrate that the inclusion of impedance volumes transforms the data representation from a structure governed primarily by seismic response behavior into a more geologically meaningful representation with stronger implications for reservoir characterization (Figures 5.2 & 5.5). As a result of this transition, lithological trends, intra-formational heterogeneities, structural discontinuities, and facies distributions within the reservoir interval are represented with improved resolution and greater consistency, leading to more reliable interpretation.

The band-limited nature of seismic amplitude data restricts the range of variance represented in multi-attribute analysis and limits the direct capture of low-frequency components associated with large-scale lithological variations and stratigraphic trends (Figures 5.2a & 5.5a). The results indicate that this limitation is significantly mitigated by incorporating seismic inversion products into the analysis. The low-frequency content introduced by inversion volumes

becomes particularly evident through the enhanced contrast between reservoir and seal units, as well as the improved differentiation between stratigraphic intervals (Figures 5.2c & 5.5c).

In addition, the representation of intra-formational heterogeneity is improved, leading to increased interpretability and geological consistency of the results. Sand-prone intervals with potentially favorable reservoir quality, which are not detectable in the attribute-only scenario, can be resolved at higher resolution in the PCA outputs when inversion products are included (Figure 5.4). These findings are consistent with previous studies demonstrating that the integration of P- and S-impedance with seismic attributes enables more reliable identification of thin gas-sand channels (Mahdy et al., 2026) and allows a more accurate and quantitative delineation of reservoir distribution through the incorporation of multiple data types (Lee et al., 2017).

The redistribution of variance axes observed in the PCA results, together with the dominant influence of impedance data on the principal components, indicates that inversion products play a major role in defining the primary directions of variance in the data space and enrich the attribute set with complementary information not sufficiently represented in seismic attributes alone (Figure 5.1). This behavior reflects a transition from a data space controlled by a limited number of dominant directions in the attribute-only scenario to a more balanced distribution of variance across principal component axes when inversion products are included, as reflected in the corresponding color palettes (Figure 5.2). This reorganization leads to a clearer representation of lithological and stratigraphic variations.

Both PCA and K-means clustering results further demonstrate that the inclusion of inversion products shifts the nature of the results from a representation governed by reflectivity and seismic signal characteristics toward a more geology-driven model based on layer properties (Figures 5.2 & 5.5). This transition significantly enhances the reservoir characterization achieved

through the applied workflow. By incorporating inversion volumes derived from the low-frequency model, physically meaningful elastic parameters are introduced into the data space alongside mathematically derived seismic attributes, increasing the sensitivity of the results to lithological variations and facies differentiation.

Seismic attributes represent different components of the seismic response (i.e., waveform geometry, continuity, phase behavior, and spectral content) and provide a high-resolution description particularly for the identification of structural geometries, reflector continuity, and local stratigraphic variations. In contrast, seismic inversion products represent quantitative elastic properties (i.e., impedance, density, and Vp/Vs ratio) that directly contribute to reservoir characterization by providing quantitative information on subsurface properties that cannot be resolved solely from amplitude-derived attributes. In particular, P- and S-impedance exhibit strong relationships with porosity and lithological variations, making them critical parameters for reservoir characterization (Chopra & Marfurt, 2024).

The integration of these two data groups within a multi-attribute analysis framework enables a combined evaluation of the high-resolution seismic response information provided by attributes and the broader-scale geological trends captured by inversion volumes. As a result, complementary information across different spectral scales is represented within the same data space, stabilizing spatial variability and producing more coherent spatial patterns (Figures 5.2 and 5.5). Similarly, previous studies in which seismic attributes are used together with inversion-derived elastic parameters, although adopting different methodological approaches, demonstrate that this integration improves the characterization of reservoir properties and facilitates more accurate delineation of reservoir distribution (Calderon & Castagna, 2007; Harryandi & Tura, 2019; Lee et al., 2017).

Overall, both methods successfully meet the objectives of the multi-attribute analysis. Despite the use of a high-dimensional and potentially redundant input dataset, PCA provides robust results both in terms of dimensionality reduction and in revealing significant variance within the data. Consistent with previous studies emphasizing its ability to organize the data space into a more structured form (Coléou et al., 2003; Kim et al., 2019; La Marca, 2020; Roden et al., 2015; Verma et al., 2025), the applied reservoir characterization workflow achieves a representative description of the input datasets within the first two principal components.

On the other hand, an important limitation arises in the PCA analysis regarding the representation of input datasets by the principal components. With the inclusion of inversion products, the increase in the number of inputs prevents all attributes from being adequately represented within the first two principal components (Figure 5.1). In the inversion-included scenario, Peak Frequency and Cosine of Instantaneous Phase are not represented within the first two principal components. This observation indicates that, with increasing input dimensionality, the number of principal components should be expanded to allow each attribute to contribute to the variance structure. However, this was not implemented in order to preserve the consistency of the comparative analysis and to avoid introducing bias in favor of the inversion-included scenario.

The K-means clustering method significantly improves the interpretability of the results by partitioning the datasets into discrete facies groups (Figure 5.5). The findings are consistent with previous studies demonstrating that K-means can effectively identify similar patterns within multi-attribute datasets (Lee et al., 2017). However, prior studies also emphasize that clustering results are sensitive to the selected attribute set and do not define explicit geological relationships between clusters (Chopra & Marfurt, 2014; La Marca, 2020; Troccoli et al., 2022; Zhao et al.,

2015), highlighting the need for careful interpretation. Therefore, the K-means results are not treated as a final classification on their own but are evaluated in conjunction with the geological interpretation derived from the PCA analysis.

The low-frequency-dominated nature of inversion products is considered as a potential limitation, as it may suppress attribute contributions and restrict the representation of high-frequency information in the analysis results. However, the findings do not indicate any clear evidence supporting such a limitation. In other words, neither PCA nor K-means clustering results suggest that the attribute-only scenario provides a superior representation compared to the inversion-integrated case. On the contrary, the inclusion of inversion products allows the high-resolution information provided by seismic attributes to be expressed within a more consistent representation (Figure 5.3). In this context, while seismic attributes retain their primary role in capturing structural and textural features, impedance data introduces an additional dimension of variation associated with elastic properties into the data space. This results in a representation that is not controlled by a single data type but rather defined by a more balanced contribution of multiple physical and spectral components.

Another potential limitation identified in the analysis relates to the model-based filtering and regularization inherent in the inversion process, which may unintentionally suppress certain high-frequency geological signatures (i.e., fracture-related scattering, small-scale heterogeneities, and thin-bed effects) by treating them as noise. This suggests that inversion products may be less sensitive to discontinuities. However, the comparative analysis indicates that both scenarios are able to resolve structural discontinuities with comparable clarity (Figure 5.2), and that these features are represented as distinct clusters in K-means results (Figure 5.5). This observation

highlights that seismic attributes preserve high-frequency signatures associated with fractures and discontinuities, thereby playing a complementary role to inversion volumes.

A notable effect observed in the analysis results is the difference in noise levels between the two scenarios. Seismic responses associated with geological features typically exhibit lateral continuity, whereas anomalies that show discontinuity and randomness in terms of amplitude and phase are largely interpreted as noise. The mathematical operations used to generate seismic attributes (e.g., first- and second-order derivatives, Hilbert transform, gradient and curvature operators, coherence and similarity calculations) tend to amplify high-frequency components of the noise present in the seismic amplitude data, causing the derived attributes to exhibit elevated noise. This is reflected in the attribute-based scenario, which exhibits higher noise levels in the results (Figures 5.2a-b & 5.3a-b). This behavior is also related to the inherently redundant and noise-prone nature of seismic data (Coléou et al., 2003).

In contrast, seismic inversion products are generally less sensitive to random noise, as they are generated using model-based constraints, multi-stack averaging, regularization, statistical smoothing, and filtering. As a result, the inclusion of inversion products introduces smoother and laterally continuous geological trends into the data space (Figures 5.2c-d & 5.3c-d). Low-frequency geological variations become more pronounced, while the influence of high-frequency noise components is reduced, leading to an overall decrease in noise levels. Consequently, the resolution and interpretability of the results are significantly improved.

In this context, a further important aspect to consider in multi-attribute analysis is the presence of shared noise components among attributes derived from the same seismic amplitude volume. When multiple attributes are generated from a single data source, the noise inherent in the amplitude data can be repeatedly represented across the attribute set, effectively becoming a

shared component and increasing redundancy within the dataset. This may lead to relationships appearing in the data space that do not correspond to real geological features and may instead be interpreted as meaningful patterns. Such artificial relationships are commonly described in literature as spurious correlations, particularly when large numbers of attributes are evaluated or when attribute selection is not constrained by physical relationships (Chopra & Marfurt, 2005; Kalkomey, 1997). The scattered patterns and locally inconsistent color distributions observed in the attribute-only scenario are indicative of this effect (Figure 5.3a-b). In contrast, the inclusion of seismic inversion products not only expands the input dataset but also introduces additional variability, reducing the likelihood that repeated noise components from amplitude-derived attributes dominate the variance structure of the data space. As a result, the overall reliability of the analysis is improved.

Despite these advantages, seismic inversion products are associated with several limitations that need to be considered. First, inversion results are highly model-dependent, as they rely on a low-frequency model constructed through interpolation of well logs and guided by geological assumptions, and their reliability is directly controlled by the accuracy of the LFM (Mahdy et al., 2026). In cases where well distribution is insufficient, lateral trends are not adequately captured, the stratigraphic framework is incorrectly defined, or inappropriate geological constraints are applied, the resulting inversion volumes may not represent true subsurface conditions. In addition, seismic inversion is an ill-posed and non-unique problem, meaning that multiple impedance solutions can produce the same seismic amplitude response depending on the selected model and parameters. This non-uniqueness arises from the strong influence of the initial model and geological constraints on the inversion results (Sokolov et al., 2021). Finally, in early exploration stages, well control is typically limited or absent, making the

generation of reliable inversion products challenging. As a result, inversion volumes are generally obtained during the development or production phases, which delays their availability (Maas, 2025). Therefore, early-stage reservoir characterization largely relies on seismic amplitude data and attributes derived from it.

From this perspective, a comprehensive evaluation of the performance of a seismic attribute-only analysis approach is also enabled. The results indicate that the attribute-only approach provides a reliable and practical solution, particularly for identifying overall trends, delineating structural features, and performing a first-order assessment of reservoir heterogeneity.

However, considering the improvements achieved through the inclusion of inversion products, it becomes clear that the low-frequency content they provide plays a critical role in multi-attribute analysis. In this regard, even in the absence of inversion products, incorporating alternative data types or derived attributes that can capture similar low-frequency trends may enhance the performance of the attribute-only approach. Although such approaches cannot fully replace inversion products, they enable the integration of broader-scale geological trends into the data space and therefore represent a promising direction for improving the quality of early-stage reservoir characterization.

Although the applied workflow primarily focuses on the integration of low-frequency components, it also suggests that the deliberate inclusion of inputs representing high-frequency content may further improve the performance of the workflow. From this perspective, enriching the analysis with data types that provide information at different scales is expected to advance reservoir characterization. In particular, the inclusion of inputs enriched in high-frequency content, which are typically underrepresented in seismic data, is anticipated to enhance the

geological detail of the results, increase sensitivity to fine stratigraphic features, and improve overall resolution, interpretability, and reliability. These considerations point to a promising direction for future research.

In addition, an important advantage of multi-attribute analysis for reservoir characterization is highlighted. Methods such as PCA and K-means clustering, which enable a more advanced analysis of relationships among datasets, also facilitate a more geologically consistent representation of the results. The findings indicate that, through the selection of appropriate input combinations, the resulting representation space can be guided to become sensitive to specific facies contrasts. This demonstrates that multi-attribute analysis extends beyond a purely descriptive approach and can be used to design analyses that are responsive to targeted geological features through the deliberate selection and organization of input data. Enriching the spectral and geophysical content of the input dataset plays a critical role in improving the accuracy and reliability of reservoir characterization. Within this framework, multi-attribute analysis can be regarded as an active design process that enables the construction of data representations aligned with reservoir characterization objectives.

7. Conclusion

The impact of integrating seismic inversion products into multi-attribute analysis is evaluated through a controlled comparative framework based on two scenarios: (1) an attribute-only dataset derived from seismic amplitude data and (2) a combined dataset incorporating both seismic attributes and inversion products. For each scenario, the input datasets were evaluated under identical analytical conditions using PCA and K-means clustering, ensuring a consistent basis for comparison.

The results demonstrate that the inclusion of P- and S-impedance leads to a transition in the data representation from a structure governed by seismic amplitude response to a more geologically meaningful organization. This transformation improves the differentiation between reservoir and seal units, as well as the delineation of reservoir boundaries, lithological trends, intra-formational heterogeneities, structural discontinuities, and facies distributions. In addition, reservoir facies architecture is more distinctly resolved, with sand-prone intervals that are not observable in the attribute-only scenario becoming identifiable in the PCA results and represented as coherent clustering patterns in the K-means outputs. A reduction in noise levels, together with an increase in effective vertical and lateral resolution, further enhances the interpretability of the results and supports more reliable geological interpretation. These improvements enable a detailed and consistent characterization of the complex reservoir architecture of the Mangahewa Formation in the Pohokura Field.

The comparative framework also provides important insights into the performance of the attribute-only approach. While this scenario effectively captures large-scale structural trends and discontinuities, its ability to resolve heterogeneity zones and delineate reservoir boundaries

remains limited. Nevertheless, amplitude-derived attributes alone continue to provide reliable information, particularly in cases where inversion products are not available.

Beyond these observations, the two-scenario analytical framework enables broader methodological insights into multi-attribute analysis. Although the applied methods play an important role in transforming the dataset into representations aligned with workflow objectives, the success of the analysis is primarily controlled by the characteristics of the input data. In particular, the spatial extent of the dataset, the correlation structure between variables, their representation within dominant variance components, and their geological relevance collectively determine the quality of the results.

The improvements achieved through the inclusion of inversion products demonstrate that the proper organization of input data can significantly enhance results. These findings provide a basis for future work. In particular, the integration of high-frequency components into the attribute set, similar to low-frequency components, is expected to further improve the effectiveness of multi-attribute analysis. Furthermore, the deliberate design of input datasets tailored to specific facies contrasts is expected to yield more targeted and discriminative results.

Taken together, multi-attribute analysis is not merely a data integration approach, but a design-oriented process in which the selection and organization of input datasets fundamentally control the quality of the results.

8. References

Calderon, J. E., & Castagna, J. (2007). Porosity and lithologic estimation using rock physics and multi-attribute transforms in Balcon Field, Colombia. *The Leading Edge, 26*(2), 142-150. https://doi.org/10.1190/1.2542439

Chopra, S., & Marfurt, K. J. (2005). Seismic attributes - A historical perspective. *Geophysics, 70*(5), 3SO-28SO. https://doi.org/10.1190/1.2098670

Chopra, S., & Marfurt, K. J. (2014). Churning seismic attributes with principal component analysis. In *SEG Technical Program Expanded Abstracts 2014* (pp. 2672-2676). Society of Exploration Geophysicists. https://doi.org/10.1190/segam2014-0235.1

Chopra, S., & Marfurt, K. J. (2024). *Essentials of seismic attributes and impedance inversion*. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560804062

Coléou, T., Poupon, M., & Azbel, K. (2003). Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge*, 22(10), 942-953. https://doi.org/10.1190/1.1623635

Greenberg, M. L., & Castagna, J. P. (1992). Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications. *Geophysical Prospecting, 40*(2), 195-209. https://doi.org/10.1111/j.1365-2478.1992.tb00371.x

Harryandi, S., & Tura, A. (2019). Sweet spot identification through seismic inversion and multiattribute transform: A case study of the Niobrara and Codell unconventionals. *Interpretation, 7*(4), T887-T898. https://doi.org/10.1190/INT-2018-0175.1

Higgs, K. (2001). *Geological reservoir characterisation study of Pohokura Field*. Fletcher Challenge Energy Limited. Ministry of Economic Development, New Zealand. Unpublished Petroleum Report PR3809. Available via the New Zealand Petroleum and Minerals Geodata Catalogue. https://www.nzpam.govt.nz/maps-geoscience/geodata-catalogue

Hossain, S., & Rahman, N. (2025). An integrated petrophysical and rock physics characterization of the Mangahewa Formation in the Pohokura field, Taranaki Basin. *Scientific Reports, 15*(1), 4983. https://doi.org/10.1038/s41598-025-88639-4

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441. https://doi.org/10.1037/h0071325

Kalid, N. Z. A., & Hamzah, U. (2014). Sequence stratigraphy of the Miocene, Pohokura field, Taranaki Basin, New Zealand. *AIP Conference Proceedings*, 1614(1), 637-643. https://doi.org/10.1063/1.4895277

Kalkomey, C. T. (1997). Potential risks when using seismic attributes as predictors of reservoir properties. *The Leading Edge*, 16(3), 247-251. https://doi.org/10.1190/1.1437610

Kim, Y., Hardisty, R., & Marfurt, K. J. (2019). Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale. *Interpretation*, 7(3), SE281-SE297. https://doi.org/10.1190/INT-2018-0246.1

King, P. R., & Thrasher, G. P. (1996). *Cretaceous-Cenozoic geology and petroleum systems of the Taranaki Basin, New Zealand* (Institute of Geological & Nuclear Sciences Monograph 13). Institute of Geological & Nuclear Sciences Limited.

La Marca, K. (2020). *Seismic attribute optimization with unsupervised machine learning techniques for deepwater seismic facies interpretation: Users vs machines* (Master's Thesis, University of Oklahoma). SHAREOK Institutional Repository. https://shareok.org/items/2de83777-843d-4ea0-a0c7-332351c7fc0a

Leal, J., Jerónimo, R., Rada, F., Viloria, R., & Roden, R. (2019). Net reservoir discrimination through multi-attribute analysis at single sample scale. *First Break, 37*(9), 77-86. https://doi.org/10.3997/1365-2397.n0058

Lee, J., Byun, J., Kim, B., & Yoo, D.-G. (2017). Delineation of gas hydrate reservoirs in the Ulleung Basin using unsupervised multi-attribute clustering without well log data. *Journal of Natural Gas Science and Engineering, 46*, 326-337. https://doi.org/10.1016/j.jngse.2017.08.007

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489

Maas, M. V. R. (2025). *Application of post-stack seismic attributes, well dynamic data, and machine learning for carbonate reservoir facies and productivity prediction for upstream optimization* (Doctoral dissertation, University of Oklahoma). SHAREOK Institutional Repository. https://shareok.org/items/1590da77-df4e-4949-8378-37df27c9920f

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297). University of California Press. http://projecteuclid.org/euclid.bsmsp/1200512992

Mahdy, A., Helmi, A., Helaly, A. S., & Mahmoud, A. M. E. (2026). A physics-driven workflow for gas-sand identification in Pliocene turbidites using pre-stack inversion and seismic attributes, offshore Egypt. *Scientific Reports, 16*(1), 1341. https://doi.org/10.1038/s41598-025-31461-9

Ministry of Business, Innovation and Employment. (2025). *Energy in New Zealand 2025*. https://www.mbie.govt.nz/assets/energy_in_new_zealand_2025.pdf

Murray, D. (2000). *Pohokura-2 well completion report*. Fletcher Challenge Energy Taranaki. Ministry of Economic Development, New Zealand. Unpublished Petroleum Report PR2516. Available via the New Zealand Petroleum and Minerals Geodata Catalogue. https://www.nzpam.govt.nz/maps-geoscience/geodata-catalogue

New Zealand Petroleum and Minerals. (n.d.). Geodata Catalogue. Ministry of Business, Innovation and Employment. https://www.nzpam.govt.nz/maps-geoscience/geodata-catalogue

New Zealand Petroleum and Minerals. (2014). *New Zealand petroleum basins (2014/2015 revised edition)*. Ministry of Business, Innovation and Employment. https://www.nzpam.govt.nz/nz-industry/nz-petroleum/resource-potential

Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution - III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, 187*, 253-318. https://doi.org/10.1098/rsta.1896.0007

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572. https://doi.org/10.1080/14786440109462720

Radwan, A. A., Nabawy, B. S., Kassem, A. A., & Elmahdy, M. (2023). An integrated workflow for seismic interpretation, petrophysical and petrographical characterization for the clastic Mangahewa reservoir in Pohokura gas field, Taranaki Basin, New Zealand. *Geoenergy Science and Engineering, 229*, 212117. https://doi.org/10.1016/j.geoen.2023.212117

Robinson, P. H., & King, P. R. (1988). Hydrocarbon Reservoir Potential of the Taranaki Basin, Western New Zealand. *Energy Exploration & Exploitation*, *6*(3), 248-262. http://www.jstor.org/stable/43754538

Roden, R., Smith, T., & Sacrey, D. (2015). Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation*, 3(4), SAE59-SAE83. https://doi.org/10.1190/INT-2015-0037.1

Shell Exploration NZ Ltd. (2012a). *Pohokura Merge 3D (Marine/TZ/Land) data processing report.* Ministry of Economic Development, New Zealand. Unpublished Petroleum Report PR4544. Available via the New Zealand Petroleum and Minerals Geodata Catalogue. https://www.nzpam.govt.nz/maps-geoscience/geodata-catalogue

Shell Exploration NZ Ltd. (2012b). *Pohokura 3D seismic simultaneous inversion (from reprocessed data 2012).* Unpublished Technical Report. Available via the New Zealand Petroleum and Minerals Geodata Catalogue. https://www.nzpam.govt.nz/maps-geoscience/geodata-catalogue

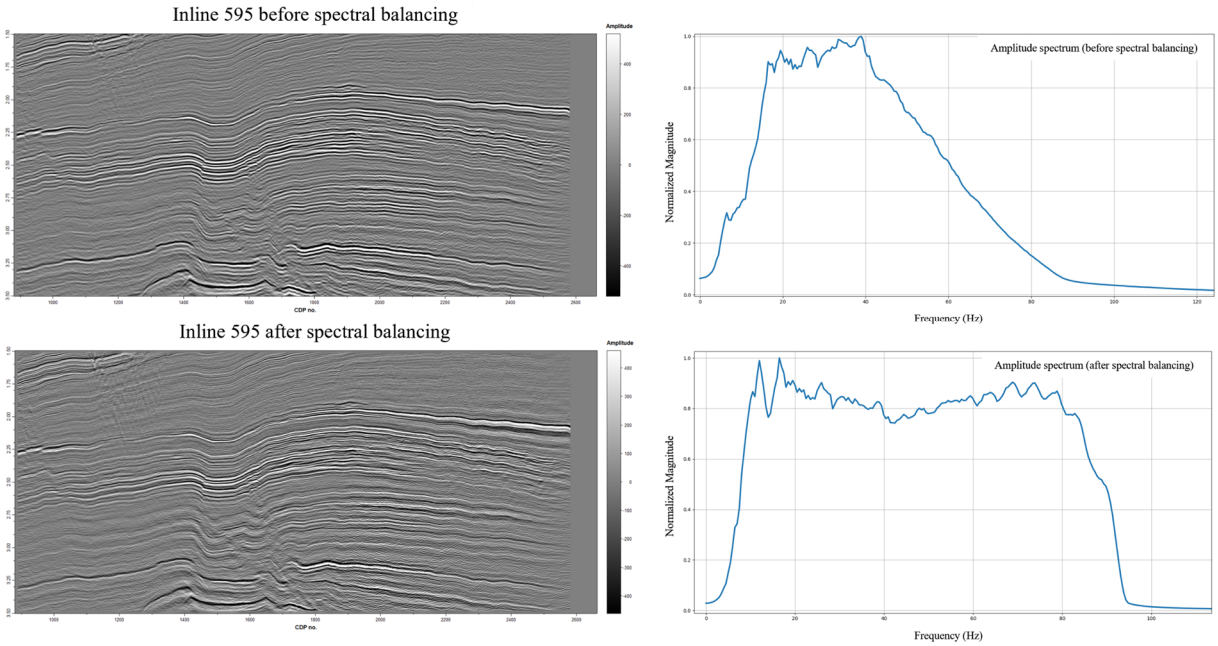
Sokolov, A., Schulte, B., Shalaby, H., & van der Molen, M. (2021). Seismic inversion for reservoir characterization. In E. Onajite (Ed.), *Applied techniques to integrated oil and gas reservoir characterization* (pp. 329-351). Elsevier. https://doi.org/10.1016/B978-0-12-817236-0.00013-3

Troccoli, E. B., Cerqueira, A. G., Lemos, J. B., & Holz, M. (2022). K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics, 198*, 104555. https://doi.org/10.1016/j.jappgeo.2022.104555

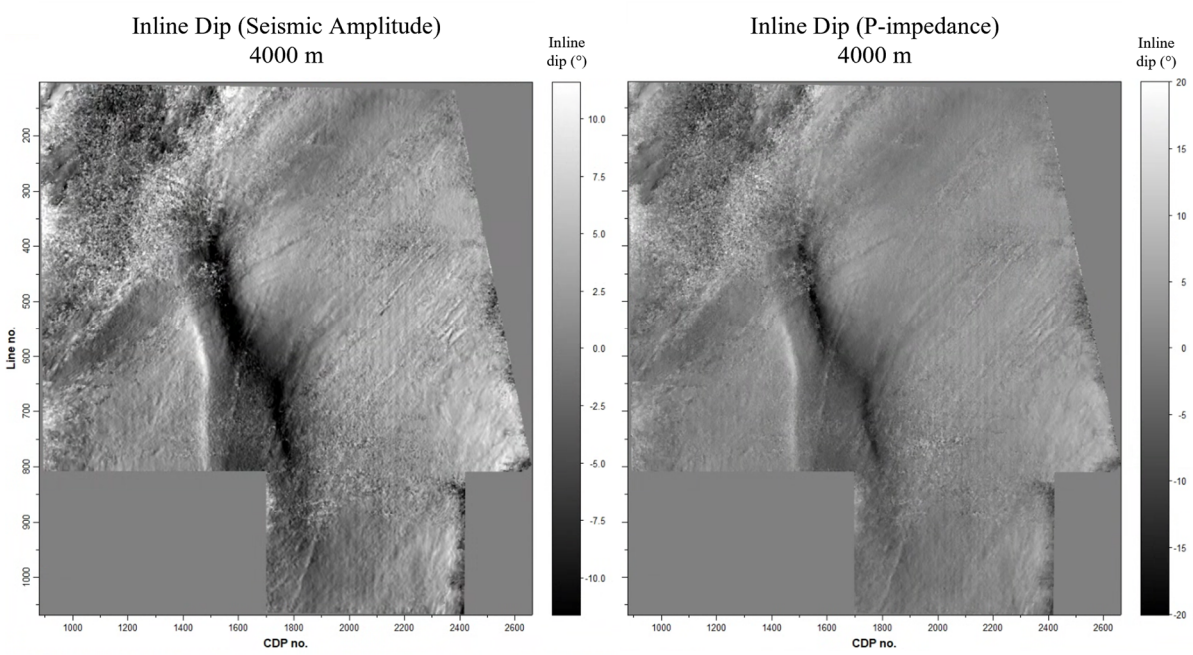
Verma, N., Kant, R., Maurya, S. P., Kumar, B., Singh, A. P., Hema, G., Singh, R., Singh, K. H., & Sarkar, P. (2025). Seismic inversion based on principal component analysis and probabilistic neural network for prediction of porosity from post-stack seismic data. *Earth Science Informatics, 18*(1), 98. https://doi.org/10.1007/s12145-024-01504-2

Zhao, T., Jayaram, V., Roy, A., & Marfurt, K. J. (2015). A comparison of classification techniques for seismic facies recognition. *Interpretation*, 3(4), SAE29-SAE58. https://doi.org/10.1190/INT-2015-0044.1

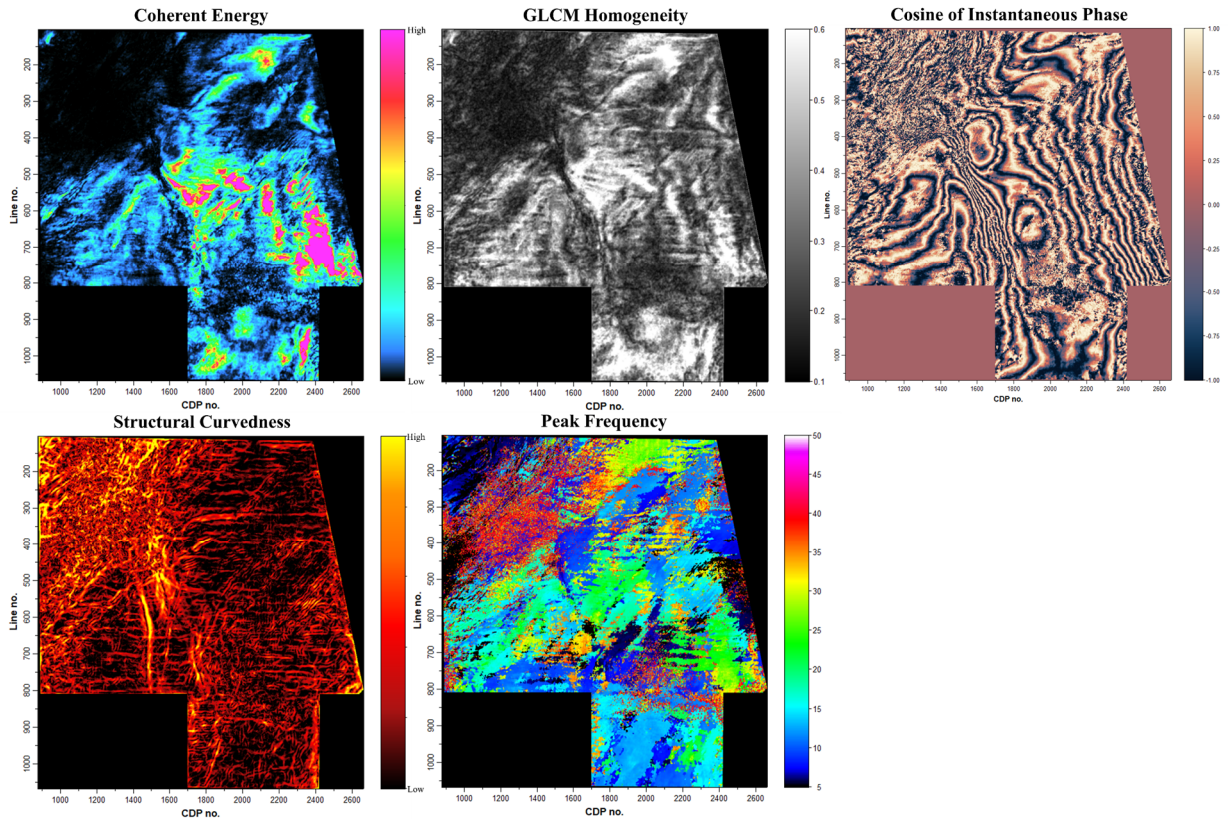
9. Appendix



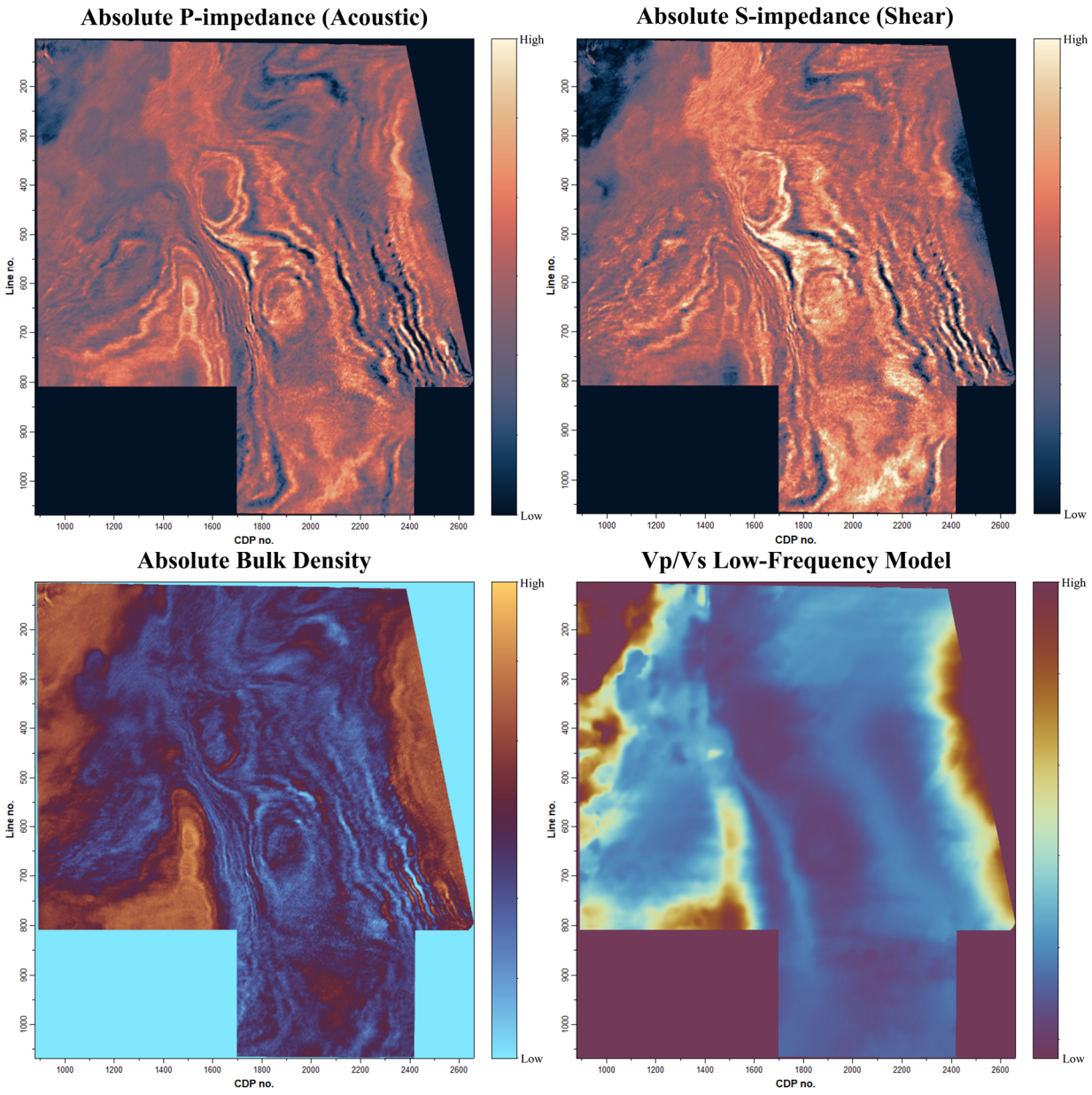
Appendix A: Comparison of seismic data along Inline 595 before and after spectral balancing, together with their corresponding amplitude spectra. Spectral balancing enhances the high-frequency content of the seismic signal, resulting in improved reflector continuity and vertical resolution.



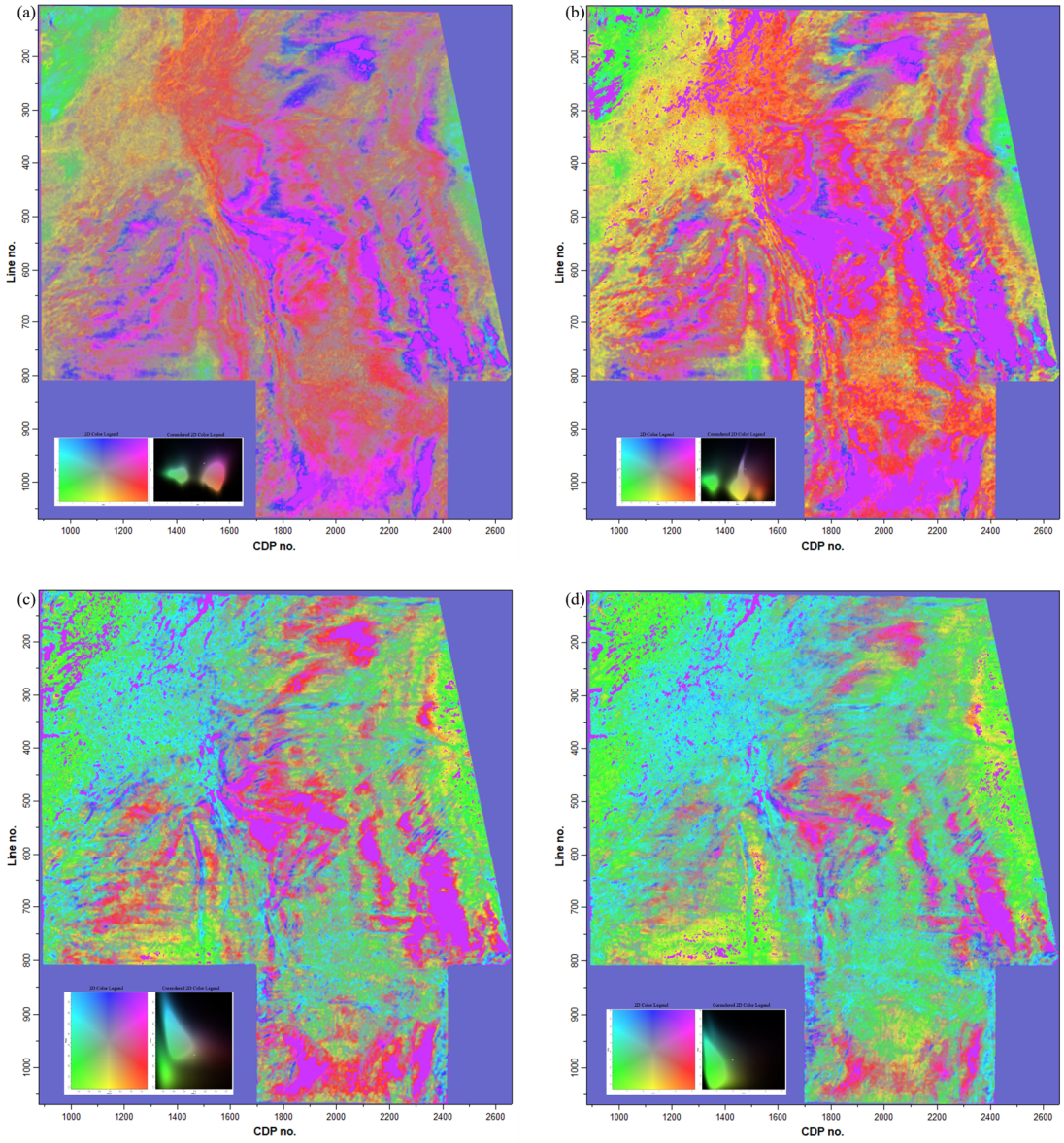
Appendix B: Comparison of inline dip attributes computed from the seismic amplitude volume and the P-impedance volume at 4000 m. The impedance-derived dip exhibits clearer and more continuous structural trends, indicating more reliable dip estimation.



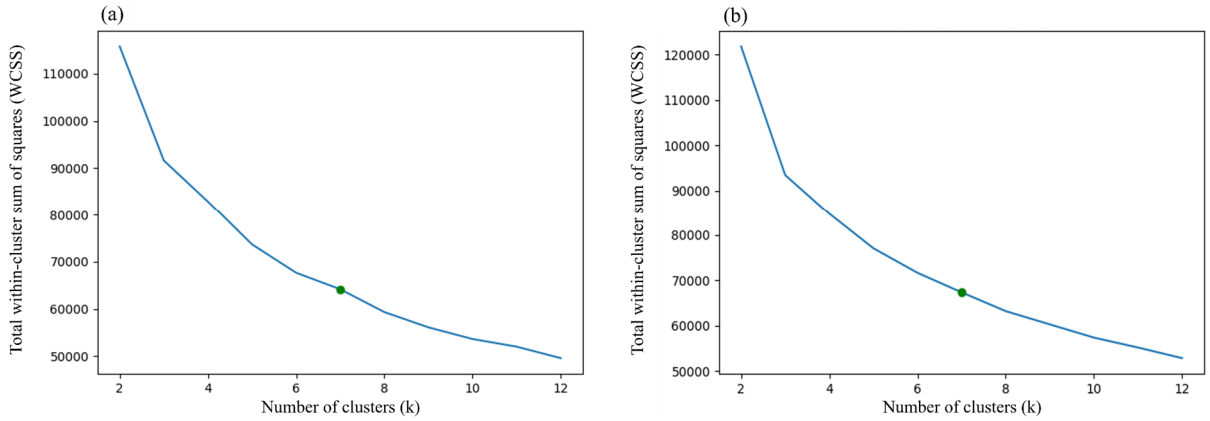
Appendix C: Seismic attributes used in the multi-attribute analysis, including coherence, GLCM homogeneity, cosine of instantaneous phase, curvedness, and peak frequency, extracted at a depth of 4000 m.



Appendix D: Candidate input volumes used in the inversion-based scenario, including P-impedance, S-impedance, density, and the V_p/V_s low-frequency model (LFM), extracted at a depth of 4000 m. Different combinations of these inputs were tested to determine the most suitable input set for the analysis.



Appendix E: Comparison of PCA results obtained using a fixed attribute set combined with different inversion volumes, extracted at a depth of 4000 m. The tested input combinations are **(a)** P- and S-impedance with LFM, **(b)** P- and S-impedance with density, **(c)** LFM only, and **(d)** density only. Only representative combinations from the tested cases are shown for illustration.



Appendix F: Elbow plots showing the variation of within-cluster sum of squares (WCSS) as a function of the number of clusters (k) for two different input datasets **(a)** attribute-only scenario **(b)** inversion-included scenario. In both cases, the rate of error reduction decreases beyond $k \approx 7$, indicating an inflection point interpreted as the optimal number of clusters.