

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

EVALUATION OF MACHINE LEARNING ALGORITHM ACCURACY IN ENERGY
EXPLORATION USING PHYSICS DERIVED SYNTHETIC MODELS.

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
Alexandro Vera Arroyo
Norman, Oklahoma
2024

EVALUATION OF MACHINE LEARNING ALGORITHM ACCURACY IN ENERGY
EXPLORATION USING PHYSICS DERIVED SYNTHETIC MODELS.

A DISSERTATION APPROVED FOR THE
SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Junle Jiang

Dr. Matthew Pranter

Dr. Brett M. Carpenter

Dr. Deepak Devegowda

There is no end to education. It is not that you read a book, pass an examination, and finish with education. The whole of life, from the moment you are born to the moment you die, is a process of learning.

Jiddu Krishnamurti

ACKNOWLEDGEMENTS

I want to express my heartfelt gratitude to the individuals whose unwavering support, guidance, and encouragement have been instrumental in my journey through the completion of my PhD program and the writing of this dissertation.

First and foremost, I am deeply thankful to my advisor, Heather Bedle, whose expertise, mentorship, and continuous support have been invaluable throughout this process. Her courage, words, and unwavering optimism rescued me multiple times. I am also grateful to Junle Jiang for his guidance and assistance. I appreciate his patience, his technical lessons, and his motivation to improve the quality of my work, also for welcoming me during his seminars.

To Erika, my wife, my deepest love and gratitude for your time and the sacrifices you have made. I made it through alive and somehow sane thanks to her, and Erika, you know it's true, now, and then, it's all because of you. My deepest gratitude to my brother Gerardo, for his steadfast optimism, unending support and understanding during the ups and downs of this academic journey. Brother, we were lucky in being handsome but not rich.

I am grateful to my family, Margarita and Jaime Vera, Julio, Carlos, and Patricia Arroyo for their love, encouragement, and belief in my abilities. To my Father who always warned me about the perils of the PhD journey, yet he invited me to take it with courage and optimism. His bedtime story about the phrase "Alea Jacta est" in the Roman times always inspired me, and today after 7 years of his loss I can tell him, I did it, I didn't go back even to gain momentum. I would love to think he would be proud. Thanks to my Mother, although, I lost her when I was 5 years old, I keep this phrase from her "You will never be perfect, but every day you are perfectible".

I am indebted to my friends at OU, for everything, from sharing a meal to a vacation adventure, Sidi, Issa Al Aamri, Jose, Raymond, Deepankar, Ashutosh, Jose Luis, Manoj, Ganiyat, Jose Pedro and his mother, Edimar, Emily, Thank you. Additionally, I am thankful to Daiane Cardoso, Karen Leopoldino, Anna, Peiyao Li, Yuwei Wang, Rodrigo and his wife Delfina, Ahmet, April, Jake, and Emilio for their friendship and camaraderie. A special thanks to all the participants of the OU UWC football league, where I always felt welcome.

To my friends in Mexico, namely Andrei, Humberto Samuel, Gerardo Garcia, David, Christian, Norma, Antonio Cervantes, and Hamlet, your support and encouragement have meant the world to me during the hard times. I am grateful for all the laughs, memes, discussions, beers, and whiskeys we shared. The next ones are on me.

I would also like to acknowledge the AASPI consortium and extend my gratitude to Kurt Marfurt, Satinder Chopra, and Hector La Madrid for their encouraging words.

A special mention goes to our administrative staff, particularly Rebecca Fay, whose help and guidance have made many things possible. I am also thankful to Leah, Ashley, and Ginger for their assistance, patience, and support.

Thank you to each and every person mentioned above for being part of this journey and for contributing to my academic and personal growth. Your support has been truly appreciated.

Finally, for anyone who reads until here, and most likely for me who in a few years will read this thesis with nostalgia; resilience, resist, patience will polish you.

Table of Contents

ACKNOWLEDGEMENTS	v
LIST OF FIGURES.....	x
LIST OF TABLES.....	xx
ABSTRACT.....	xx
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: GENERATIVE ADVERSARIAL NETWORKS TO ASSIST SUBSURFACE GAS STORAGE CHARACTERIZATION USING GEOSTATISTICAL SEISMIC INVERSION- AN EXAMPLE FROM RAY REEF, MICHIGAN BASIN, USA.	5
ABSTRACT	5
INTRODUCTION.....	6
GEOLOGIC BACKGROUND	7
PREVIOUS CHARACTERIZATION OF RAY REEF	11
DATASET DESCRIPTION	14
METHODOLOGY	16
RESULTS.....	35
DISCUSSION	50
CONCLUSION	54
DATA AVAILABILITY.....	55
ACKNOWLEDGMENTS	56
CHAPTER 3: IN-DEPTH ANALYSIS OF SELF-ORGANIZING MAPS FOR CHARACTERIZING AMPLITUDE PLAYS IN SILICICLASTIC RESERVOIRS WITH ROCK PHYSICS-DERIVED 3D SYNTHETIC MODELS. *	57
ABSTRACT	57
INTRODUCTION.....	57
DATA AND WORKFLOW	60
SYNTHETIC MODEL AND ATTRIBUTE ANALYSIS RESULTS.....	69
APPLYING SYNTHETIC RESULT INSIGHTS TO STYBARROW DATA	84
DISCUSSION	92
CONCLUSION	95
ACKNOWLEDGMENTS	96

CHAPTER 4: SEAL AND RESERVOIR RISK EVALUATION USING HIERARCHICAL CLUSTERING ANALYSIS WITH SEISMIC ATTRIBUTES IN NORTHWESTERN AUSTRALIA. *	97
ABSTRACT	97
INTRODUCTION.....	98
WORKFLOW AND METHODS	107
RESULTS	118
DISCUSSION	127
CONCLUSION	130
ACKNOWLEDGMENTS	131
DATA AND MATERIAL AVAILABILITY.....	131
DECLARATION OF COMPETING INTERESTS	131
CHAPTER 5: ASSESSING MACHINE LEARNING ACCURACY IN PREDICTING IN-SITU GAS SATURATION AND LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS TO UNDERSTAND ITS UNCERTAINTY: A CO₂ EXAMPLE FROM KEVIN DOME. *	132
ABSTRACT	132
INTRODUCTION.....	133
METHODS	138
RESULTS	145
DISCUSSION	150
CONCLUSION	153
ACKNOWLEDGMENTS	154
DATA AND MATERIAL AVAILABILITY.....	154
CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS.....	155
APPENDIX	158
APPENDIX A: A short note on geostatistical inversion	158
APPENDIX B: GTM methodology for unsupervised seismic clustering and the result of synthetic data.....	164
APPENDIX C: When the levee breaks: Visualization of a deepwater crevasse splay in Late Cretaceous in Exmouth Basin Offshore Western Australia. *	169
ABSTRACT	169
INTRODUCTION.....	170
METHODS	176

RESULTS	177
ALTERNATIVE INTERPRETATIONS	184
CONCLUSION	185
ACKNOWLEDGMENTS	186
DATA AND MATERIALS AVAILABILITY	186
<i>APPENDIX D: Contrasting faulting styles of salt domes and volcanoes: Can unsupervised learning techniques differentiate fault styles? *</i>	187
ABSTRACT	187
INTRODUCTION	188
RESULTS	204
VOXEL AND FAULT PATCH INTERPRETATION	207
MODERN ENVIRONMENT COMPARISON	210
ALTERNATIVE INTERPRETATIONS	212
DISCUSSION	213
Conclusions	214
ACKNOWLEDGMENTS	217
<i>APPENDIX E: Elastic properties of an idealized sandstone</i>	218
<i>APPENDIX F: Brief documentation on the clustering selection methods</i>	220
REFERENCES	222

LIST OF FIGURES

Figure 2.1. Location of the study area including the distribution of Niagaran reefs in Michigan. The contrasting reef density between the Northern Reef Trend (Gill 1973, 1977a, 1977b) and the Southern Reef Trend is primarily a consequence of the tilted basin. The overlapped elevation map mimics paleo topography during the time of deposit and shows higher structural elements in the north. The structural liniments show a relatively stable basin.

Figure 2.2. Panel a) A schematical view of Reef internal architecture tied with the relative sea level curve. Facies, fauna, and lithologies found in the reef are described in Panel b). (Figure modified from Gill, 1973; Rine, 2016).

Figure 2.3. Panel a) shows the porosity distribution per facies and Panel b) shows the permeability distribution per facies.

Figure 2.4. Panel a) on the left shows the map view of the Ray Reef top (A-2 Carbonate) with a distribution of borehole data and the gas storage performance clustering. Panel b). Isometric view of the presence of boreholes and how these boreholes cover the reef structure thoroughly.

Figure 2.5. Workflow to train a generative adversarial network from reservoir images.

Figure 2.6. Panel a) Borehole cross section showing gamma ray, permeability, porosity, productivity index calculation both at borehole scale and seismic scale (averaging information in blocks of 16 ft), discretized productivity index classifying the productivity rock type and the facies described from core data. b) map view of the cross section and the storage performance per well, i.e. which borehole is storing more gas per month.

Figure 2.7. Scheme of geostatistical seismic inversion methodology from partial stacks to impedance versions. Panel a) Shows the starting point of the methodology elastic logs and from borehole data and the seven partial angle stacks. Panel b) Lastly the variograms assist the geostatistical extrapolation using Kriging from borehole data to generate the priori probability. Panel c) shows the outputs of the geostatistical seismic inversion, a posterior probability of reservoir, p-impedance, s-impedance, and density.

Figure 2.8. Panel a) shows the 3D data points of the p-impedance, s-impedance, and bulk density. Panel b) shows the histogram of each elastic parameter.

Figure 2.9. Panel a) shows the p-impedance histogram, Panel b) show the s-impedance distribution and panel c) shows the bulk density distribution. Color triangles indicate the position of the P10, Mean and P90 values of the elastic parameters. Blue points show the distribution of the saved realizations to be used in the facies and fluid probability method.

Figure 2.10. Panel a) shows the cross-plot in V_p/V_s – p -impedance domain of the three productivity index zones in the reservoir, and Panel b) shows the theoretically probability distribution assigned to each reservoir types. The probabilities distributions are used to estimate the probability of each facies to be linked to the productivity rock type.

Figure 2.11. Scheme showing the limited reservoir models available in seismic inversion reservoir characterization. Normally only one FFP realization is used to define the reservoir fluids, but we seek to generate as many samples as possible to train the GANs.

Figure 2.12. Panel a) shows the 3D view of the Ray reef structure from the A-2 carbonate to the Lockport/Guelph formation interpreted on 3D seismic and exaggerated vertically 10 times. Panel b) shows the facies and fluid probability coming from the mean elastic realization of the geostatistical seismic

inversion at the shallowest reservoir position. Panel c) Shows the time slice at 460 ms of the Ray reef from the same FFP volume and d) Shows the time slice at 490 ms of the FFP volume. These images work as an example of the images used to train the GANS.

Figure 2.13. Generator and Discriminator start from the latent space. We keep the weights of the generator updating with the discriminator. The schematics of GANs generate synthetic images from random noise and samples with clear labels that constrain the discriminator.

Figure 2.14. Synthetic gather with permeability curve to show the highest permeability associated with a negative reflection at near angles and changing to slightly positive with higher angles.

Figure 2.15. Panel a) East-West borehole permeability/porosity cross-section Panel b) South-north permeability/porosity cross-section within the Ray Reef. In both sections, the reef stratigraphy is highlighted with a red polygon, reefs showing the variability of flow capacity with permeabilities up to 1 Darcy and porosities above 10%. The south and north borders exhibit lower permeabilities and porosities.

Figure 2.16. Panel a) Full seismic stack with borehole rock-types across Ray reef. Panel b) p-impedance with borehole rock -types showing the high p-impedance layers outside the reservoir corresponding to anhydrite layers. Inside the reservoir the core reef shows the lower p-impedance and best reservoir properties. Panel c) shows the random section on panels a and b, with the time slice of the p-impedance. The time slice of the p-impedance shows a good distribution with no local anomalies.

Figure 2.17. Panel a) displays the s-impedance inversion result, anhydrite layers observe a high shear impedance. The reef's core is better described in the shear impedance section where the high shear impedance represents vertical bioherm growths. The best fluid reservoir facies (gray) correlate with intermediate values of shear impedance. On the other hand, panel b) displays the VpVs ratio inversion result. Red arrows highlight the reef's core, higher VpVs ratio agrees with the AVO class observed on the synthetic data. Gray facies correlate well with the reservoir intervals. The VpVs inversion includes a hard constrain to prevent smaller values than 1.42 and thus keeping the elasticity theory. Panel c) shows the map view of the shear impedance inversion to monitor for abnormal values or artificial effects.

Figure 2.18. Panel a) displays the bulk density resulted from the geostatistical, lower bulk density values within the reservoir are displayed in black. The white transition helps to visualize the fluid reservoir limits from the reservoir rock. Panel b) showing the residuals serves as a quality control of the seismic inversion, the objective of the inversion is to generate a set of elastic parameters that creates a synthetic seismic that minimize the misfit with the observed data. In panel b, we observe small amplitude and low frequency residuals except beneath Gray formation, the lack of borehole control increases the misfit function and thus the residuals are larger. Panel c) shows the map view of the bulk density to monitor for abnormal values or artificial effects.

Figure 2.19. Panel a) Acoustic impedance histogram normally distributed with a well-defined standard deviation, Panel b) shear impedance histogram, with a narrow uncertainty, VpVs ratio histogram exhibit a quasi-bimodal distribution due to the hard constraints imposed on the VpVs ratio. Panel d) shows the seismic residual histogram to show the small residuals, this histogram is slightly right-skewed suggesting the need of increasing the low-frequency values to reduce the positive residuals. Panel e) the cross-plot of the shear and acoustic impedance colored coded by the inverted bulk density. Panel f) is the acoustic impedance versus VpVs ratio, a template suggested by Ødegaard and Avseth (2003) to correlate the rock physics template to hydrocarbon and porosity effects. g) Defining the prior probability density functions (PDFs) for each reservoir facies. VpVs ratio versus p-impedance cross plots augments the fluid effects.

Figure 2.20. Panel a) Seismic amplitude overlapped by the mean input used to generate one of the FFP results, which is the posteriori probabilities of the Bayesian inference. This figure shows the reef core as the most productive interface. Panel b) shows the map view at 470 ms of the reservoir probability. It is clear that the area within the red square can be correlated with the reef's migration from south to north.

Figure 2.21. 3D View of the captured geobodies taking 75% as the threshold on the fluid probability. The geobodies show the impermeable layers and the porosity patches.

Figure 2.22. Panel a) shows the time slice possible reservoir realizations including the fluid effect. These images are used as training for the generator, and as a discriminator. Panel b) shows the first iteration (epoch # 1), where there is only random noise to simulate the model.

Figure 2.23. Panel a) is the GANs result for 40 epochs, each epoch takes ~ 12 minutes to complete its simulation, Panel b) shows the Losses versus epochs for both the discriminator and the generator, in theory, the smaller the loss, the closer the generation is to the original data. Panel c) shows the score versus epoch plot. For example, in panel a) red circles correlates well with the most probable reservoir in the FFP result. While white circles show the heterogeneities of the reservoirs.

Figure 2.24. Time slice of the Ray Reef at 470 milliseconds with the different seismic attributes used in the self-organizing mapping. Panel a) shows the average frequency, b) dip azimuth, c) dip magnitude, d) energy ratio similarity, e) envelope, f) GLCM Energy, g) GLCM Entropy, h) GLCM homogeneity, i) inst. Cosine of phase, j) inst. Frequency, l) relative acoustic impedance, m) Curvature K1, n) Curvature K2, o) Inst. Frequency, p) Corendering of dip azimuth and dip magnitude with Sobel similarity.

Figure 2.25. Comparing previous inversion results Panel a) and Panel d) show the acoustic impedance inversion and seismic residual (observed – synthetic) on crossline 126 around borehole R-117. Panel b) and e) shows the current acoustic impedance inversion and seismic residuals. It is notable the small-scale residuals on the zone of interest from A-2 Carbonate and Gray formation. Beneath the Gray formation due to the lack of borehole control the residuals magnitude is larger. In Panel c) we present the acoustic impedance time slice at 470 ms.

Figure 3.1. Location map of datasets used in this study, the Stybarrow oil field located 50 km Northwest of Exmouth on the Exmouth sub-basin. Four boreholes are available with a complete suite of borehole wireline logs. The field is produced in siliciclastic systems with a well-documented history of production.

Figure 3.2. Well-tie of STY-1 borehole showing the polarity of the reservoir rock and the AVO/AVA effect. Panel a) the extracted wavelet at the borehole location. Panel b) shows the observed seismic overlaid with the litho-types, and c) the synthetic seismic calculated from the convolution of the wavelet and the impedance log. Panel d) shows the cross-correlation of the observed dataset with the synthetic seismic, e) the p-impedance profile recorded at the borehole, and f) A synthetic gather is calculated to show the in-situ AVO response.

Figure 3.3. Observed seismic before (Panel a) and after post-processing (Panel b) across the reservoir, see the increased resolution within the reservoir close to the borehole, highlighted by the red circle. Panel c) shows the RMS amplitude extraction over the reservoir interface before processing, and Panel d) The RMS extraction after post-processing.

Figure 3.4. Proposed workflow to evaluate the accuracy of self-organizing maps (SOMs). The acronym RPT stands for Rock Physics Template, and EDA is short for Exploratory Data Analysis. Petroelastic seismic inversion is used as a tool to compare the accuracy, particularly on the observed dataset. The workflow seeks to evaluate the sensitivity of SOMs to reservoirs that exhibit an amplitude seismic response.

Figure 3.5. Synthetic elastic and petrophysical properties of the three reservoirs proposed. These plots are pseudo boreholes. Panel a) Acoustic impedance, b) Vp/Vs ratio, c) Bulk Density, d) water saturation and the schematic of each fluid (Red represents wet gas, green is 32 ° API oil, and Blue is brine). Panel e) is the total porosity (blue) and the defined reservoir space, f) Pressure and temperature plots included in the Fluid substitution model for each reservoir.

Figure 3.6. 3D Synthetic elastic models, pane a) synthetic Inline (IL), crossline (XL), and depth slice of a) P-Impedance, b) VpVs ratio.

Figure 3.7. Synthetic Inline (IL) overlayed with the borehole reservoir model. a) Near, b) Mid, and c) Far angle stack., d) Fullstack overlayed with borehole reservoir type (red is gas and green is light oil). The tuning effect is present in each reservoir stage.

Figure 3.8. Exploratory data analysis summary, panel a) shows the histogram of the seismic attribute Table 2, panel b) shows the Manhattan correlation of seismic attributes, to show which attributes are correlated negatively (red) or positively (blue). The shape of the circle or ellipse shows the Gaussian distribution.

Figure 3.9. InLine of the synthetic model showing amplitude and texture attributes and the reservoir labels. a) Hilbert transform, b) Model labels, c) Amplitude volume technique (AVT), and d) GLCM energy.

Figure 3.10. PCA of the seismic attributes and the projection of each seismic attribute to the orthogonal components (Dimension 1 and Dimension 2). Panel a) Shows the contribution to Dimension 1—panel b) % contribution of seismic attributes to the second largest component. Panel c) Contribution of each seismic attribute represented as vectors in the PCA1 and PCA2 dimensions.

Figure 3.11. Labels for the synthetic model and possible clusters to be defined using SOM in each panel. The different colors aim to show different geobodies. Panel a) shows the ideal case, where SOMs define the ten different hydrocarbon properties included in the model. Panel b) shows a second scenario where SOMs cluster four different clusters (three different reservoirs and a cluster for shale). Panel c) shows a third scenario with only two clusters (reservoir rock and shale).

Figure 3.12. Reservoir classification and attribute importance using clusters of two, three, four, or five. Panel a) shows the reservoir separation using two clusters, and panel b) Shows three clusters with brine points overlapping in Cluster 1. In more than three clusters, the misclassification increases. Panel c) Attribute importance for the first four clusters, Cluster,2, Cluster 3, Cluster 4, and Cluster 5.

Figure 3.13. Seismic inversion results, panel a) p-impedance, panel b) VpVs, and panel c) Density. Panel d) shows the mid stack to keep for context. Panel e) RPT cross-plot of p-impedance, VpVs, and density as color code of all reservoirs.

Figure 3.14. Self-organizing mapping of three reservoirs. The top section (panel a) is the map projection of the two SOM latent axes at the time of the first gas reservoir. b) map projection of the two SOM latent axes at the time of the second oil reservoir. Panel c) 3D view of the SOMs and Panel d) 3D view of the geobodies captured by the latent space and the accurate labels to evaluate the clustering accuracy.

Figure 3.15. Displays seismic inversion results for the Stybarrow field. It includes panels for a) inverted acoustic impedance, b) VpVs, c) total stack, d) inverted bulk density, e) RPT template, and f) the captured reservoir geobody overlapping seismic amplitude extraction on the Macedon Fm.

Figure 3.16. Extracted geobodies from SOM using 12 clusters. The geobody anomaly is 21 % larger than the seismic inversion. Additionally, the SOMs clustered four different elements that haven't been produced/drilled or tested.

Figure 3.17. Cross-correlation of the three-reservoir calculated from the petro-elastic seismic inversion and the SOMs showing the spatial variability of the SOMs characterization.

Figure 4.1. Location of the two seismic surveys available to evaluate hierarchical clustering analysis in the classification of reservoir sandstones. The contrasting drilling activity results from the oil and gas fields located in the Exmouth sub-basin. In contrast, the Canning 3D dataset, and the eastern portion of the North Carnarvon Basin (NCB) is a frontier exploration area with two boreholes drilled within a 5,160 km² survey.

Figure 4.2. Available seismic and borehole data in Northwestern Australia to study the risk of sand presence. Panel a) is the original seismic data associated with the Stybarrow 3D seismic cube. Panel b) is the regional location map b.1) Seafloor structural map interpreted on the seismic data. Panel b.2) shows the seafloor structural map interpreted in the Canning 3D seismic. We limit our study laterally and the cyan polygon shows the extend of the seismic data used in this study. We interpreted the Albian-Aptian unconformity -a regional unconformity- to delimit our study vertically. The Albian-Aptian unconformity is the interval where the petroleum systems are in the NCB. Our zone of interest will include sediments from the Albian-Aptian unconformity to the Late Triassic sandstones.

Figure 4.3. Our proposed workflow has four steps, data input including 3D seismic amplitude and borehole data. Data preprocessing where seismic and borehole data are conditioned and transformed into seismic attributes and rock-types. Our main process, which includes, Exploratory data analysis (EDA) and Principal component analysis (PCA). Finally, the results, where we present the 3D volume of sand and shales and analyze the lateral seal risk. The arrows show the direction of the workflow.

Figure 4.4 Panel a) Stybarrow enhanced seismic data and wireline logs from the borehole inside the Area of interest. Gamma-ray and resistivity are overlayed to indicate shales and sandstone presence. Panel b) Canning 3D enhanced seismic data and wireline logs from the Delambre-1. Panel c) and Panel d) compare amplitude distribution against frequency before and after data enhancement.

Figure 4.5. Seismic attributes from Canning 3D. In amplitude dependent attributes, which relay on an window we use the same window (33 ms) in both seismic volumes. Fault probability using deep learning algorithm in Canning 3D used a batch of 64, while in Stybarrow used 32.

Figure 4.6. Panel a) Box-plots of the seismic attributes after normalization. Panel b) Pearson's correlation that show the statistical correlation of the seismic attributes. Panel c) Density plots of the seismic attributes. Box-plots of GLCM energy and total energy have a strong contrast, which contribute greatly to cluster differences.

Figure 4.7. Eigenvectors of the seismic volume, we observe that the envelope attribute shows the highest contribution, followed by the GLCM energy, Peak CWT magnitude, Sweetness, Variance, Total energy, and Sobel similarity at 6 Hz.

Figure 4.8. Panel a) Sketch to show HCA clustering; the schematics exaggerate the clustering to reduce the variability. Panel b) Example of links and the different levels, and how the similarity matrix assists the clustering.

Figure 4.9. Canning seismic amplitude and the Levitt-1 borehole are used to sketch an implementation of our methodology. The analysis of the lateral seal risk using shale-gauge and fault probability is a seismic scale approximation of the risk of two lithologies are in contact among a fault plane.

Figure 4.10. HCA cluster results in Stybarrow 3D to show the geological-like clusters and a comparison of multispectral similarity. Panel a) 3D view of two-time slices, the Barrow and Macedon Formation. Panel b and c) show the map view of the Formations shown in Panel a). The map view helps to show the sandstone distribution. Panel d) shows the multispectral similarity. i.e., the Sobel similarity voices blended in RGB. The calibrated clusters' detectability and separability help define sandstone presence, and the rock type definition is sharper than the Sobel Similarity blend.

Figure 4.11. Stybarrow-1 data, showing the petrophysical evaluation and rock type from the mineral evaluation. The NTG of Stybarrow-1 from 2,000 m TVD to 2,400 m is 40% Also, the rock types at seismic scale and its correlation with the HCA cluster. Cluster 1 shows excellent correlation with the seismic scale sandstone. Cluster 5 and 4 have the largest misclassification.

Figure 4.12. Seal risk analysis, Panel a) manual fault interpretation on the amplitude data and the fault probability overlapped to show the similarities between the interpreter's perception of a fault and the fault probability, red and green arrow shows the amplitude characterization Macedon Formation—panel b) Map view of the fault probability on a time slice and interpenetrated fault planes, cyan arrow is indicating the NE-SW fault plane on panels c and d. Panel c) shows HCA clusters transformed into lithologies across the reservoir; the green and red arrow now shows the Macedon Formation transformed into a sandstone cluster, and red are the limits of the lithology capture to be multiplied by the fault probability; the fault plane is in the center in black—panel d) Lateral seal risk on the Fault NE-SW.

Figure 4.13. Panel a) A 3D perspective around the Delambre-1 borehole shows the lithology distribution from the calibrated HCA cluster. Panel b) RGB color mix used in Panel c). Panel c) shows multi-spectral similarity exhibiting structural and stratigraphic features. Panel c) A map view shows the channelized features where sandstone dominates the distribution. HCA Clusters improve the interpretation in contrast with multi-spectral similarity.

Figure 14. Levitt-1 data, showing the petrophysical evaluation and rock type from the mineral evaluation. The NTG of Levitt-1 from 2,200 m TVD to 4,900 m is 60% Also, the rock types at seismic scale and its correlation with the HCA cluster. Claystones and silty sandstones shows the largest misclassification. Cluster 1 and Cluster 3 shows a good calibration with sandstones and shales. The high NTG in the Levitt-1 structure is promising yet the highly faulted nature presents a risk on the lateral the sealing.

Figure 4.15. Canning seal risk analysis depicts a high risk in Canning seals. Panel a) displays HCA classification already transformed into rock types intersected with fault probability, delineating lithology computation within the fault zone. The dashed black line to the green line is the zone where the computation of lithologies with the fault plane takes place, and the green and red arrows show the left and right-hand sides of the central fault (cyan arrow). The solid black line corresponds to the next panel. Panel b) illustrates sandstone distribution in the Lower Cretaceous via map view, revealing NE-trending clinoforms and geofoms characterized by hundreds of meters of height sandstones over tens of kilometers (magenta and green arrows) indicative of a continental deposit. Panel c) projects HCA lithologies, emphasizing more sandstone-related clusters than shale clusters (black arrows). Panel d) shows seal risk probability on the fault plane, indicating a high sealing risk in the Lower and Middle Jurassic deposits and traps due to sandstone juxtaposition.

Figure 5.1. Location map and spatial distribution of available data. Panel a) Montana geological map and the documented faults from the Geological survey of Montana. Panel b) Borehole distribution around the CO₂ sequestration project, 212 boreholes with at least triple combo wireline logs. Panel c) Spatial distribution of 35 boreholes with enough information to estimate petrophysical properties (Total and effective porosity, vertical permeability and, water saturation), including water saturation, mineralogy volumes (Calcite, Shale, Dolomite, Anhydrite).

Figure 5.2. Arbitrary seismic section on a depth domain pre-stack migration including seismic interpretation of geological formations and faults. The high-dip angle faults display the structural complexity in the area and the risk associated with fluid injections. Two reservoir flags (green and red) show the impossibility in associate injection intervals to seismic amplitude.

Figure 5.3. Our proposed Methodology, to include explainable machine learning in understanding prediction models. While Random Forest have plots to interpret the prediction and we keep a blind test borehole to see how the model performs with a blind borehole. CNN only evaluates the accuracy. However, it is important not to trust only the accuracy or loss function of the CNN prediction. Wireline data is pre-processed analyzed for missing data, augmented, and normalized. We consider subject-matter expert data, the inclusion of Sonic travel time, calcite volume, anhydrite volume and shale volume, total and effective porosity, in addition to gamma-ray, shallow and deep resistivity, density and neutron density.

Figure 5.4. Panel a) Danielson-22-17 and Panel b) Wallawein-22-1 boreholes with petrophysical evaluation. The evaluation includes lithology and fluid effects. The template includes gamma-ray and water saturation, porosity, and permeability. Volume of Illite, anhydrite, calcite, dolomite, quartz, and irreducible water. The tadpoles coming from the FMI interpretation helps generate the stereonet and analyze fracture azimuth (SW-NE 40 °) presented in panel c) and panel d)

Figure 5.5. Sw predicted vs Sw interpreted using RF (Panel a) and RNN (Panel b).

Figure 5.6. Top plot, LIME analysis plots for RNN and random forest. Bottom panel shows the important features for the random forest model.

Figure 5.7. Geostatistical models in 3D using Sw_RNN and Sw_RF in top and middle panel. Panel a) shows the water saturation estimated using RNN 3D grid. Panel b) shows the water saturation estimated using random forest 3D grid, and Panel c) shows the spatial correlation performed at each grid cell, where the value of the cell for both models is correlated. The difference can be considered as a measure of uncertainty. Observe how inside the two zones of interest (black arrows) Bakken to Potlatch Formation and Nisuk to Upper Duperow Formation the differences are larger, casting doubt on the model.

Figure 5. 8. Uncertainty plot of gas saturation. Panel a) shows the initial kernel density estimate. After assuming a prior model, the slope and intercept model parameters to the most likely linear regression, based only on the L2 norm minima of the error vector over all sample data. Noise is noise added to the synthetic data. Red and cyan arrow show the Bayesian uncertainty.

Appendix figures

Figure A.1. The scheme of seismic amplitude distribution as a probability distribution combined with the geological parameters.

Figure A.2. Geostatistical seismic inversion workflow on Jason Geoscience workbench 10.2.

Figure A.3. Changing the reservoir shape and layer distribution impacts heavily the inversion calculations.

Figure B.1. An Example of Non-Linear Transformation from Data Space (3-Dimensions) to latent Space (2- Dimensions) by embedding a manifold in data space, where $r_1, r_2, \dots, r_n$ denotes the points in latent space, $y = W.\phi(r)$ represents the mapping function.

Figure B.2. Visualization of the GTM classification of the reservoirs. In this view the highest hydrocarbon saturation of the first reservoir is classified with a large contrast with the shale background.

Figure B.3. A 3D perspective on the GTM clustering of the different reservoirs. The anomaly of the first reservoir is exaggerated vertically and beneath the wedge casting doubt on the true extension of the anomaly.

Figure B.4. Map view comparing GTM and SOM clustering classification of the synthetic data.

Figure C.1. The study area, denoted by the red polygon, represents the Exmouth Basin, the zone of interest, and the available borehole information.

Figure C.2. Vertical seismic sections. (a) Northeast–southwest section that shows (1) the Windalia Radiolarite horizon, a fossil that delimits the Mesozoic age; (2) the slope of the Tertiary Basin; and (3) the tectonic inversion and fault intensity on the Mesozoic. (b) The study area (red polygon) representing the Exmouth Basin, the zone of interest, and the available borehole information.

Figure C.3. (a) Seismic section parallel to the levee break and well-1 to calibrate the clay content and porosity. Inside the dashed black box, the chaotic reflectors of the FLT contrast with the sigmoid pattern northwest denoted by green arrows. (b) Acoustic impedance section showing the levee walls; the chaotic reflection patterns are transformed into intercalations of shales and sandstones.

Figure C.4. (a) LIDAR elevation model of the Mississippi River embayment overlapping a satellite image showing the elevation difference that allows the sediments to spill and breach the levee (<10 m). The levee breaks continuously deposit mud sediments during floods. (b) Illustration modified from Sprague et al. (2005) showing the erosion and deposition ratio as a slope function. We include the existence of preexisting faulting as a factor that contributes to canyon breaches.

Figure C.5. The high-resolution bathymetry shows a high-angle slope around the Arguello terrace. The seismic section S-S' shows the seismic profile associated with the canyons. The red box on profile C shows the feeder of the crevasse splay for this system (the green box) and it also is on the profile. The black arrows show the splay's beginning, and the red arrows show a slump on the profile, most likely due to the seismic activity in the area. Letters on the map are related to the profiles.

Figure C.6. (a) Schematic depiction of principle depositional elements in deepwater settings; the red square shows the study area of our FLT. (b) A constant time slice with the surface of interest denoted with opacity to show the area of improvement by using a stratal slice. (c) The surface of interest; the red arrow shows the canyon walls and points out the direction of the sediment supply. The black arrows and dashed red lines emphasize the levee breach. According to the sketches, at this location of the canyon, sediments ought to be contained. (d) The amplitude stratal slice of the canyon reflector. The crevasses splay signature is enhanced, which invites flattening of the seismic volume and the follow-up attribute calculation. The blue line is the direction of the seismic section in the following panels. (e) Seismic section parallel to the canyon shown in a 3D view to enhance the canyon's slope and the levee's breach. A black circle shows the area of interest on the next panel. (f) The interesting interval and the black surface's chaotic reflectors. Cross confirmed by red, and blue lines represent the location of the black circle. The red line has a thickness of 85 m and the blue line represents 200 m.

Figure C.7. (a) Multispectral decomposition of seismic amplitude showing braided channels associated with the splay. (b) Multispectral coherence corendered with amplitude exaggerating the break of the levee and the braided nature of the crevasses splay. The yellow and white arrows show features of interest and open different questions for interpretation.

Figure C.8. The amplitude section flattened to exaggerate the incisions created by the can- yon and the chaotic fills limited by strong parallel reflections. Dashed discontinuous lines are to showcase the termination of the fillings against the reflectors. The dashed black box corresponds to the levee break, both in the section and on the map.

Figure C.9. (a) Seismic amplitude corender with K1 and K2. (b) Horizon slice extraction of two attributes: flatness and curvedness; most of the extraction is white, but the red value of flatness shows the center of the feeder channels. (c) Acoustic impedance and coherency corendered to represent the crevasse splay, and high values of impedance indicate shales. (d) Envelope corendered with Sobel similarity to show the canyon walls and the edge of the crevasses splay. High-amplitude values are most likely associated with sandstones.

Figure C.10. CNN fault prediction showing the fault intensity of the rift stage. This fault system hardly continues the stability of the drift of the Tertiary.

Figure C.11. Voxel representation of the facies that delineate the submarine can- yon, the crevasses splay, and the sand overbank. The facies are well identified using seismic attributes and geobody extraction methods.

Figure D.1. Panel a) Zoom of the location of the datasets used to analyze volcanic intrusions. (d) Zoom of the Gulf of Mexico ocean bottom to show the location of the seismic dataset. Example A is survey B-38-94-LA-3D and Example B is survey B-69-94-LA-3D, both surveys are not used in the process but help as examples of salt intrusions geometries.

Figure D.2. Panel a) High amplitude chaotic reflections inside the Kora volcano, side chamber also present. Panel b) Null contrast with high amplitude on top of the salt dome, a parallel reflector crosscuts the salt dome and mimics the ocean floor. The effect is known as bottom simulating reflector or BSR. The BSR appears in porous rock, indicating the salt thickness delineating the salt dome feeder.

Figure D.3. Panel a) Salt dome is associated with a complex fault system; the ocean bottom reflects the nuance of faulting around the salt dome. Internal reflections inside the salt dome are null or the result of multiples in the data. The strata around the salt dome, particularly to the left, show an amplitude increase due to the differential compaction. Panel b, on the other hand, regardless of the larger accommodation space due to the half-graben present in the area, exhibits no differential compaction, probably due to the usual compaction trend.

Figure D.4. a) High-low dip angles on top of the salt dome, faults are inferred, and exhibit a low amplitude contrast. The lower limit of the salt dome is unclear. b) A complex salt sheet with two different structural heights affected the strata above it with complex faults. Some of them continue to the ocean floor. The salt sheet thickness can be hypothesized from 0.5 to 0.8 seconds. c) two different salt shapes, a salt dome affects the ocean floor to the left, and it can be observed with a possible salt feeder. To the right, a salt canopy with a possible salt welded fault, the effect of the salt's arbitrary shape can also be observed on the strata deformed above. The ocean floor is almost undisturbed, except for the structural difference of the salt canopy. d) The most evident effect on the seafloor in GoM B the normal faults due to collapse (synthetic and antithetic faults) on the sediment above the salt dome. An evident seafloor deformation can show that the fault has similar dip magnitudes but opposite dip azimuth.

Figure D.5. The seismic appearance is associated with volcanic intrusions. Panel a) Kora arbitrary seismic line across the Kora intrusion. Yellow arrows highlight a few sedimentary structures. In particular, the thalweg – the central portion of the channel of most extensive channel present in the system correlates with a fault beneath it (channel centers are most likely to occur above a fault?). Panel b) Nimitz seismic section across the volcanic intrusion. Yellow arrows show different volcanic intrusions. The red rectangle is a zoom for the same seismic section to clearly show the faults above the volcanic intrusion.

Figure D.6. 3D projection of a self-organizing map with a 6x6 topology showing clearly the nodes associated with faults. Mainly for these datasets, nodes, or neurons 11-12 and 17-18 have the highest probability associated with faults.

Figure D.7. Panel a) Time slice of the SOM fault classification associated with a salt dome in the Gulf of Mexico. (b) Horizon slice keeping the SOM nodes associated with faults opaque and corendering with Sobel similarity. On both slices, radial faults are evident. Panel (c) time slice of Self-Organizing Map on a different dataset on the Gulf of Mexico; in particular, the GoM B dataset confirms the salt intrusion on the seafloor. The ocean bottom collapsed and subserved deepwater sediment distribution. Panel (d) shows the horizon slice with better visualization of the radial faults, particularly the center of the fault collapsing.

Figure D.8. a) Time slice of the SOM fault classification associated with a volcano in the New Zealand basin. b) Horizon slice with the same SOM fault classification. Radial faults are present in both slices, yet more evident on the horizon slice. Panel c) Time slice of Self-Organizing Map on the Kora volcano. Panel d) Horizon slice shows the radial pattern with a predominant fault on the southeast portion of the volcano cinder.

Figure D.9. The same dataset, GoM B presented in two different perspectives. Panel a) Voxel extraction of the fault system associated with selected SOM neurons. Panel b on the right shows the versatility of automatic fault extraction; using the selected SOM neurons, the fault planes are captured for the entire system and area of study. In Panel b) Fault planes are filtered. The filter removes low fault dip angle fault planes and faults with azimuth parallel to NE-SW. the high fault dip magnitude angle and similar fault dip azimuth are preserved.

Figure D.10. a) Voxel extraction of the fault system associated with selected SOM neurons in a volcanic environment and the horizon slice with the SOM extraction b) demonstrates the versatility of automatic fault extraction. We capture the fault planes for the entire system using the selected SOM neurons, and using automatic fault extraction, we can extract specific values for each fault plane. For example, the top of the volcano with faults with a dip higher than 70 degrees is kept visible.

Figure D.11. a) Mount Taranaki overlaid with dip azimuth and hillshade and elevation, an image from the Landsat 8 and its sensors. A hyperbola separates the mount as it does in the Kora volcano. The radial faults present contrast with the drainage patterns very subtle. b) Kora volcano has a fault dividing the caldera next to the central cinder cone. The radial fault patterns are like the Taranaki Mount and the salt intrusion with collapsing blocks near the center of the intrusion. c) Zorghos salt intrusion and the fault trend mirror the salt intrusion on the Gulf of Mexico with a central portion collapsed by salt, an aperture, and a system of antithetic and synthetic faults. d) The salt dome triggers the seafloor collapse, which is dynamic; while the overburden increases around the salt diapir, it is constantly uplifted.

Figure D.12. Mars intrusion, local Mars time 15:23 acquired on 22 November 2006. Flight altitude 285.8 km. Solar latitude of reference 139 °, northern summer. Each pixel represents 57.2 cm, and the north is up. The geo (mars) morphology of the features inside the black box highlighted by the hillshade invites interpretation. In the blue box, the hillshade attribute helps to exaggerate the presence of a channelized system, or canyons maybe? Inside the yellow rectangle which intrusion is more likely? The pockmarks show the presence of craters from an asteroid impact. An alternative interpretation could be gas escaping the subsurface.

Figure E.1. a) Bulk's modulus versus porosity distribution for the minerals mixed in Table 2. The selected distribution is the modified Hashin-Shtrikman model. b) Shear modulus distribution against porosity. c) The result of fluid substitution for the given different fluid cases.

Figure E.2. Comparison of three clustering evaluation techniques. These three techniques yield a different suggestion for the number of clusters in the dataset. The red square is where the derivative of the elbow plot shows a constant. The elbow plot suggests 12 to 14 clusters, and the rest of the plots suggest 2 to 3 clusters.

LIST OF TABLES

Table 2.1. Partial stack derived from the RPM forward modeling. The combination of these partial angle stacks yielded the most significant reservoir detection and correlated better with the observed gathers.

Table 3.1. Key Stybarrow field metrics depict the properties of the reservoir and its performance.

Table 3.2. Seismic attributes found in the literature review were used for reservoir characterization and fluid discrimination (Excluding AVO attributes and inversion attributes).

Table 3.3. Seismic attributes suggested by the PCA analysis and the display of the presence of different fluids in different reservoir rocks.

Table 3.4. Reservoir characterization correlations of each method, the petroelastic seismic inversion from the SSI, correlate most with the actual reservoir models.

Table 4.1. Petrophysical parameters to discretize the wireline log data that reads different parameters of the rocks into rock types.

Table 5.1. Accuracy comparisons between the different quality/quantity datasets (Tier 1 represents more borehole data and less and raw wireline logs, Tier 2 represents less borehole data but more and better wireline logs). Simulation ranking # 3, random forest on Tier 1 has an ambiguous R2, since it is a statistical measure of the goodness of fit of a regression model, having a 0.5 is deceiving.

Table A3.1. Minerals and percentages that constitute the model sandstone and the elastic moduli of each mineral are used to calculate dry and saturated velocities as a function of porosity.

ABSTRACT

Oil and gas exploration and production struggles with meeting energy needs while minimizing environmental impact. Amidst artificial intelligence (AI) and machine learning (ML) to expedite energy exploration and minimize risks associated with energy investments through enhanced predictive accuracy. The application of AI/ML holds promise in expediting the identification of geological facies linked to reservoir rock and delineating seismic faults at seismic scale or generated 3D velocity models based on image and not forward modeling. However, before fully harnessing AI/ML capabilities, it's imperative to rigorously assess its ability to address these challenges with precision and confidence. This study aims to evaluate how machine learning can augment geoscience practices by enhancing accuracy, managing a multitude of 3D seismic attributes, and overcoming limitations in real-world interpretation.

The work is divided into four chapters, each exploring different algorithms and methodologies. Chapter 2 uses generative adversarial networks (GANs) with geostatistical seismic inversion to characterize the Ray Reef gas storage area and generate alternative reservoir models. Chapter 3 compares unsupervised machine learning techniques with established geophysical methods for capturing reservoir facies and fluid effects using seismic attributes, validated on synthetic and field data. Chapter 4 employs unsupervised machine learning to classify lithologies and estimate fault probability in a large offshore seismic dataset, evaluating lateral seal risks. Finally, Chapter 5 delves into explainable machine learning (LIME) to understand uncertainties in 3D subsurface water saturation modeling, comparing random forest and recurrent neural network predictions.

The combined findings of this research underscore the significant potential of machine learning techniques to enhance various aspects of oil and gas exploration and production

workflows. By leveraging the power of generative adversarial networks, the study demonstrates their effectiveness in generating alternative reservoir models, providing valuable insights for reservoir characterization and risk assessment. Furthermore, the application of unsupervised learning algorithms showcases their capability in identifying reservoir facies and fluid effects from seismic data, offering a powerful tool for reducing interpretation uncertainties. The integration of explainable machine learning techniques, such as LIME, sheds light on the decision-making process of complex models, enabling a better understanding of subsurface water saturation modeling and associated uncertainties. Collectively, the results highlight the transformative impact of machine learning in geoscience, paving the way for more efficient and informed decision-making processes in oil and gas operations while mitigating environmental risks.

Throughout the dissertation, the importance of integrating domain knowledge, rock physics principles, and synthetic data generation is emphasized to validate machine learning models and ensure physically meaningful results. The work highlights the potential and limitations of these techniques, advocating for a balanced approach that combines data-driven methods with domain understanding in geoscience applications.

CHAPTER I: INTRODUCTION

Oil and gas exploration and production are notoriously boom and bust. Today's economy demands access to reliable energy sources at the same time it favors an energy transition, and in major cities around the world the inhabitant's perception condemns oil and gas activities. The interplay between politics, energy and human behavior is complex to explain. Machine learning and deep learning (artificial intelligence in general) boom and bust is probably easier to explain, with excessive optimism often leading to cycles of hype and disappointment.

While machine learning became a separate field of study in 1956 at summer conference. It has been christened as a method in which machines would be as intelligent as humans within a generation. Nonetheless, since 1956 machine learning has had winters and springs, and not until the past two years when large language models have reignited artificial intelligence, however this time, key scientific figures have spoken against its unchecked development.

In this dissertation, I use machine learning and deep learning as tools, evaluating their accuracy while recognizing their value in analyzing large data sets. However, at this point, the Moravec's paradox still limits machine learning applications in geosciences – machine learning excels in finding correlations within datasets, but it often fails to understand the most valuable inputs and the meaning of the results. This is not solely a problem of machines; Archie's equation in estimation of water saturation, derived from resistivity measurements of Gulf of Mexico sandstone plugs, has been used for over 70 years and with time, the physics of the problem had to catch with empirical formula. Furthermore, Archie's equation is useful in carbonate rocks if the formation is relatively homogeneous and rock and fluid properties are well understood.

The chapters that follow address and contrast the use of machine learning and deep learning algorithms, from generative adversarial networks in the second chapter to recurrent neural

networks in the last chapter. In my evaluation, I use well-cemented physics concepts in wave propagation and rock physics to generate synthetic models, testing the machine learning algorithm validating the outputs from in machine learning.

In Chapter 2, I used geostatistical simultaneous seismic inversion to characterize the Ray Reef gas storage area in terms of p-impedance and V_p/V_s ratio. I then generated 100 realizations of the reservoir with different facies, fluid probabilities, and elastic property variations. These reservoir scenarios were used to train a generative adversarial network (GAN), allowing the trained GAN model to generate additional synthetic reservoir scenarios. While this approach enables the creation of alternative reservoir models from a single experimental dataset, it raises questions about how much we can trust these GAN-generated "fake" data, as they are produced under specific conditions and epistemic uncertainties.

Building on this, Chapter 3 focuses on evaluating the accuracy of unsupervised learning techniques in capturing reservoir facies and fluid effects using seismic attributes. I created a 3D synthetic seismic dataset with two common lithologies (shale and sandstone) and three fluid zones, using rock physics modeling and Gassmann fluid substitution. This allowed me to test and compare unsupervised machine learning methods against a well-established geophysical approach, both on the synthetic data and on an offshore reservoir in Australia. The goal was to develop a novel analysis technique to assess the capabilities of unsupervised learning in delineating reservoir facies from seismic data.

Chapter 4 demonstrates the benefits of using unsupervised learning to classify lithologies over large surface areas where borehole control is limited. Specifically, I applied unsupervised learning to the Canning 3D seismic dataset offshore Australia, which covers an area of 3,527 km² but has only two boreholes. Additionally, this chapter estimates fault probability and combines the

lithology classification and fault mapping results to evaluate lateral seal risk by assessing the contact between permeable lithologies on either side of the estimated fault planes. The results were calibrated using nearby field data with abundant borehole control.

In Chapter 5, I explore the use of explainable machine learning (XAI) to understand the uncertainties in 3D subsurface water saturation modeling. I combined borehole data and seismic data to create a petro-facies model and then employed local interpretable model-agnostic explanations (LIME) to identify the most important inputs for predicting water saturation. This allowed me to compare the interpretability of random forest regression and recurrent neural network predictions, as the basis for random forest is more intuitive, while neural networks can be obscure. Explainable machine learning is important to understand the predictions, as the underlying basis of random forest regression is more intuitive, while neural networks can be more obscure. The use of local interpretable model-agnostic explanations (LIME) helps remove the ambiguity, allowing us to better comprehend the "black box" between the inputs and outputs, and identify the dragons, as the old saying goes: "There are inputs, there are outputs, and between, there be dragons."

These chapters exemplify when machine learning is overreaching, highlighting the need to consider how confident we can be about its results, and when those results can be deemed meaningful. Importantly, the work explores methods to enhance our understanding of machine learning outputs.

Ultimately, this dissertation underscores that a complete analysis of the appropriate tools to use when addressing geophysical problems requires a careful decision-making process that weighs the risks and benefits. While machine learning and deep learning offer powerful capabilities for analyzing large datasets and identifying correlations, their results must be evaluated

and interpreted within the context of established physical principles and domain knowledge. Blindly trusting these techniques without rigorous validation and a deep understanding of their inner workings can lead to erroneous conclusions or misinterpretations, particularly in complex domains like geosciences.

The core theme of this dissertation is the importance of integrating domain-specific knowledge and physical principles into machine learning workflows in geosciences. By generating synthetic models based on well-established rock physics and wave propagation concepts, and by employing explainable AI techniques, this work aims to ensure that machine learning models not only achieve high predictive accuracy but also provide interpretable and physically meaningful results, bridging the gap between data-driven models and domain understanding.

CHAPTER 2: GENERATIVE ADVERSARIAL NETWORKS TO ASSIST SUBSURFACE GAS STORAGE CHARACTERIZATION USING GEOSTATISTICAL SEISMIC INVERSION- AN EXAMPLE FROM RAY REEF, MICHIGAN BASIN, USA.

**To be submitted to Expert Systems with Applications, an Elsevier Journal.*

ABSTRACT

Defining the reservoir extensions and the most suitable units to storage or product fluids from is crucial in energy industry. To accomplish the definition, it is necessary to estimate the boundaries of the reservoir and the inner flow barriers, as well as rock types with favorable conditions, i.e. good porosity, good permeability. As data is gathered, diverse methodologies aim to minimize uncertainty as reservoir engineers plan based on initial findings and update storage projections. However, gas storage in carbonate rocks poses a multifaceted challenge due to various factors impacting capacity. Many techniques have been devised to reduce uncertainty regarding the porosity-permeability relationship, yet it persists as a complex issue. Generative adversarial networks and other diffusive models are an active area of research to bypass the computational time it takes to make a reservoir model or in other areas of application a starting velocity model of the subsurface. In both examples it requires a starting point to start the learning and generation of the machine learning models. Here, a reservoir characterization methodology integrates rock physics modeling, geostatistical seismic inversion to generate an image of the subsurface, and generative adversarial networks to evaluate seismic attributes and identify high porosity-permeability facies in a Paleozoic bioclastic accumulation in the Michigan Basin.

Initially, a rock physics model (RPM) analyzes the elastic response per litho-type and productivity facies on Ray Reef, a mature carbonate field. Subsequently, geostatistical inversion delineates the extension of litho-types with optimal storage performance. Lastly, generative adversarial networks (GANs) produce scenarios representing the non-linear relationship between porosity and storage capacity within the reservoir. Results are validated against available borehole

data and blind borehole tests to gauge GANs accuracy. A practical data integration approach is employed to identify zones within the reservoir with the most probable productivity index, thereby reducing uncertainty. In conclusion, the combination of geostatistical inversion and GANs yields a reliable spatial distribution of gas storage performance within the reservoir based on established gas storage profiles.

INTRODUCTION

Gas storage helps locally by facilitating energy during low temperatures at high latitudes, and globally, it will also contribute to the atmospheric CO₂ reduction quota. Ideally, the amount injected should be preserved and extracted integrally. Nonetheless, ambiguities in the subsurface model can result in differences that require careful monitoring. These differences are mitigated with careful reservoir characterization analysis and are abated as the model integrates more information over time (Caetano, 2012; Adegbite et al., 2021; Avseth and Lehocki, 2021). This paper addresses the gap by integrating geostatistical elastic seismic inversion and generative adversarial networks to make a 3D storage capacity model based on pixel-based simulation on the Ray Reef complex. This allows the transition from deterministic porosity-permeability models, often an extrapolation from borehole data (Haynie, 2009; Caetano, 2012; Rine et al., 2017; Corte et al., 2020), to storage models based on elastic properties.

The study area selected to demonstrate this workflow is a depleted gas field located in Macomb County, Michigan, 25 miles northeast of Detroit, Michigan (Figure 2.1). The structure has an area of 2 square miles and a 580–600-foot thickness. Ray Reef is located on the Southern Reef Trend, characterized by its lower reef density in contrast with the Northern Reef Trend (Armstrong and Carter, 2010; Rine, 2015; 2017; Nadhim, 2019) in the Michigan Basin.

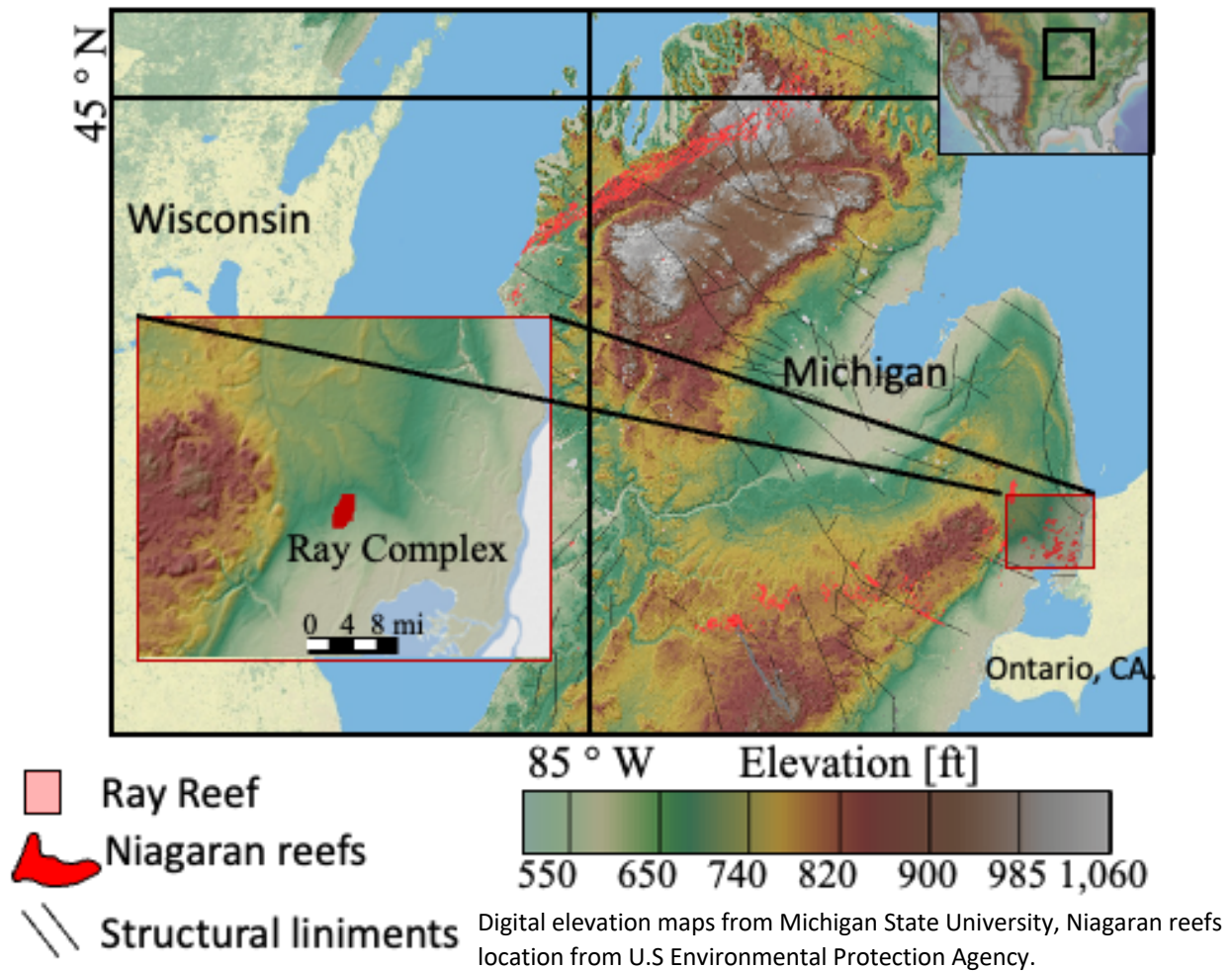


Figure 2.1. Location of the study area including the distribution of Niagaran reefs in Michigan. The contrasting reef density between the Northern Reef Trend (Gill 1973, 1977a, 1977b) and the Southern Reef Trend is primarily a consequence of the tilted basin. The overlapped elevation map mimics paleo topography during the time of deposit and shows higher structural elements in the north. The structural liniments show a relatively stable basin.

GEOLOGIC BACKGROUND

The Ray Reef is one of the 1,100 pinnacle reefs formed during the Silurian in the Michigan Basin (Wold, 2008; Trout, 2012; Silver et al., 2021). These reefs were deposited during the Niagaran and mid-late Silurian mainly in two bands: Northern and Southern Reef Trends. Both trends are at the limits of the basin and the result of the failed Midcontinent Rift (Armstrong and Carter, 2010). At the time of deposition, the basin was situated south of the equator (latitudes

around 20-25° S), with water depths ranging from a few feet to 200 feet at the depocenter of the basin. Equatorial warm and calm waters encouraged bioaccumulation and reef growth alongside sediments and pre-existing structural heights. However, rapid relative sea-level changes at different episodes promoted vertical growth and controlled sediment deposition, creating a well-defined reservoir architecture (Figure 2.2a) and different scour and exposure surfaces that serve as flow barriers and control fluid mobility within the reef.

Ray Reef's shape is strongly related to two factors: (i) paleowind direction and (ii) relative-sea level changes. While paleowind direction shaped the steep southern flanks (~60°) in contrast with the gently dipping north flank (~15°), fast relative-sea level changes triggered the reef growth in the vertical direction compared to the horizontal (Rine et al, 2017; Grammer et al., 2018). The reef developed predominantly in water depths less than 50 ft (Jodry, 1969). However, Wold (2008) observed grainstone deposited between 3 and 33 ft and aerial exposed surfaces. Other observed facies, such as stromatoporoid, were found in the core description, and these are typically deposited in water depths closer to 200 ft. In general, the lithologies and facies found within the reef are summarized in Figure 2.2b.

The Ray Reef is covered entirely by the carbonate and evaporite deposits of the Salina Group, which completes the sequence of deposits and reservoir architecture presented in the model (Figure 2). While these sequences are essential from a geological standpoint, a more detailed understanding of the diagenesis is required to predict reservoir performance, especially the best flow zones (Ahmed, 1946). Therefore, the classification herein proposed is based on productive rock types discretized using the productivity index calculation. Ray Reef's shape is strongly related to two factors: (i) paleowind direction and (ii) relative sea-level changes. While the paleowind

direction shaped the steep southern flanks ($\sim 60^\circ$) in contrast with the gently dipping north flank ($\sim 15^\circ$), fast relative sea-level changes triggered the reef growth in the vertical direction compared to the horizontal. The reef developed predominantly in water depths less than 50 feet (Jodry, 1969). However, Wold (2008) observed grainstone deposited between 3 and 33 feet and aerial exposed surfaces, indicating periods of shallower water depths and subaerial exposure. Other observed facies, such as stromatoporoids, were found in the core description, and these are typically deposited in water depths closer to 200 feet, suggesting periods of deeper water conditions. In general, the lithologies and facies found within the reef are summarized in Figure 2b and reflect the varying depositional environments and sea-level fluctuations during the reef's growth.

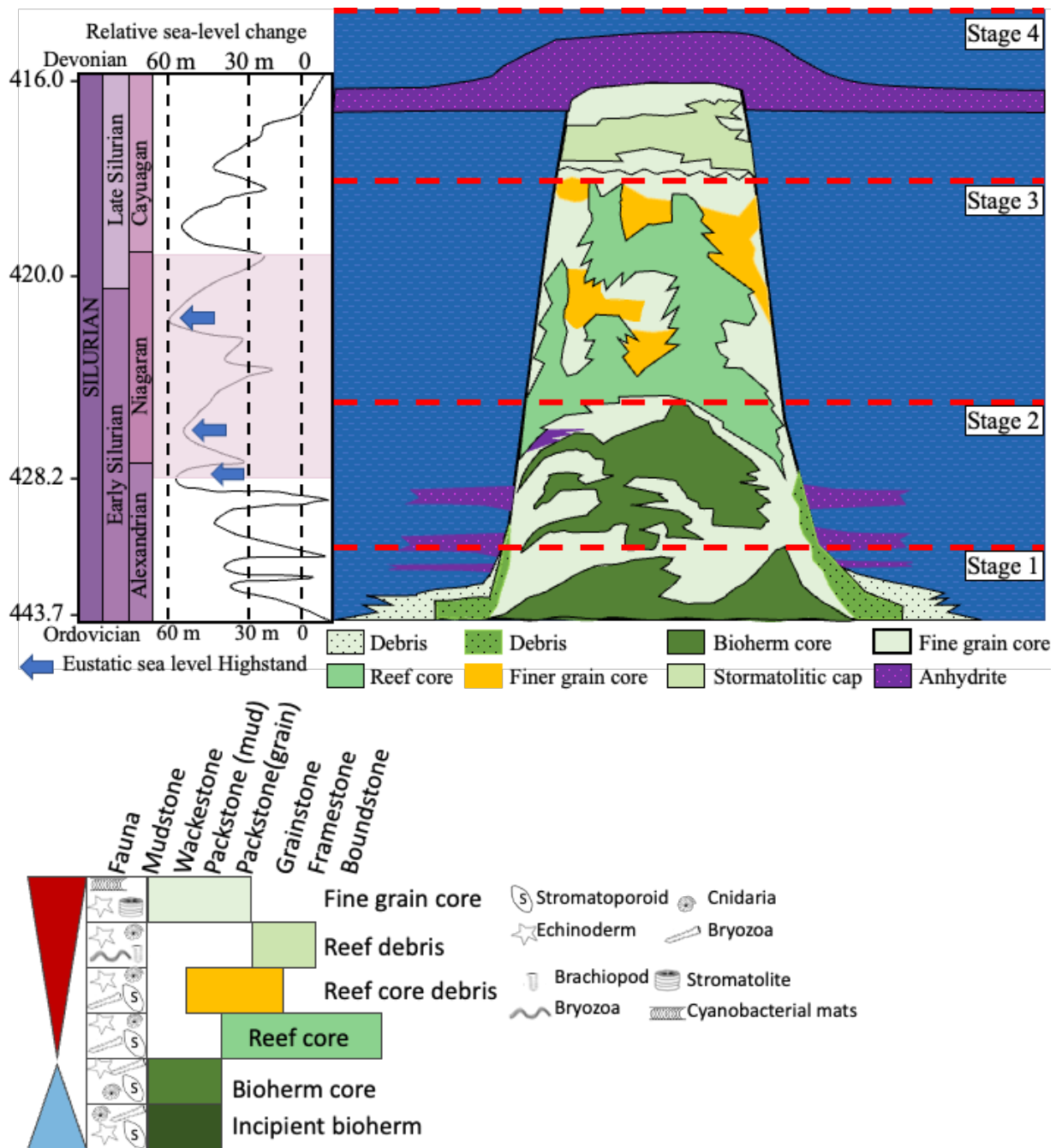


Figure 2.2. Panel a) A schematic view of Reef internal architecture tied with the relative sea level curve. Facies, fauna, and lithologies found in the reef are described in Panel b). (Figure modified from Gill, 1973; Rine, 2016).

Ray Reef has proven to be a prolific gas reservoir, with nearly 36 BcF produced out of the initially estimated 50 BcF. Presently, operations include 49 gas storage wells and 17 observation wells. Over the last five decades, Ray Reef has undergone extensive characterization through 3D

geological models (Harmon et al., 1978; Howell et al., 1999; Armstrong and Carter, 2010; Rine, 2016; Rine et al., 2018; Silver et al., 2021) continually updating lithofacies, reservoir architecture, permeability, and porosity models.

PREVIOUS CHARACTERIZATION OF RAY REEF

The most recent reservoir characterization studies have advanced significantly in high vertical resolution, employing 3D geostatistical modeling to address facies distribution, permeability, and porosity (Mesolella et al., 1974; Haynie, 2009; Rine, 2015; Rine et al., 2017; 2020). These studies have shown marked improvements from earlier works (Sears and Lucia, 1979; Wold, 2008) to the latest findings by Rine (2020). The advancements are attributed to enhanced core description techniques and the application of Sequential Gaussian Simulation (SGS), which effectively incorporates constraints such as sequence boundaries serving as permeability barriers.

Moreover, Rine (2020) integrated the facies model with production history data, providing further insights into reservoir behavior. Porosity and permeability averages across these studies exhibit similar ranges: porosity ranging from 2.5% to 30% with a mean of 12%, and permeability following a logarithmic distribution from 0.0001 mD to 1000 mD, with a mean of 100 mD. Importantly, where borehole data is available, distinct porosity and permeability distributions are observed for each facies (Figure 2.3). This comprehensive characterization enhances our understanding of Ray Reef's reservoir properties and aids in effective reservoir management and production optimization.

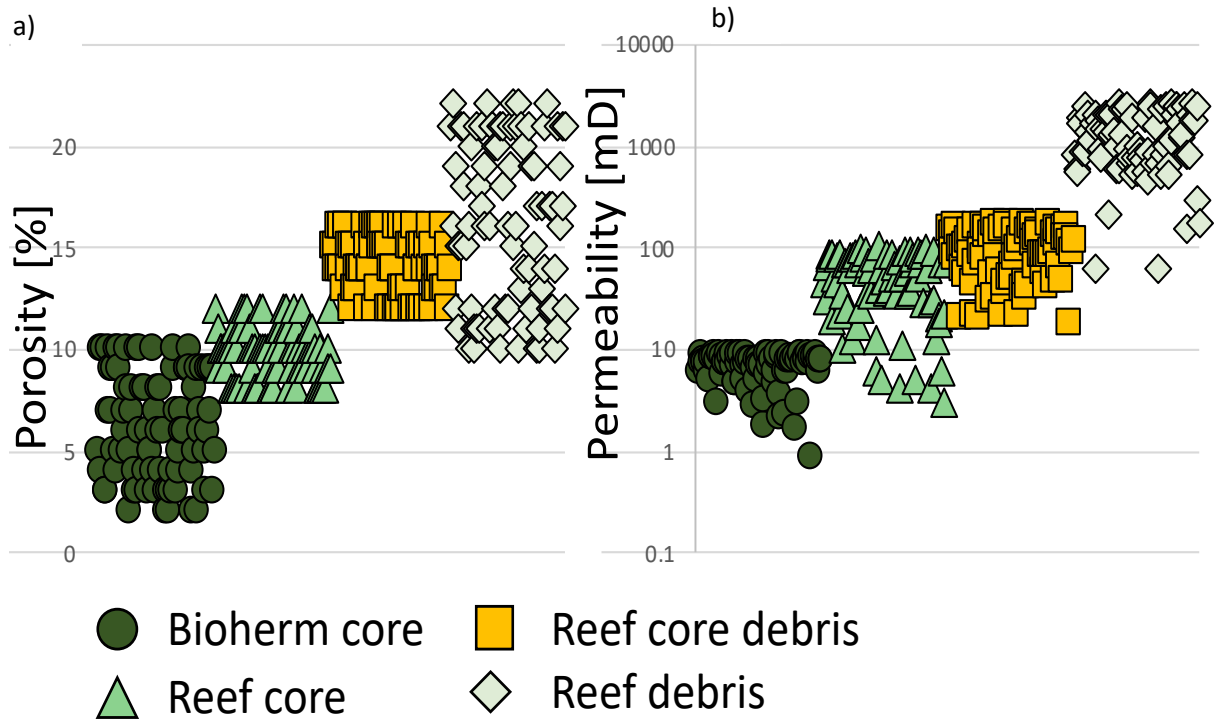


Figure 2.3. Panel a) shows the porosity distribution per facies and Panel b) shows the permeability distribution per facies.

In contrast, Silver et al. (2021) introduced flow units derived from a non-linear correlation between porosity-permeability and seismic attributes. Employing generative topographic mapping (GTM), a semi-supervised technique, they addressed the question of whether a feasible relationship exists between flow units and seismic data (Prasad et al., 2003; Daley et al., 2006; Barajas-Olalde et al., 2019; Carcione et al., 2020; Adegbite et al., 2021). The answer, however, proved elusive, as Silver et al., (2021) encountered challenges in creating a high-resolution flow facies model. Similarly, Buist et al., (2021) utilized self-organizing maps (SOMs) to delineate seismic facies correlated with higher permeability observed through borehole data, and Sams et al. (2017) showed the viability of seismic inversion on a deeper carbonate buildup. Maurya and Sarkar (2016) compared the post stack inversion onshore, finding geostatistical inversion the most accurate and most complex method.

In Buist et al. (2021), the SOMs incorporated eight seismic attributes to explore sub-seismic reef reservoirs in the Michigan Basin, our attempt uses fifteen seismic attributes, 6 non-linearly correlated. In this study, a combination of geostatistical seismic inversion (GSI) and generative adversarial networks was employed to pinpoint the most productive zones within the reef. The productivity index, akin to Al-Jawad (2018), measures a borehole's capacity for production or injection based on pressure and flow rate. While seismic data does not directly sample flow zone indicators, it captures changes in elastic parameters (p-impedance, s-impedance, and density). Therefore, rock-physics models facilitated the connection between the productivity index and seismic data.

GSI was harnessed within the hydrocarbon patch to generate highly detailed models of flow zones, necessitating dual or triple parameters such as p-impedance, s-impedance, and density to discern the facies of interest (Caetano, 2012; Sams et al., 2017; Avseth et al., 2021). The final step involved training and implementing a machine learning algorithm, specifically GANs. Machine learning has seen exponential growth in the past decade, with deep learning and shallow learning being two primary classifications based on the number of layers. While shallow learning comprises only a few convolutional layers, deep learning may feature 24, 48, or 64 layers, depending on the network architecture. GANs, a deep learning technique, generate multiple realizations of the reservoir based on image training (Goodfellow, 2016; Goodfellow et al., 2016) and integrate the outcomes into the productivity index with the highest likelihood. This combined methodology suggests borehole connectivity calibrated with borehole history, enhancing our understanding of reservoir behavior, and aiding in optimal reservoir management.

DATASET DESCRIPTION

Seismic data gathered at a 1 ms sampling rate (500 Hz Nyquist frequency) became available in 2019, covering an area of 61 mi² with a bin size of 82x82 ft and a recording time of 2 seconds. This seismic acquisition effectively captured the reef reservoir, which holds 90% of its capacity, equivalent to 64.5 BCFs. To process this data, ray tracing calibrated with borehole data facilitated the stacking of time Kirchhoff migrated gathers into seven partial angle stacks.

Moreover, the study area benefits from data obtained from over 80 boreholes. Among these, 27 boreholes are equipped with core plugs, allowing for laboratory measurements of permeability and porosity within the reef. The remaining boreholes are equipped with wireline logs, providing data such as gamma-ray (GR), resistivity, and spontaneous potential (SP). A small subset of boreholes (5%) also includes compressional wireline logs and density measurements at the same intervals. The distribution of this available information is illustrated in Figure 2.4.

The scarcity of wireline logs is mitigated largely through the utilization of a rock-physics model, which incorporates a comprehensive geological understanding of the mature gas field. This approach ensures a more comprehensive characterization of the reservoir despite limited wireline log availability.

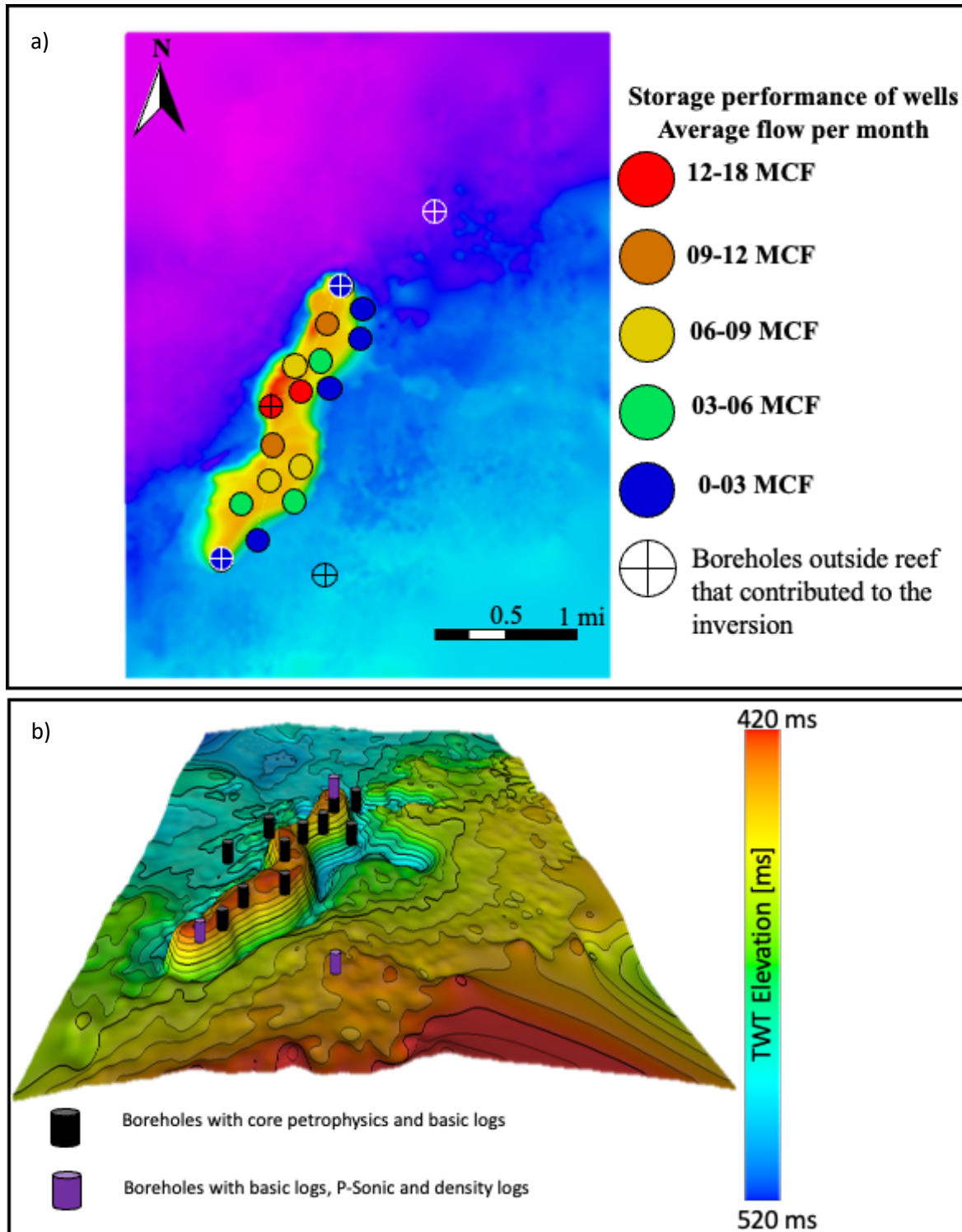


Figure 2.4. Panel a) on the left shows the map view of the Ray Reef top (A-2 Carbonate) with a distribution of borehole data and the gas storage performance clustering. Panel b). Isometric view of the presence of boreholes and how these boreholes cover the reef structure thoroughly.

METHODOLOGY

The generative adversarial network is a pixel-to-pixel generator, and it trains a convolutional neural network to replicate the targeted pixel distribution. The target is defined during the training and the network is as accurate as the target input. While we can use seismic attributes or sequential Gaussian simulation to generate an image of the reservoir, in one hand, seismic attributes are susceptible to different properties and calculations, on the other hand sequential gaussian simulation requires a careful data integration and is dependent on the spatial data density. Therefore, we choose geostatistical seismic inversion to generate an accurate map of the reservoir which is used to train the generative adversarial networks. At the core of this workflow lies the integration of borehole data with seismic amplitude information. Various methods have been developed to alleviate the scale problem (Kallweit and Wood, 1982; Avseth et al., 2005; Chopra et al., 2006; Ba et al., 2014; Nanda, 2016; Hong Bing et al., 2019; Tian et al., 2022). However, the vertical resolution of seismic data at this depth is 16 feet, using Rayleigh's criteria. The definition of the vertical resolution impacts the productivity index definition. Since the physical properties below 16 ft are unrelated to seismic amplitude, the productivity index calculated at each borehole is upscaled to 16 ft, and a productivity facies below this threshold is averaged.

The proposed workflow for correlating generative adversarial networks to simulate reservoir distribution from seismic data is outlined in Figure 2.5, beginning with data loading. After interpretation of the reservoir and major geological formations, the workflow comprises three main stages and each stage has important steps:

- Stage 1: Productivity Index calculation
 - At the borehole level, the productivity index is calculated where data is available. The main drivers are permeability and porosity measurements.

- S-wave estimation, AVO/AVA analysis, and partial angle stack definition
- Stage 2: Geostatistical inversion
 - Calculation of an accurate image of the reservoir. Here, we choose to use geostatistical seismic inversion to calculate the facies and fluid probability, both methods being Bayesian.
- Stage 3: Generative adversarial network
 - The reservoir volume is sliced vertically, and the sliced images of the reservoir are used as training data for the generative adversarial network.

An alternative approach is to use seismic attributes. Nonetheless, seismic attributes show an ambiguous response to fluid and lithology of the subsurface. Buist et al. (2022) and Maas et al. (2023) used unsupervised learning to define reservoir properties on carbonate rock. However, we choose to use a seismic inversion process to calculate p-impedance, s-impedance, VpVs ratio, and density. This approach provides physical properties linked to the four main properties of the reservoir: fluid saturation, fluid type, lithology, and porosity. These four volumes are entered into a seismic reservoir engine. We choose the facies and fluid probability as a Bayesian inference method of the reservoir probability. Arevalo-Lopez et al. (2016) used petro-elastic seismic inversion, where the elastic properties are used to calculate water saturation. Alternatively, we generated an image of the reservoir using self-organizing maps from seismic attributes. However, it was not used for training the generative adversarial network. The combination of geostatistical seismic inversion and facies and fluid probability reduces the uncertainty in reservoir presence and distribution. This structured approach facilitates the extraction of valuable insights from the integrated data, enabling a comprehensive understanding of reservoir characteristics and behavior.

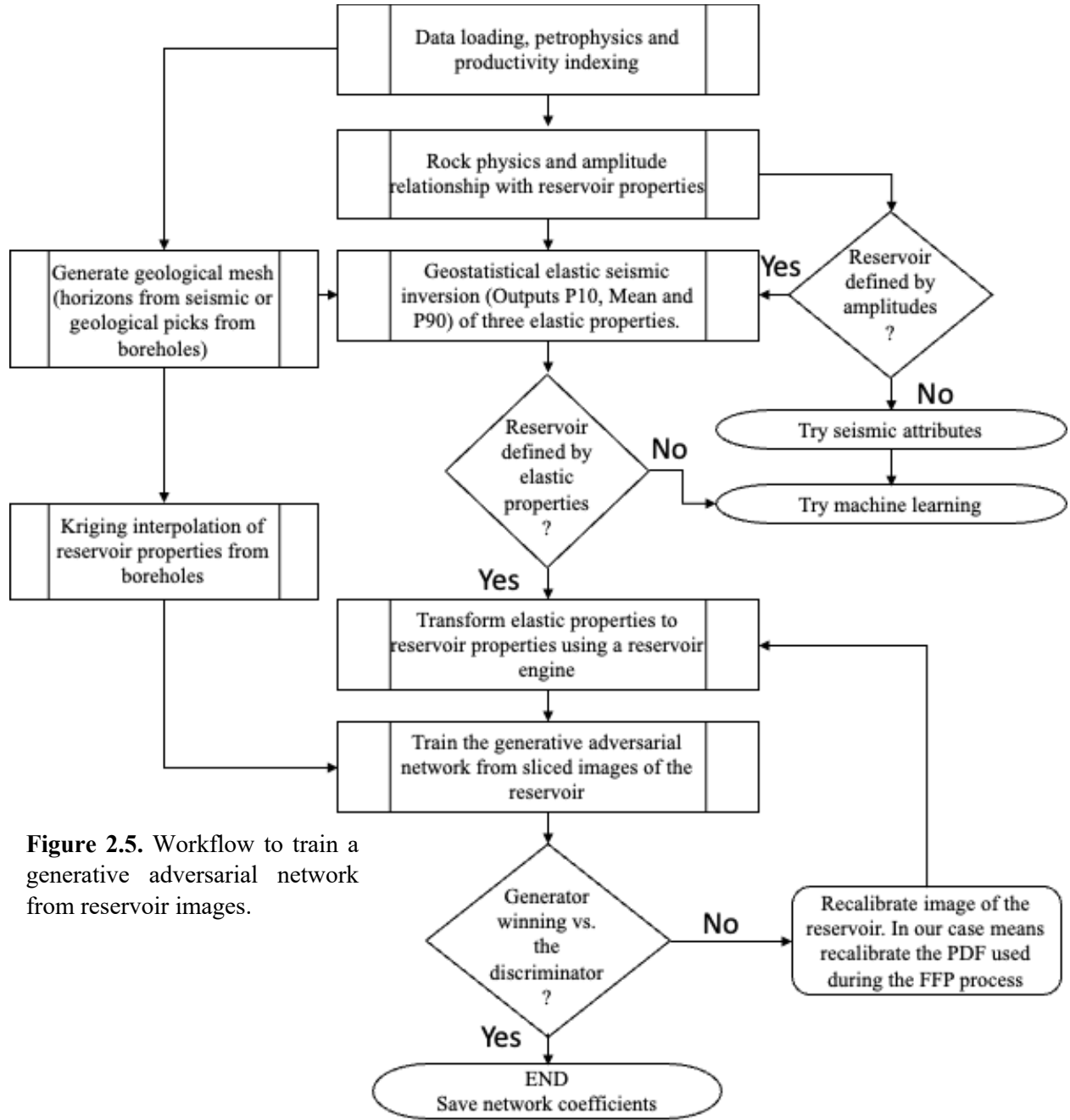


Figure 2.5. Workflow to train a generative adversarial network from reservoir images.

Stage 1: Productivity index calculation

The first stage of the workflow deals with translating geological heterogeneities (rock types, porosity, water saturation and permeability) into productivity index curves. The productivity index or injectivity index is presented in equation 1. The calculation considers a null skin factor a

fluid viscosity of 0.11 cp. The $\left(\frac{r_e}{r_w}\right)$ ratio comes from the original reservoir appraisal plans (Mesolella et al., 1974), this ratio is 400 [ft/ft]. The permeability (k) and rock type thickness (h) come from borehole data, which distribution is presented in Figure 2.4b.

$$PI = \frac{0.78 \times 10^{-3} * k * h}{v * B * (\ln\left(\frac{r_e}{r_w}\right) - 0.5 - S)} \text{ Eq. 1}$$

Where:

k = permeability, mD r_e = External borehole radius, h = lithology thickness, ft
ft.

v = fluid viscosity, cp r_w = Wellbore radius, ft. PI = productivity index

S = Skin factor = 0

The ratio or r_e/r_w can also be substituted with pressure drawdown to get more accurate results, the pressure drawdown is reached once the bottom-hole pressure has flowed at a stable rate for a sufficient period (Ahmed, 1946; Asquith and Krygowski, 2006; Tiab and Donaldson, 2012; Al-Jawad and Saleh, 2020). This study is limited by the seismic vertical resolution, in this seismic study at this depth the resolution is 16 ft (4.88 m) and productivity index is upscaled to a minimum thickness of 16 ft. The resulting averaged productivity index from borehole scale to seismic scale is presented in Figure 2.6.

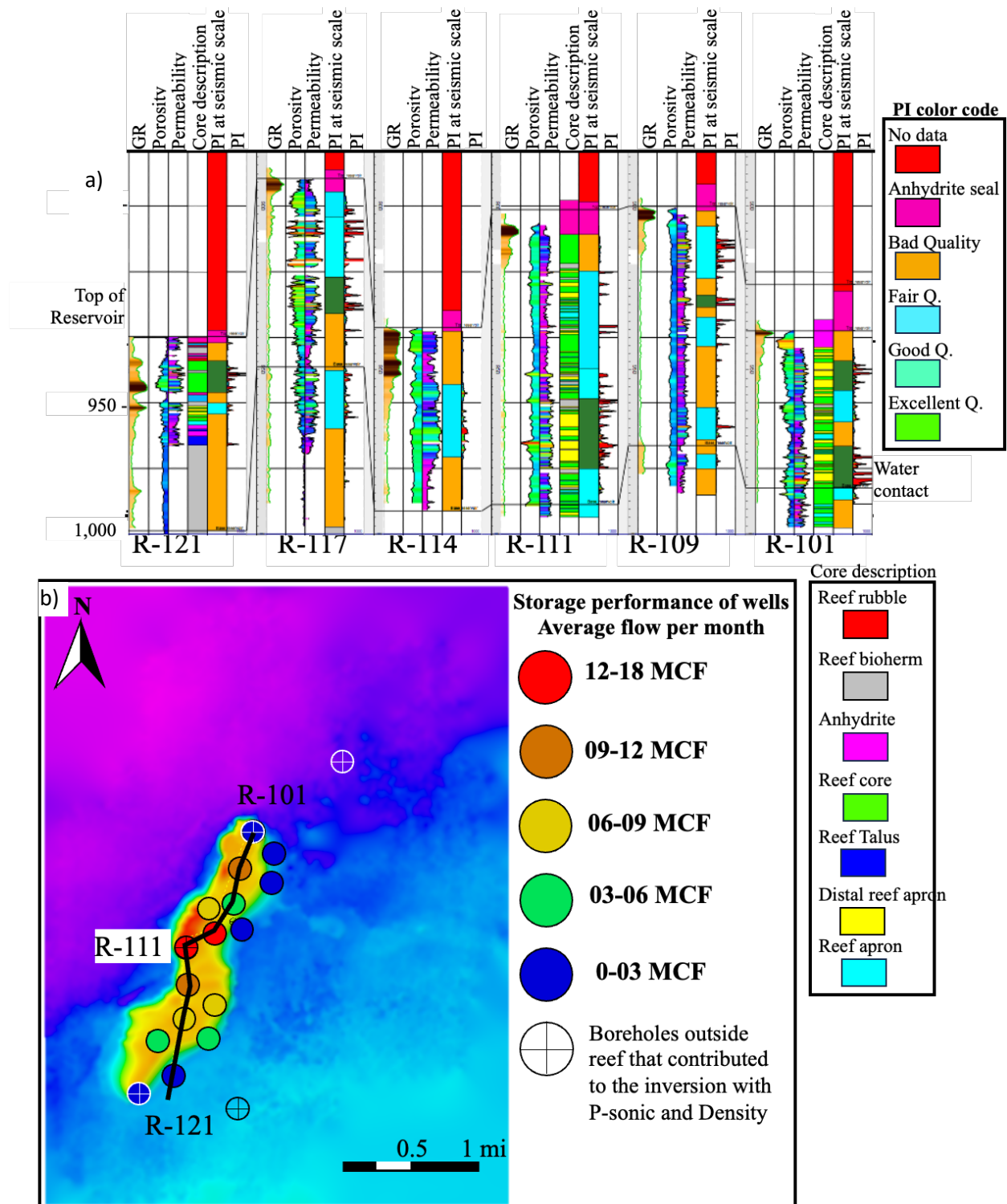


Figure 2.6. Panel a) Borehole cross section showing gamma ray, permeability, porosity, productivity index calculation both at borehole scale and seismic scale (averaging information in blocks of 16 ft), discretized productivity index classifying the productivity rock type and the facies described from core data. b) map view of the cross section and the storage performance per well, i.e. which borehole is storing more gas per month.

S-wave estimation, AVO/AVA analysis, and partial angle stack definition

Having defined the scale of what the seismic resolution can define and discretized the productivity index calculation we need to link the parameters related to the pre-stack data. For an acoustic case, i.e. the zero-offset stack, it is only required one acoustic parameter, the p-impedance or often named as acoustic impedance. In pre-stack data, three parameters are needed to calculate the variability of amplitude versus angle or offset (AVA/AVO); 1) p-impedance, 2) s-impedance or VpVs ratio (depending on the algorithm parametrization, and 3) bulk density. Since none of the boreholes have s-wave wireline log, we need to predict the s-wave log using a rock physics model or use an existing equation (Xu et al., 2007; Akhundi et al., 2014; Amini., 2018; Arevalo-Lopez., 2018).

Arevalo-Lopez (2018) showed the importance of s-wave estimation and its impact on AVO/AVA analysis, we decided to create a shear wave estimation equation using a rock physics model. The model required accurate petrophysics model including the definition of the mineralogy and lithology, effective porosity, and water saturation. Since the reservoir is primarily carbonated, the dominant minerals are calcite and clay minerals. Traces of pyrite and anhydrite are also included in the petrophysical and rock-physics model. The results of the petrophysical models are calibrated with core descriptions from the wells. Since velocity and density are a function of porosity, mineralogy, and fluid content, RPM uses effective porosity, gas saturation, and mineral analysis to calculate shear wave velocity. Silver et al. (2021) used Xu-White to calculate the s-wave but is insufficiently accurate and the cross-plots resulting in that calculation are physically inaccurate. Additionally, the AVO response we estimate from the synthetic gathers is a class III, not observed on the gathers and not expected since this is a carbonate rock not a poorly

consolidated siliciclastic. Prasad (2003) showed the variability of elastic properties of carbonate rocks as a function of porosity and permeability, and the expected AVO classes (I and IV).

Our s-wave estimation allowed us to check two important features. First and most important, the calculated s-wave allowed us to make a synthetic gather using an elastic forward modeling (Castagna and Backus, 1993; Aki and Richards, 2002; Hall and Wang., 2009) at each borehole and to perform an amplitude versus angle analysis on the synthetic gather at the borehole location. The analysis correlates as expected with the observed class IV in the observed seismic gathers. This correlation allowed us to study the observed response and link the amplitude variations to petrophysics (Chi and Han, 2009; Gegenhuber et al., 2015). The second feature to analyze using the synthetic gathers is the optimum angle stack distribution to keep the amplitude variability meaningful to fluid and lithology effects. The suggested partial angle stack definition to keep the amplitude variability meaningful is displayed in Table 2.1.

Angle Stack name	Angle range of each partial stack [°]	Dominant frequency [Hz]
Near	2-9	80
Near-mid	7-17	72
Mid	12-19	56
Mid-far	17-24	45
Far-mid	22-29	38
Far	27-34	33
Ultra-far	32-39	27

Table 2.1. Partial stack derived from the RPM forward modeling. The combination of these partial angle stacks yielded the most significant reservoir detection and correlated better with the observed gathers.

Stage 2: Geostatistical seismic inversion

In the second stage of the workflow, we expand on the methodology presented in Penna and Moreira-Lupinacci (2024). We choose geostatistical seismic inversion to generate the most accurate seismic reservoir characterization image. In our case, in contrast with Penna and Moreira-Lupinacci (2024) we use rock types, and we use 50 realizations of the elastic parameters instead of the mean realization. More details on the geostatistical seismic inversion are found on Doyen

2007, chapter four. The geostatistical seismic inversion seeks to link the storage history with the seismic amplitude. Geostatistical inversion considers the Bayesian theorem and Markov chain Monte Carlo simulation (Gilks et al., 1995; Torres-Verdín et al., 1999; Avseth et al., 2005; Caetano, 2012; Côte et al., 2019; Avseth and Lehocki, 2021) to find the most likelihood elastic parameters to estimate a synthetic seismic and minimize productive facies related to the observed amplitude. Geostatistical inversion is widely used to improve vertical resolution and investigate thin layers. Rayleigh's criterion sets the vertical seismic resolution to $\lambda/4$; nonetheless, geostatistical seismic inversion improves the seismic resolution to $\lambda/16$ (Chopra and Castagna 2014; Efnik et al., 2016; Côte et al., 2019; 2020). Keeping in mind that the average dominant frequency is 70 Hz at this depth, wavelength (λ) is an average of 64-92 ft, which results in a vertical seismic resolution of 16-23 ft and 8-12 ft for the geostatistical seismic inversion vertical resolution.

A scheme of the workflow is depicted in Figure 2.7 where the inversion steps considered are as follow: (1) the seven partial stacks from Table 1 and their amplitude and phase-dependent wavelets, (2) Low-frequency models built using kriging extrapolation biased to the facies present in the reef and well-logs filtered at seismic frequency. (3) Initial deterministic seismic inversion as a trend. (4) borehole calibration of the calculated impedances and the recorded impedance. (5) Probability density functions (PDF) building per each production rock type correlated to p-impedance and VpVs ratio, both at borehole scale and seismic scale. (6) Markov chain simulation using the deterministic models and the PDFs of each flow zone.

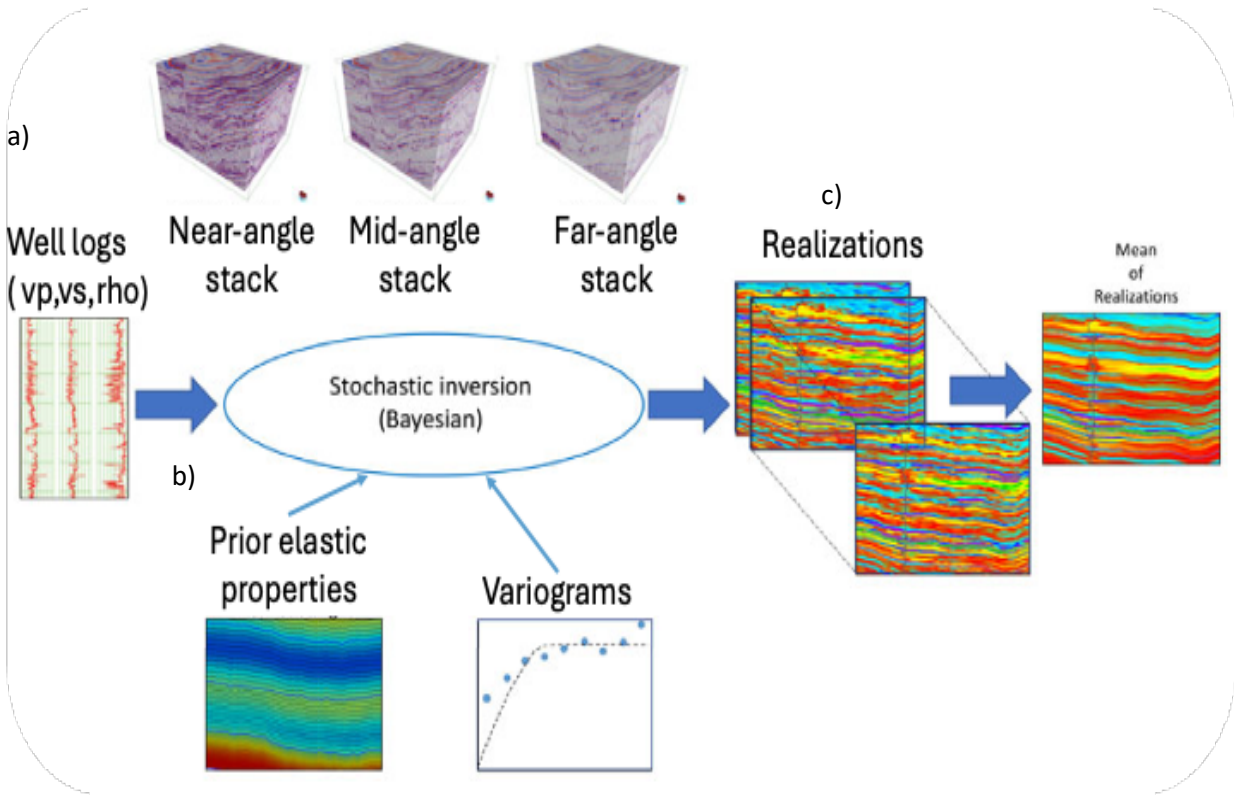


Figure 2.7. Scheme of geostatistical seismic inversion methodology from partial stacks to impedance versions. Panel a) Shows the starting point of the methodology elastic logs and from borehole data and the seven partial angle stacks. Panel b) Lastly the variograms assist the geostatistical extrapolation using Kriging from borehole data to generate the priori probability. Panel c) shows the outputs of the geostatistical seismic inversion, a posterior probability of reservoir, p-impedance, s-impedance, and density.

More details on the geostatistical seismic inversion can be found on Appendix A, as well as the p-impedance, s-impedance and bulk density volumes that are used as priori data for the geostatistical seismic inversion. The geostatistical seismic inversion solves a 3D probabilistic density function of seismic realizations to minimize the misfit function using hard constraints. Figure 2.8 shows the 3D PDF and the kernel of the solution.

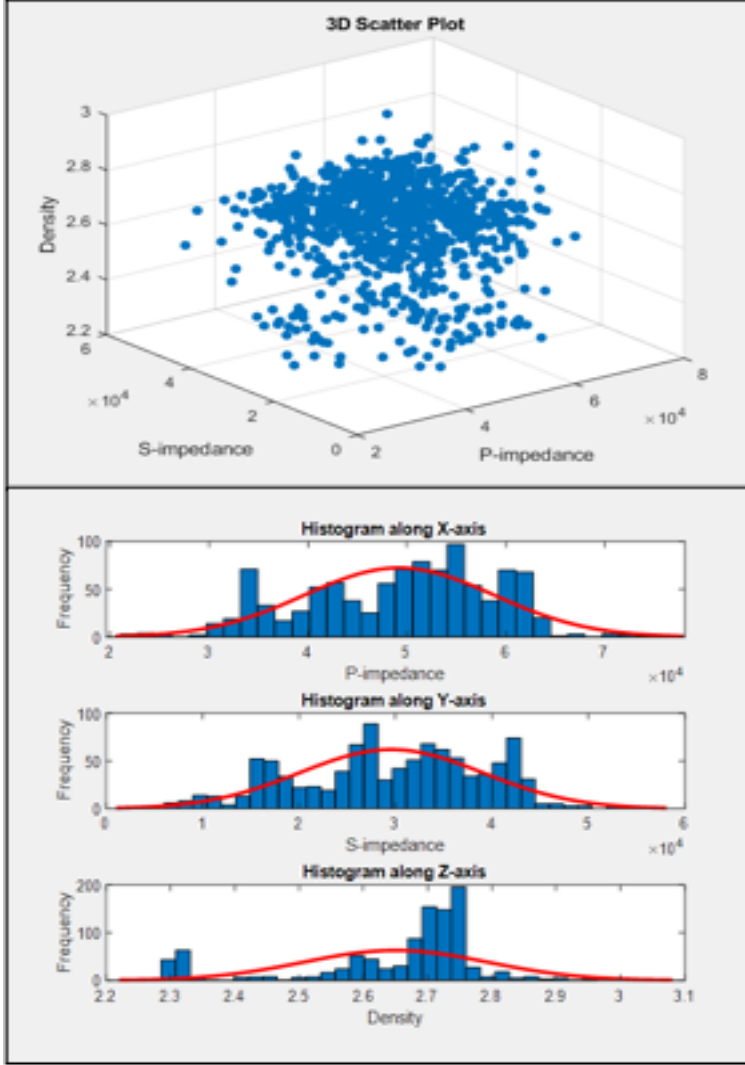


Figure 2.8. Panel a) shows the 3D data points of the p-impedance, s-impedance, and bulk density. Panel b) shows the histogram of each elastic parameter.

Since the goal of this research is provide with the best training possible for the generative adversarial network, we decided to export the maximum number of realizations (50) of the elastic parameters (p-impedance, s-impedance and, density). In a traditional seismic reservoir characterization, the mean elastic parameters are used for the reservoir engine. Nonetheless we export the 50 possible realizations to use them on the next step, the facies and fluid probability. Figure 8 shows the schematically which values are exported from the inversion process, to visualize the augmentation process. In Figure 2.9 each dot represents one of the realizations exported and used in the facies and fluid probability method.

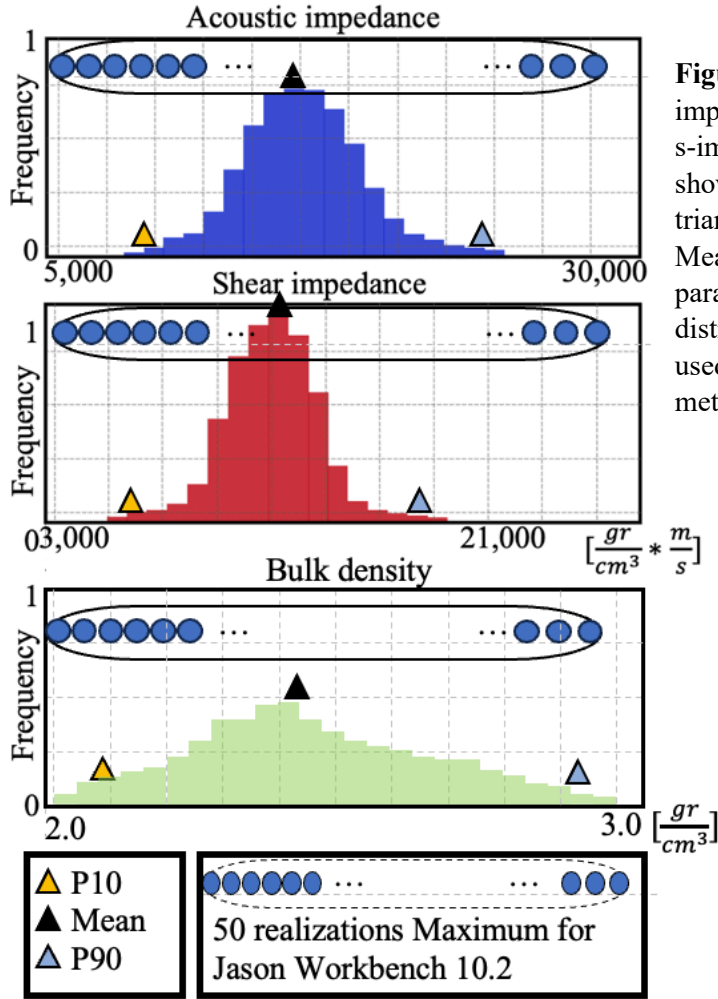


Figure 2.9. Panel a) shows the p-impedance histogram, Panel b) show the s-impedance distribution and panel c) shows the bulk density distribution. Color triangles indicate the position of the P10, Mean and P90 values of the elastic parameters. Blue points show the distribution of the saved realizations to be used in the facies and fluid probability method.

The Bayesian inversion merges the probability density of elastic properties into a global probability density function that represents the posterior PDF to minimize the seismic and prior information.

In summary geostatistical inversion described the amplitude distribution spatially using the information provided by the petrophysics and the rock-physics models. Geostatistical inversion uses Markov chain Monte Carlo sampling, which generates many forward simulations constrained by the low-frequency model and hard constraints. The multiple realizations of the elastic properties are linked to the facies and fluid probability that best describe the observed data and boreholes.

Facies and fluid probability from GSI

Facies and fluid probability use the Markov Chain Monte Carlo to evaluates all the possible realizations from each unit reservoir type PDF and the defined uncertainties. The prior probabilities are given by the borehole correlation with rock types and elastic properties. Using the outputs of any elastic seismic inversion can be used in a facies, fluid, and probability, resulting in one realization of the reservoir. This algorithm is an alternative to a petro-elastic seismic inversion, in both methods the goal is to estimate the kernel distribution associated to the target fluid. We choose the realizations estimate the prior probability density functions per reservoir facies. Estimating the probability for each facies and fluid on the seismic inversion uses the Bayesian probability theory to create probability volumes indicating the likelihood that particular facies is present at a data point in the seismic survey considering the uncertainties from the elastic properties.

The uncertainties are reduced by building a PDF based on the facies or density. The PDFs help to estimate the prior probabilities, if the overlap of the properties is larger the uncertainties are larger. The Bayes' rule is used in the calculation of the posterior probability of each selected facies as eq 2:

$$P(facies_i|input) = \frac{P(input|facies_i)*P(facies_i)}{\sum[(input|facies_i)*(facies_i)]} \quad \text{eq 2.}$$

Where:

$P(facies_i|input)$ is the posterior probability. i = the index for particular facies

$P(input|facies_i)$ is the selected Pdf. $input$ = The inversion results

$P(facies_i)$ is the prior probability

The FFP methodology requires the prior probability per facies and the PDF associated to these discrete values. In this case, in eq.2, i goes from 0 to 6 representing the discrete facies or fluid lithologies (seal =0, bad =1, fair =2, regular=3, good =4, excellent =5). Figure 2.10 shows the parameterizations of the elastic PDF, we choose the V_p/V_s ratio p -impedance. This is a crucial step in the methodology since we are looking to correlate elastic parameters to reservoir properties not to elastic attributes. In the data augmentation, the PDFs remain but the density of the elastic realization changes, modifying the FFP output. While it would be ideal to have a PDF for each discrete facies, PDF A clusters the excellent, good, and regular facies. The PDF A is the reservoir area accounts for low p -impedances and low V_p/V_s ratio. PDF B clusters fair and bad reservoirs, where low p -impedance values account for good porosity without any fluid effect. PDF C encircles bad rock types and seals.

There is a natural gap in observation resolutions, at borehole scale (full resolution) there is a good clustering of the following rock types: excellent, good, and regular reservoir facies. None and bad rock types are well cluster at the high-impedance values. At seismic scale i.e. one point of data every 8 ft. The three PDFs distribution are preserved at both scales and are kept during the FFP process.

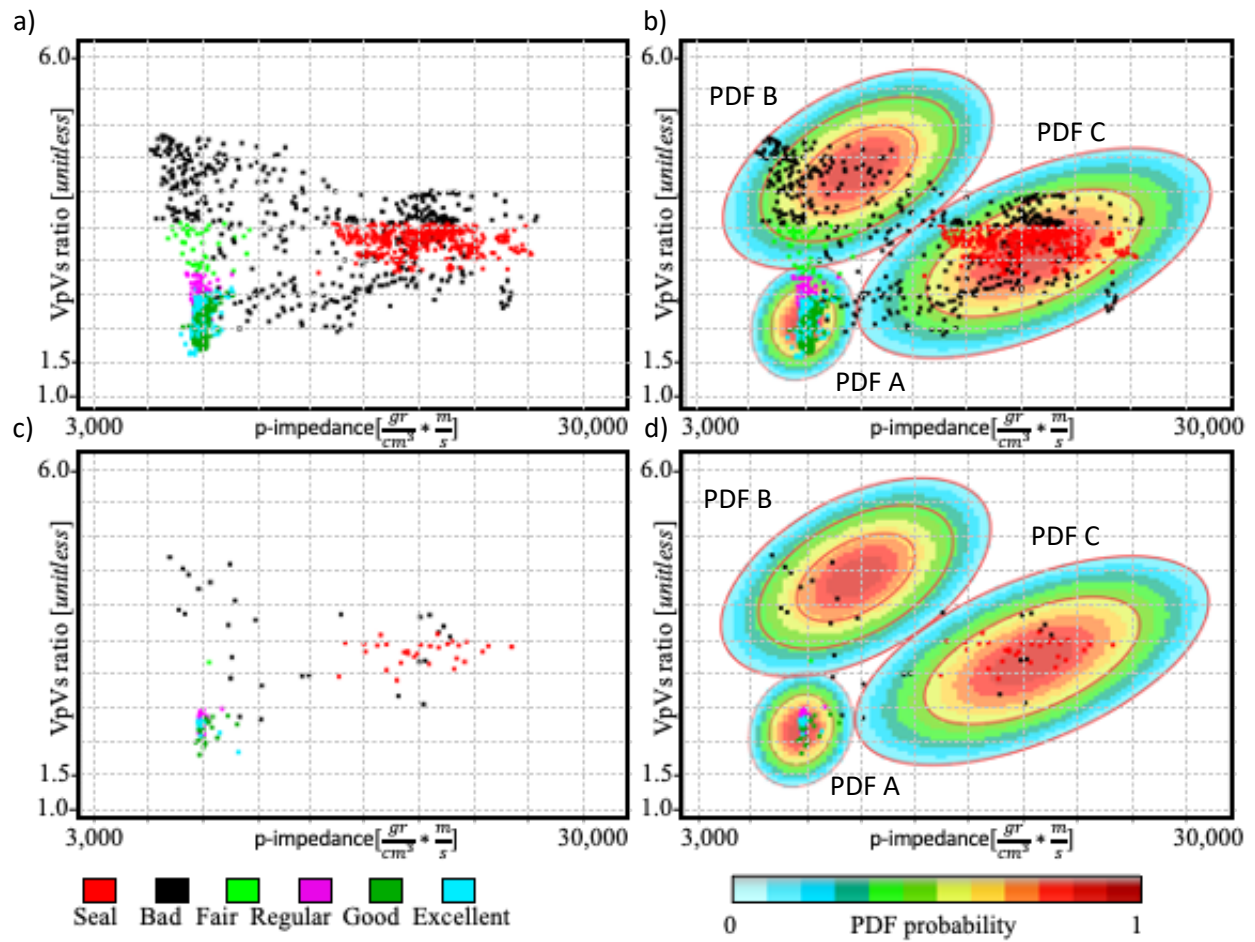


Figure 2.10. Panel a) shows the cross-plot in VpVs – p -impedance domain of the three productivity index zones in the reservoir, and Panel b) shows the theoretically probability distribution assigned to each reservoir types. The probabilities distributions are used to estimate the probability of each facies to be linked to the productivity rock type.

In Figure 2.9 and Figure 2.10 we defined the priori distributions to generate the posteriori distribution of the reservoir with FFP. PDF A, B and C are kept during the integration of the 50 p-impedance and VpVs ratio results 50 realizations. This means that in total there will be 50 FFP volumes that used from the P2 realization of elastic parameters to the P98 realization of the elastic parameters. Figure 2.11 summarizes how these 50 realizations are done sequentially to calculate the 50 FFP volumes.

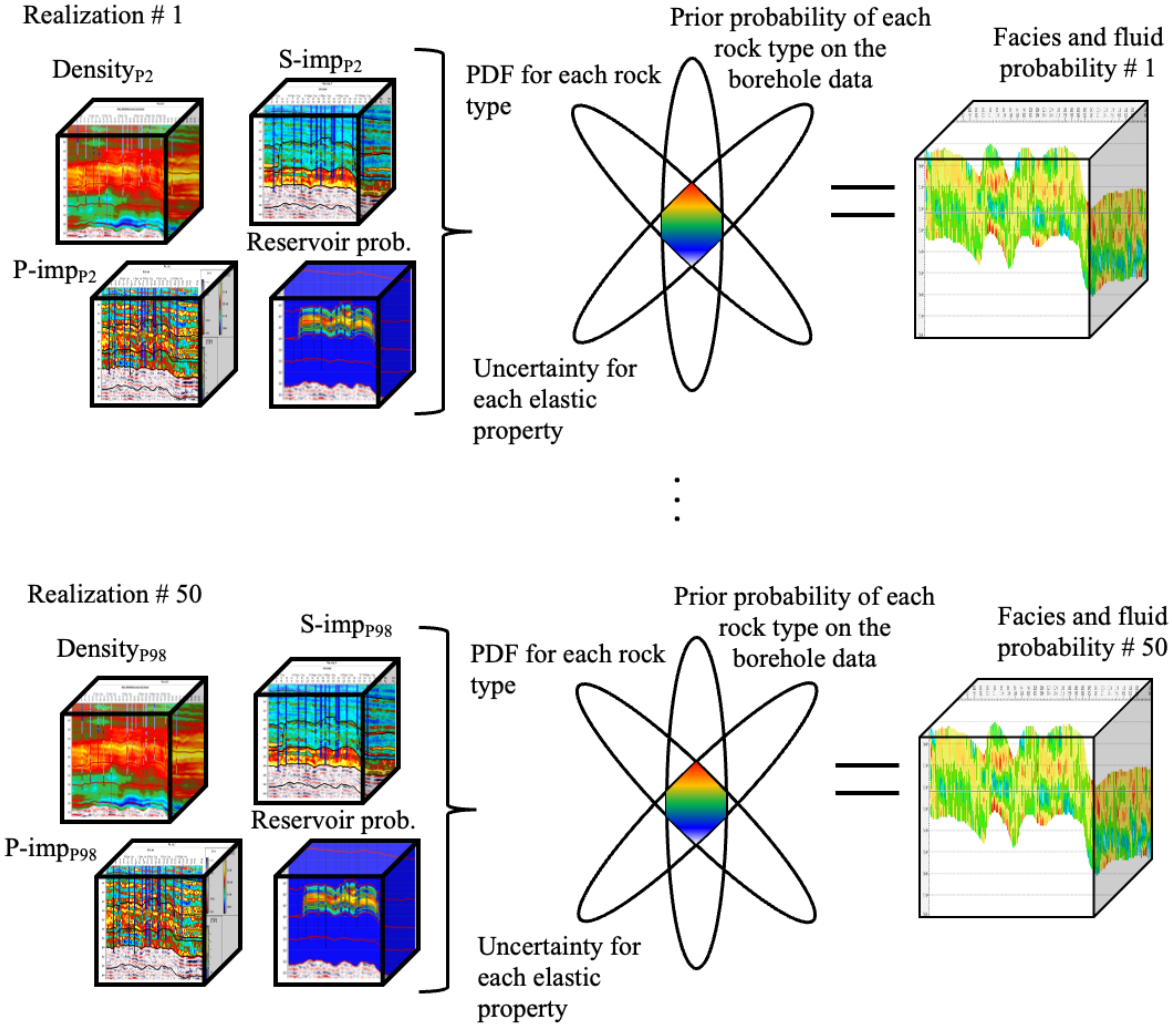


Figure 2.11. Scheme showing the limited reservoir models available in seismic inversion reservoir characterization. Normally only one FFP realization is used to define the reservoir fluids, but we seek to generate as many samples as possible to train the GANs.

The generation of reliable images of the reservoir is important to enter not only correct information to the generative adversarial network but also a vast amount of training data. Alternatively, image processing techniques offer augmentation processes such as rotation, shearing, stretching, and squeezing of the image. Nonetheless, we need to keep the pixel position and intensity for an accurate training of the generator. To achieve this, Figure 2.12 shows how the FFP slices of the 50 FFP volumes are entered into the GANs training.

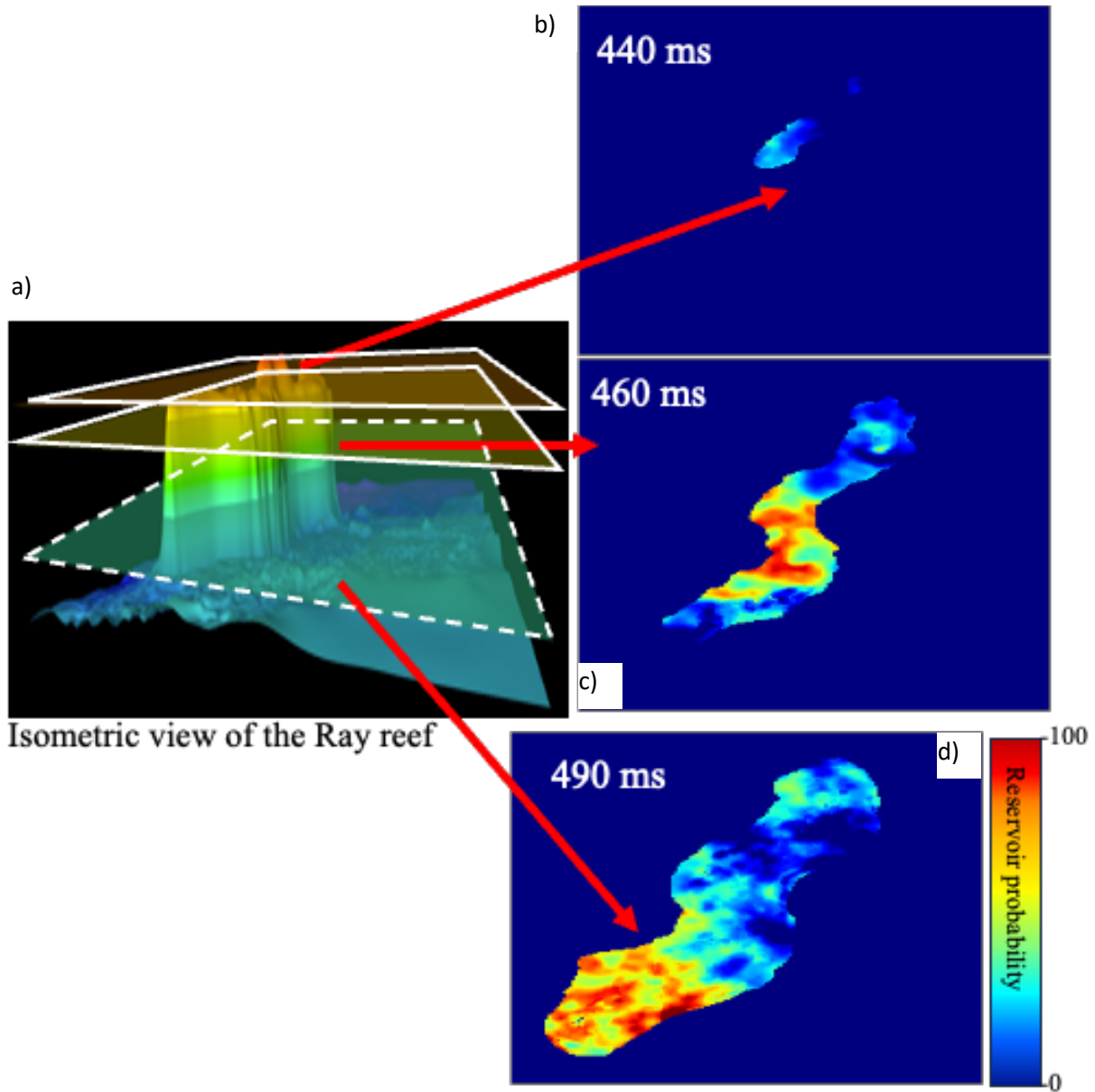


Figure 2.12. Panel a) shows the 3D view of the Ray reef structure from the A-2 carbonate to the Lockport/Guelph formation interpreted on 3D seismic and exaggerated vertically 10 times. Panel b) shows the facies and fluid probability coming from the mean elastic realization of the geostatistical seismic inversion at the shallowest reservoir position. Panel c) Shows the time slice at 460 ms of the Ray reef from the same FFP volume and d) Shows the time slice at 490 ms of the FFP volume. These images work as an example of the images used to train the GANS.

Take the mean FFP realization for example, reservoir thickness is in average 68 ms (reservoir thickness average 203 ft) which translate in 68 samples of the mean FFP realization of the reservoir. Now, if we extrapolate the available FFP we have 3,400 (68 samples x 50 facies and fluid probability cubes) 2D maps of the reservoir that will serve to train the GANs. In summary the facies and fluid probability cubes are transformed into a training dataset used as input for the GANs. Other methods for labeling an image of the reservoir are welcomed to train the GANs.

Stage 3: Generative adversarial networks

GANs is a pixel-to-pixel prediction that considers the color intensity and distribution of each pixel. GANs, developed in 2016 (Goodfellow, 2016, Goodfellow et al., 2016), consists of two neural networks - a generator and a discriminator - that are trained simultaneously in a competitive process. Both generator and discriminator are multi-layer perceptron (MLP). It learns to generate new data with the same statistics as the training set. In the image generating task deep CNN (convolutional neural network) is used as generator. It uses deep learning convolution neural networks to generate new data. In this task generator is used to generate reservoir images from given reservoir sketch and discriminator is used to check whether a generated image is “real” or “synthetic”. The network is trained to minimizing generator loss and maximizing discriminator loss.

The generator model is CNN model that starts with a random data, which for the image context resembles noisy pixels. On the other part of the deep learning architecture, the discriminator classifies the model [1 real, 0 fake]. As the number of epochs increases, the generator has access to both real images and created (fake) images as input for training with real image data as real data (label 1) and fake data with class (label 0). At each epoch the discriminator has to

classify between real and fake images (generated images), at this point the two neural networks become an optimization problem, because as the epochs increases, the generator consistently beats the discriminator, thus the need to train the discriminator. We train the discriminator by giving the generated images a class label 1, becoming not generated figure but part of the input, with this the discriminator has an augmented database to compete against the generator.

We use a sigmoid function as a loss function, it determines, how close the generated image is from the true image. The objective function of GAN network is a min-max optimization problem given by the equation:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_{data}(z)} [\log(1 - D(G(z)))]$$

In our context, we used GANs to generate multiple realizations of the reservoir facies/fluids probability models that were initially derived from the geostatistical seismic inversion. By training on the real inversion results, the GANs produces augmented synthetic realizations representing the potential distribution of reservoir properties. This allowed evaluating a range of scenarios rather than just a single deterministic model. Our GANs can work with other reservoir models, in this case the reservoir static model is taken the facies and fluid probability (FFP) realization. The architecture used for reservoir characterization GANs included 48 layers for the generator and discriminator to train both models simultaneously. Figure 2.13 shows the schematic of how GANs process images and their architecture.

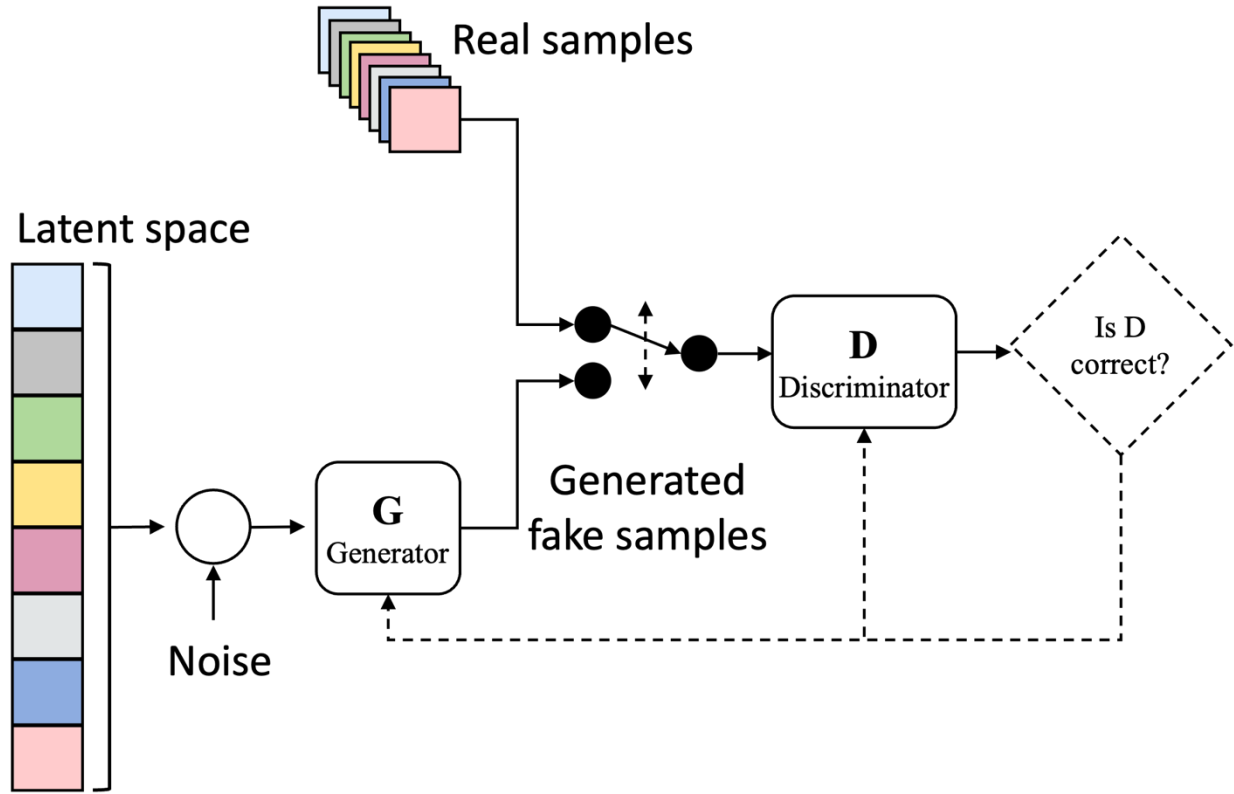


Figure 2.13. Generator and Discriminator start from the latent space. We keep the weights of the generator updating with the discriminator. The schematics of GANs generate synthetic images from random noise and samples with clear labels that constrain the discriminator.

Multiple realizations of the facies and fluid probability (FFP) are exported as maps or horizontal slices. The slices have a label for each one of the flow units, which will guide the reservoir detection by GANs. The GANs algorithm will produce a posterior probability distribution $P(W|R)$ over the network weights (W) from these reservoir maps (R).

Having defined the fluid reservoir using the PDF shown above, each time or depth slice of the reservoir with a defined threshold probability becomes a training image for GANs process. We selected 75 % of probability to define what is a fluid reservoir. Our GANs concept is cemented in the concept that if the generator surpasses the discriminator, each output becomes a realization of the fluid reservoir with a 75% probability of gas. We selected cross-entropy as the loss function to

evaluate the two neural networks and SoftMax as the activation function. Given the limitation in computational resources, we estimate GANs using 20 and 40 epochs.

Lastly, the self-organizing maps experiment in a nearby reef complex by Buist et al. (2021) suggested the probability of linking the best carbonate facies to permeability. We wanted to compare the fluid probability map with the self-organizing mapping using seismic attributes. We use seismic trace and geometric attributes (White, 1991; Marfurt and Chopra, 2007), and followed Matos and Marfurt, (2010) self-organizing methodology. Self-organizing map from Kohonen (1990) in seismic interpretation creates a matrix from the normalized seismic attributes to train the SOM algorithm. During training, the SOM iteratively organizes the input samples onto a 2D map, where similar samples are clustered together, and dissimilar samples are mapped farther apart. The two axes are correndered in a color space to find the features of interest clustered.

RESULTS

Presenting the productivity index estimated from GANs requires showcasing concatenated results. The results in this study are separated into three sections; first, at the borehole scale, where petrophysics, rock-physics, and production analysis are used to generate an S-wave velocity used in the inversion, synthetic gathers, and productivity index. Secondly, the geostatistical inversion shows where the horizontal and vertical distribution of the storage is captured by the elastic property's realizations. Thirdly, in the GANs stage, the multiple realizations of the inversion are used to generate training images used by the GANs generator to create synthetic samples with the same probability of having a highly productive index across the reef reservoir. Finally, the results show an agreement between the highest productive index found at each well and in the reef and its continuity, highlighting the correlation between reservoir connectivity and storage capacity.

Stage 1: Productivity Index and shear wave model results

Shear wave modeling is the starting point of the results since the productivity index is a function of two elastic two properties (VpVs ratio or s-impedance and p-impedance). Not only does the RPM model shear wave, but it also allows the linking of porosity, permeability, and mineral content to P-wave velocity in boreholes where compressional velocity is not recorded. Figure 2.14 shows the share wave analysis and the synthetic gather created from the modeled logs. In Figure we observe an AVA class IV associated with a stiff matrix with high gas content, as expected for a gas storage carbonate reservoir (Ostrander, 1984). The synthetic gather also suggests having more than 25° degrees to get an amplitude effect.

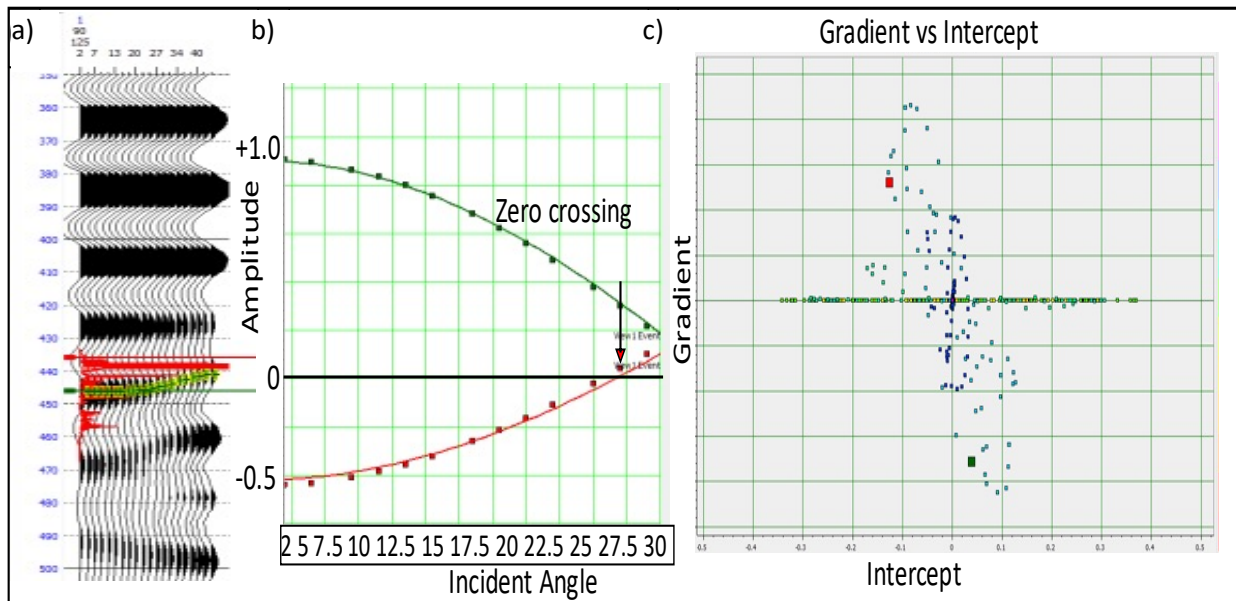


Figure 2.14. Synthetic gather with permeability curve to show the highest permeability associated with a negative reflection at near angles and changing to slightly positive with higher angles.

Class IV is associated with a stiff matrix with high gas content, as expected for a gas storage carbonate reservoir. The synthetic gather also suggests that the data needs to have more than 25° degrees to observe an amplitude effect. The following result is the definition of the productivity index and its relationship with the elastic properties of the carbonates at log scale and seismic

resolution. The permeability and porosity profiles across the reservoir show variability between the perpendicular cross-sections (Figure 2.14).

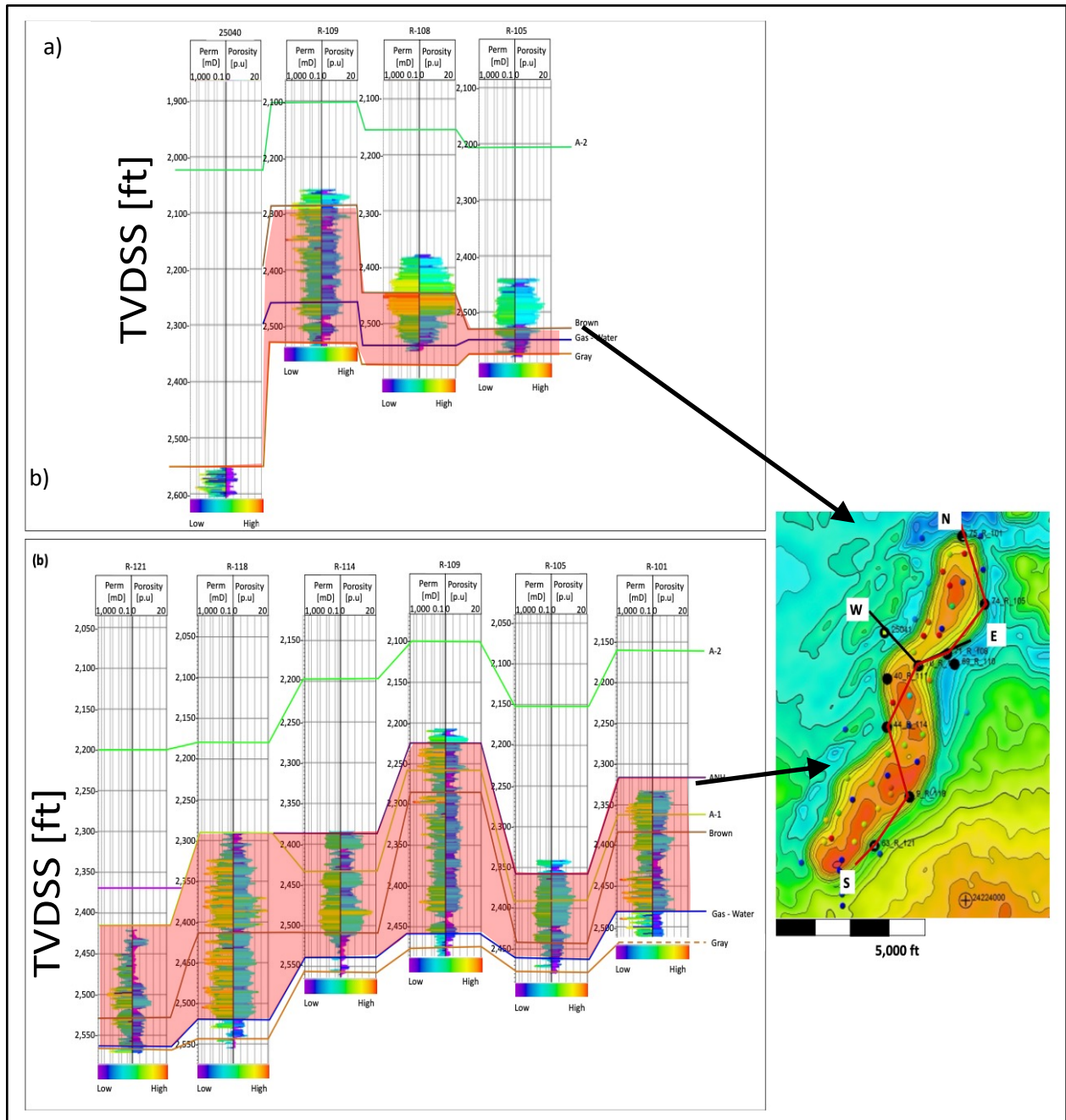


Figure 2.15. Panel a) East-West borehole permeability/porosity cross-section Panel b) South-north permeability/porosity cross-section within the Ray Reef. In both sections, the reef stratigraphy is highlighted with a red polygon, reefs showing the variability of flow capacity with permeabilities up to 1 Darcy and porosities above 10%. The south and north borders exhibit lower permeabilities and porosities.

Stage 2: Geostatistic inversion results

Using stochastic seismic inversion, the heterogeneous nature of the reservoir is captured, both vertically and laterally, with an increased resolution compared with a deterministic inversion (Appendix Figure A4). We gain a better definition of the core reef and its contrast with the bioherm reef. The facies associated with the core reef are better connected and exhibit a low impedance layer. Figures 2.16 thru 2.18 illustrate the geostatistical pre-stack seismic inversion products (p-impedance, s-impedance, V_p/V_s ratio, and density) or posteriori mean realizations, in present the posteriori elastic values in Appendix A.

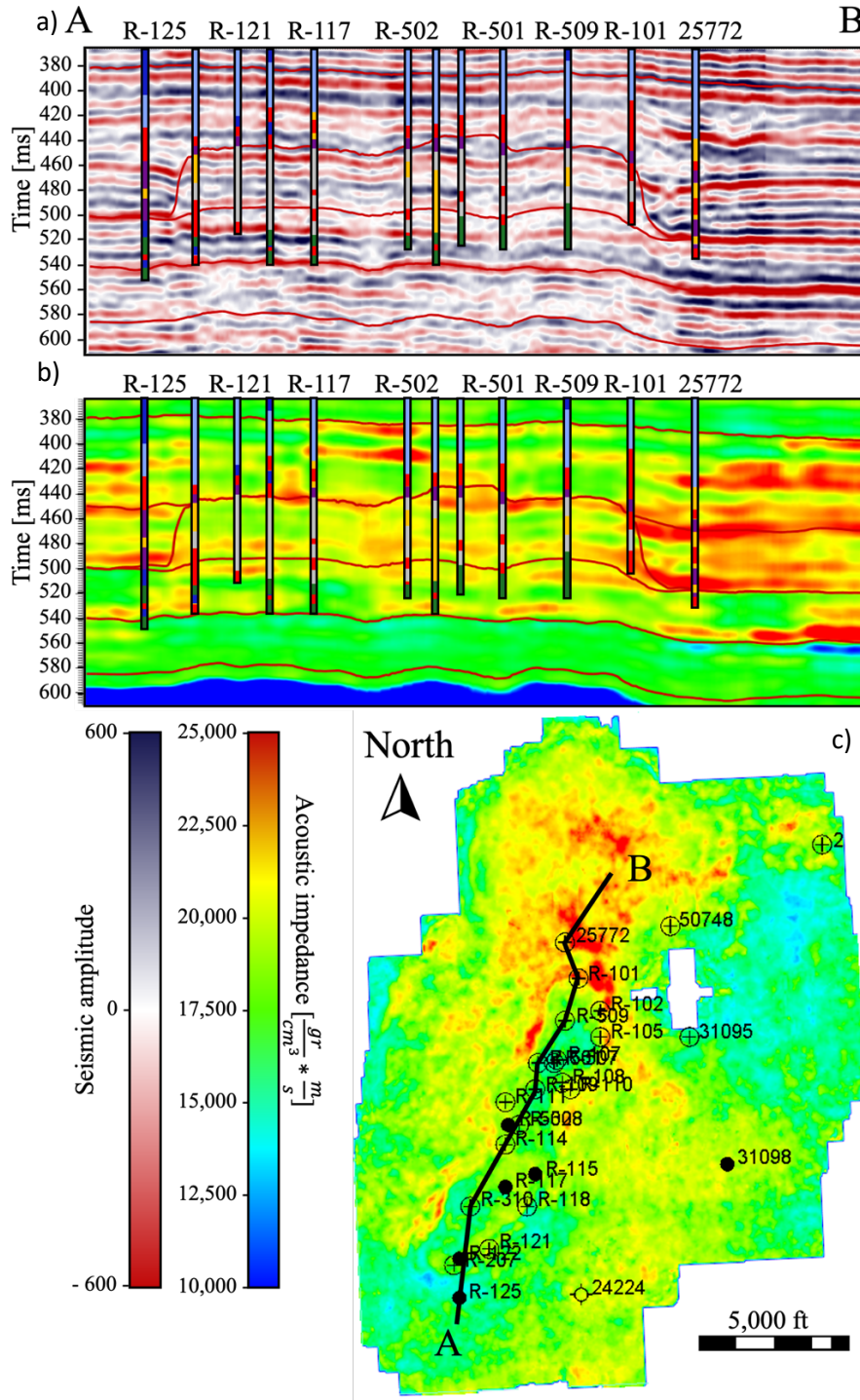


Figure 2.16. Panel a) Full seismic stack with borehole rock-types across Ray reef. Panel b) p-impedance with borehole rock-types showing the high p-impedance layers outside the reservoir corresponding to anhydrite layers. Inside the reservoir the core reef shows the lower p-impedance and best reservoir properties. Panel c) shows the random section on panels a and b, with the time slice of the p-impedance. The time slice of the p-impedance shows a good distribution with no local anomalies.

Figure 2.17 shows s-impedance and VpVs ratio inversion results, in these plots the fluid effects become more evident, and the reef core growth can be delineated (red arrows). These plots are important to use a probabilistic density function to separate the fluid effect from the lithology and porosity.

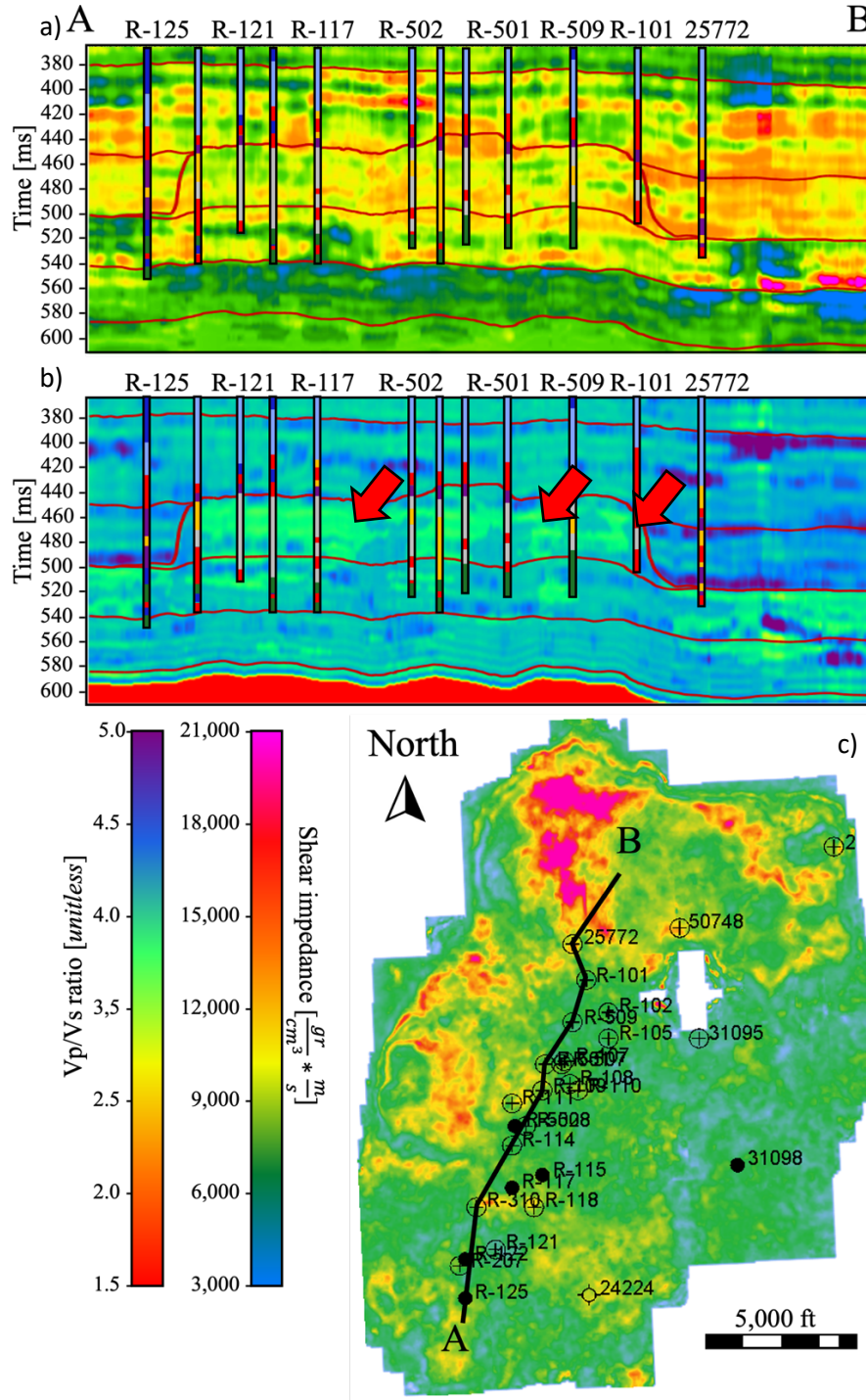


Figure 2.17. Panel a) displays the s-impedance inversion result, anhydrite layers observe a high shear impedance. The reef's core is better described in the shear impedance section where the high shear impedance represents vertical bioherm growths. The best fluid reservoir facies (gray) correlate with intermediate values of shear impedance. On the other hand, panel b) displays the Vp/Vs ratio inversion result. Red arrows highlight the reef's core, higher Vp/Vs ratio agrees with the AVO class observed on the synthetic data. Gray facies correlate well with the reservoir intervals. The Vp/Vs inversion includes a hard constrain to prevent smaller values than 1.42 and thus keeping the elasticity theory. Panel c) shows the map view of the shear impedance inversion to monitor for abnormal values or artificial effects.

Figure 2.18 displays the mean posteriori bulk density realization and the residuals as a first quality control. Density is high-frequency filtered to have a smooth density transition, useful to construct the probability density functions. The low frequency residuals show the good match between true and synthetic seismic.

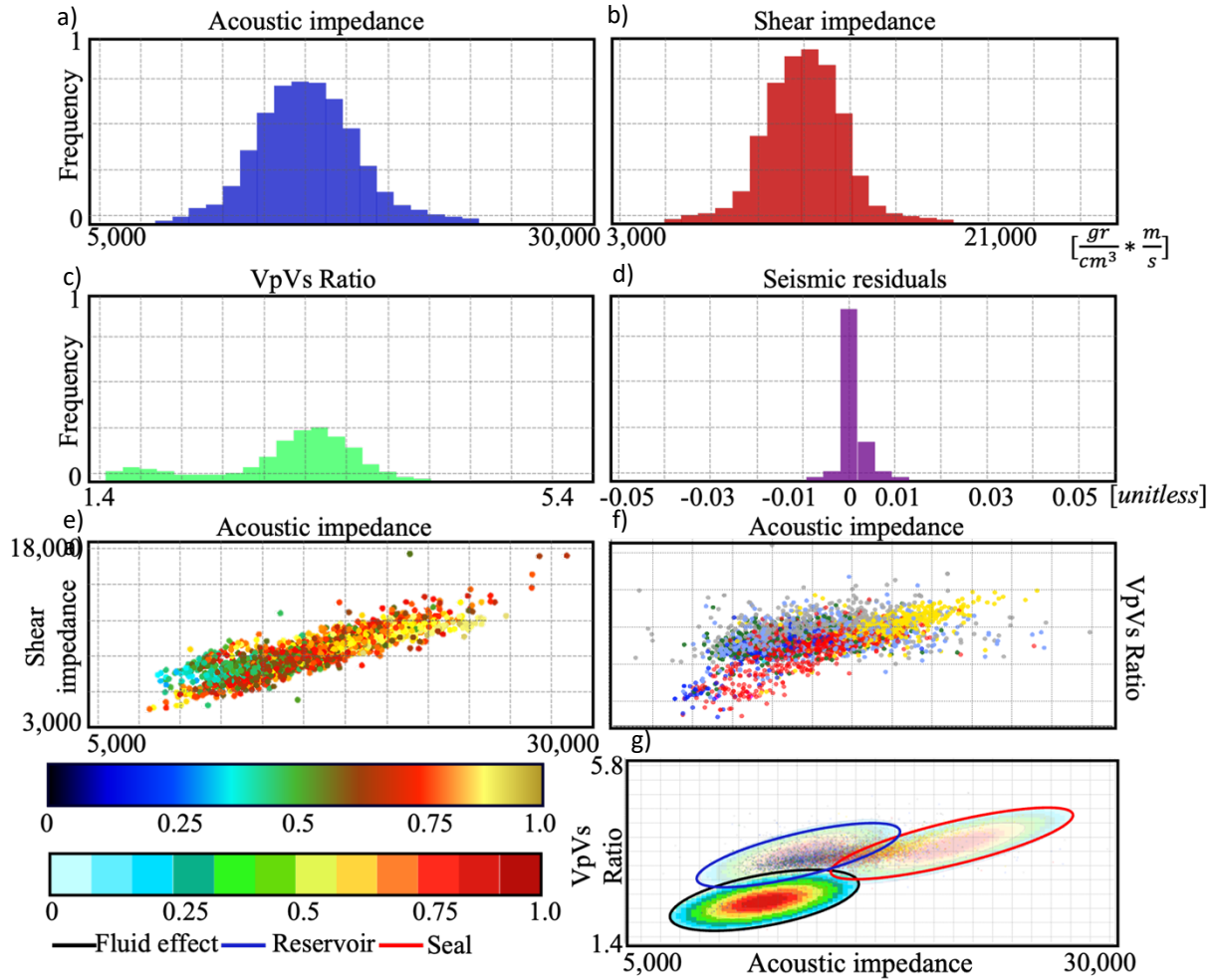
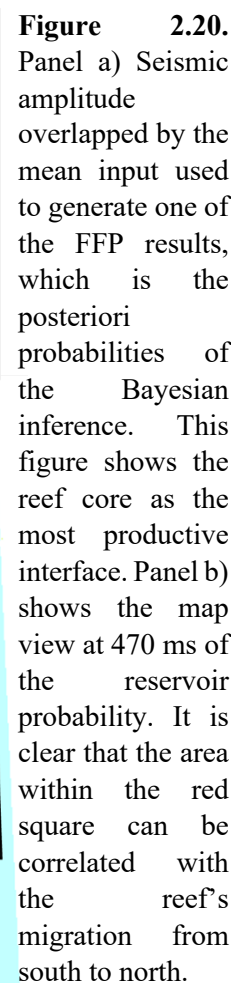


Figure 2.19. Panel a) Acoustic impedance histogram normally distributed with a well-defined standard deviation, Panel b) shear impedance histogram, with a narrow uncertainty, VpVs ratio histogram exhibit a quasi-bimodal distribution due to the hard constraints imposed on the VpVs ratio. Panel d) shows the seismic residual histogram to show the small residuals, this histogram is slightly right-skewed suggesting the need of increasing the low-frequency values to reduce the positive residuals. Panel e) the cross-plot of the shear and acoustic impedance colored coded by the inverted bulk density. Panel f) is the acoustic impedance versus VpVs ratio, a template suggested by Ødegaard and Avseth (2003) to correlate the rock physics template to hydrocarbon and porosity effects. g) Defining the prior probability density functions (PDFs) for each reservoir facies. VpVs ratio versus p-impedance cross plots augments the fluid effects.

The probability density functions (PDF) built from the VpVs ratio and p-impedance volumes increased the delineation of the reservoir architecture (Hong Bing et al., 2019). Figure 2.20 shows the reservoir probability estimated from the facies fluid probability (FFP) between the Gray formation and A-2 Carbonate formation. We limit the FFP estimation to keep the results relevant to the reef and avoid the non-unicity.



43

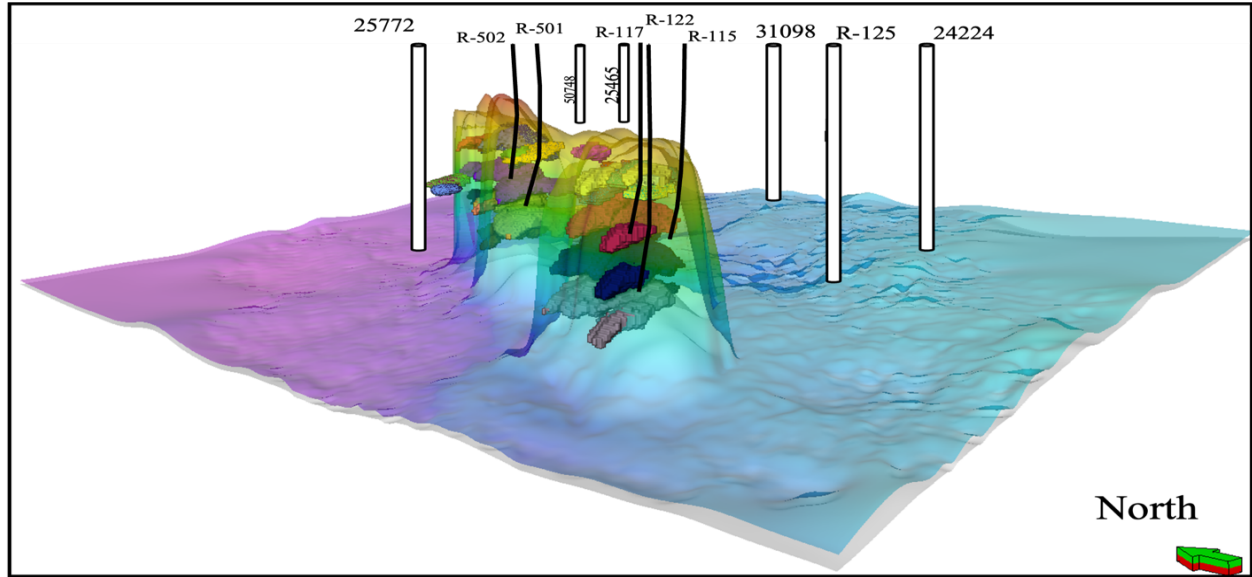


Figure 2.21. 3D View of the captured geobodies taking 75% as the threshold on the fluid probability. The geobodies show the impermeable layers and the porosity patches.

We evaluate the accuracy of the volumetric estimation we use the volumetric method for original gas in place (OGIP). This formula uses static geological data to determine the volume of the pore space of the reservoir and using a gas formation volume factor we can estimate the OGIP.

$$G = \frac{V_{grv}(\overline{h_n}/\overline{h_g})\overline{\phi} \overline{S_g}}{\overline{B_g}}$$

Where:

• G is the original-gas-in-place of the reservoir, • h_g is the average gross reservoir thickness, ft. SCF.

• V_{grv} is the gross rock volume, ft³.

• S_g is the average initial gas saturation, fraction.

• $\overline{h_n}/\overline{h_g}$ is the net-to-gross thickness ratio, • $\overline{\phi}$ is the average reservoir porosity, fraction. fraction.

• h_n is the average net reservoir thickness, ft.

• $\overline{B_g}$ is the average formation volume factor of the gas at initial conditions, ft³/SCF.

The estimated OGIP is 69 BcF in contrast with the 64 BcF reported in the storage report, an 8% difference. At this point we have characterized the reservoir statically, using the

geostatistical inversion outputs we produced three scenarios; P10, Mean and P90, and selected the mean (most likely scenario). Next, we evaluate using generative adversarial networks to build multiple reservoir models. Similar to Monte-Carlo simulations, where the output of a function is evaluated multiple times (realizations) as a function of a specific parameter (sandstone porosity for example in the water saturation calculation) starting with a normal probability distribution (scenario 1 = lowest porosity in the histogram, last scenario is the highest porosity possible). The Monte Carlo simulations are possible scenarios based on true data. Similar, the geostatistical inversion, which uses a Bayesian framework outputs, multiple probabilities of the reservoir, in static reservoir characterization, we usually only consider the P10, Mean and, P90. We export the byproducts of the geostatistical inversion from less likely to most likely generating 99 probability cubes of the reservoir, sampled a 1 ms and with a reservoir thickness of 80 ms makes a total of $80 \times 99 = 7,920$ reservoir scenarios, which represents the lowest probability presence of the reservoir at the top of reservoir (440 ms) to the most probable reservoir presence at 520 ms. allows us to generate multiple captions of the reservoir. We generated a labeled dataset of 7,920 of the same reservoirs to train a generative adversarial network that can produce at a fraction of the time multiple reservoir simulations. Moreover, the trained weights of the network, can take seismic images and generate a reservoir probability map at a fraction of a time.

Stage 3: Generative adversarial networks results

Using GANs in reservoir characterization involves training the generator network using images of the reservoir to create synthetic data. This synthetic data can then fill in gaps in the real data, augment the existing data, and provide a more comprehensive understanding of the subsurface properties. Additionally, it creates another dimension to evaluate the possible distribution of the reservoir. In GANs the initial generator (epoch 1) is the random noise with a

similar normal distribution than the training images. In Figure 2.22 we present the input and the random noise distribution of the generator. From epoch 1, the generator attempts to overcome the discriminator based on the model provided.

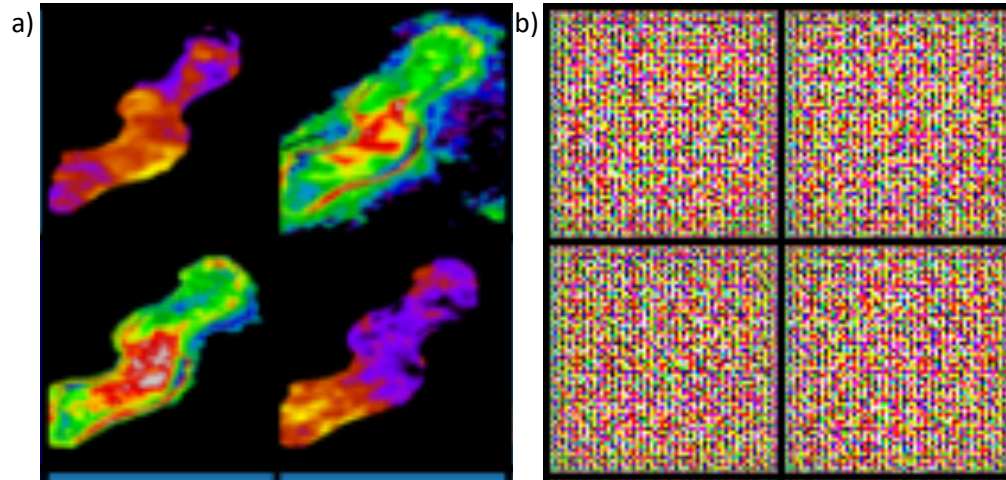


Figure 2.22. Panel a) shows the time slice possible reservoir realizations including the fluid effect. These images are used as training for the generator, and as a discriminator. Panel b) shows the first iteration (epoch # 1), where there is only random noise to simulate the model.

From the true reservoir images and the random noise presented in Figure 2.22 the generator after twelve epochs is able to overcome the discriminator. Figure 2.23 shows the result of the generative model for the fortieth epoch.

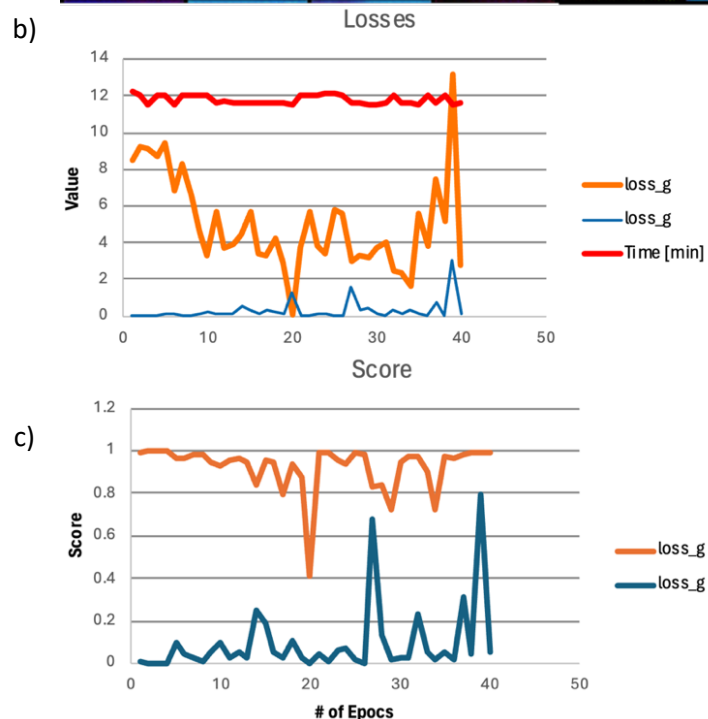
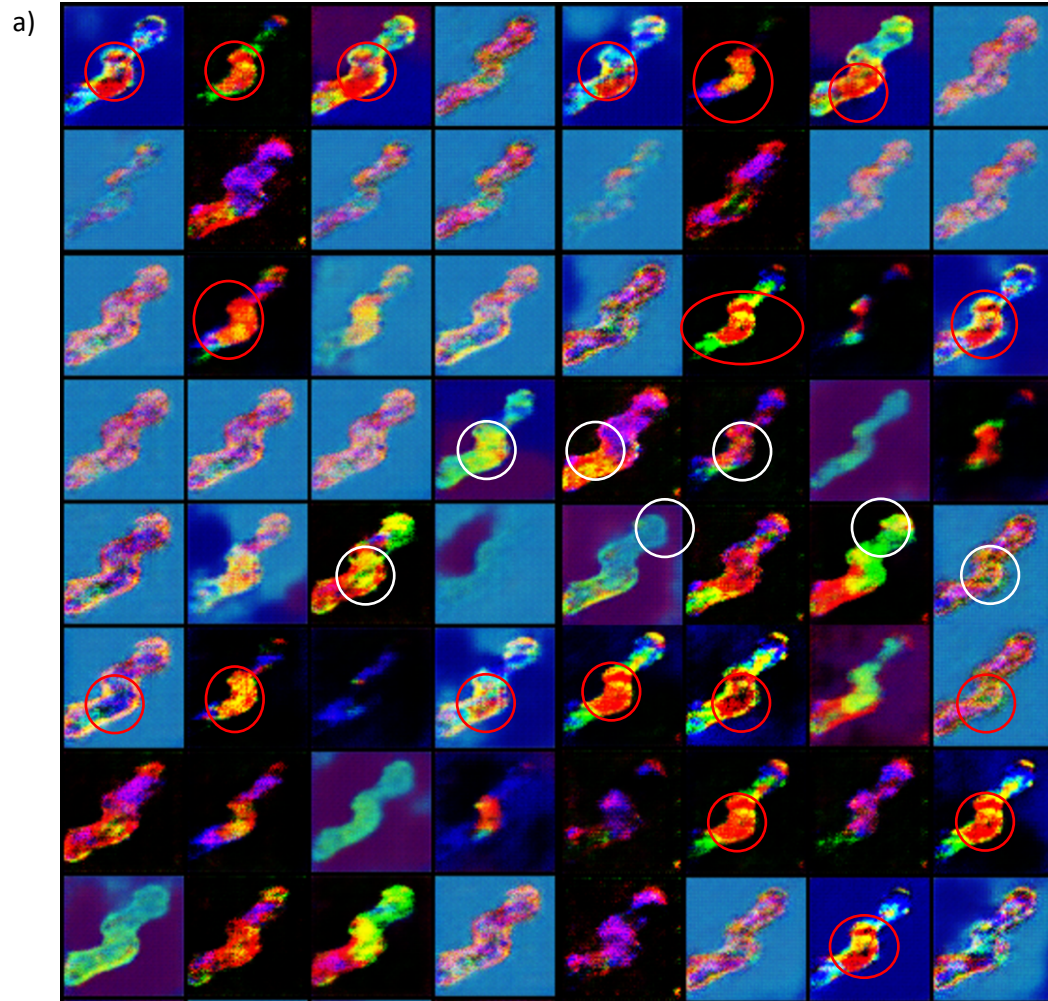


Figure 2.23. Panel a) is the GANs result for 40 epochs, each epoch takes ~ 12 minutes to complete its simulation, Panel b) shows the Losses versus epochs for both the discriminator and the generator, in theory, the smaller the loss, the closer the generation is to the original data. Panel c) shows the score versus epoch plot. For example, in panel a) red circles correlates well with the most probable reservoir in the FFP result. While white circles show the heterogeneities of the reservoirs.

Alternatively, to the time-consuming seismic inversion, self-organizing mapping (SOM) of the reservoir was used in the Puttygut Reef reservoir, using seismic attributes to capture the permeability and porosity relationship with seismic attributes. In our dataset, self-organizing map with the selected seismic attributes performs excellent in capturing the reef pinnacle and the internal reservoir architectures (Figure 2.24). The zoom on the SOM figure shows the reef core (white circle), contrasting with the reef apron (red circle) and reef talus (black circle). Nonetheless, the two SOM axes and its 2D projection in the color bar are a mathematical concept that represent the SOMs latent space, which is the projection of the manifold in N-dimensional attribute space (N= number of seismic attribute). The two SOM axes are correndered in a cross-plot and are projected in a 4096 colors bar (64 x 64 colors) and the combination of colors makes SOM axis projection interpretation highly qualitative and requires careful control to avoid the non-unicity of the axis projection. On the other hand, the vertical resolution blends storage intervals, making it difficult to assess the individual volumes where gas is accumulated. Facies and fluid probability output for the reservoir has a specific value where we can apply an additional threshold of what constitutes a reservoir; we calibrated with borehole data that the probabilities above 75% isolate the storage units. In contrast, it is highly interpretative to link a range of colors to a specific range of permeability and porosity, which adds extra uncertainties to the reservoir definition and modeling.

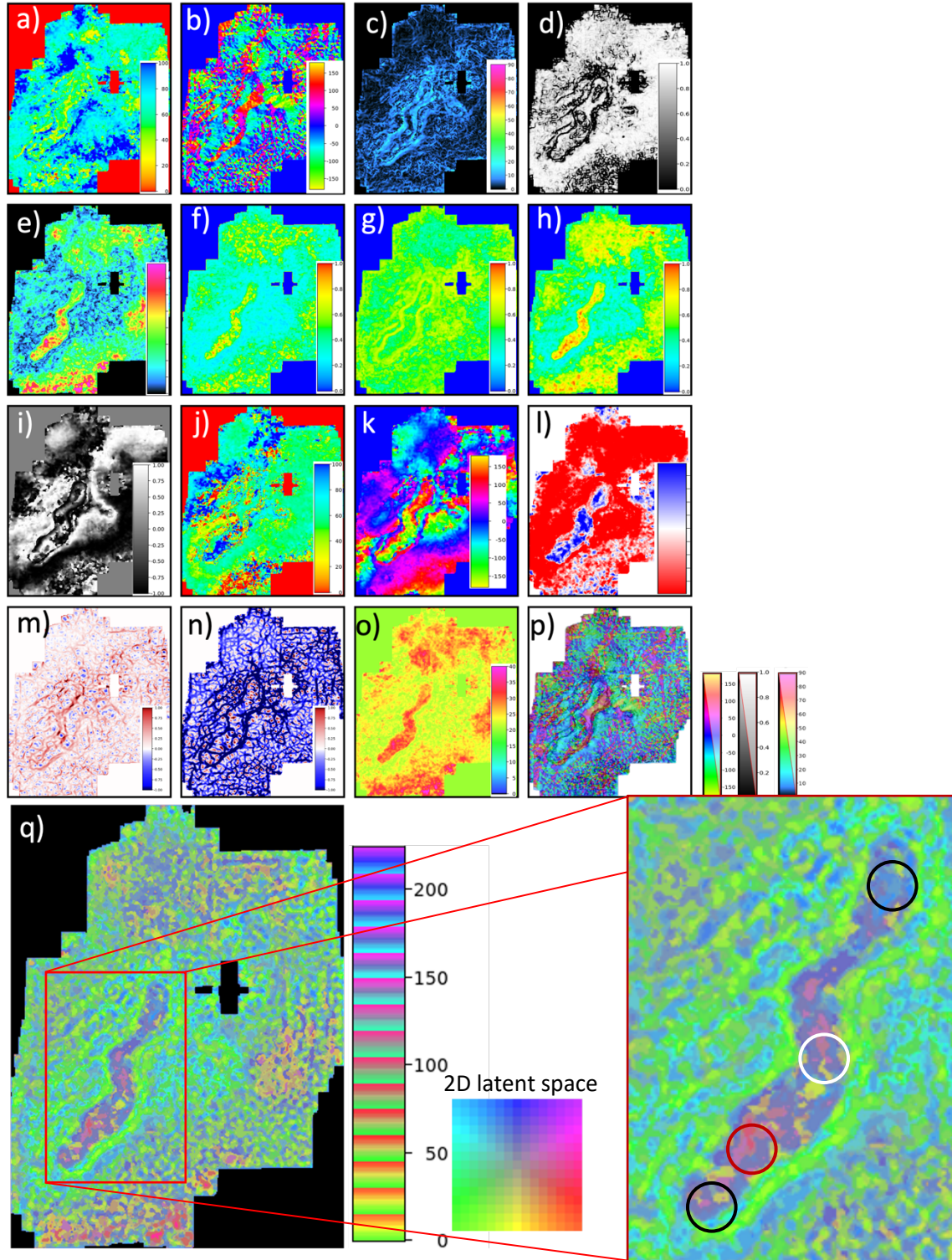


Figure 2.24. Time slice of the Ray Reef at 470 milliseconds with the different seismic attributes used in the self-organizing mapping. Panel a) shows the average frequency, b) dip azimuth, c) dip magnitude, d) energy ratio similarity, e) envelope, f) GLCM Energy, g) GLCM Entropy, h) GLCM homogeneity, i) inst. Cosine of phase, j) inst. Frequency, l) relative acoustic impedance, m) Curvature K1, n) Curvature K2, o) Inst. Frequency, p) Corendering of dip azimuth and dip magnitude with Sobel similarity.

DISCUSSION

Wold (2008), Haynie (2009) and, Silver et al., (2021) characterized the Ray Reef reservoir based on borehole data and seismic data, the later. The main difference between these three studies and our study is the small-scale heterogeneities captured in Figure 23. In this workflow, the facies and fluid probability results capture both the low porosity dolomites within the bioherm interval and the fluid effect. Moreover, our study captures the heterogeneities in the reservoir due to compaction and other diagenetic alterations (Haynie, 2009; Trout, 2012), which is accurately defined by the elastic inversion contrast the deterministic inversion attempts (Silver et al., 2021), and the machine learning characterization of a nearby reef (Buist et al., 2021). The geostatistical inversion benefits from increasing the borehole density available. While Avseth et al., (2010) predicted s-wave and p-wave velocity from porosity, lithology, and permeability, we use 24 boreholes, building a rock-physics model to augment the elastic data available at each borehole, in contrast with Silver, 2021. Figure 2.25 shows the contrast between the p-impedance from Silver, 2021 and our study and the inversion residuals. In the p-impedance section (cross-line 126) of both studies the main difference is strong wavelet imprint difference and the unbalanced high p-impedance value distribution as noted with the circle in panel a). The main differences between these p-impedance section is the number of boreholes used during the simultaneous seismic inversion when building the low-frequency model. A critical aspect on the quality control of a seismic inversion output is the seismic residuals (observed seismic – synthetic seismic), which are extremely large and with a strong phase effect (goes from negative to positive as each seismic reflector) highlighted by the black square in panel d). The goal of the residuals is to show how the synthetic seismic calculated from the parameters matches phase and amplitude of the observed seismic, as in green square of panel e). While the model base inversion is strongly dependent on

borehole control, the residuals increases where borehole data is unavailable, at the base of the Gray formation as showed with the magenta square in panel e). Nonetheless, the high residuals outside of the zone of interest are inconsequential, and the low frequency, low amplitude residuals within the reservoir and at the overburden are crucial to validate the inversion results.

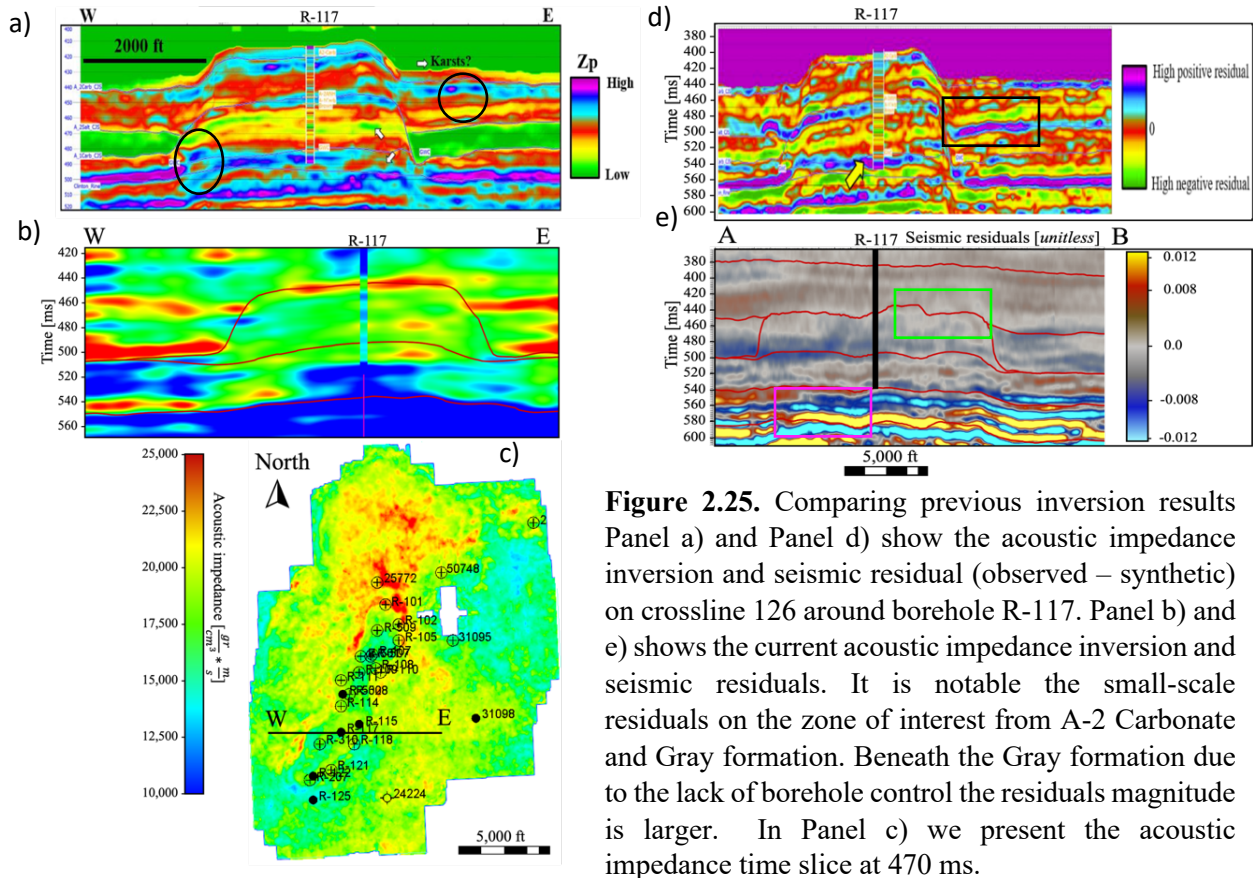


Figure 2.25. Comparing previous inversion results Panel a) and Panel d) show the acoustic impedance inversion and seismic residual (observed – synthetic) on crossline 126 around borehole R-117. Panel b) and e) shows the current acoustic impedance inversion and seismic residuals. It is notable the small-scale residuals on the zone of interest from A-2 Carbonate and Gray formation. Beneath the Gray formation due to the lack of borehole control the residuals magnitude is larger. In Panel c) we present the acoustic impedance time slice at 470 ms.

In this workflow, the improved seismic inversion allows for a more accurate reservoir characterization reflected on the volume of gas in place percentage difference between the reported injected gas and the stochastic inversion is 8 %. The seismic inversion is parametrized in terms of p-impedance, s-impedance, and bulk density, where V_p/V_s ratio is estimated from ratio of p-impedance by s-impedance ($V_p\rho/V_s\rho$). Estimating lambda-rho and mu-rho ($\lambda-\rho$ $\mu-\rho$) requires a non-linear transformation of the s-impedance and p-impedance increasing the intrinsic errors. In

contrast with the previous seismic inversion interpretations (Silver et al., 2021) we used the rock physics template from Hermana et al. (2016), which considers VpVs ratio -p-impedance. The template showed on Figure 2.12 improves the separability of the reservoir rock and reservoir fluids, allowing a clear PDF definition.

Comparing the distribution in the previous studies we successfully defined the best layers for injection and production (macro and mega pore throats), separating the vertical barriers. Additionally, our findings indicate that the core of the pinnacle and the inflection to north are the best reservoir rocks. Our hypothesis is that as the pinnacle migrated north due to the sea-level rise and the strong winds, creating the sigmoid shape of the reef, it favored erosion and diagenesis alterations that enhanced the porosity and permeability. The area and productivity index calculated from FFP and Figure 17 matched the boreholes' production index and the production history. This result shows that boreholes R-508, R-509, and R507 are better ranked and have the most considerable probability and thickness of macro and mega pore throats. Moreover, this map also shows the high porosity but most likely limited permeability of the bioclastic accumulations of the windward talus. However, the large porosity needs to show promising perspectives for developing the rims of the Ray Reef (Zheng, 1999).

Using generative adversarial networks to characterize subsurface information produces sketches (Song et al., 2021) to improve the visualization and enhanced the understanding of facies distribution. The progressive trained generator creates smooth images of the reservoir probability and shows as a viable option for future GANs-related research. Reservoir characterization is an essential field in petroleum engineering that involves modeling the physical properties of subsurface reservoirs to optimize oil and gas production. One of the critical challenges in reservoir

characterization is obtaining accurate and comprehensive data about the subsurface properties, which can be difficult and expensive to obtain through traditional means. The use of GANs in reservoir characterization is still a relatively new field, and many challenges need to be addressed. One challenge is the limited availability of high-quality subsurface data for training the GANs models. Another challenge is the difficulty in validating the accuracy of the synthetic data generated by the GANs models.

Despite these challenges, there have been several successful applications of GANs in reservoir characterization. For example, Wang et al. (2020) used GANs to generate synthetic seismic data for subsurface imaging and found that the synthetic data improved the accuracy of the subsurface model. Li et al. (2021) used GANs to generate synthetic borehole log data to characterize carbonate reservoirs. They found that the synthetic data provided a more comprehensive understanding of the subsurface properties. Our study contextualizes the data integration and presents GANs as a novel simulation for this mature gas storage reservoir.

Our facies fluid probability and GANs results can guide enhanced recovery process, injecting in the areas where both FFP and GANs consistently show a high reservoir probability. Injecting, for example, CO₂ in better connected intervals allows CO₂ percolate in larger rock porous area similar to what Harrison (2020) found. Toelle (2019) characterizes the Ray reef using seismic data aiming to delineate strategies for enhanced recovery, our FFP results, compares with the porosity map presented, particularly at the core of the reef.

The self-organizing map results presented in Figure 22 exhibit the facies variability within the core of the reef, similar to the geological variability in Buist et al. (2021). Nonetheless, the non-unicity makes the clustering map a qualitative interpretation technique and its integration to a

fluid simulation software to model the fluid distribution is no straightforward. We compared the results with facies map, and it mixes bioherm facies with core reef facies.

CONCLUSION

Our proposed methodology applied in this mature gas field highlights how the geostatistical seismic inversion better defines reservoir heterogeneities and thin flow zones. We benefit from having denser well control to build accurate rock physics models and calibrate the elastic inversion, contrasting with a previous study that had sparser well data. Contrasting the volumetric estimation of gas in place from the stochastic approach serves as a quality control of volume of the available storage space (8% difference from reported injected volumes) it is an improvement over a past machine learning study on a nearby reef and other seismic inversion which the individual injection volumes. At this point the seismic reservoir characterization is to the leading edge of technology in the industry. Generating a fluid and facies probability of a reservoir is an industry standard. However, using generative adversarial network to make maps of the reservoir probability is at this point an academic research question to estimate the value of it in such a workflow.

In conclusion, the utilization of geostatistical seismic inversion methodologies stands out as paramount, with particular emphasis on the pivotal role of rock-physics, which transcends mere novelty to become indispensable in seismic inversion processes. The merging of geostatistical seismic inversion with generative adversarial networks facilitates the production of dependable 3D models delineating gas storage capacity and productivity index distribution within reservoirs, thereby encapsulating the spatial distribution of optimal storage and production zones. Our geostatistical seismic inversion bests the deterministic inversion by enhancing vertical resolution,

thus enabling a more precise delineation of reservoir heterogeneities, and fostering correlation with established gas storage profiles derived from well data.

The integration of generative adversarial networks into reservoir characterizations represents a growing frontier capable of generate multiple simulations at a fraction cost of Monte-Carlo simulations of the reservoir. GANs, proficient at synthesizing data similar to actual datasets, serve to bridge gaps in existing information, thereby delivering a more holistic comprehension of subsurface reservoirs. While acknowledging existing challenges and the need to impose physical constraints in the model, GANs is a valuable tool for reservoir characterizations.

GANs afford the capability to generate myriad realizations of reservoirs, each embodying varying probabilities of high productivity index occurrences, thereby endowing an additional dimension to assess potential reservoir distributions. The volumetric estimation derived from stochastic inversion exhibits a mere 8% deviation from reported injected gas volumes, thereby attesting to the methodology's accuracy. The integration of rock physics modeling, seismic inversion techniques, and machine learning methodologies such as GANs presents a pragmatic approach to data integration, thereby facilitating the characterization of complex reservoirs and mitigating uncertainties inherent in storage capacity estimation.

DATA AVAILABILITY

Dataset presented in this article needs to be requested from authors, for authorization to share from data owners.

ACKNOWLEDGMENTS

We thank SLB, Matlab, CGG and AASPI (University of Oklahoma) for the software used for, seismic visualization and borehole integration, rock physics models, synthetic data generation, seismic inversion, and attribute. Additionally, thanks to the AASPI group at OU and SEG organization for their financial support.

CHAPTER 3: IN-DEPTH ANALYSIS OF SELF-ORGANIZING MAPS FOR CHARACTERIZING AMPLITUDE PLAYS IN SILICICLASTIC RESERVOIRS WITH ROCK PHYSICS-DERIVED 3D SYNTHETIC MODELS. *

*This chapter is under review in the SEG & AAPG journal, *Interpretation*, as a general paper submission

ABSTRACT

Estimating rock properties from seismic data is crucial for quantifying the extent of hydrocarbon reservoirs. However, this process is often time-consuming and requires the integration of multiple data sources. In contrast, combining seismic attributes using self-organizing maps (SOMs) offers a rapid and successful approach to defining seismic facies. The accuracy of SOMs to characterize siliciclastic reservoirs is tested against a 3D synthetic seismic dataset and a real seismic dataset. A combination of seismic attributes and best practices in exploratory data analysis are used to evaluate the reliability of the self-organizing map to differentiate reservoir rock (lithology) from the fluid response. The SOMs reservoir classification is contrasted against a 3D synthetic seismic, which includes three different reservoirs, each with three different fluid amplitude responses (Gas – Oil- Brine) that honor a theoretical rock physics response and are labeled to evaluate the accuracy and sensitivity of SOMs. Additionally, SOM reservoir classification accuracy is compared against a petroelastic simultaneous seismic inversion from the Stybarrow seismic dataset. We find that the self-organizing mapping of siliciclastic reservoirs using seismic attributes improves the lithological discrimination but not the fluid detection.

INTRODUCTION

Unsupervised machine learning has gained widespread recognition and popularity for its ability to handle vast and complex datasets effectively. In recent years, substantial efforts have been devoted to refining data classification strategies. Self-organizing maps (SOMs), originally developed by Kohonen (1990; 2013) for financial data clustering, have found expanded utility in

the field of seismic interpretation. Particularly in conjunction with seismic attributes, as demonstrated by Coléou and Azbel (2003), De Matos et al. (2007), Roy et al. (2010), Zhao et al. (2015), Hussein et al. (2021a), La Marca et al. (2020, 2021, 2022). SOMs have not only proven successful in characterizing seismic facies but have also been applied to characterizing hydrocarbon reservoirs (Roden et al., 2015; Hussein et al., 2021b). Moreover, the combination of self-organizing maps and seismic attributes has shown promise in delineating subtle amplitude variations related to hydrocarbons, both in offshore 2D seismic lines (Chenin & Bedle, 2022) and onshore carbonate reservoirs (Buist & Bedle, 2021) as well as offshore carbonate reservoirs (Maas et al., 2023).

In a significant advancement, Chenin and Bedle (2020), followed by Lubo-Robles et al. (2023), proposed using a suite of seismic attributes and SOMs as a tool for gas hydrate detection. However, questions remain regarding the accuracy of these results and whether they are contingent on input parameters or consistently produced by the clustering algorithm. Addressing these inquiries is of utmost importance.

Since Coléou and Azbel's pioneering work in 2003, the availability of seismic attributes within interpretation software has grown exponentially (Chopra & Marfurt, 2007; Barnes, 2006, 2016; Lubo-Robles et al., 2023). As the number of attributes has increased, dimensionality reduction techniques have become essential to select the most informative inputs, saving computational time and redundancies.

Exploratory Data Analysis (EDA) plays a pivotal role in machine learning and seismic attribute interpretation, providing valuable insights into subsurface structures and unlocking the potential of seismic data. EDA enables visualizing seismic attributes' distribution, variability, and

patterns. Techniques such as histograms, scatterplots, and box plots help identify trends, anomalies, and potential outliers, highlighting geological areas of interest and data quality issues. Furthermore, EDA facilitates the identification of associations and correlations between seismic attributes and geological features. By exploring spatial relationships and cross-plots, meaningful connections can be discovered, such as the relationship between amplitude attributes and reservoir lithology or fluids. EDA also aids in feature selection, determining cluster numbers, and identifying data redundancies, setting the stage for other dimensionality reduction techniques. Principal Component Analysis (PCA) is selected to reduce data size and redundancies, thus saving computational time, and simplifying complex datasets.

Self-organizing mapping is an appealing unsupervised clustering algorithm due to its minimal reliance on a priori information, limited only to determining the number of clusters. However, it remains sensitive to input data and is influenced by the number of realizations (Vatanen et al., 2015). The original SOM formulation, rooted in heuristic gradient-descent procedures, has evolved to enhance thresholding in self-organizing machines (Ross-Ashby, 1962; Keller, 2008). SOMs have been successfully applied to identify geological features, including seismic facies identification using a wide range of seismic attributes, such as phase, frequency, and texture attributes (Roy et al., 2014; Roden et al., 2015; Zhao et al., 2016; Lubo-Robles & Marfurt, 2017; Ha et al., 2021). These algorithms have also been instrumental in studying subtle and complex fault systems (Hussein et al., 2021a; Perico et al., 2022) and have found diverse applications in geophysics, including seismic inversion, reservoir characterization, and earthquake clustering analysis (Gholami et al., 2017; Guo et al., 2021; Ren et al., 2020; Zhang et al., 2022).

Conversely, the application of simultaneous AVO inversion or simultaneous seismic inversion (SSI) for characterizing siliciclastic reservoirs and estimating petroelastic responses associated with various hydrocarbons is well-documented (Grana & Della Rossa, 2010; Zhang et al., 2012; Arévalo-López et al., 2016; Wollner et al., 2017a; Yenwongfai et al., 2017). SSI proves effective in reservoir prospecting, particularly in data-scarce scenarios, including situations where a rock physics model (Avseth, 2010) is unavailable. Furthermore, it is a valuable tool in reservoir characterization during field appraisal to mitigate risks in business decisions (Kelly et al., 2001; Skidmore et al., 2001). Unlike deterministic or model-based seismic inversion, SSI requires only a single initial estimate of the earth's property models, which are iteratively adjusted to align with observed seismic data while maintaining sparsity and geological constraints.

DATA AND WORKFLOW

Data Description

The dataset with partial angle stacks and borehole wireline information available to assess the accuracy of UML is in the Exmouth-Basin. Stybarrow field was discovered in 2003, and the reservoir corresponds to high-quality but friable sandstone rocks deposited in the Early Cretaceous (Macedon Formation). The Stybarrow (Figure 1) reservoir is a structural trap limited by a NE-SW trending tilted fault system. The Stybarrow field was discovered in 2004 with the drilling of Stybarrow-1 on the continental shelf at 825 meters of water depth and found the Macedon sandstones at 2,230 meters depth. The field was decommissioned in June 2015 (Wilkinson, 2015; Cavcic, 2022). The initial reservoir characterization calculated hydrocarbon reserves close to 80 MMBE (Duncan et al., 2013; Wilkinson, 2015). In 2015, 64 MMBOE were recovered using eight production boreholes and two water injectors connected to a floating production storage and offloading vessel (FPSO).

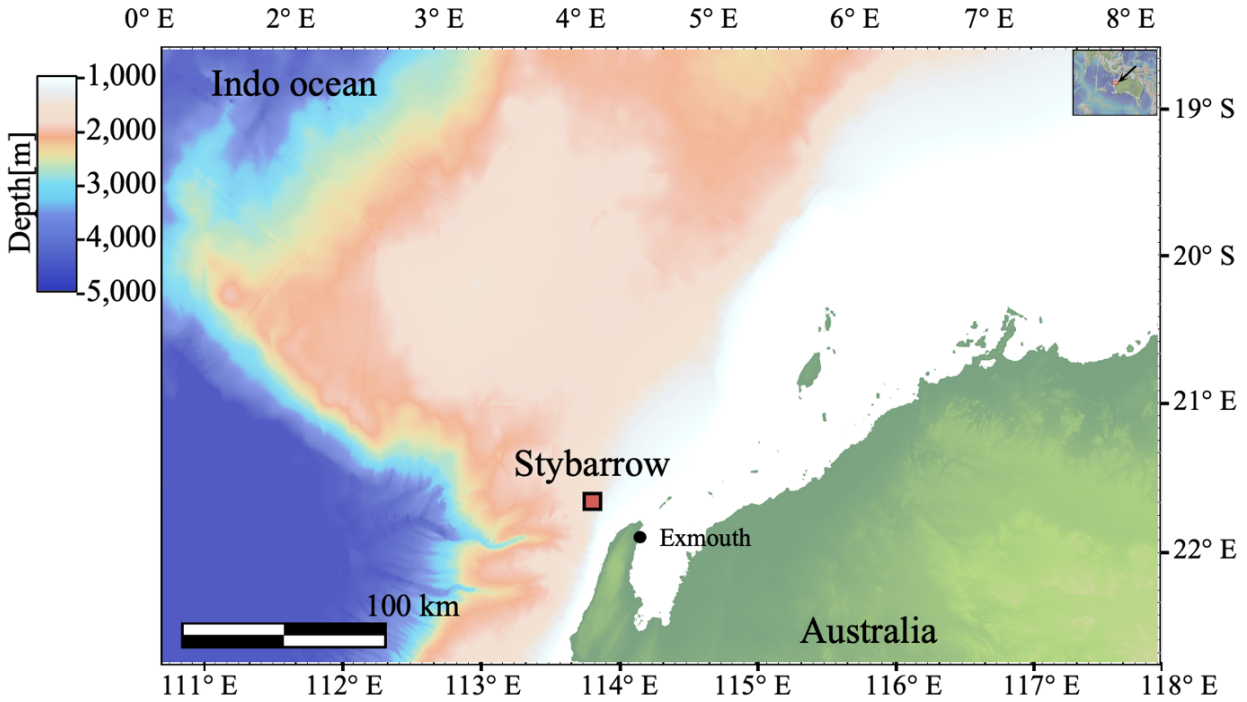


Figure 3.1. Location map of datasets used in this study, the Stybarrow oil field located 50 km Northwest of Exmouth on the Exmouth sub-basin. Four boreholes are available with a complete suite of borehole wireline logs. The field is produced in siliciclastic systems with a well-documented history of production.

The Stybarrow field dataset has four partial angle stacks (near, mid, far, and ultra-far), processed in 2003 and sampled at three milliseconds (3 ms), four exploratory boreholes, and a detailed petrophysical core analysis (Locke, 2004; Arevalo Lopez et al., 2016). The datasets have enough information to conduct SSI and SOMs. The production history of the boreholes is available to expand more on the performance of each borehole (Wilkinson, 2015). The critical reservoir parameters of the Stybarrow field are presented in Table 3.1.

Reservoir metrics	Units	Stybarrow
Depth	Meters [m]	2,230
Hydrocarbon density	API°	22
Average porosity	[%]	18
Average Water Saturation	[%]	25
Gas-oil-ratio	[sm ³ / sm ³]	50-72
Pressure gradient.	PSI/meter	1.18
Temperature	° Celsius	65
Area	Kilometers [Km ²]	7
Hydrocarbon column height	Meters [m]	90

Net pay	Meters [m]	35
Net-to-gross	[p.u]	0.9
Porosity range	[p.u]	12-30
Permeability range	[mD]	10-4,000
Oil formation volume factor	[rb/stb]	1.25
Recovered volume	MBOE*	64

Table 3.1. Key Stybarrow field metrics depict the properties of the reservoir and its performance.

The selected dataset has four partial angle stacks and a fullstack of the Stybarrow field, near [3-13°], mid [10-23°], far [20-30°], and ultra-far [30-41°]. The seismic exhibits a European polarity convention, a dominant frequency of 35 Hz, and an amplitude response as a function of angle (AVA) Figure 3.2. Additionally, the reservoir shows inner heterogeneities that impact the extension and performance of the reservoir (Ementon et al., 2004; Hill et al., 2008; Duncan et al., 2013; Halpin et al., 2013), adding an extra layer of complexity for the SOMs to classify.

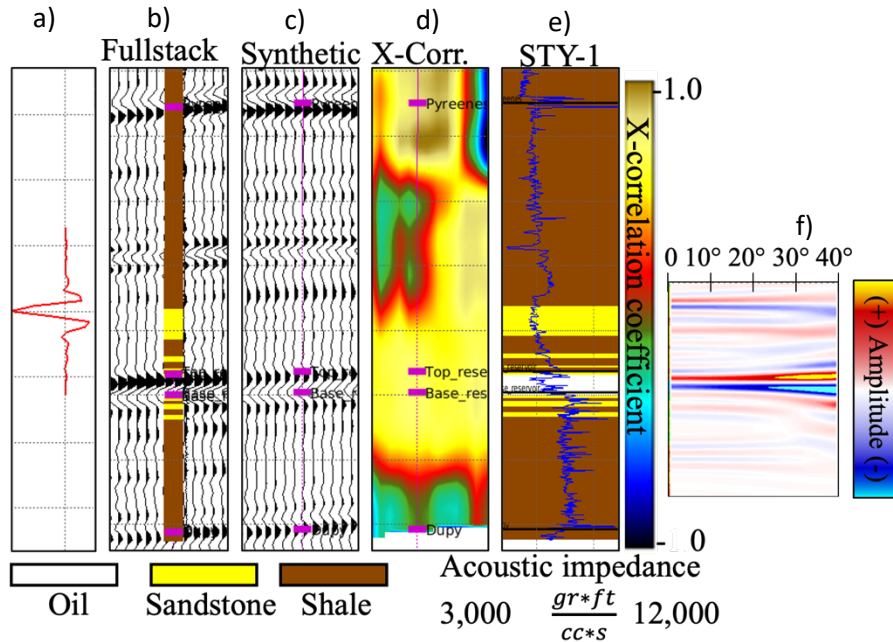


Figure 3.2. Well-tie of STY-1 borehole showing the polarity of the reservoir rock and the AVO/AVA effect. Panel a) the extracted wavelet at the borehole location. Panel b) shows the observed seismic overlaid with the litho-types, and c) the synthetic seismic calculated from the convolution of the wavelet and the impedance log. Panel d) shows the cross-correlation of the observed dataset with the synthetic seismic, e) the p-impedance profile recorded at the borehole, and f) A synthetic gather is calculated to show the in-situ AVO response.

Figure 3.3 illustrates a seismic line traversing the reservoir along with borehole data comprising gamma-ray and resistivity measurements. This figure demonstrates the enhancement

of hydrocarbon response achieved through structural-oriented filtering and spectral balancing techniques, aimed at improving the seismic signal-to-noise ratio and bandwidth (Fehmers & Höcker, 2003). These improvements in seismic data quality are particularly beneficial for structural and stratigraphic interpretation, especially in the calculation of seismic attributes. The application of root-mean-squared (RMS) amplitude extraction at the reservoir's top (Macedonian formation) reveals that the relative amplitudes remain consistent, while the seismic anomaly exhibits enhanced horizontal resolution.

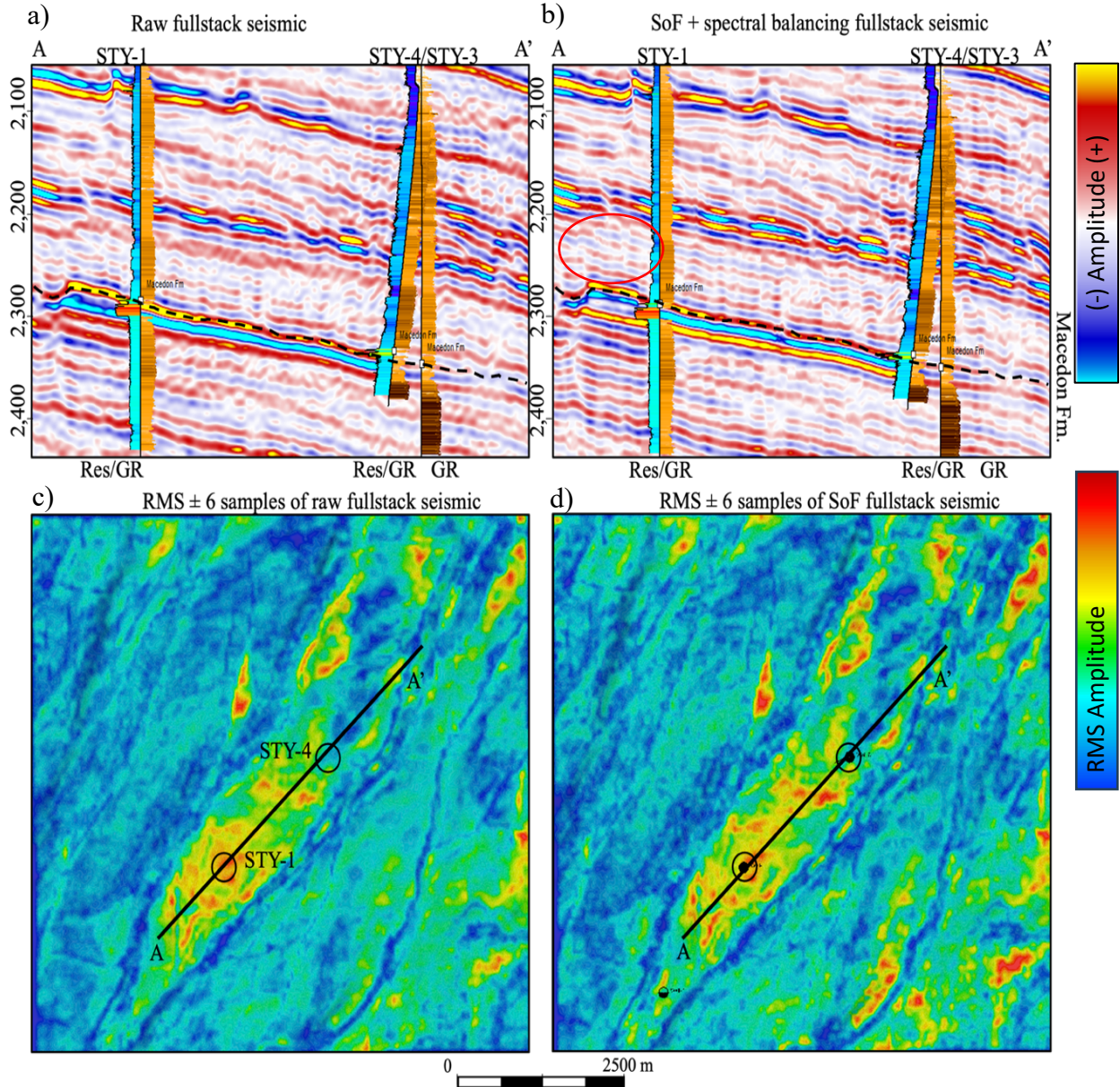


Figure 3.3. Observed seismic before (Panel a) and after post-processing (Panel b) across the reservoir, see the increased resolution within the reservoir close to the borehole, highlighted by the red circle. Panel c) shows the RMS amplitude extraction over the reservoir interface before processing, and Panel d) The RMS extraction after post-processing.

WORKFLOW

Seismic interpreters work in multidisciplinary groups that agglomerate multiscale and multidisciplinary data to solve ill-posed multivariate problems – the process is called integration. Machine learning helps analyze, cluster, and interpret information. While SOMs expedite the

processes of data integration, facies prediction, and seismic interpretation, the issue of result reliability remains partially unresolved. To address this concern, it is essential to validate SOMs using a synthetic model, as suggested in this study. The methodology employed to assess the accuracy of SOMs and compare them with reservoir facies prediction is illustrated in Figure 3.4.

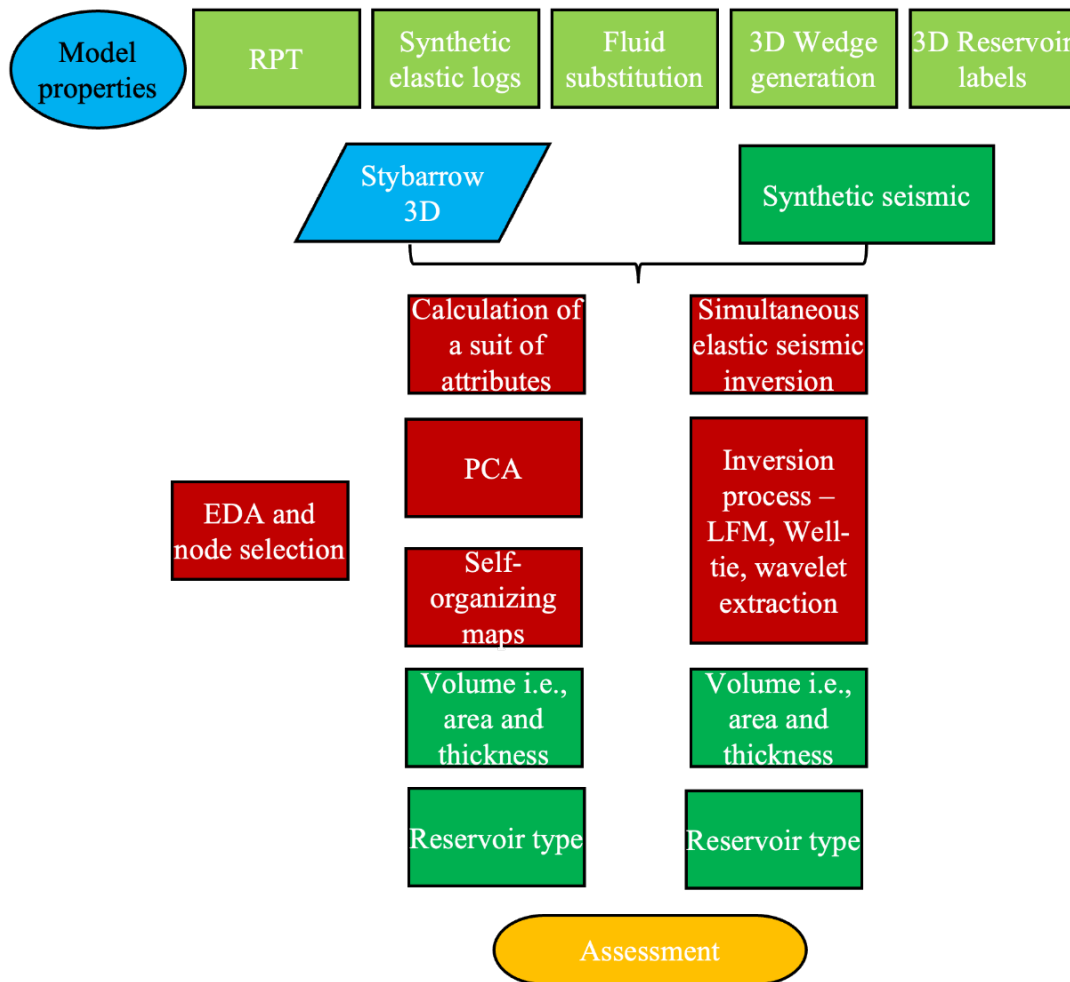


Figure 3.4. Proposed workflow to evaluate the accuracy of self-organizing maps (SOMs). The acronym RPT stands for Rock Physics Template, and EDA is short for Exploratory Data Analysis. Petroelastic seismic inversion is used as a tool to compare the accuracy, particularly on the observed dataset. The workflow seeks to evaluate the sensitivity of SOMs to reservoirs that exhibit an amplitude seismic response.

The methodology presented above has three main subdivisions. In the first stage of the methodology, rock physics modeling is used to generate reliable pre-stack synthetic seismic data. The 3D synthetic seismic exhibits different amplitude responses associated with reservoir fluids at different conditions. In the second stage, an observed post-stack seismic volume joins the synthetic

pre-stack seismic to conduct two parallel methodologies: i) The self-organizing map to characterize the reservoirs on the synthetic data and the Stybarrow field. ii) The petroelastic seismic inversion, where reservoir properties are extracted from p-impedance and VpVs volumes. In the third stage, the results are compared, highlighting the differences and advantages of each methodology. Below, we describe these methods and the reasoning behind incorporating them into the analysis.

Self-organizing maps (SOM) of seismic datasets

Self-organizing maps (SOMs) are an unsupervised machine learning technique developed by Kohonen in 1982. SOMs have been widely used for seismic facies identification (Roy et al., 2014; Roden et al., 2015). The key idea is to map high-dimensional input data onto a lower-dimensional manifold, usually a 2D grid while preserving topological relationships. This mapping clusters similar input data points together on the grid.

The SOM training iteratively updates grid point weights to match the input data distribution. A random input is presented in each training epoch closest grid point and neighbors to match that input better. These adjustments diminish with distance from the selected point. After training, the input space is effectively mapped to the 2D grid, positioning similar inputs nearby. Resulting clusters can be visualized with colors, where similar hues indicate similar seismic responses or facies (Zhao et al., 2017, 2018).

From the literature review, the elbow plot (% of variance) calculated suggests using 12 clusters. SOMs require grid size for initialization, so a 12x12 grid is selected, with each cell representing a cluster. A 64x64 grid is also tested, effectively projecting 4096 clusters.

Petroelastic seismic inversion

Previous studies (Fatti et al., 1994; Helland-Hansen et al., 1997; Torres-Verdín et al., 1999; Ma, 2001; Ma, 2002; Quackenbush et al., 2006; Barajas-Olalde et al., 2019) have routinely employed petroelastic seismic inversion techniques for hydrocarbon volumetric estimations. Uri et al. (2017) presented a methodology for correlating seismic inverted elastic properties with hydrocarbon reserves on siliciclastic datasets.

Petroelastic seismic inversion begins with a simultaneous seismic inversion on angle stacks to calculate p-impedance, s-impedance, and density from reflection data. A rock physics model is applied at the borehole location to establish correlations between the elastic parameters and total porosity, clay content, and water saturation. In the case of the synthetic dataset, the reservoir's volume (both area and thickness) is estimated through a cross-plot of values associated with Rock Physics Templates (RPT). For the Stybarrow field, borehole elastic properties from four boreholes are plotted, and the area of the RPT cross-plot associated with hydrocarbon-bearing zones is mapped onto the seismic inversion volumes.

The p-impedance, s-impedance, and density inverted models result from meticulous parameterization and utilizing the three partial angle stacks with their corresponding wavelets. In the synthetic dataset, the low-frequency models are derived by smoothing p-impedance, s-impedance, and density volumes. In contrast, for the Stybarrow dataset, the low-frequency models are constructed using borehole data and a smoothly varying velocity model. In the synthetic dataset, a merge cut-off frequency of 5 Hz is applied, and the wavelets are scaled as follows: near (0-15°), mid (15-30°), and far (30-45°). The wavelet for each partial stack (near, mid, far, ultra-far) is extracted using the observed dataset using the borehole locations and well-tie.

Volume (area and thickness) and reservoir type comparison

Our next step involves transforming and comparing the effectiveness of SOMs and seismic inversion in identifying different reservoir types and quantifying the disparity in hydrocarbon volume estimation.

In the synthetic case, we evaluate the accuracy of the results of petroelastic seismic inversion for synthetic and Stybarrow datasets and the self-organizing mapping of the reservoirs by contrasting the outcomes with the actual reservoir model and measuring the degree of correlation for the reservoirs and fluids in the reservoir.

In the results of petroelastic seismic inversion for Stybarrow datasets and the self-organizing mapping of the reservoirs, we examine geobodies. We employ the area and thickness (true-vertical thickness) of the extracted geobodies, to estimate the initial volume of in-place oil (STOIIP) using both methods and measure the difference against the decommissioning report.

We adopt the same reservoir parameters for both methods and utilize the initial volume-in-place formula (STOIIP) for the observed reservoir and seismic dataset. We substitute the area and thickness values derived from each technique and compare them with the total hydrocarbon volume reported in the decommissioning report.

$$STOIIP_{inv\ or\ SOMs} = \frac{7758 * Area_{inv\ or\ SOMs} * h_{inv\ or\ SOMs} * \theta_{avg} * S_o}{B_{oi}}$$

Where:

$STOIIP_{inv\ or\ SOMs}$ = stock tank oil in place of the petro elastic inversion or SOMs, depending on which area and thickness is used. (STB)

$Area_{inv\ or\ SOMs}$ = Area of the petro elastic inversion or SOMs (Acre)

$H_{inv\ or\ SOMs}$ = Thickness of the petro elastic inversion or SOMs (feet)

θ_{avg} = Average porosity (25%) from borehole histogram (%)

So = Oil saturation (80 %) from borehole data

Boi = Oil formation volume factor (1.25 rb/stb)

The documented volume of hydrocarbon produced by the end of 2015 is 64 MMBOE. The original estimation was 60-90 MMBOE (Wilkinson, 2015).

In the following sections, we build upon the insights from the literature review, which highlights best practices involving seismic attributes and machine learning. We leverage Exploratory Data Analysis (EDA) and Principal Component Analysis (PCA) in our methodology to select relevant seismic attributes, as identified in the literature, for characterizing hydrocarbons in siliciclastic reservoirs using self-organizing maps (SOMs). Additionally, we present the established methodology for characterizing hydrocarbon presence in sandstone as a benchmark to evaluate the accuracy of SOMs on real-world datasets. This includes introducing the observed dataset and detailing the reservoir parameters. Through a comparative analysis, we assess the performance of SOMs against simultaneous seismic inversion. The following section provides a comprehensive exposition of our proposed methodology.

SYNTHETIC MODEL AND ATTRIBUTE ANALYSIS RESULTS

Average sandstone mineral composition and elastic moduli (Schon, 2004) create the dry-rock elastic response using the modified Hashin-Shtrikman bounds as a function of total rock porosity (Mavko et al., 2009; Bosch et al., 2014). The rock physics model is recommended when linking rock properties with seismic amplitudes (Castagna et al., 1985; 1993; Avseth, 2010; Avseth and Lehocki, 2021). The surrounding rock is a shale with constant elastic moduli. For more information about how the elastic properties are defined as a function of porosity, the reader is invited to Appendix E.

The fluid substitution of gas, hydrocarbon, and brine on the dry-rock framework allows the generation of the three reservoirs (Kumar, 2006). Porosity changes are already included in the dry-rock elastic response, and the fluid elastic properties follow Batzle and Wang (1992). Figure 3.5 synthesizes the elastic and petrophysical properties of the reservoir models.

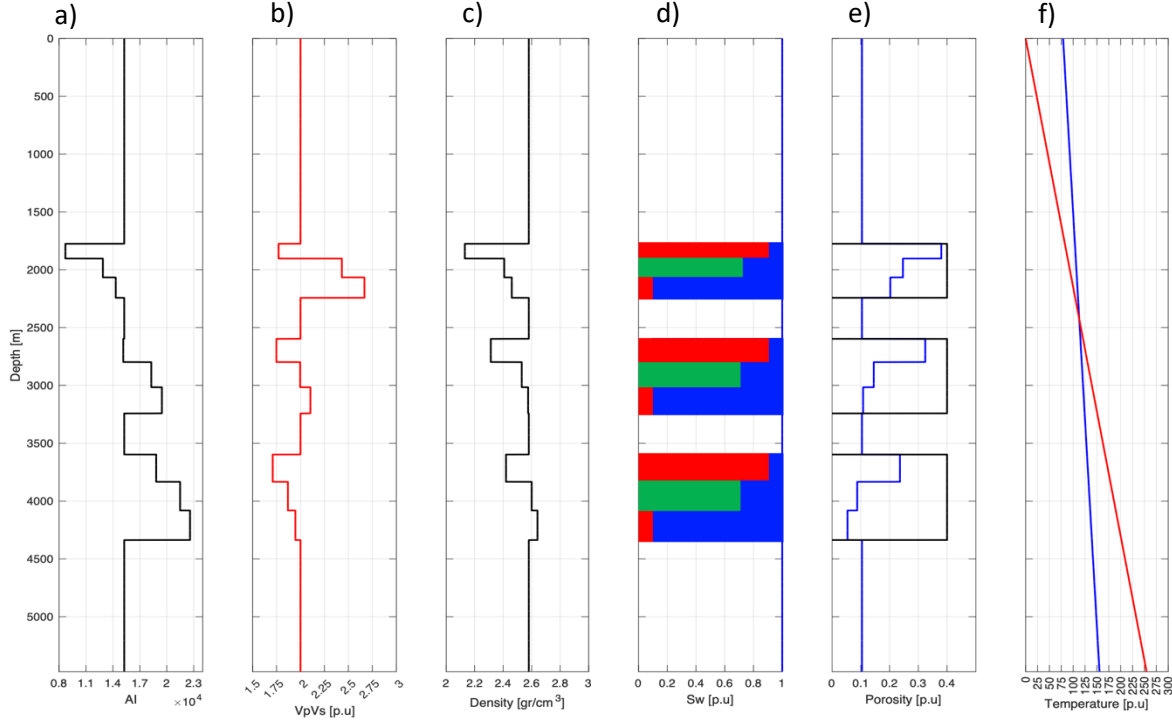


Figure 3.5. Synthetic elastic and petrophysical properties of the three reservoirs proposed. These plots are pseudo boreholes. Panel a) Acoustic impedance, b) Vp/Vs ratio, c) Bulk Density, d) water saturation and the schematic of each fluid (Red represents wet gas, green is 32 ° API oil, and Blue is brine). Panel e) is the total porosity (blue) and the defined reservoir space, f) Pressure and temperature plots included in the Fluid substitution model for each reservoir.

Figure 3.5 is paramount in the methodology; it shows the properties' variability and is the target of SOM classification. The synthetic model with defined parameters defines amplitude anomaly to each fluid response and defines the volume labels for each reservoir and lithology. The boundary between lithology sandstone and shale is the first objective we seek to resolve using SOMs. Nonetheless, the main objective is to define fluid contacts between the three reservoirs using SOMs. Additionally, the rock physics model used to generate the sandstone reservoirs and fluid response brings transparency to the seismic modeling. The seismic trace in its purest form

can be expressed as the convolution of a wavelet with the reflectivity series from the earth and random noise. The petroelastic model is extrapolated (Figure 3.6) and uses a Backus average (Backus, 1962) of the P-velocity to generate a time-to-depth relationship to transform the models into the time domain and perform the convolution with each wavelet to generate the pre-stack synthetic model with an angle distribution from 0 to 45°.

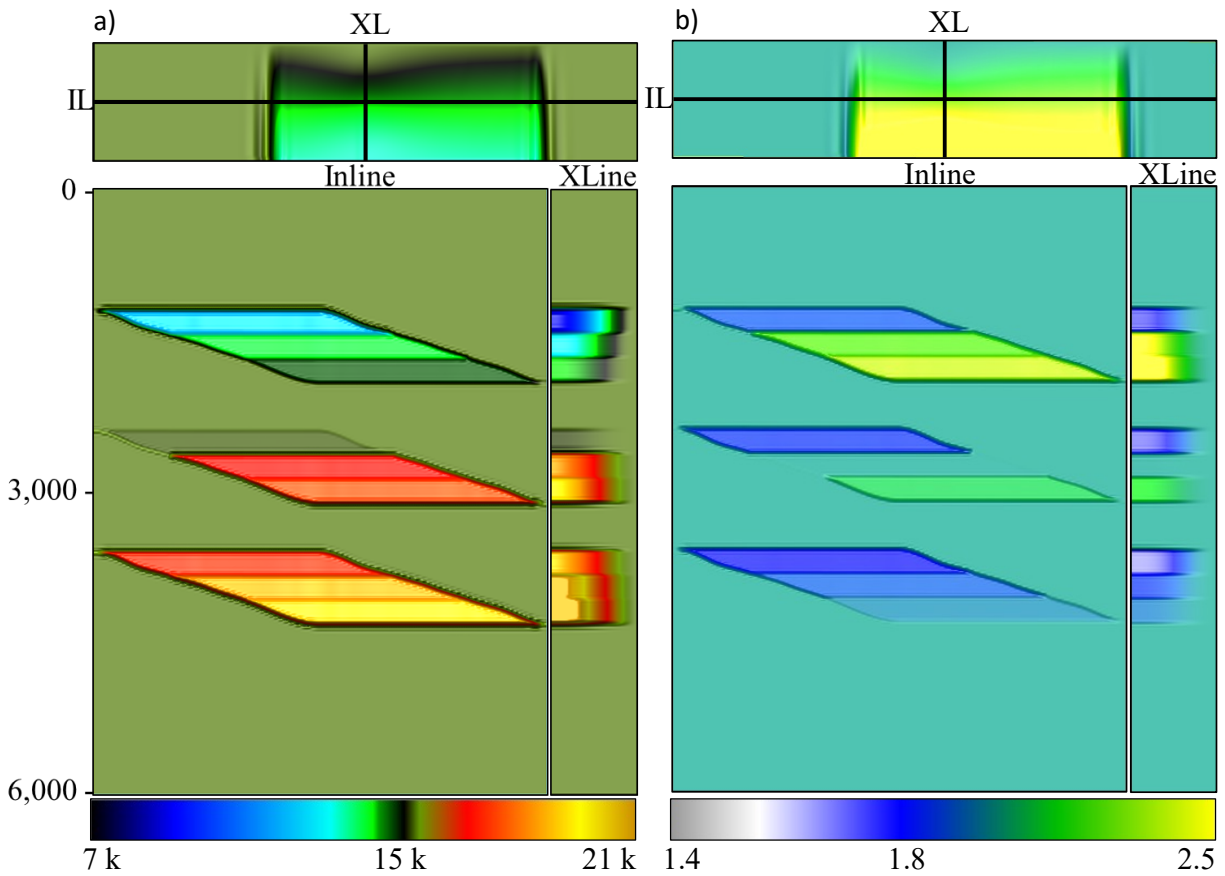
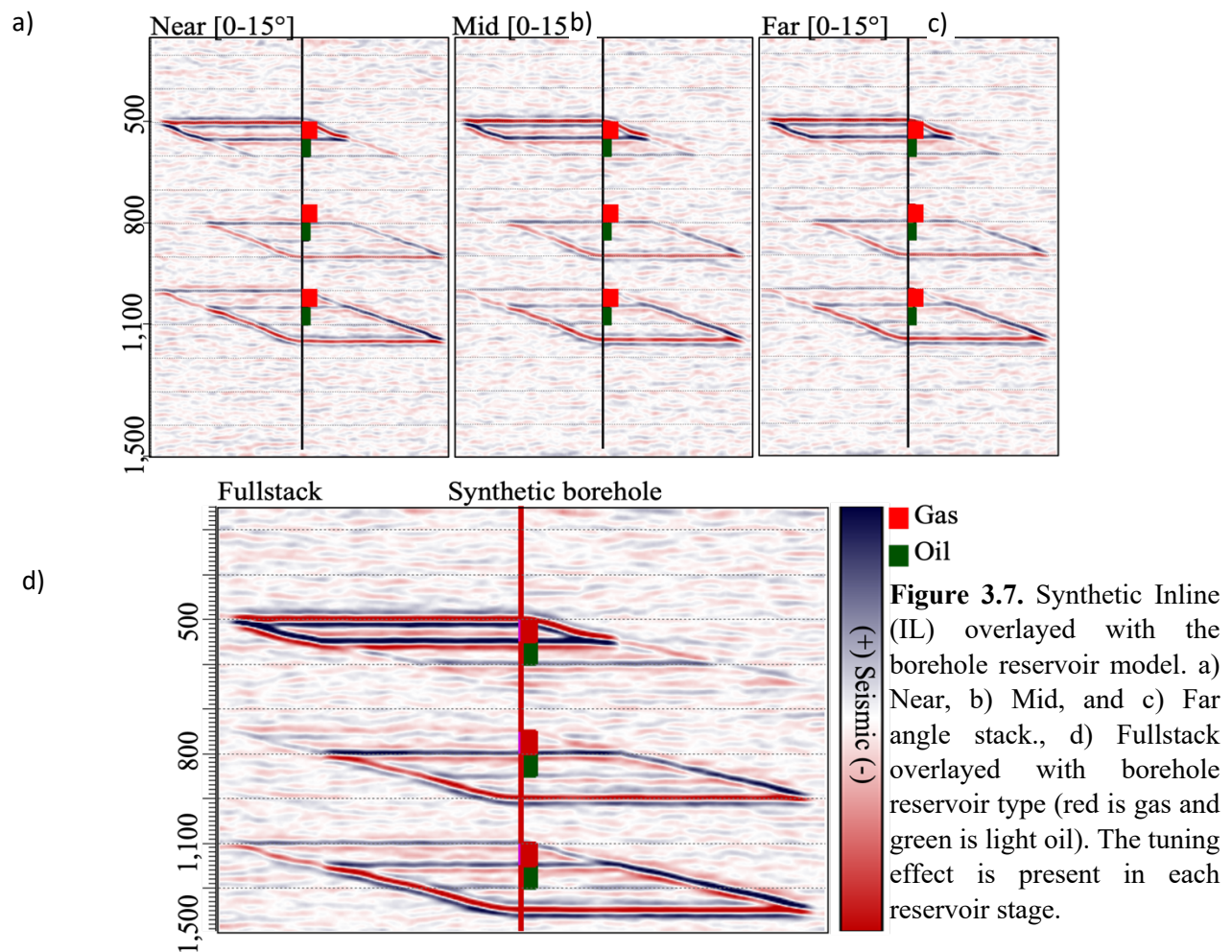


Figure 3.6. 3D Synthetic elastic models, pane a) synthetic Inline (IL), crossline (XL), and depth slice of a) P-Impedance, b) VpVs ratio.

Three partial angle stacks are created from the pre-stack seismic using the relationship between AVA reflection coefficients for PP reflectivity (r_{pp}) and elastic parameters, as described by the Knott-Zoeppritz equations (Sheriff, 1995; Knott, 1899). Absorption mechanisms are disregarded, resulting in a laterally and vertically constant wavelet. Spectral noise is introduced

into each partial angle stack. Spectrally shaped noise is preferred over random amplitude noise for real-world conditions (Hayes, 1996; Oppenheim & Schaffer, 2010). Spectrally shaped noise is typically generated by applying a filter to white noise. This filter shapes the frequency components of the noise according to the desired spectral characteristics (Bendat & Piersol, 2010).

The petroelastic seismic inversion process requires the generation of these partial angle stacks, and the suite of attribute calculations used for the SOMs classification is derived from the fullstack. The full stack is calculated from the geometric mean of the partial angle stacks (Figure 3.7). The geometric mean is less sensitive to extreme values since it involves multiplication. It measures the central amplitude that reflects the AVO effect between partial stacks (Ross, 2009).



Generation and selection of a suite of attributes

With the improved resolution seen in Figure 3.3 and the synthetic data presented in Figure 3.7, our next task is to select the seismic attributes that contribute to the SOM clustering of the reservoir. The following attributes were computed on enhanced seismic data of the Stybarrow field and the synthetic data. Chopra and Marfurt (2005) classified more than 50 seismic attributes, including meta-attributes, i.e., attributes computed from others. Barnes (2007) already showed attribute redundancy; therefore, our literature review suggests using the seismic attributes in Table 2 to perform the EDA and PCA to select the subset of attributes with the highest importance to conduct the reservoir clustering using SOMs.

Attribute name	Definition	Interpretation	Example of usage
Amplitude volume technique (AVT)	Average of the square amplitudes as a function of frequency content.	Highlight anomalous amplitudes	Vernego & Trincherro, 2015
Aberrrancy	Third derivative of structure, the rotation of the coordinate system aligns with the local reflector dip and azimuth.	Detection of subtle changes in dip or azimuth.	Roden et al., 2015
Attenuation	Relative frequency attenuation across the evaluation point.	Indirect indicator of gas accumulation or lithologies that entrap energy.	Chenin & Bedle,2020;2022
Average frequency	Frequency content beyond the maximum energy of the envelope.	Highlights unconsolidated sands due to hydrocarbon content.	Chenin & Bedle,2020;2022
Chaos	Lack of organization in dip and azimuth of the seismic data	Highlight directly positions of reflector disruption due to structure o stratigraphy	Chopra and Marfurt. 2016
Coherent Energy-weighted gradient	Is the sum of the energy of the corresponding balanced traces.	Similarity weighted by the inline and crossline dip	Chenin & Bedle,2020;2022, Lubo-Robles et al., 2023
Cosine of phase	Real part of the instantaneous phase angle at each point of the trace	Exaggerate stratigraphical changes, alignments, or misalignments	Chenin & Bedle,2020;2022

Curvatures (K1 & K2)	The reason for the change of direction of a curved surface	Highlights the geometric behavior of event alignments.	Roden et al., 2015
Energy ratio similarity	Discriminate between seismic reflections associated with different subsurface features	Valuable for characterizing seismic anomalies, with specific frequencies.	Mass et al., 2023
Envelope	Measure of the instantaneous energy of the analytic seismic trace, independent of phase	Provides information on intensity of reflections.	Chenin & Bedle,2020;2022, Mass et al., 2023
GLCM Energy	Measure of textural uniformity in the data.	Combined with similarity attributes help delineate lithological changes.	Lubo-Robles et al., 2023
GLCM Entropy	Measure of seismic amplitude randomness along the structural dip	Fan Sand deposits and other siliciclastic architectures.	Lubo-Robles et al., 2023
GLCM Homogeneity	Measure of seismic amplitude smoothness along the structural dip	Overbank and splay deposits and other siliciclastic architectures.	Ha et al., 2021, Lubo-Robles et al., 2023
Hilbert transform	Derivative of the real part of the Hilbert transform	Highlight amplitude changes as a function of the total energy	Chenin & Bedle,2020;2022
Instantaneous frequency	Product of the Hilbert transform is the frequency of the seismic trace at each point	Attenuation and layer thickness	Chenin & Bedle,2020;2022
Instantaneous Q	Ratio of frequency to the amplitude decay	Gas detection	Chenin & Bedle,2020;2022
Interval P-velocity	Square root of two elastic modulus and the bulk density ³	Lithology and fluid indicator, useful for show porous lithology.	Mass et al., 2023
Peak CWT magnitude	The magnitude of energy above the bandwidth	Highlight frequency content associated with geological features	Lubo-Robles et al., 2023
Relative acoustic impedance	Continuous integration of the original seismic trace with the subsequent application of a low-cut filter	Highlight changes in lithology or fluid content, risk assessment of tuning.	Chenin & Bedle,2020;2022
Sobel Similarity	3D Enhanced similarity of seismic trace on a vertical and lateral window	Channels, faults, and other discontinuities	Roden et al., 2015
Sweetness	Ratio of the envelope to the square root of the instantaneous frequency	Sandstone versus shale discriminator	Chenin & Bedle,2020;2022

Thin bed	Time derivative of the phase function, difference of instantaneous frequency and time averaged frequency.	Thin bed indicator that highlights interference zones in phase.	Roden et al., 2017,
Total Energy	Total energy in the analysis window	Useful to delineate high amplitude independently of the frequency content	Lubo-Robles et al., 2023

Table 3.2. Seismic attributes found in the literature review were used for reservoir characterization and fluid discrimination (Excluding AVO attributes and inversion attributes).

Exploratory data analysis (EDA), Attribute selection, and grid size estimation for SOMs

From the seismic attribute list above, only a P-wave velocity is the measure of a physical property indirectly, and only a few of the attributes have histograms with similar data range—nonetheless, the extracted attributes from the seismic data aid in interpreting and characterizing subsurface properties. EDA should be the initial step in understanding the input space (number of attributes) and their statistical distribution.

The EDA analysis (histogram distribution, Pearson’s correlation) uses the complete data space. Figure 3.8 shows the data input's statistical distribution and presents the need for data normalization. Figure 3.8a presents the histograms of the seismic attributes, The histograms show the overlapping of data distribution, and suggest the need to normalize the dataset. Figure 3.8b is Pearson’s correlation, which shows the attribute correlation. We remove highly correlated attributes to eliminate redundancy and improve computational cost. Strong negative and strong positive correlation of attributes indicates the redundancy of attributes.

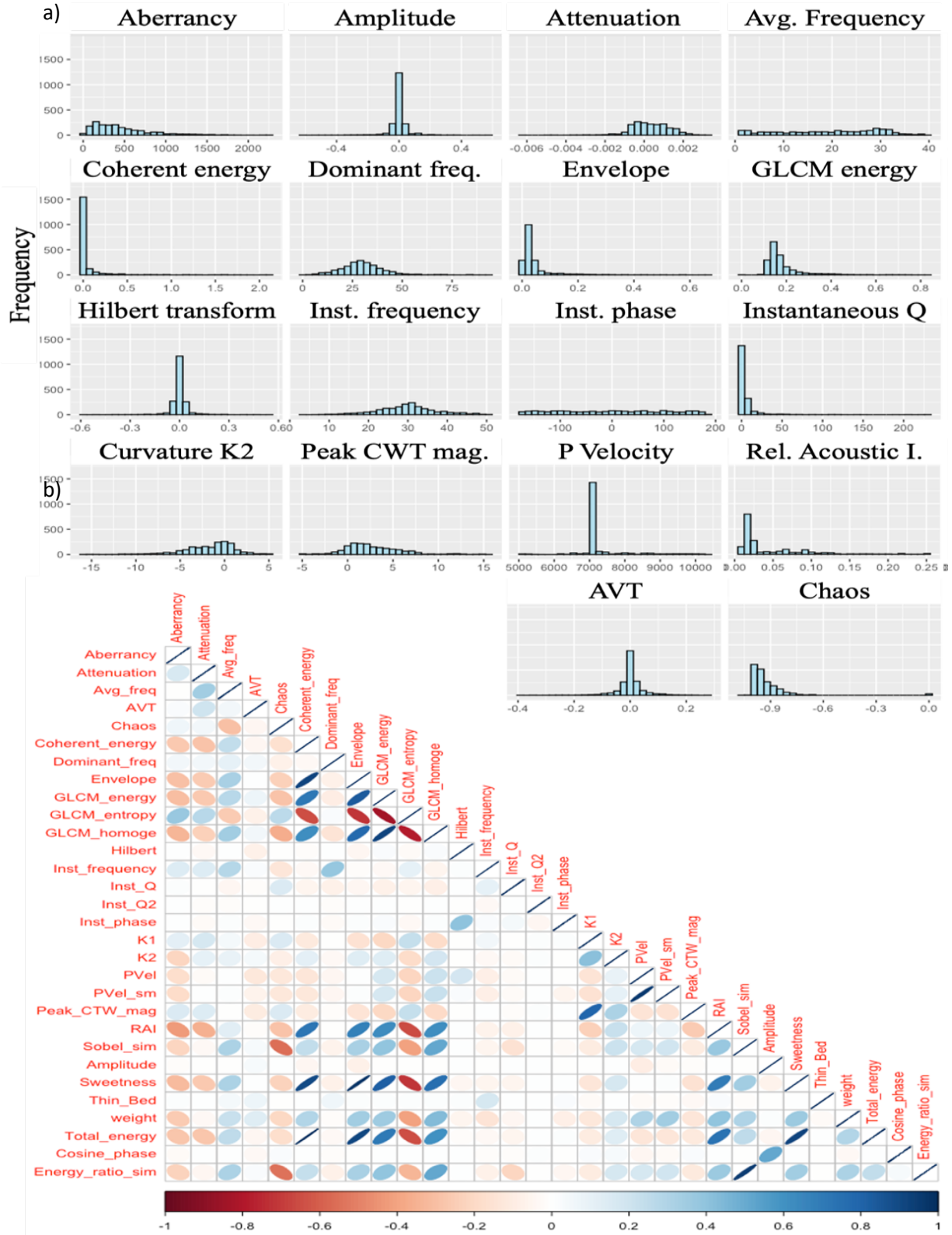


Figure 3.8. Exploratory data analysis summary, panel a) shows the histogram of the seismic attribute Table 2, panel b) shows the Manhattan correlation of seismic attributes, to show which attributes are correlated negatively (red) or positively (blue). The shape of the circle or ellipse shows the Gaussian distribution.

In Figure 3.9, we present examples of seismic attributes in the synthetic model, including the Hilbert transform, AVT (amplitude volume technique), and GLCM (Gray-Level co-occurrence matrix) energy, along with the seismic model and reservoir labels. These attributes showcase various aspects of the reservoirs in the model. Both the Hilbert transform and GLCM energy exhibit sensitivity to spectral noise, but they excel at highlighting reservoir edges more effectively than AVT.

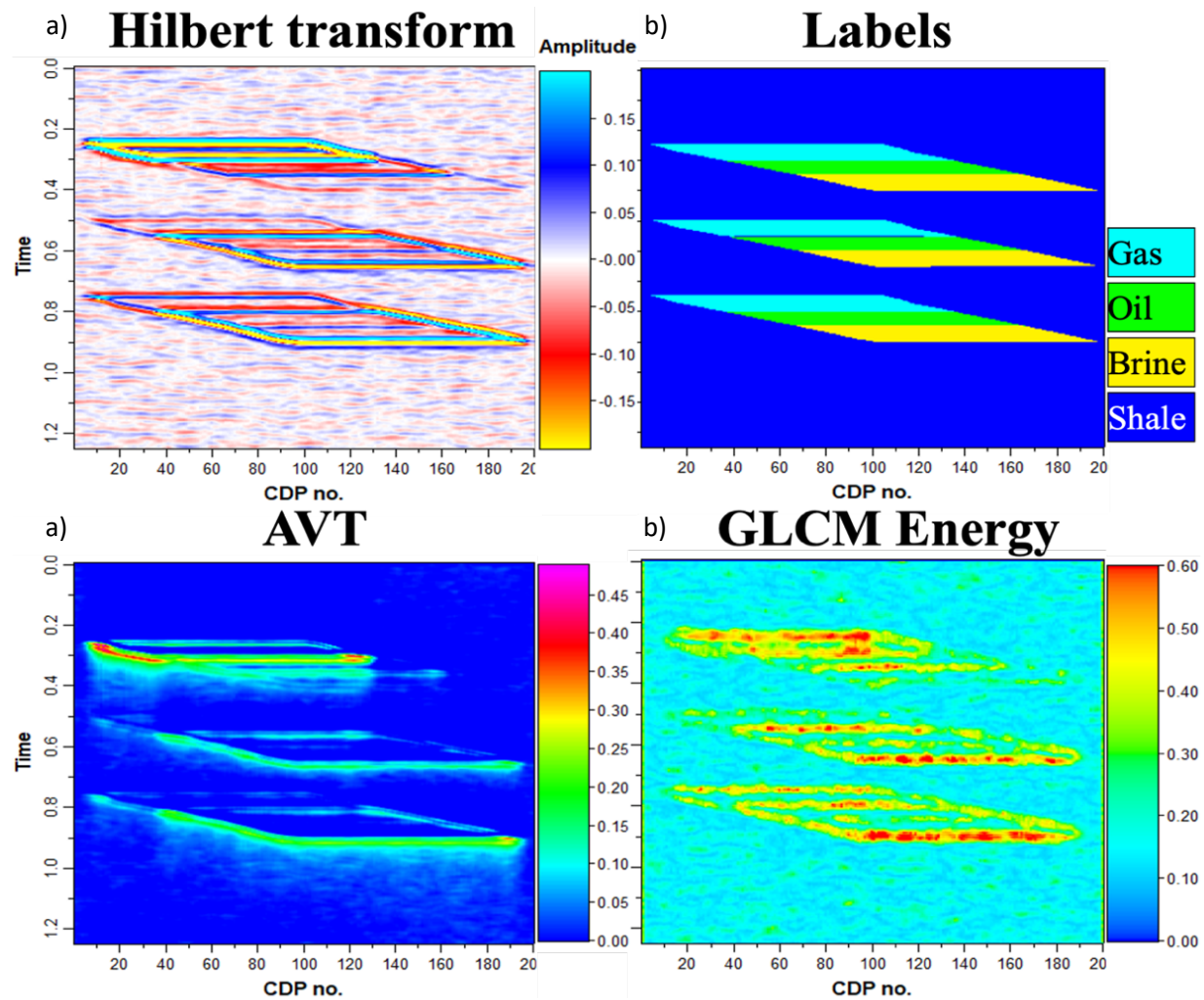


Figure 3.9. InLine of the synthetic model showing amplitude and texture attributes and the reservoir labels. a) Hilbert transform, b) Model labels, c) Amplitude volume technique (AVT), and d) GLCM energy.

We employ a methodology for selecting a meaningful set of seismic attributes, which involves conducting Principal Component Analysis (PCA) to assess the significance of each

attribute. PCA is a mathematical technique that constructs linear combinations of attributes to capture the directions of maximum variance containing independent information. In the realm of seismic attribute analysis, PCA finds widespread utility in tasks such as reservoir characterization, lithology discrimination, assessing low-gas saturation, and seismic inversion (Taner et al., 1979; Coléou et al., 2003; Chopra & Marfurt, 2007; Avseth et al., 2005; Chenin & Bedle, 2022).

By applying PCA to the entire attribute set, we effectively reduce the number of variables representing more than 95% of the independent information down to just seven key attributes. The first principal component aligns itself with the attribute exhibiting the highest variability, while the subsequent orthogonal components encapsulate the remaining covariance. Figure 3.10 illustrates the contributions to the first and second dimensions, shedding light on the significance of each attribute.

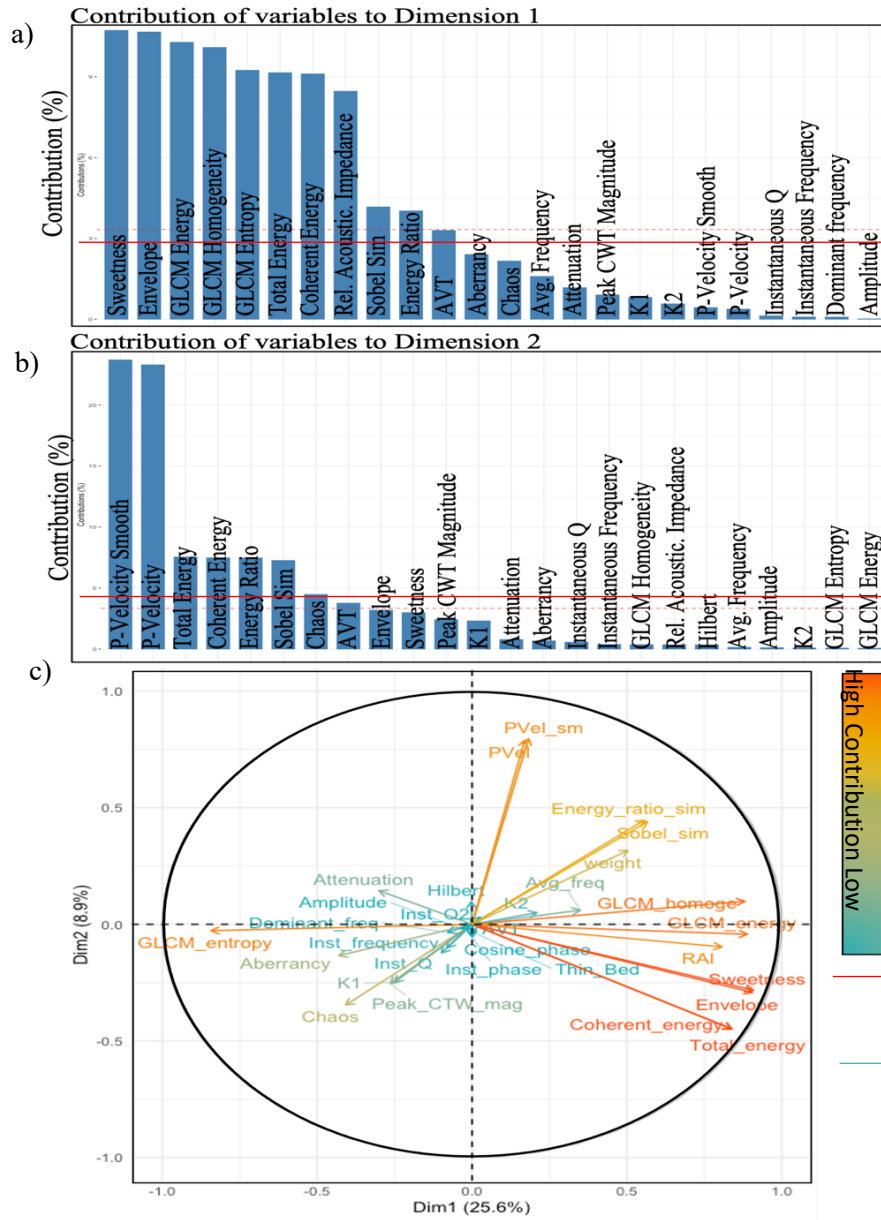
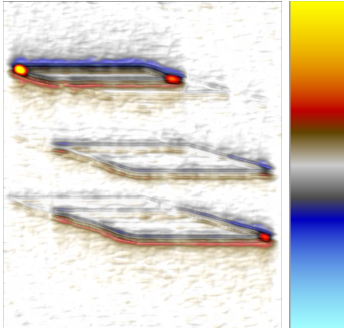
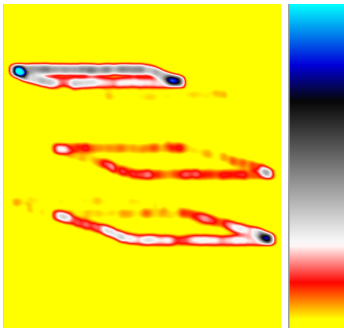
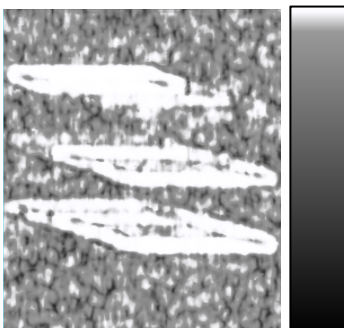


Figure 3.10. PCA of the seismic attributes and the projection of each seismic attribute to the orthogonal components (Dimension 1 and Dimension 2). Panel a) Shows the contribution to Dimension 1—panel b) % contribution of seismic attributes to the second largest component. Panel c) Contribution of each seismic attribute represented as vectors in the PCA1 and PCA2 dimensions.

From our PCA analysis, we identified seven attributes that describe the synthetic seismic and amplitude response. We carefully chose uncorrelated attributes based on their significant contributions in Figure 3.10 to guide our cluster selection and grid size determination.

For instance, attributes like envelope, sweetness, and relative acoustic impedance (RAI) are strongly correlated. However, regarding defining lithology sharply, RAI and Envelope stand out. On the other hand, Instantaneous frequency excels at distinguishing between sand and shale,

which aligns with our primary lithological interests. Therefore, we prefer using sweetness over Envelope and RAI. A similar consideration applies to the P-velocity model. While a high-frequency velocity model is available as an attribute to match the full waveform inversion (FWI) velocity model (Mass et al., 2023), we opt for a smooth P-velocity model. This choice is made because even though the synthetic model's frequency content could be realistic, it may introduce a significant bias. You can find the complete list of selected attributes in Table 3.3, sorted alphabetically for your reference and the wedge model to reference the fluid contacts.

Attribute	Contribution (%)	Attribute response to fluid
AVT	15%	
Coherent energy	20%	
Energy ratio similarity	16%	

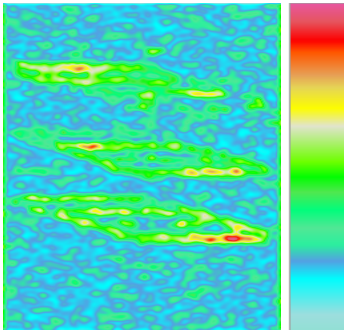
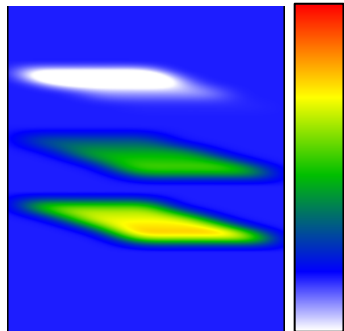
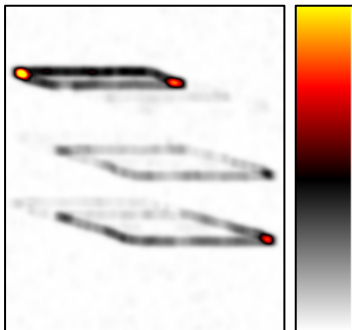
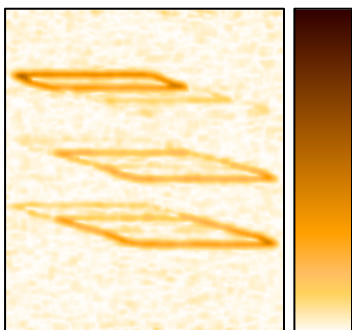
GLCM Energy	22%	
P Velocity (smoothed)	46%	
Total Energy	20%	
Sweetness	25%	

Table 3.3. Seismic attributes suggested by the PCA analysis and the display of the presence of different fluids in different reservoir rocks.

Once the list of attributes sensitive to variation of fluids and its representation in seismic amplitude is defined, seismic attributes are calculated for both datasets, the Stybarrow and Synthetic wedges. For the synthetic case, the cluster selection is unequivocal; there are three scenarios: one where SOMs clusters highlight all reservoirs and fluid types (Figure 3.11a). A second where four main clusters (Shale, Reservoir #1, Reservoir #2, and Reservoir #3) are defined (Figure 3.11b). The third scenario is where SOMs clusters highlight all reservoirs as a single cluster (Figure 3.11c).

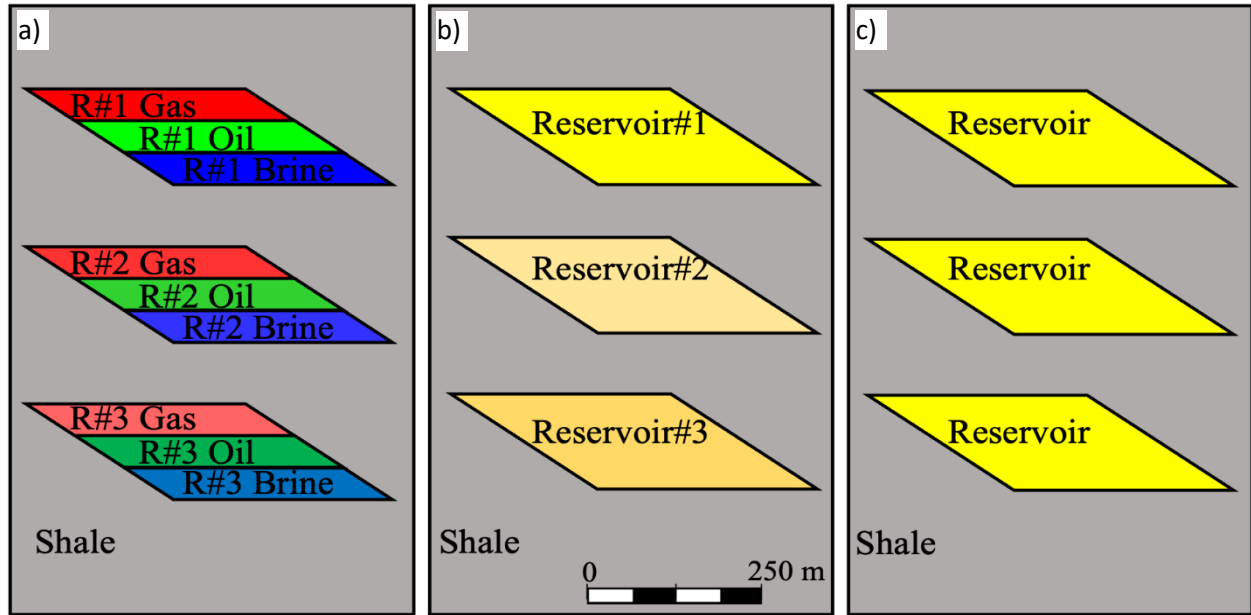


Figure 3.11. Labels for the synthetic model and possible clusters to be defined using SOM in each panel. The different colors aim to show different geobodies. Panel a) shows the ideal case, where SOMs define the ten different hydrocarbon properties included in the model. Panel b) shows a second scenario where SOMs cluster four different clusters (three different reservoirs and a cluster for shale). Panel c) shows a third scenario with only two clusters (reservoir rock and shale).

Utilizing a synthetic model enhances our grasp of clustering by enabling precise input control. For instance, through attribute calculation, Principal Component Analysis (PCA), and Exploratory Data Analysis (EDA), we identify relevant seismic attributes while also acknowledging their limitations in characterizing reservoirs II and III due to reduced amplitude

contrast. Determining the optimal cluster count is a crucial aspect of unsupervised learning. To assist interpreters when dealing with datasets with unknown clusters,

We employ heuristic tools to define the cluster count and assess cluster efficiency. In contrast to La Marca et al. (2023) and Maas et al. (2023), we decided for the statistical gap and silhouette plot over the elbow method. The silhouette plot suggests using two clusters, while the statistical gap indicates a decrease in cluster efficiency after three clusters. Figure 3.12a displays the initial clustering using the silhouette plot and statistical gap output, revealing increased misclassification beyond three clusters. Figure 3.12b presents a radial plot of each attribute's importance in cluster selection. Detailed definitions of these heuristic methods, along with plots for the elbow method, silhouette plot, and statistical gap analysis, can be found in Appendix B.

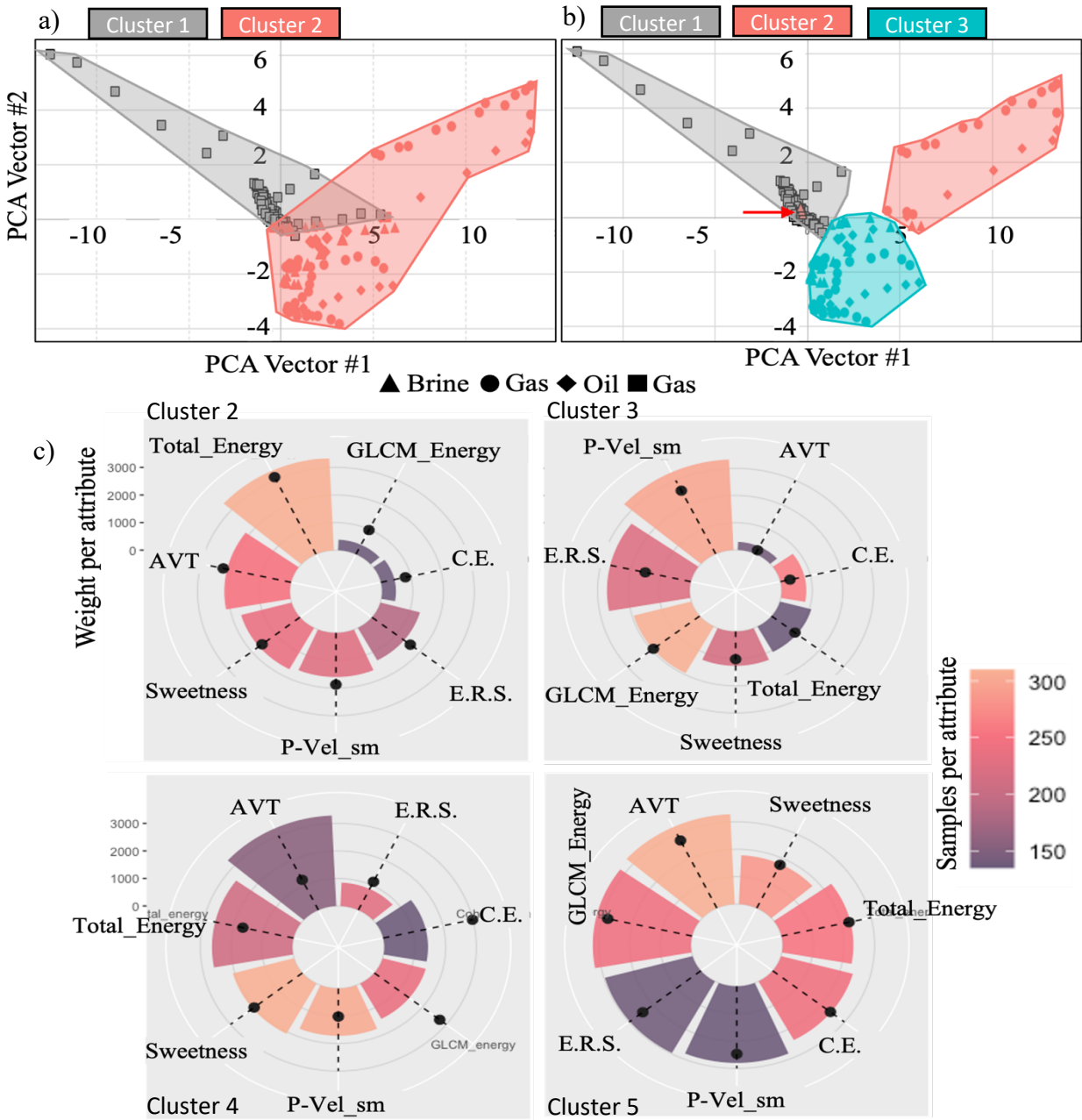


Figure 3.12. Reservoir classification and attribute importance using clusters of two, three, four, or five. Panel a) shows the reservoir separation using two clusters, and panel b) Shows three clusters with brine points overlapping in Cluster 1. In more than three clusters, the misclassification increases. Panel c) Attribute importance for the first four clusters, Cluster,2, Cluster 3, Cluster 4, and Cluster 5.

APPLYING SYNTHETIC RESULT INSIGHTS TO STYBARROW DATA

Both methods are compared using the optimum parameters available. First, we present the results for the synthetic case (Simultaneous seismic inversion and SOMs), followed by the results

of the Stybarrow dataset. This approach allows us to comprehensively gauge the performance and disparities between the two methods in detecting reservoirs and estimating hydrocarbon volumes.

Results of the Synthetic dataset

Petroelastic seismic inversion outputs p-impedance, VpVs ratio, and density to use the rock physics template RPT and correlate the hydrocarbon presence to elastic properties using borehole data. The first volumes to present are the synthetic models. Figure 3.13 presents P-Impedance, VpVs ratio and density, and the RPT colored by density to show the hydrocarbon response.

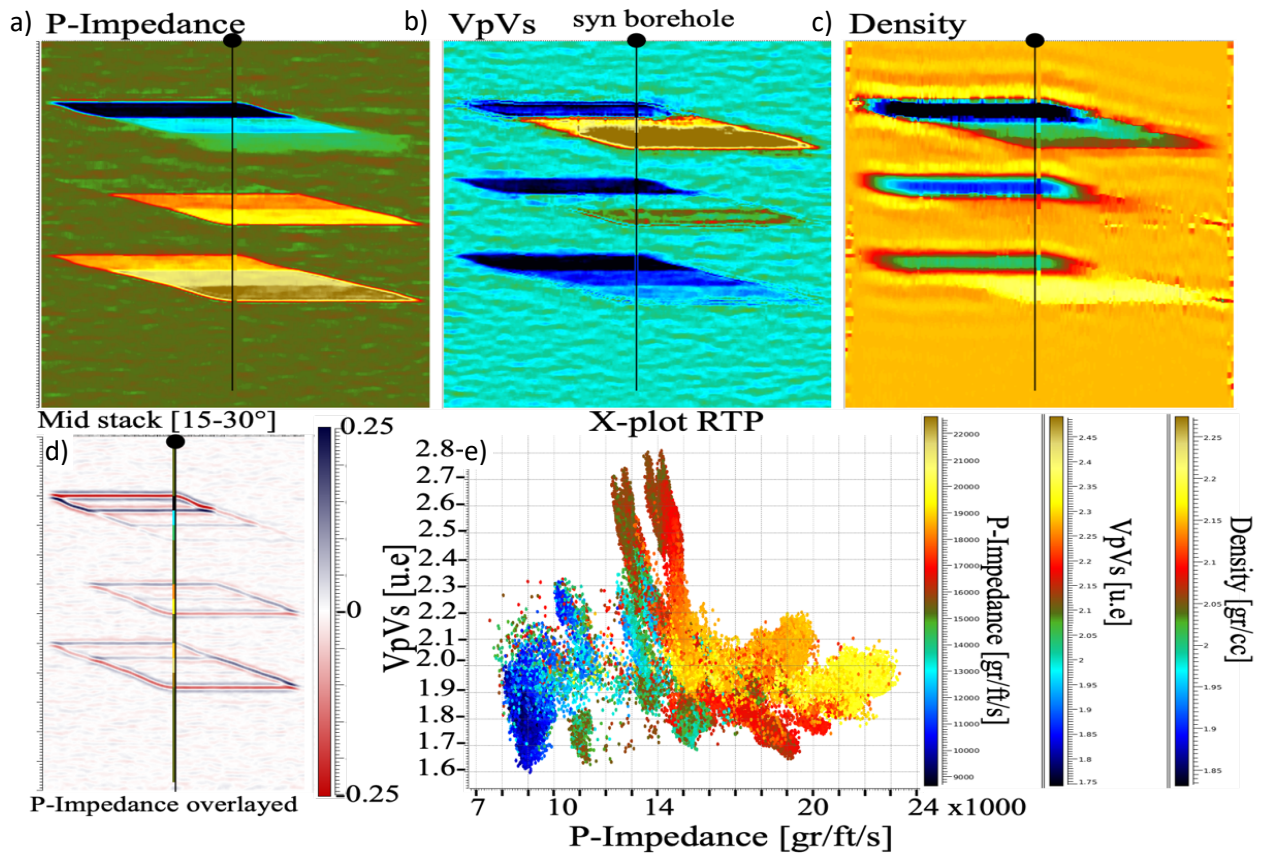


Figure 3.13. Seismic inversion results, panel a) p-impedance, panel b) VpVs, and panel c) Density. Panel d) shows the mid stack to keep for context. Panel e) RPT cross-plot of p-impedance, VpVs, and density as color code of all reservoirs.

The correlation (simple correlation coefficient) between the seismic volumes presented in Figure 3.6a (the true p-impedance model) and Figure 3.13a (the inverted impedance model) stands

at 96%. Additionally, the average correlation of reservoirs (as depicted in Figure 3.11a) with the estimated reservoir using the rock physics template (RPT) exceeds 90%. The complete list of reservoir correlations is provided in Table 3.4. These variations in correlation arise from several factors, including:

- Using a single vertical (time) gate for estimating the wavelet.
- Differences in parameterization for specific reservoir types lead to deviations in amplitude anomalies.
- Considerations related to wavelet scaling and the accuracy of the low-frequency model.

It's worth noting that estimating seismic inversion one reservoir at a time typically results in higher correlations, which aligns with the standard workflow for estimating various amplitude-dependent reservoirs. Notably, any remaining differences in correlation are attributed to noise introduced into the partial angle stacks.

The result of self-organized mapping of the synthetic reservoir shows the inaccuracy of mapping amplitude responses for different fluids is presented in Figure 3.14. Additionally, we try generative topographic maps, another unsupervised classification and the results of this classification and methods are presented in Appendix B of this thesis.

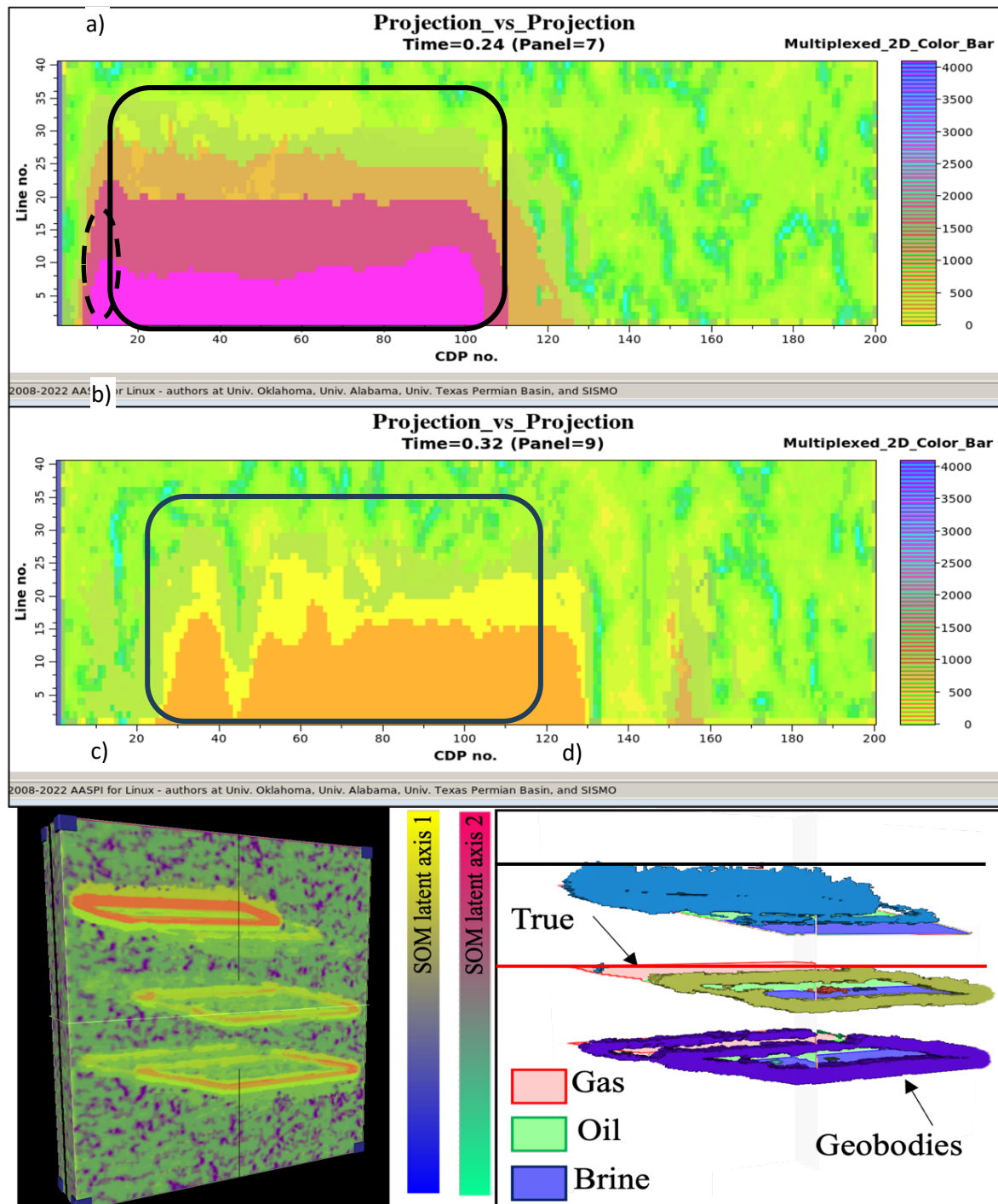


Figure 3.14. Self-organizing mapping of three reservoirs. The top section, panel a) is the map projection of the two SOM latent axes at the top of the first gas reservoir, its position corresponds to the black line in panel d). Black box delineates the area of the 80% gas saturation line, and black ellipse show the wedge border and its amplification due to amplitude effects. In both areas there is a mismatch with the true extent of the reservoir. b) Map projection of the two SOM latent axes at the position of the top second reservoir (red line in panel d). Black box shows the extent of the 80 % gas saturation, and the mismatch is exaggerated. Panel c) 3D view of the SOMs and Panel d) 3D view of the geobodies captured by the latent space and the accurate labels to evaluate the clustering accuracy.

Stybarrow dataset results

Naturally, the seismic inversion of the observed datasets requires extra calibration and careful petrophysics and rock physics analysis. The detailed steps of the seismic inversion are out of the scope of this work. Figure 3.15 shows the fullstack, p-impedance, VpVs ratio, and inverted bulk density of the Stybarrow field. The RPT of the borehole response and seismic volumes is also presented to show the correlation of the reservoir with the elastic properties.

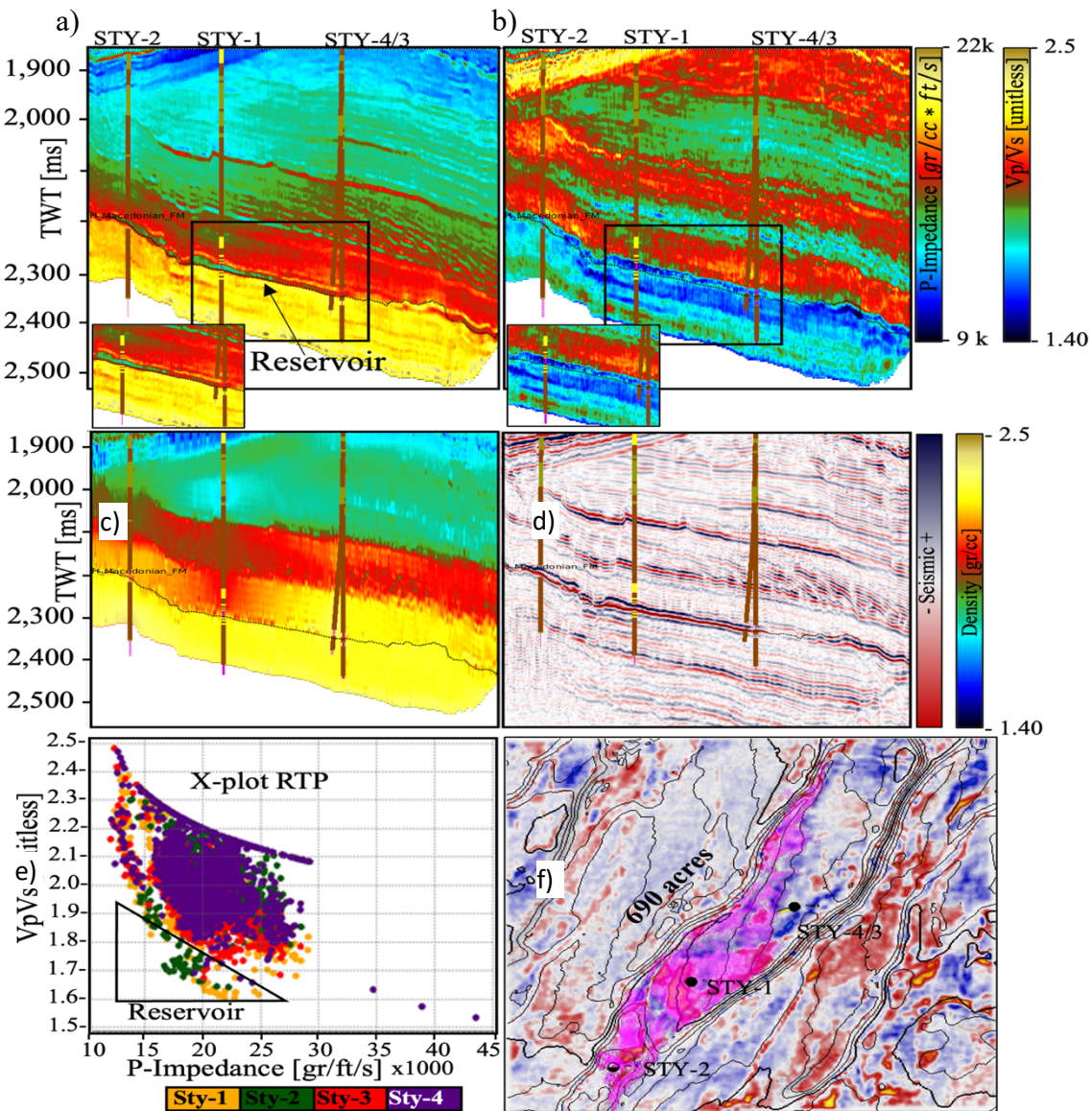


Figure 3.15. Displays seismic inversion results for the Stybarrow field. It includes panels for a) inverted acoustic impedance, b) VpVs, c) total stack, d) inverted bulk density, e) RPT template, and f) the captured reservoir geobody overlapping seismic amplitude extraction on the Macedon Fm.

The SOMs of the Stybarrow data show primarily a hydrocarbon response entangled with the lithological response (K-means and generative topographical mapping are also available but on a different project). Figure 3.16 shows the 3D projection of the latent spaces, the distance vector, and the volume of these geobodies. The window is restricted between the top of the Pyrenees Fm, and the base of the Macedon Fm, shifted 102 ms. Focusing on the SOM analysis is a sound practice to avoid misclassifications. A second test of the SOM shows how useful it is for classifying a stratigraphical response.

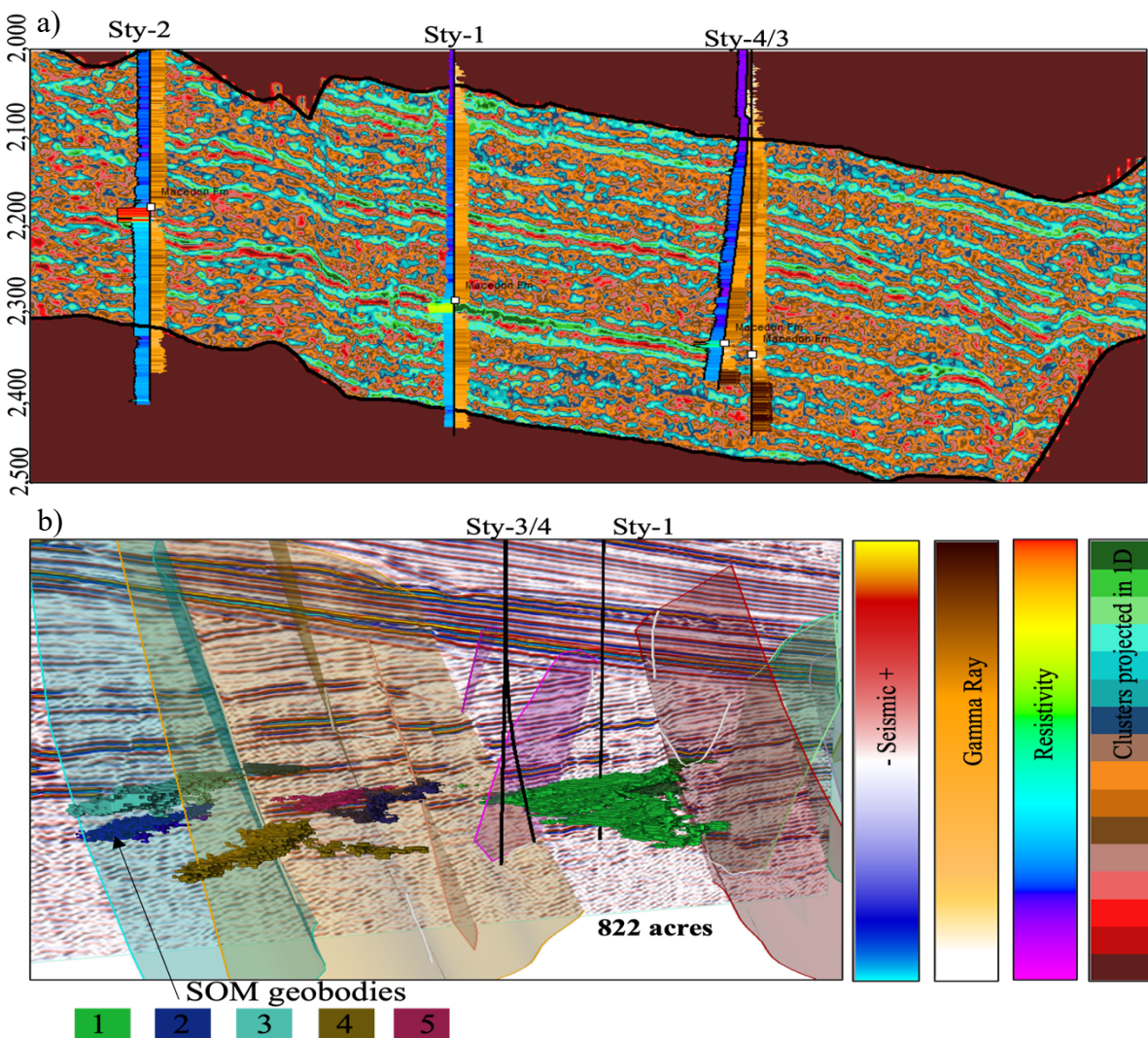


Figure 3.16. Extracted geobodies from SOM using 12 clusters. The geobody anomaly is 21 % larger than the seismic inversion. Additionally, the SOMs clustered four different elements that haven't been produced/drilled or tested.

Comparison of the accuracy of both methods on the datasets

Three different self-organizing mappings are implemented to study the accuracy of SOMs (available in supplemental information). The self-organized mapping of the synthetic reservoir shows the inaccuracy of mapping amplitude responses for different fluids. Moreover, the self-organizing mapping of the wedges is less continuous than the seismic inversion and, therefore, less accurate. The effect of seismic noise in the self-organizing mapping and seismic attributes is more evident than in the seismic inversion, suggesting that SOMs are more susceptible to noise. Table 3.4 summarizes the correlations of the two methods with the actual reservoir labels and where the most significant misfit arises in characterizing the synthetic dataset.

Method	Reservoir 1	Reservoir 2	Reservoir 3	Average	Largest Misfit
SOMs	78 %	65 %	61 %	68 %	R1_B_1 and R2_G1
SSI	96 %	89 %	88 %	91 %	R3_O_2

Table 3.4. Reservoir characterization correlations of each method, the petroelastic seismic inversion from the SSI, correlate most with the actual reservoir models.

Spatially, the correlation of the reservoir characterized by the seismic inversion shows an excellent correlation for the first reservoir and is suitable for the second reservoir, only acceptable for the third reservoir (Figure 3.17a). In contrast, the spatial correlation of SOMs is acceptable for the first reservoir, inaccurate for the second reservoir, and acceptable for the third attribute.

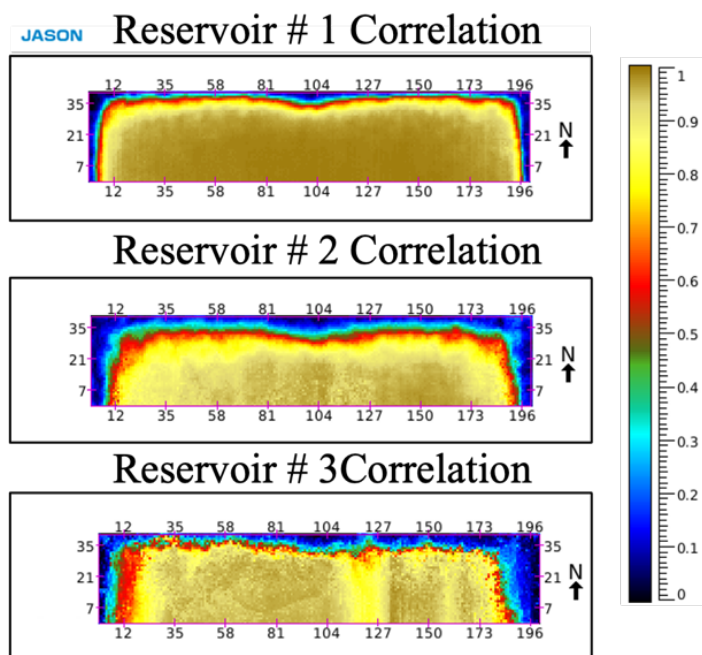


Figure 3.17. Cross-correlation of the three-reservoir calculated from the petro-elastic seismic inversion and the SOMs showing the spatial variability of the SOMs characterization.

SOMs capture more reservoir rock inside and outside the reservoir block (bounded by faults). The possibility of SOMs entangling the reservoir response with rock response increases, which accounts for the area difference. Discarding the geobodies outside the block (geobodies 2 to 5), geobody # 1 from SOMs has an area of 2,400 acres versus the 1,980 acres calculated using seismic inversion.

Reservoir thicknesses are also different, although the differences include the wavelet effect; SOM geobodies are expected to be thicker than seismic inversion geobodies. The main reservoir from borehole information shows three flow nits that averaged a total thickness of 33 m (108 ft), with a seismic resolution ($\lambda/4$) of 28 m. We used the velocity model to transform each point of the SOMs geobody into reservoir thickness, which has an average of 44 m (~147 ft), and the average reservoir thickness from the inversion is 36 m (~121 ft).

The % difference in area and thickness is close to 21 %; the initial volume estimation using SOMs will be larger than seismic inversion. STOIP for the SOMs volumetric and Inversion

volumetric are calculated using the following $\phi=18\%$, $S_o=75\%$, $B_{oi}=1.25$ and rounded to the next integer:

- 1) $STOIIP_{SOMs}=100$ MMBOE (MMBOE = millions of barrels of equivalent oil)
- 2) $STOIIP_{Inv} = 69$ MMBOE

The area difference is 15.8%, and the thickness difference is 17.7%, the initial volume of oil in place difference is 31 %. Comparing the difference in the synthetic case where the accuracy difference is above 20 %, there is a mix of error sources using SOMs.

DISCUSSION

This study evaluates the accuracy of self-organizing maps for siliciclastic reservoir characterization. We rigorously test the accuracy of SOMs using three different amplitudes versus angle (AVA) anomalies, across over 28,000 traces in 3D, escalating on the Roy et al. (2010) model. This allows for an in-depth analysis of the strengths and weaknesses of unsupervised learning. Results reveal that SOMs have a maximum accuracy of 78% and a minimum of 61%. On average, over 20% lower accuracy than seismic inversion methods in distinguishing between reservoir rock properties and pore fluid effects when relying solely on seismic amplitude and seismic attributes for clustering. This highlights the inherent limitations in the SOM approach for characterization based on amplitude attributes alone.

The synthetic seismic modeling methodology allows us to assess the machine learning algorithm's performance precisely. Building upon recent work by La Marca et al. (2023), which used synthetic modeling for unsupervised lithological classification of deepwater channels, our study goes a step further. We investigate the accuracy of Self-Organizing Maps (SOM) in estimating reservoir properties using seismic attributes sensitive to amplitude changes. Our model considers rock physics relationships encompassing mineralogy, porosity, temperature, pressure,

fluid saturations, and fluid types, resulting in distinct Amplitude Versus Angle (AVA) responses for various wedge models. This approach enables insightful testing and analysis of SOMs performance and accuracy. Moreover, comparing SOMs and petroelastic seismic inversion results on this meticulously modeled dataset provides a quantified measure of a significant performance gap. Insights gained from synthetic testing are invaluable for guiding applications to real seismic data and enhancing validation.

High-quality seismic and proper preprocessing are critical prerequisites for both SOMs and seismic inversion approaches. However, SOMs demonstrate substantially greater sensitivity to data issues like missing traces or high noise levels as compared to inversion techniques. This aligns with findings from Vantanen et al. (2015) on the impacts of missing data on SOMs performance. Nonetheless, broad bandwidth and high signal-to-noise ratios benefit either approach, provided data quality is adequate for calculating the seismic attributes used in the clustering analysis. The selection of seismic attributes from the numerous possibilities helps provide the most relevant inputs for clustering. From our literature review, we reduced the list of seismic attributes from 24 to 7 using exploratory data analysis and principal component analysis. The selection of P-velocity and GLCM energy attributes agrees with Maas et al. (2023). Figure 12b shows the attribute's importance in each cluster. Cluster three has P-velocity as the attribute that contributes the most to its classification, and at the same time, using three clusters shows the greatest distance between reservoir rock and shale.

For the Stybarrow field example, SOMs successfully clustered reservoir rock properties but failed to distinguish the all-important pore fluid effects reliably. This led to an inaccurate larger reserve estimation compared to the seismic inversion informed by borehole data. In contrast, Maas et al. (2023) and Hussein (2021) demonstrated SOM's capabilities for reservoir discrimination

using seismic attributes. However, our analysis indicates porosity contrasts strongly influenced their delineation. Well-established workflows have demonstrated superior results from simultaneous seismic inversion for concurrently characterizing lithology and fluids, given the availability of borehole log data (Wollner et al., 2017; Barajas-Olalde et al., 2019). The transition to leveraging inverted elastic properties for petrophysical analysis has expanded capabilities beyond direct hydrocarbon indicators to include reservoir architecture insights.

The simultaneous seismic inversion of the wedge models included a single borehole and its extracted wavelet, are contingent on precise low-frequency modeling (beyond the scope of this work) and accurate wavelet estimation. To enhance wavelet accuracy, calibration with a borehole is recommended. For the Stybarrow, our petroelastic inversion results for the Stybarrow dataset align with previous research (Arevalo-Lopez et al., 2016; Duncan et al., 2013) in terms of anomaly size and reservoir heterogeneity (Hill et al., 2008; Duncan et al., 2013). The inverted p-impedance and V_p/V_s ratio correlates well with Arevalo-Lopez et al. (2016), bolstering confidence in the inversion and its outputs. Reservoir area and thickness estimations rely on cross-plot values from borehole data.

Self-organizing maps emerge as a viable alternative to seismic inversion for identifying prospective areas in geophysical exploration contexts with sparse borehole data. However, for detailed reservoir characterization, particularly when elastic wireline logs are available, seismic inversion or its derived attributes outperform SOMs in terms of accuracy, leveraging rock physics trends (Avseth & Lehocki, 2016).

Our study offers a concise overview of SOMs' applicability for reservoir characterization, delineating their strengths, weaknesses, and limitations related to resolution and the ability to distinguish between lithologic and fluid effects. We compare SOMs with simultaneous seismic

inversion to quantify these differences and present alternative cluster selection options for more informed decision-making.

CONCLUSION

This study explores unsupervised machine learning, specifically self-organizing maps, for reservoir characterization, and we note that proper preprocessing and data quality are paramount for SOM success. Thoughtful attribute selection using exploratory data analysis and principal component analysis provides optimal inputs. However, conventional practices, like the elbow plot for cluster estimation, prove ill-suited for seismic data. Alternative techniques offer improved, more geologically reasonable cluster estimates.

In this case, SOMs depend primarily on geometric attributes, constraining the detection of amplitude anomalies linked to fluids. This contrasts with other studies successfully using SOMs for geometric lithologic classification. Preprocessing and multiple attribute calculations also make SOMs more time-intensive. Random decimation accelerates clustering but can degrade results. Missing data and noise disproportionately impact SOMs compared to inversion. While SOMs show promise, results reveal key limitations in distinguishing reservoir rock from pore fluids. Analysis of synthetic seismic datasets exposes inherent inaccuracies in SOMs, with over 20% lower accuracy versus seismic inversion. This is attributed to reliance solely on geometric attributes and seismic amplitude for clustering.

The analysis of three distinct wedge datasets using SOMs revealed inherent inaccuracies, particularly when a larger number of clusters is used. Additionally, synthetic testing demonstrated that geobodies tend to cluster in the presence of a gas, oil, and brine mixture, with a significant accuracy difference observed when compared to simultaneous inversion methods. Specifically, a discrepancy exceeding 20% was evident in both synthetic and observed datasets. In contrast,

simultaneous seismic inversion workflows offer well-defined inputs, and a feasibility study becomes a crucial element in delineating the study's limitations—a step that is notably absent in SOMs.

To conclude, while unsupervised machine learning, specifically SOMs, holds promise for geological feature characterization, our research underscores the need for a careful approach. Addressing pre-processing, attribute selection, and data quality challenges is critical for successfully applying these methods.

ACKNOWLEDGMENTS

We want to thank the Australian Governmental Department of Geoscience for providing the survey HCA2000A 3D, borehole data including, wireline logs and associated reports required for this study. AASPI, CGG, Geophysical Insights, MathWorks, and SLB are greatly acknowledged for providing the following software to the University of Oklahoma and the School of Geoscience: AASPI version 2023, Jason 10, Paradise 3.2.3, Matlab 2023, Techlog 2019 and, Petrel G&G software v.2022, respectively. Thanks to S. Chopra, H. Arevalo Lopez, and J. Dvorkin for the valuable discussion.

CHAPTER 4: SEAL AND RESERVOIR RISK EVALUATION USING HIERARCHICAL CLUSTERING ANALYSIS WITH SEISMIC ATTRIBUTES IN NORTHWESTERN AUSTRALIA. *

*This chapter submitted to the *Journal of Applied Geophysics*, in November 2023.

ABSTRACT

Assessing the presence and quality of reservoir rocks and their sealing capacity is crucial for various applications, including hydrocarbon, geothermal, and CO₂ sequestration projects. Typically, exploration geoscientists rely on seismic attributes and borehole logs into interpretation to integrate diverse data for estimating reservoirs and seals. However, for all seismic interpreters, the process is time-consuming. In this study, we explore the application of hierarchical clustering analysis (HCA), an unsupervised machine learning technique, to streamline the integration of multidisciplinary information. While HCA and similar techniques may occasionally misclassify critical data, we demonstrate how to enhance their accuracy by carefully selecting the number of clusters and their calibration with borehole data.

The novelty of our work is the innovative transformation of HCA clusters into a 3D lithology model, which can significantly facilitate the estimation of reservoir rock and seal-rock juxtaposition risk. Using the HCA clustering hierarchy, five clusters effectively discern the presence and quality of seal and reservoir rock in two different datasets. The classification, in combination with the fault probability, addresses the seal risk in Jurassic sandstones oil and gas structural traps offshore the Northern Carnarvon Basin.

INTRODUCTION

In siliciclastic basins, accurately determining reservoir rock's presence, distribution, and quality is a critical challenge. This is often done through seismic facies identification, a technique interpreters use to define seals and reservoir rocks, which involves categorizing rock sections with similar properties and depositional geometries. An additional aspect of these analyses is understanding the juxtaposition of shale and sandstone units, particularly in highly faulted areas. The risk of leakage and fluid compartmentalization increases when sandstone reservoirs, though present, are interconnected across faults. To address this challenge, we use a calibrated hierarchical cluster analysis to evaluate reservoir rock characteristics in the Carnarvon Basin off the coast of northwestern Australia. The vertical zone of interest are the strata from the Late Cretaceous to Cretaceous to Early Jurassic. We aim to assess the reservoir rock's presence and distribution by leveraging a combination of seismic attributes, borehole data, and HCA. Two distinct datasets are utilized—one in the eastern part of the basin with a proven reservoir and another undergoing frontier exploration in the western part. This methodology allows us to enhance our understanding of subsurface conditions and mitigate risks associated with sandstone reservoirs and faulted areas while comparing the detectability of reservoir rock using unsupervised learning.

The use of seismic attributes to explore the geology subsurface has been well documented (Taner & Sheriff, 1979; Chopra and Marfurt, 2007; Barnes, 2016). Seismic attributes are quantitative measurements or characteristics derived from seismic data's amplitude, frequency, and geometric arrangement. The horizontal resolution of seismic data is complemented by the vertical resolution of the borehole, enabling the integration and calibration of seismic attributes to highlight crucial geological features such as lithological changes, indicators of porosity, and fluid presence (Roden et al., 2015; 2017b).

Seismic attributes can also aid in fault detection, identifying structural features, and enhancing seismic interpretation (Marfurt, 2018). For example, seismic data conditioning aims to enhance the signal/noise ratio, removing coherent and incoherent noise. Footprint removal filters and structurally oriented filters have helped improve the quality and interpretability of seismic attributes. Spectral balance stabilizes the frequency distribution and improves the and unsupervised learning classification (Qi et al., 2016; Lubo-Robles et al., 2023; Chopra et al., 2023) and it also is helpful for improving fault definition (Qi et al., 2017; Qi et al., 2019),. Additionally, using unsupervised clustering accelerates subsurface characterization and data interpretation and helps integrate extensive seismic data. In one case, Zhao et al. (2015) used self-organizing maps (SOM) for unsupervised seismic facies analysis and demonstrated mapping facies associations using multiple seismic attributes as input vectors. In another example, Wallet and Hardisty (2019) compared K-means, Gaussian mixture models, and SOMs for unsupervised facies classification using seismic and bypassing the borehole wireline data entirely.

Many researchers are actively use seismic attributes and machine learning to discriminate fluids and their effects on reservoir rocks (Roden et al., 2017a; Tahmasebi et al., 2017; Verma et al., 2020). Lubo-Robles et al. (2023) characterized gas hydrates using total energy, peak frequency, peak magnitude, Gray-Level co-occurrence matrix (GLCM) attributes, dip deviation, coherence, and self-organizing maps. Maas et al. (2023) characterized light oil in porous and permeable subsalt carbonates in Brazil, expanding the list of seismic attributes and machine learning algorithms to test. Ballinas et al. (2023) claimed to have differentiated low-gas saturation (LGS) and commercial gas saturation by using spectral shape attributes, independent frequency magnitudes, and statistical attributes. Their work applied supervised learning algorithms - decision trees (DT), random forest (RF), and k-nearest neighbors (KNN) - to separate the difference between low gas

saturation, commercial gas saturation, and free oil in sandstones. Li et al. (2023) used coherent energy, curvedness, peak spectral magnitude, and frequency as geologic constraints and a transformation of the seismic data to train a neural network to predict seismic facies.

With the presence of large amounts of data, from seismic attributes to borehole information, methods are needed to allow an efficient analysis of these multidimensional datasets. Hierarchical cluster analysis is an unsupervised machine learning algorithm for grouping similar data points into clusters based on their similarity (Jain et al., 1999; Everitt and Dunn, 2001), and has been previously employed in seismic analysis. Emerson et al. (2020) evaluated hierarchical clustering techniques for seismic facies analysis in the context of seismic facies definition. However, the lack of borehole data prevented the contextualization of the calculated clusters. HCA can be employed to identify patterns and relationships within seismic attributes and borehole data, ultimately aiding in identifying reservoir rocks. HCA benefits from a dimensionality reduction like ICA or PCA (Lubo-Robles et al., 2019). In contrast with DT, RF, and KNN, HCA requires only one hyperparameter, the number of clusters. La-Marca et al. (2023) used the elbow method; however, agglomerated or poorly sorted clusters hinder the elbow method calculation (Schubert et al., 2023). Our selection of gap statistics is a novel approach in seismic facies automatic interpretation to define the optimum number of clusters heuristically.

There are multiple factors that need to be considered when performing reservoir risk analyses. Lateral seals formed by shale units, play a critical role in preventing the migration of hydrocarbons out of reservoirs (Yielding et al., 1997). Shale distribution and thickness are key factors in evaluating the sealing capacity of potential hydrocarbon traps during lateral seal analysis (Downie et al., 2019). A higher shale gauge ratio, indicating increased contrast across a shale-sand

interface, signifies greater sealing capacity, preventing fluid migration like oil or gas. This ratio is integral to lateral seal analysis in seismic exploration, evaluating the proportion of shale within a fault adjacent to a sandstone, thereby preventing fluid migration across the fault. The interpretation of faults in seismic data has evolved from manual fault picking to deep-learning fault prediction (Wu et al., 2019). Our selection of the fault probability Wu et al. (2019) shows the fault existence, dip magnitude, and dip azimuth of faults between the Aptian-Jurassic intervals. The skeletonization of the fault probability (Qi et al., 2017) assists in estimating lateral seal analysis and defining fault planes. The fault probability herein is calculated using principal component structure-oriented filtering to reject noise and sharpen edges in seismic amplitudes, and the fault probability.

To address these issues of understanding reservoir risk, we propose a novel calibrated HCA to address reservoir rock and shale definition. The HCA is calibrated to petrophysical borehole data. Additionally, we present a lateral seal analysis using fault probability. We reduce the attribute selection by using exploratory data analysis and principal component analysis. The number of clusters is selected using borehole data and the gap statistics method. The datasets available include an oil and gas producer field, the Stybarrow, and a frontier exploration area, the Canning 3D (Figure 4.1). The well-defined reservoir rocks in the Stybarrow seismic cube are Lower Cretaceous fluvial-deltaic Barrow Group sandstones. The reservoir rocks in the Canning 3D are Lower Cretaceous basin floors and turbidites in the Barrow Group, Upper Jurassic shoreface Jansz sandstones, and Upper Triassic shallow marine-deltaic Brigadier Formation. In both seismic studies, the traps exhibit a structural imprint.

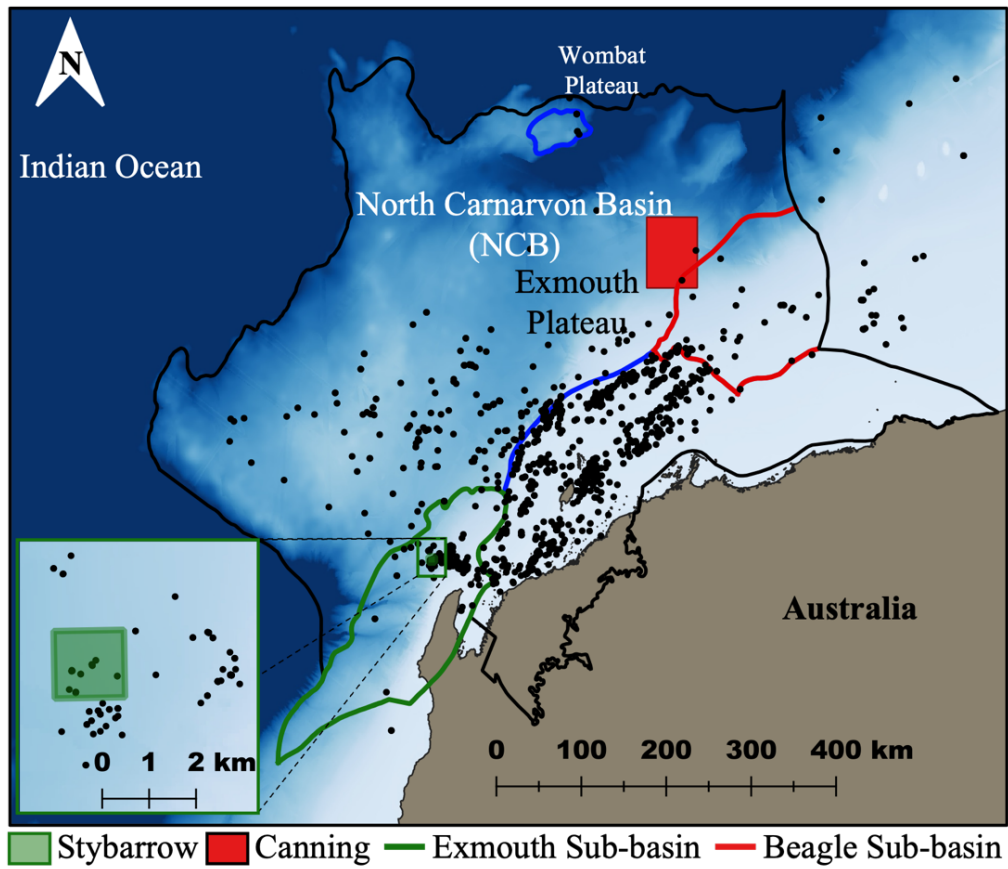


Figure 4.1. Location of the two seismic surveys available to evaluate hierarchical clustering analysis in the classification of reservoir sandstones. The contrasting drilling activity results from the oil and gas fields located in the Exmouth sub-basin. In contrast, the Canning 3D dataset, and the eastern portion of the North Carnarvon Basin (NCB) is a frontier exploration area with two boreholes drilled within a 5,160 km² survey.

GEOLOGY OF THE NORTH CARNARVON BASIN

The North Carnarvon Basin is part of the greater West Australian continental margin and comprises several sub-basins, and is an ideal region with a plethora of data that can be used to demonstrate the usefulness of HCA for reservoir risk analysis. The Northern Carnarvon Basin covers an area of approximately 535,000 km² in water depths up to 4,500 m. It is the most prolific basin in Australia, divided into Barrow, Beagle, Dampier, Exmouth sub-basins, and the Exmouth Plateau. It contains Paleozoic, Mesozoic, and Cenozoic sediments dominated by deltaic to marine siliciclastics and shelf carbonates of the Mesozoic to Cenozoic age. Hydrocarbon reservoirs are

located between the Albian-Aptian unconformity and the Late Jurassic (Bishop, 1999; Bradshaw et al., 1998). Hydrocarbon generation, migration, and trap formation in the basin have been strongly controlled by syn-rift structuring, deposition, and post-rift reactivation. The generalized stratigraphy of the basin can be related to six tectonic phases, which, in turn, are related to the rifting phases that culminated in the continental breakup of the west Australian margin in the Middle-Late Jurassic and Early Cretaceous: 1. Pre-rift 2. Early syn-rift 3. Main syn-rift 4. Late syn-rift Barrow Delta, 5. Post-breakup subsidence, 6. Passive margin.

The Stybarrow 3D is within the Exmouth sub-basin. The Exmouth sub-basin represents a series of substantial rift depocenters where the sediment sequences are predominantly Triassic, Jurassic, and Lower Cretaceous. The Lower Cretaceous succession in the Exmouth and Barrow sub-basins is notably influenced by the Barrow Delta (Tindale et al., 1998). The reservoir rock is the Macedon sandstone deposited during the Early Cretaceous (Ementon et al., 2004; Duncan et al., 2013), with a porosity range from 12-32 % and a net-to-gross (NTG) range of 30-80%.

The Canning 3D seismic survey is predominantly on the Exmouth Plateau, a subsided continental platform characterized by a faulted, predominantly Triassic sedimentary succession. Jurassic sediments are generally either thin or absent (Tindale et al., 1998; Stagg et al., 2004; Lech, 2013). The dominant structural trend varies between the north and northeast, reflecting the interplay between the oblique extensional vectors and the pre-existing structural grain of the basement (Stagg et al., 2004). The reservoir rocks are Upper Jurassic shallow marine-deltaic sandstones with an 8-18% porosity range and a high NTG range of 60-80%. A small portion of the Canning 3D is within the Beagle Sub-basin, where the reservoir architectures are like the Exmouth Plateau, fluvial-deltaic to shallow marine sandstones of the Upper Triassic.

The targeted reservoir rocks within the Stybarrow seismic cubes consist of fluvial-deltaic Macedon Formation Sandstone (Lower Cretaceous) with an average porosity of 18%. These rocks form anticlinal structural/stratigraphic traps in the Lower Cretaceous. In contrast, the reservoir rocks in the Canning 3D are less constrained and encompass Lower Cretaceous basin-floor fans in the Barrow Group, shoreface facies in the Upper Jurassic upper-lower Jansz Sandstones, and Upper Triassic shallow marine-deltaic Brigadier. The average porosity for these formations is 12%. Potential traps within the Canning 3D include tilted Triassic fault blocks and associated drapes, forming both structural and stratigraphic traps.

DATASET DESCRIPTION

Two seismic surveys, Stybarrow 3D (55 km²) and Canning 3D (3,527 km²), are used to evaluate the use of hierarchical clustering analysis into rock-type classification and lateral seal analysis. The Stybarrow survey was processed in pre-stack migration in 2003, with a dominant frequency of 35 Hz, 3 ms sampling rate, and 2.7 seconds in two-way time (TWT) total recording. Four exploratory boreholes were drilled between 2004 and 2006, providing contemporary wireline geophysical logs, cores, sidewall plugs, and fluid samples. The wildcat -first exploratory borehole was the Stybarrow-1, finding oil-bearing sandstone at 2230-2290 m in true vertical depth (TVD). The second borehole explored the reservoir extension up-dip (Stybarrow-2), finding gas. The Stybarrow-3 aimed to find the oil-water contact; nonetheless, it drilled 3 meters of the production sandstones, Stybarrow-4 borehole, a sidetrack from Stybarrow-3 drilled up-dip to find the down-dip extension of the reservoir. The Delambre-1 borehole, drilled in 1984, offers wireline information available at the time, while the Levvit-1, drilled in 2015, provides contemporary wireline logs, cores, sidewall plugs, and fluid samples.

Additionally, the Canning 3D survey, characterized by a dominant frequency of 30 Hz, 2 ms sampling rate, and 5.2 TWT seconds of total recording, encompasses two boreholes drilled with a 35-year gap. The Stybarrow 3D and Canning 3D seismic data are presented in Figure 4.2 to present the reader with the raw seismic data, the distribution of the borehole data, and the frequency and amplitude differences between the seismic cubes and vertically. Figure 4.2a shows an arbitrary line connecting the location of the borehole data. Similarly, Figure 4.2b shows the amplitude data corresponding to the Canning 3D connecting the borehole data. We chose Stybarrow 3D, which is 65 times smaller than Canning 3D, to test the seismic attribute manually, interpret the faults in the system, and explore the sandstone juxtaposition in detail. We use machine learning, mainly unsupervised learning, to test and calibrate the findings in Stybarrow 3D and extrapolate the findings into Canning 3D.

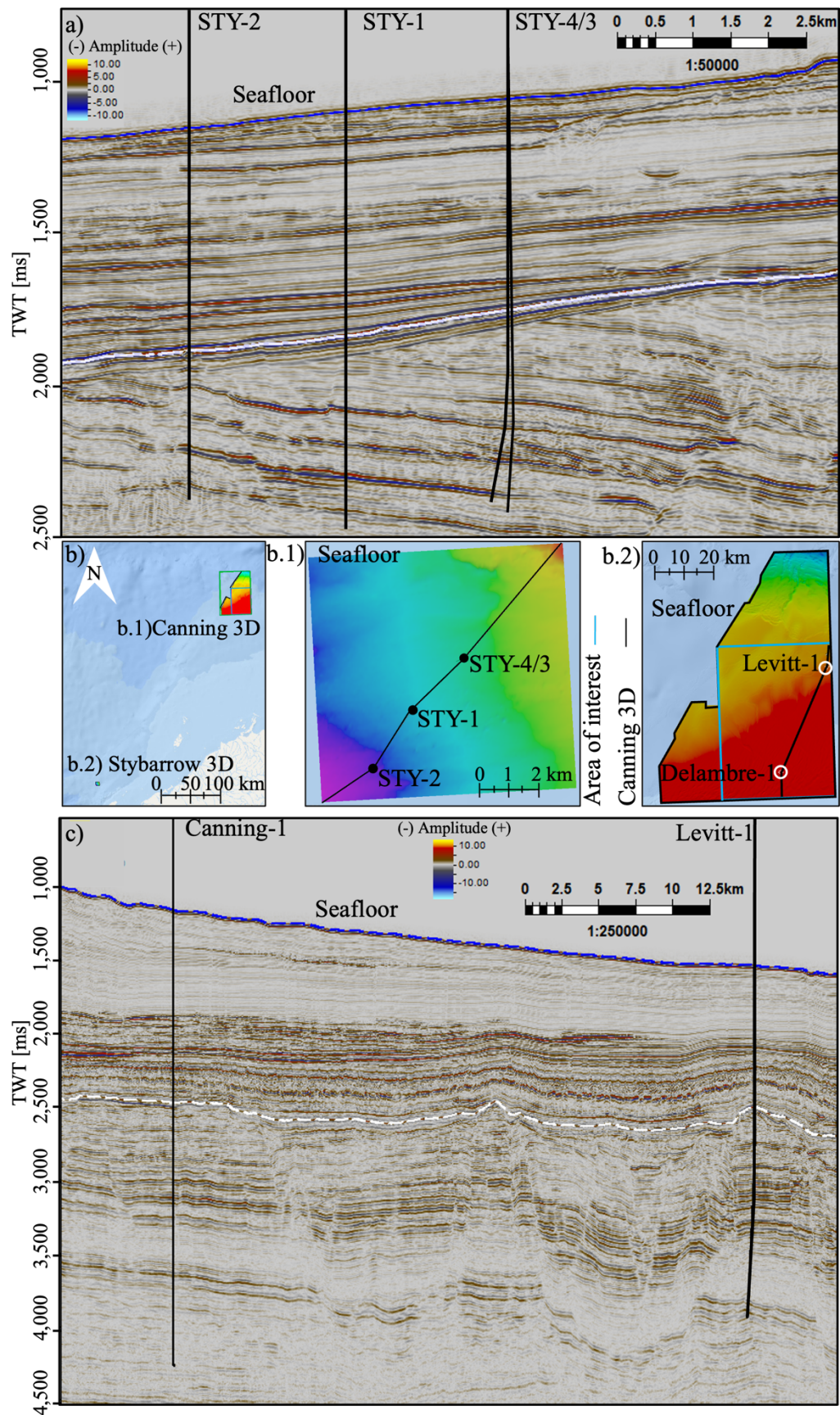


Figure 4.2.

Available seismic and borehole data in Northwestern Australia to study the risk of sand presence. Panel a) is the original seismic data associated with the Stybarrow 3D seismic cube. Panel b) is the regional location map b.1) Seafloor structural map interpreted on the seismic data. Panel b.2) shows the seafloor structural map interpreted in the Canning 3D seismic. We limit our study laterally and the cyan polygon shows the extend of the seismic data used in this study. We interpreted the Albian-Aptian unconformity -a regional unconformity- to delimit our study vertically. The Albian-Aptian unconformity is the interval where the petroleum systems are in the NCB. Our zone of interest will include sediments from the Albian-Aptian

WORKFLOW AND METHODS

WORKFLOW

Our automated facies classification expedites the delineation of reservoir rock boundaries and facilitates the analysis of lateral seal risks. The introduced workflow involves the refinement of borehole and seismic data by converting wireline logs into distinct rock types and reservoir rock properties. Seismic data enhancement encompasses the conditioning of seismic data through acquisition footprint removal, structurally oriented filtering (SoF), and spectral balancing, aiming to homogenize the frequency bandwidth. Simultaneously, borehole data is meticulously prepared by eliminating borehole and environmental effects that may influence property measurements. This preparation involves a petrophysical analysis known as well-tie, aiding in identifying the zone of interest in the seismic dataset. The transformation of wireline logs, including gamma-ray, density, neutron, resistivity, and compressional sonic logs, yields petrophysical parameters such as porosity, water saturation, mineralogy content, and rock types (claystone, sandstone, shale, silty sandstone).

Establishing a time-depth relationship during the well-tie process allows us to utilize borehole stratigraphy and formation tops to interpret seismic data. The chosen zone of interest, the regional unconformity to the Late Triassic, is applied to both the Canning 3D and the Stybarrow 3D datasets. Exploratory data analysis and principal component analysis guide determining a meaningful number of clusters during hierarchical clustering analysis (HCA). Calibration of the HCA at each borehole is a critical step, enhancing confidence and providing context to the data clusters, ultimately leading to generating a 3D sandstone and shale volume. Integration of fault probability and lithology models further facilitates lateral seal risk analysis. Figure 4.3 provides a graphical overview of the methodology.

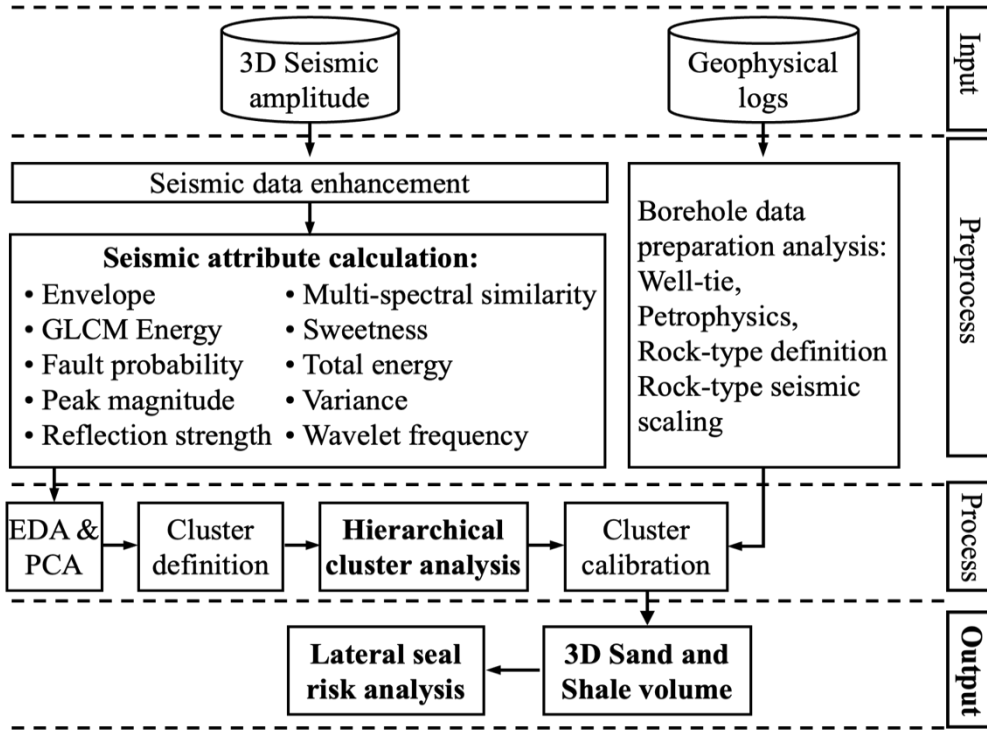


Figure 4.3. Our proposed workflow has four steps, data input including 3D seismic amplitude and borehole data. Data preprocessing where seismic and borehole data are conditioned and transformed into seismic attributes and rock-types. Our main process, which includes, Exploratory data analysis (EDA) and Principal component analysis (PCA). Finally, the results, where we present the 3D volume of sand and shales and analyze the lateral seal risk. The arrows show the direction of the workflow.

METHODS

Seismic enhancement includes the footprint removal methodology that employs Kx-Ky filters and adaptive subtraction to remove grid-like noise patterns imprinted in seismic data, improving seismic quality, following structural-oriented filtering (Fehmers and Höecker, 2003; Luo et al., 2006), which is an edge preservation approach to sharpen structural edges. The final step is spectral balancing, which is crucial for improving vertical and lateral resolution in a decade-old seismic dataset. We implement the spectral balancing method discussed by Matos and Marfurt (2014), as demonstrated by Chopra and Marfurt (2016). Having enhanced Stybarrow 3D and Canning 3D seismic data and limited the horizontal extent, we cropped vertically from the Albian-Aptian unconformity to the Late Triassic interval. In the area of interest (AOI) from the Albian-

Aptian unconformity to the Late Triassic, we calculate the seismic attributes on the workflow based on its geometrical, frequency, and amplitude response. Figure 4.4 shows the enhanced seismic data for both seismic surveys and the frequency-amplitude spectrum comparison.

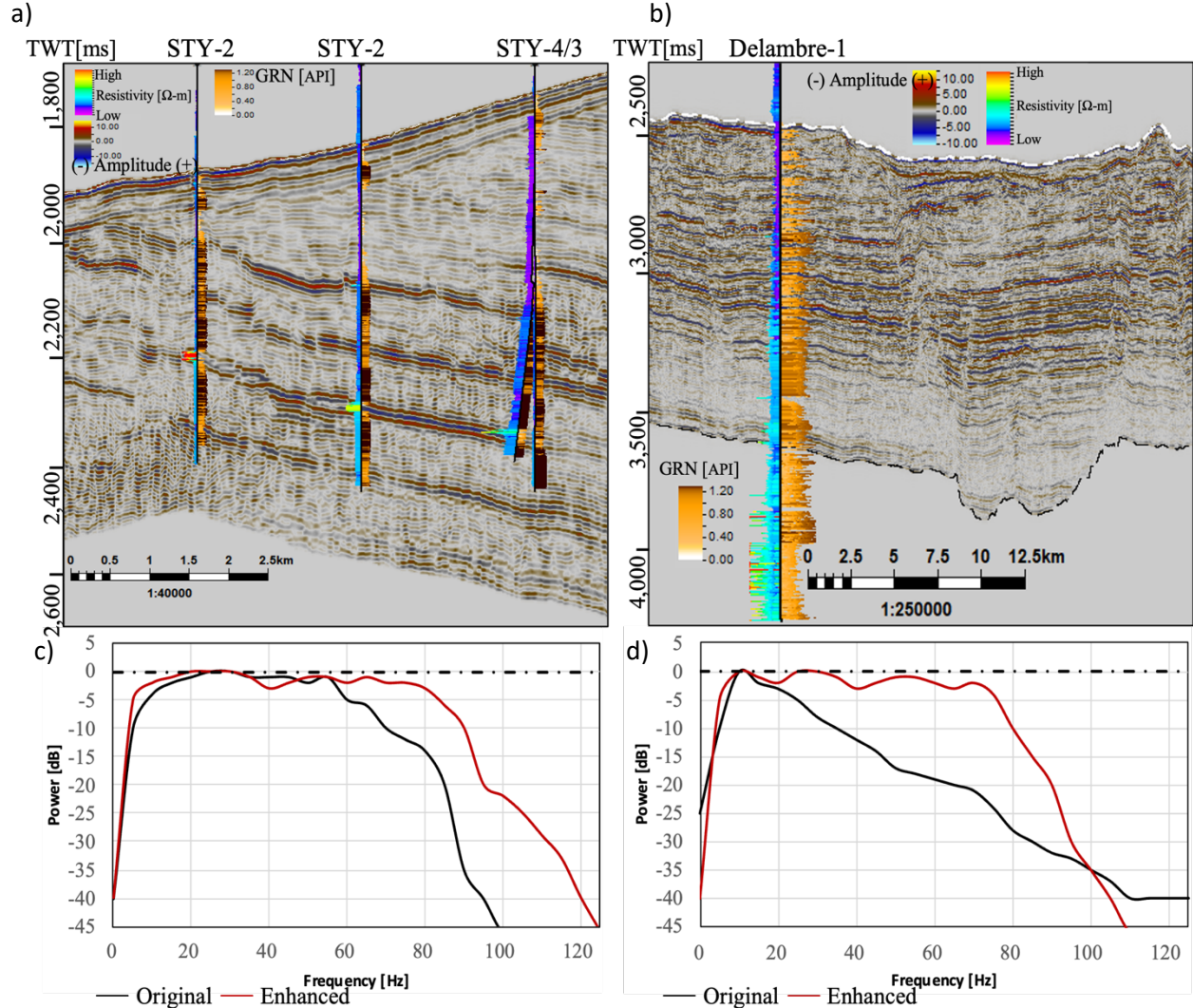


Figure 4.4 Panel a) Stybarrow enhanced seismic data and wireline logs from the borehole inside the area of interest. Gamma-ray and resistivity are overlaid to indicate shales and sandstone presence. Panel b) Canning 3D enhanced seismic data and wireline logs from the Delambre-1. Panel c) and Panel d) compare amplitude distribution against frequency before and after data enhancement.

In Figure 4.4, the enhanced vertical resolution and the vertical delimitation of the area of interest set the stage for our subsequent task: attribute calculation. We meticulously selected ten seismic attributes based on a comprehensive literature review, considering their definitions, and evaluating their effectiveness in describing various structural or stratigraphic features. The

attributes we selected include amplitude-dependent, textural, geometrical, spectral, and energy-dependent.

One of the instantaneous attributes within the amplitude-dependent category is Envelope, which is a measurement of the instantaneous energy of the analytic seismic trace, irrespective of phase. It provides valuable insights into the intensity of reflections. Reflection Strength, another instantaneous seismic attribute, quantifies the total energy of a reflection manifested in reflection amplitude. Notably, it remains independent of the imaginary part of the Hilbert transform, specifically the seismic phase. Sweetness, a derived attribute from the envelope to the square root of instantaneous frequency, proves valuable in discerning sandstone characteristics. For instance, clean sand embedded in shale exhibits a high envelope and lower instantaneous frequency, translating to higher sweetness. In contrast, low values of sweetness are associated with shale-like deposits.

Additionally, our focus extends to attributes with textural characteristics, such as the gray-level co-occurrence matrix. GLCM encompasses textural attributes like energy, entropy, and homogeneity. GLCM energy serves as a measure of textural uniformity in the data. Its application becomes apparent in distinguishing parallel reflectors associated with sandstones from chaotic, low-amplitude reflectors linked to shales.

Peak spectral magnitude and wavelet frequency are spectral seismic attributes derived from the continuous wavelet transformation (CWT). The CWT spectral decomposition aids in measuring the maximum amplitude of the wavelet transformation, offering insights into impedance changes and tuning thickness. Wavelet frequency is defined as the summation of the wavelet frequency within the signal bandwidth in the frequency domain, occurring through

complex wavelet transform in the spectral domain. This attribute provides a nuanced perspective on the frequency composition of the seismic data.

We used RGB-CMY mixers to combine the 6, 32, and 54 Hz spectral voices for three voices of the multi-spectral Sobel similarity. Sobel similarity as similarity is a geometric attribute, that measures the magnitude of inline and crossline derivatives of seismic amplitude along structural dips in different frequency bands. The color rendering of spectral voices highlights distinct scales of horizontal discontinuities.

Lastly, total energy, a measure of the sum of the energy of the weighted/balanced analytic traces in Hilbert space, is an amplitude-based seismic attribute. Meanwhile, variance, a geometric attribute, gauges the similarity of waveforms or traces adjacent over given lateral and/or vertical windows, offering insights into discontinuities related to faulting or stratigraphy. Figure 5 displays a sample of a vertical and horizontal slice of the seismic attributes used in Canning 3D and the Delambre-1 borehole to show the correlation of the gamma-ray response to each seismic attribute. Hierarchical clustering allows the combined analysis of seismic attributes to provide a more comprehensive understanding of subsurface structures and properties.

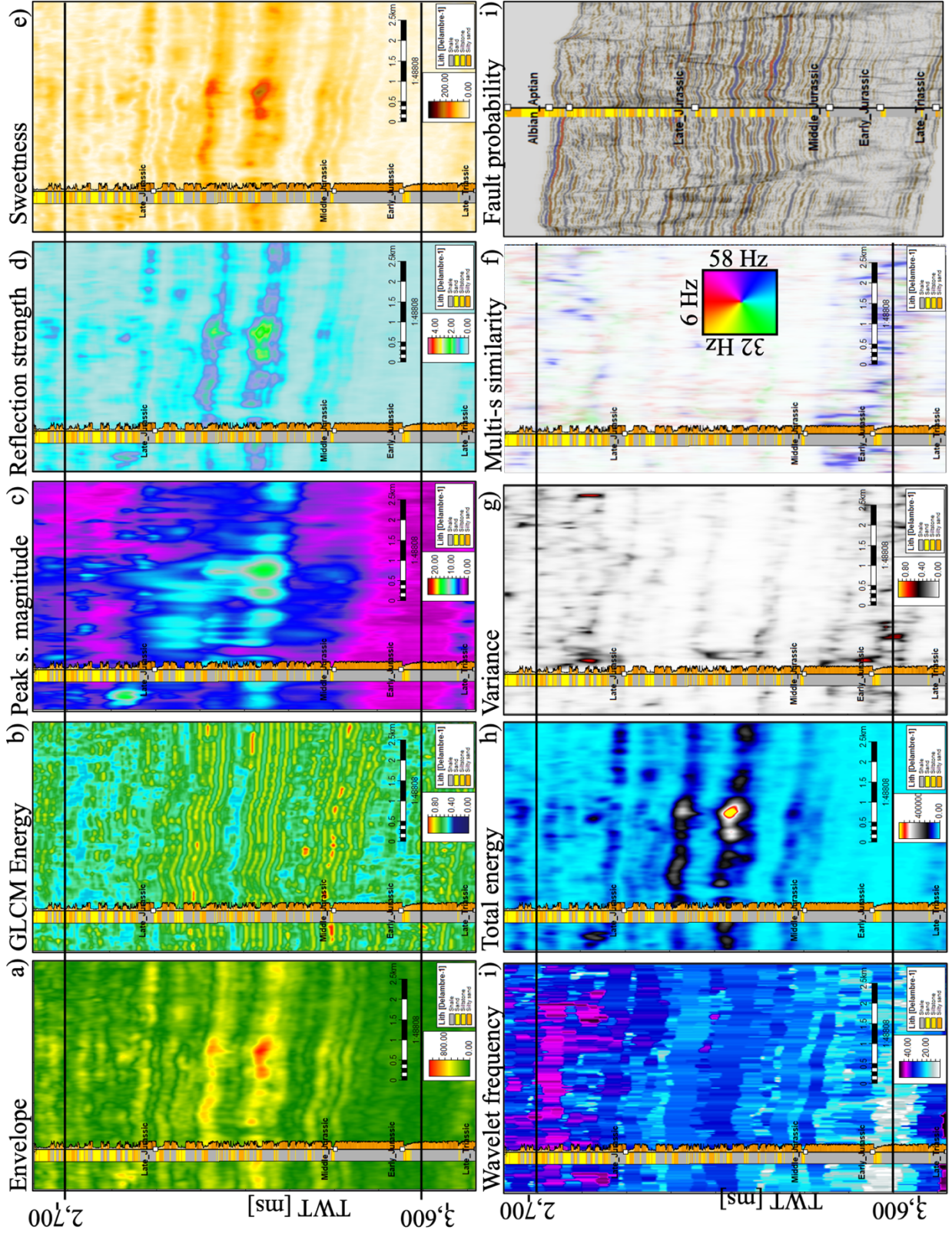


Figure 4.5. Seismic attributes from Canning 3D. In amplitude dependent attributes, which relay on window we use the same window (33 ms) in both seismic volumes. Fault probability using deep learning algorithm in Canning 3D used a batch of 64, while in Stybarrow used batch size of 32.

Exploratory data analysis (EDA), performed in both seismic volumes, showed the statistical differences among seismic attributes. In Figure 4.6, we normalized the seismic attributes using Z-score normalization for a direct comparison. Figure 4.6 shows three EDA visualization methods – box-plot, Pearson’s correlation, and density functions. Box plots indicate the statistical distribution of the dataset mean, P10, P50, P90, and the standard deviation in how the box is thinner or thicker. Thinner boxes show a minor standard deviation. Pearson’s correlation estimates the correlation between seismic attributes to keep the parameters independent. The density plots of histograms show the show statistical probability density distribution of the distribution of the seismic attributes compared to the distribution and overlapping of the seismic attributes. Ideally, box plots of seismic attributes should not overlap. Removing highly correlated avoids extra computational time.

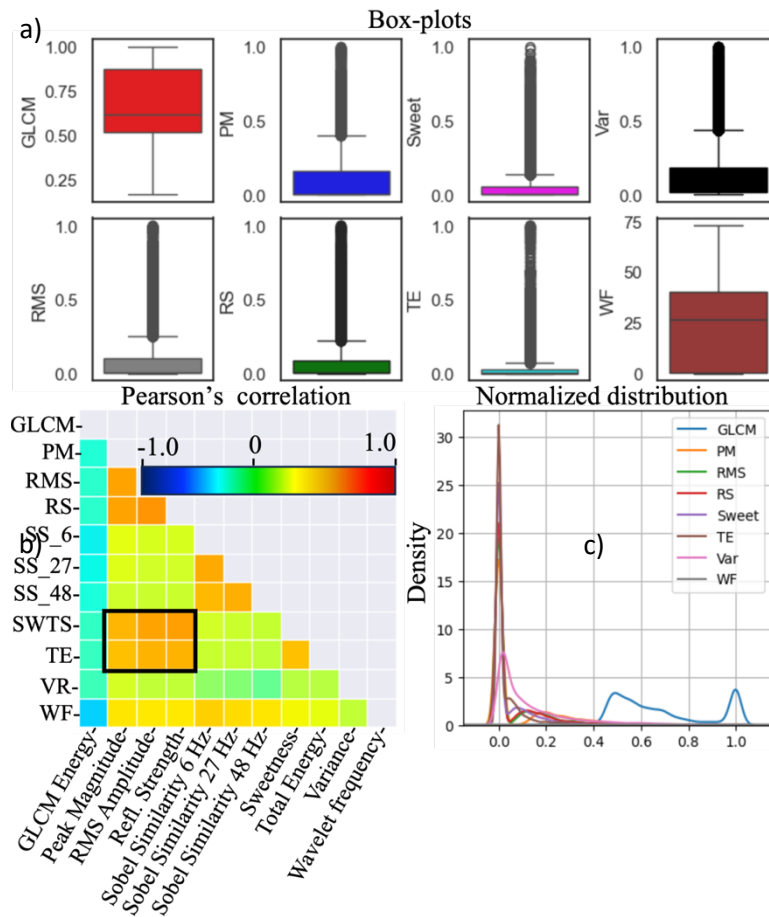


Figure 4.6. Panel a) Box-plots of the seismic attributes after normalization. Panel b) Pearson’s correlation that show the statistical correlation of the seismic attributes. Panel c) Density plots of the seismic attributes. Box-plots of GLCM energy and total energy have a strong contrast, which contribute greatly to cluster differences.

The first objective of attribute normalization is to facilitate a direct comparison across different seismic surveys. The second is eliminating scale dependency and generating a Gaussian statistic attribute response suitable for subsequent principal component analysis. Subsequently, we compute the correlation matrix for the normalized training dataset and then determine eigenvectors and eigenvalues. Figure 4.7 shows the seismic attributes exhibiting the most significant contributions within the eigenvectors for all seismic attributes.

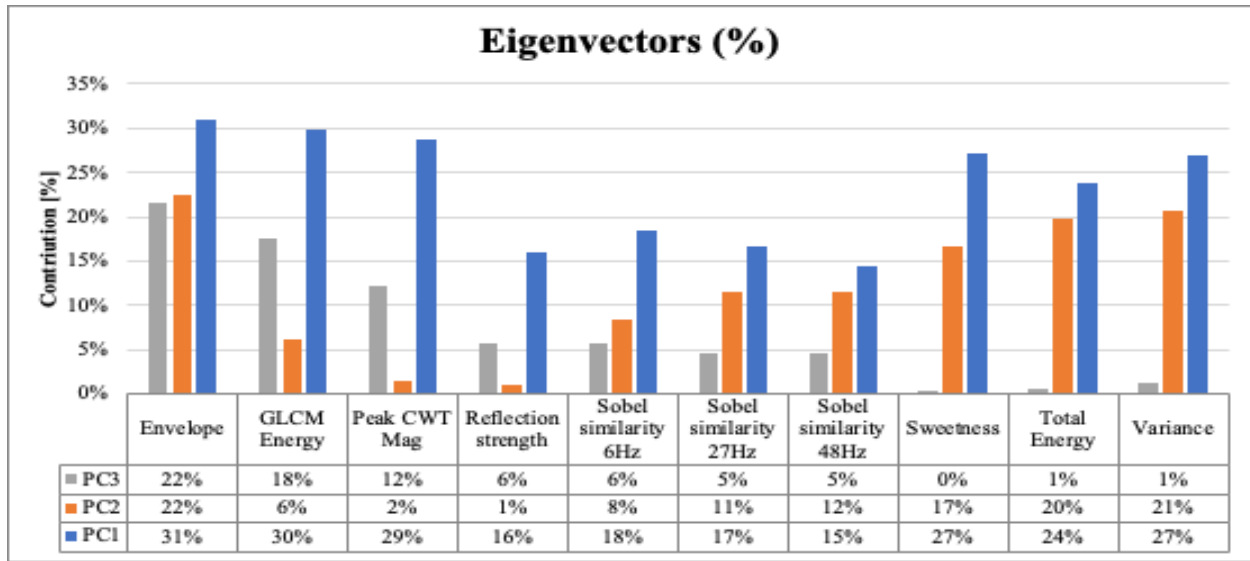


Figure 4.7. Eigenvectors of the seismic volume, we observe that the envelope attribute shows the highest contribution, followed by the GLCM energy, Peak CWT magnitude, Sweetness, Variance, Total energy, and Sobel similarity at 6 Hz.

We utilized the gap statistics method to determine the optimal number of clusters for hierarchical clustering analysis. This minimizes the difference between the data matrix and a uniform random matrix with averaged within-cluster distances (equation 4.1). K^* represents the optimal number of clusters that maximizes this gap statistic.

$$K^* = \operatorname{argmin}\{K | G(K) \geq G(K+1) - S'_{K+1}\}$$

Where:

$G(K) = \log(W_k^{unif}) - \log(W_k^{data})$ and $S'_{k+1} = S_k \cdot \sqrt{1 + 1/n}$, S_k =std. deviation of $\log(W_k^{unif})$ across n iterations.

A key novelty of our methodology is using calibrated HCA to group similar data points based on Euclidean distance. HCA creates a hierarchy of nested clusters visualized in a dendrogram. The gap statistic indicated that the data could be categorized into 5 distinct clusters. However, these clusters were not arbitrarily selected. We leveraged complete linkage, which uses the maximum distance between points, to identify sparse and representative clusters. Our selection of complete linkage uses the most significant Euclidean distance, keeping the clusters sparse and representative. In Figure 4.8, we show an example using color theory of how individual data points were hierarchically grouped into larger clusters using the complete linkage criterion. This forms a cluster tree (dendrogram) to represent the nested groupings of data points. In the workflow and methodology herein proposed, the borehole calibration and the number of clusters defined with the gap statistics define the categories in which seismic attributes are clustered.

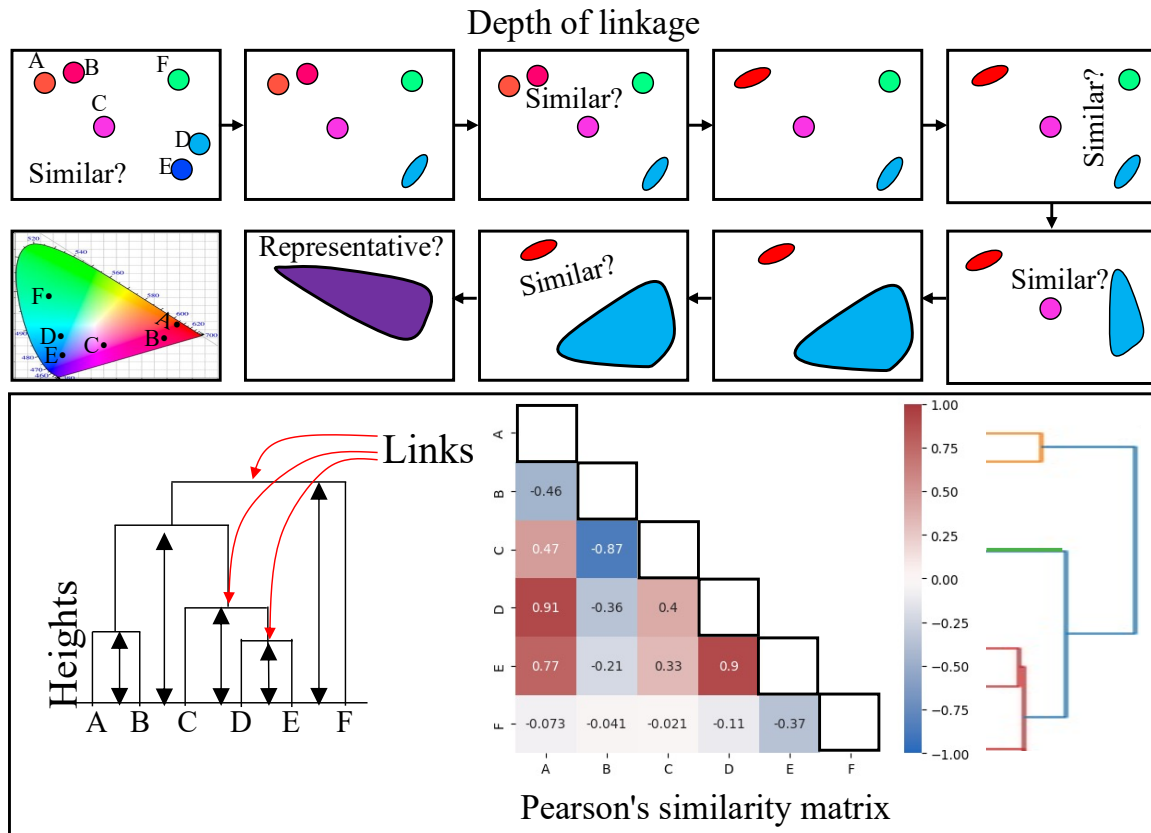


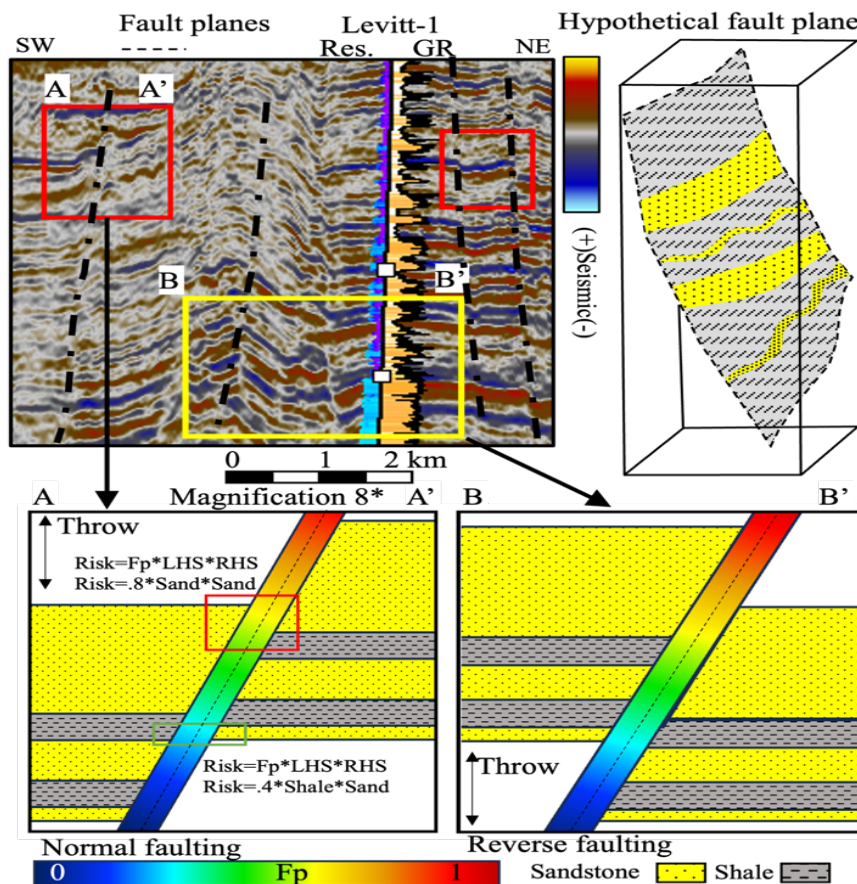
Figure 4.8. Panel a) Sketch to show HCA clustering; the schematics exaggerate the clustering to reduce the variability. Panel b) Example of links and the different levels, and how the similarity matrix assists the clustering.

Finally, the HCA-defined clusters were transformed into sand volume percentages using the actual rock types identified at the borehole locations. We defined seismic resolution as one-fourth wavelength, estimated from the dominant frequency and interval velocity logs. This resulted in seismic resolutions of 28 meters for Stybarrow and 34 meters for Canning. The most frequent rock type within a sliding window of the respective resolution was used to upscale the borehole rock types in both datasets. Our petrophysical analysis of the wireline logs, calibrated with the core data and drilling cuttings, defined four rock types using selected cutoffs (Table 4.1).

Rock type cutoffs		
Rock type	Cutoff	Calibration points
Shale	Shale >70% and PHIT <2%	Sidewall samples
Siltstone	Shale <70%, Quartz <50%, 10< PHIT < 5 %	Sidewall samples and Core
Silty sandstone	Quartz >50, Feldspars >10 Calcite >10%	Sidewall samples and Core
Sandstone	Quartz >70% and PHIT >10%	Sidewall samples and Core

Table 4.1. Petrophysical parameters to discretize the wireline log data that reads different parameters of the rocks into rock types.

Having calibrated the 5 clusters to the seismic scale rock types and visualized the cluster spatial distribution, we defined a volume of sandstones. The subsequent step involves assessing juxtaposition risk and shale-gauge ratio on faults to estimate lateral seal risk. This process includes skeletonizing the calculated fault probability to enhance fault plane definition, considering the fault dip magnitude and azimuth. The resulting 3D sand volume distribution is then multiplied by the fault probability cube. For a visual representation, refer to Figure 4.9, which illustrates the fault probability and sand-shale juxtaposition used in determining these risks.



With our refined methodology and systematic workflow, we have significantly enhanced seismic data by integrating borehole information into rock types. This integration connects the calculated HCA clusters and AOI. The primary objective of this study is to define lateral seal risk in 3D using seismic attributes and to map the 3D distribution of sandstones within AOI. This approach aims to expedite prospect definition and efficiently pinpoint areas that warrant further inspection.

In the results section, we present our findings in two segments. Firstly, the Stybarrow 3D dataset calibrates and evaluates our workflow. In interpreting the Stybarrow 3D dataset, we manually identified faults. Subsequently, the Canning 3D dataset is utilized to test our methodology rigorously.

RESULTS

We utilize the HCA to automatically group seismic facies into clusters based on their attribute data space similarities. We then calibrate and test this HCA seismic facies clustering firstly using the Stybarrow 3D dataset, taking advantage of known well control and geology in the survey to match seismic facies clusters to actual rock units and lithologies. This allows us to verify that the HCA clusters correspond to meaningful geologic units. Once the seismic facies clustering is calibrated, we then utilize the Canning 3D seismic volume as an additional test to evaluate how well the HCA seismic facies extend spatially. The Canning 3D is a frontier exploration survey where the borehole results are promising regarding the quality of the reservoir rock (good porosity). Applying the calibrated HCA clustering from Stybarrow to Canning provides an independent test to determine if consistent seismic facies are recognized and if they can be reliably correlated across the surveys. This will demonstrate both the portability and validity of the HCA

seismic facies classification. Moreover, we use lateral fault analysis to show an application of our calibrated cluster analysis.

STYBARROW RESULTS

The Stybarrow 3D serves as a valuable tool for validating the effectiveness of HCA in mapping the 3D distribution of sandstones and assessing lateral seal risk within the area of interest. The Stybarrow structure is characterized by a tilted fault block trending NE-SW, with the intersection of NNE/NE and E-W trending normal faults creating an elongated, triangular trap that forms structural closure to the southwest. Figure 4.10 comprehensively summarizes the HCA clustering, juxtaposed with multi-spectral Sobel similarity for a contrasting perspective. Two specific stratigraphic intervals, the Barrow Group and Macedon Formation, are highlighted in Figure 10a. Moreover, the correlation of Cluster 5 to NW-SE fault systems is evident (cyan arrow) is also depicted in panels a), b), and c). The selection of Cluster 5 is guided by variance and similarity attributes, as depicted in Figure 10d, through multi-spectral Sobel similarity. Examining the Barrow Group (Figure 4.10b) reveals a sandstone-rich deltaic system, emphasized by the red arrow: clusters 1 and 2 pinpoint cleaned sandstones and silty sandstones within the Barrow formation.

Turning attention to HCA clusters around the Macedon Formation (Figure 4.10c), a basin-floor fan bordered by claystone and shales across faults is evident (green arrows). The compartmentalization of turbiditic flow aligns with the geological background of the area, accentuated by Cluster 5. This submarine fan, often associated with the early stage of a lowstand system tract (LST) during falling relative sea levels, suggests deposition on the lower slope or basin, typically associated with favorable porosities. However, horizontal slices in Figure 4.10

(panels b and c) reveal Stybarrow-3 and Stybarrow-4 downdip outside the clusters associated with sandstones and silty sandstones. HCA plays a crucial role in estimating the volume or lithologies required for reserve estimations.

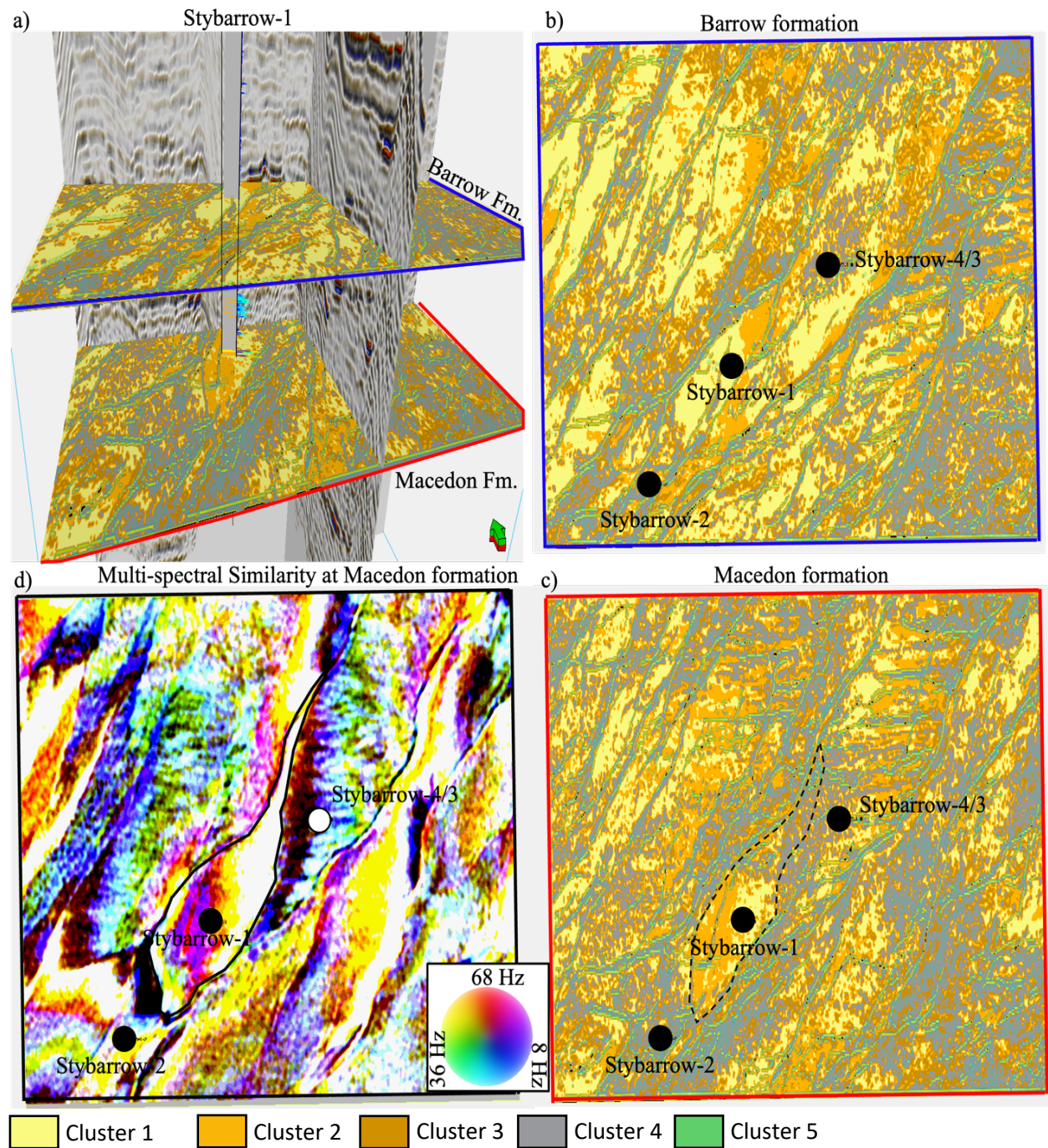


Figure 4.10. HCA cluster results in Stybarrow 3D to show the geological-like clusters and a comparison of multispectral similarity. Panel a) 3D view of two-time slices, the Barrow and Macedon Formation. Panel b and c) show the map view of the Formations shown in Panel a). The map view helps to show the sandstone

distribution. Panel d) shows the multispectral similarity. i.e., the Sobel similarity voices blended in RGB. The calibrated clusters' detectability and separability help define sandstone presence, and the rock type definition is sharper than the Sobel similarity blend.

Cluster calibration aims to reduce the ambiguity of machine-learning-generated clusters by showing their correlation with rock types. Figure 4.11 compares the borehole scale and histogram of seismic scale rock types against clusters. This histogram plot is an accuracy test where the number of lithology samples is compared with the clusters across the four boreholes.

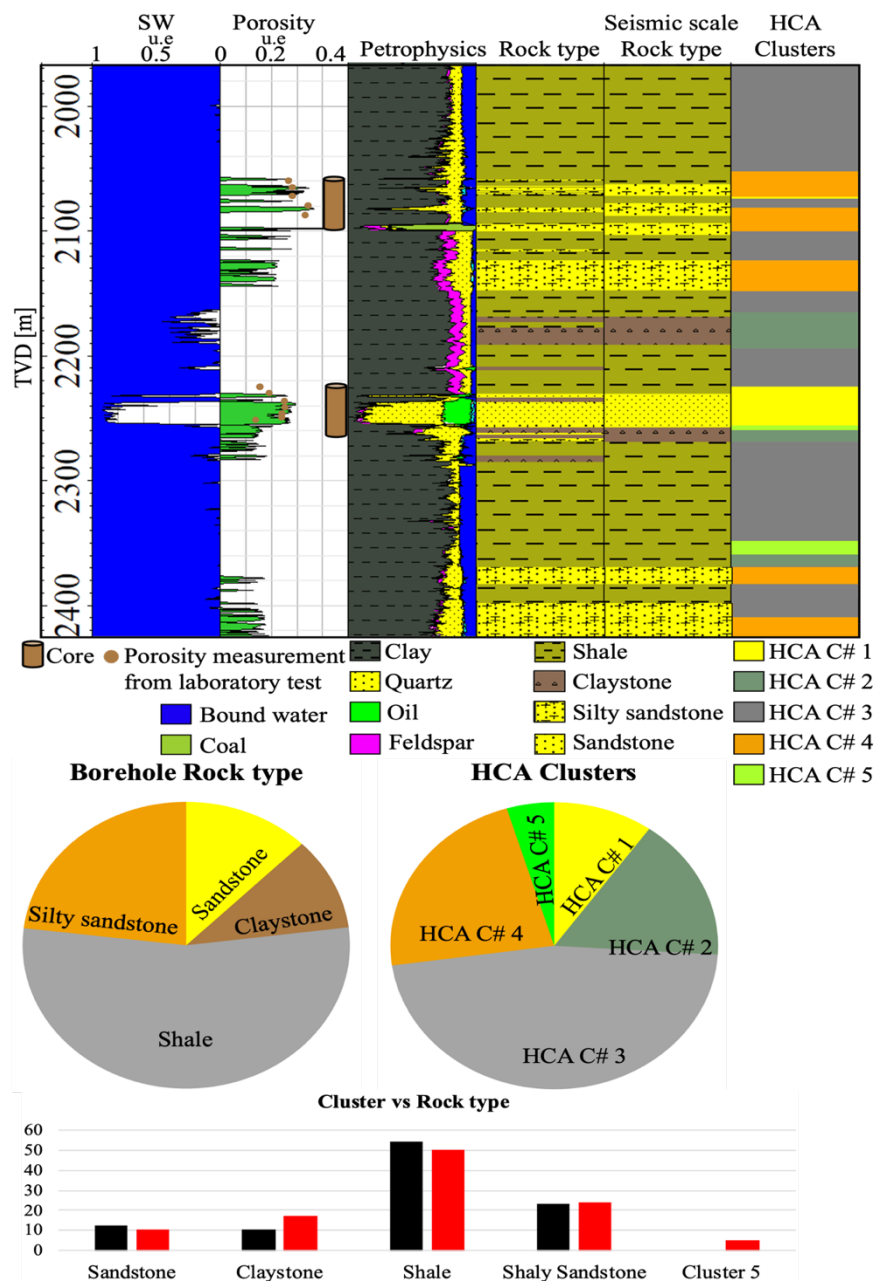


Figure 4.11. Stybarrow-1 data, showing the petrophysical evaluation and rock type from the mineral evaluation. The NTG of Stybarrow-1 from 2,000 m TVD to 2,400 m is 40%. Also, the rock types at seismic scale and its correlation with the HCA cluster. Cluster 1 shows excellent correlation with the seismic scale sandstone. Cluster 5 and 4 have the largest misclassification.

Lastly, in Figure 4.12, we show the interpreted fault framework of the interval of interest compared with the fault probability voxels to bolster the confidence in the fault probability attribute as an automatic fault extraction method. Also, the fault probability helps with the lateral seal risk when computed with the rock type correlation.

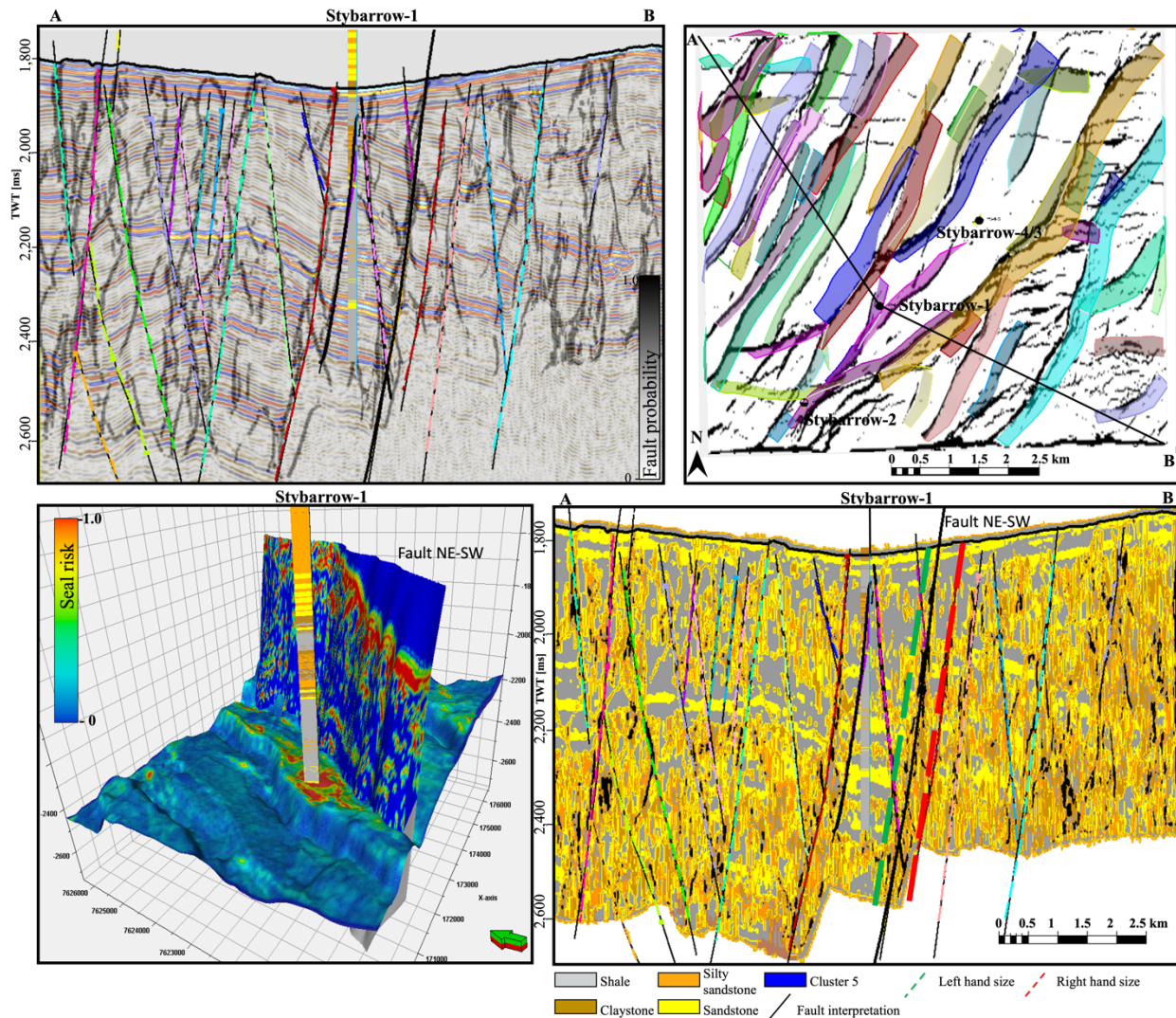


Figure 4.12. Seal risk analysis, Panel a) manual fault interpretation on the amplitude data and the fault probability overlapped to show the similarities between the interpreter's perception of a fault and the fault probability, red and green arrow shows the amplitude characterization Macedon Formation—panel b) Map view of the fault probability on a time slice and interpenetrated fault planes, cyan arrow is indicating the NE-SW fault plane on panels c and d. Panel c) shows HCA clusters transformed into lithologies across the reservoir; the green and red arrow now shows the Macedon Formation transformed into a sandstone cluster, and red are the limits of the lithology capture to be multiplied by the fault probability; the fault plane is in the center in black—panel d) Lateral seal risk on the Fault NE-SW.

CANNING 3D RESULTS

The Canning 3D dataset is a challenging area that has undertaken several phases of tectonic deformation. We used unsupervised learning to test our methodology and value in automatically mapping the reservoir rock distribution. The AOI of Canning 3D is more than 63 times larger than Stybarrow, and detailed mapping is very time-consuming. Looking closer into the HCA clusters, cyan arrows reveal meandering channels and a complex fault system in the interval between the Cretaceous and Middle Jurassic (Figure 4.13). Purple arrows on the panels b) and c) show deltaic systems. Given that we transformed HCA clusters into lithologies, and their deposition shows good agreement, the cluster is validated. We contrast with spectral decomposition (RGB) color blending to see an improvement.

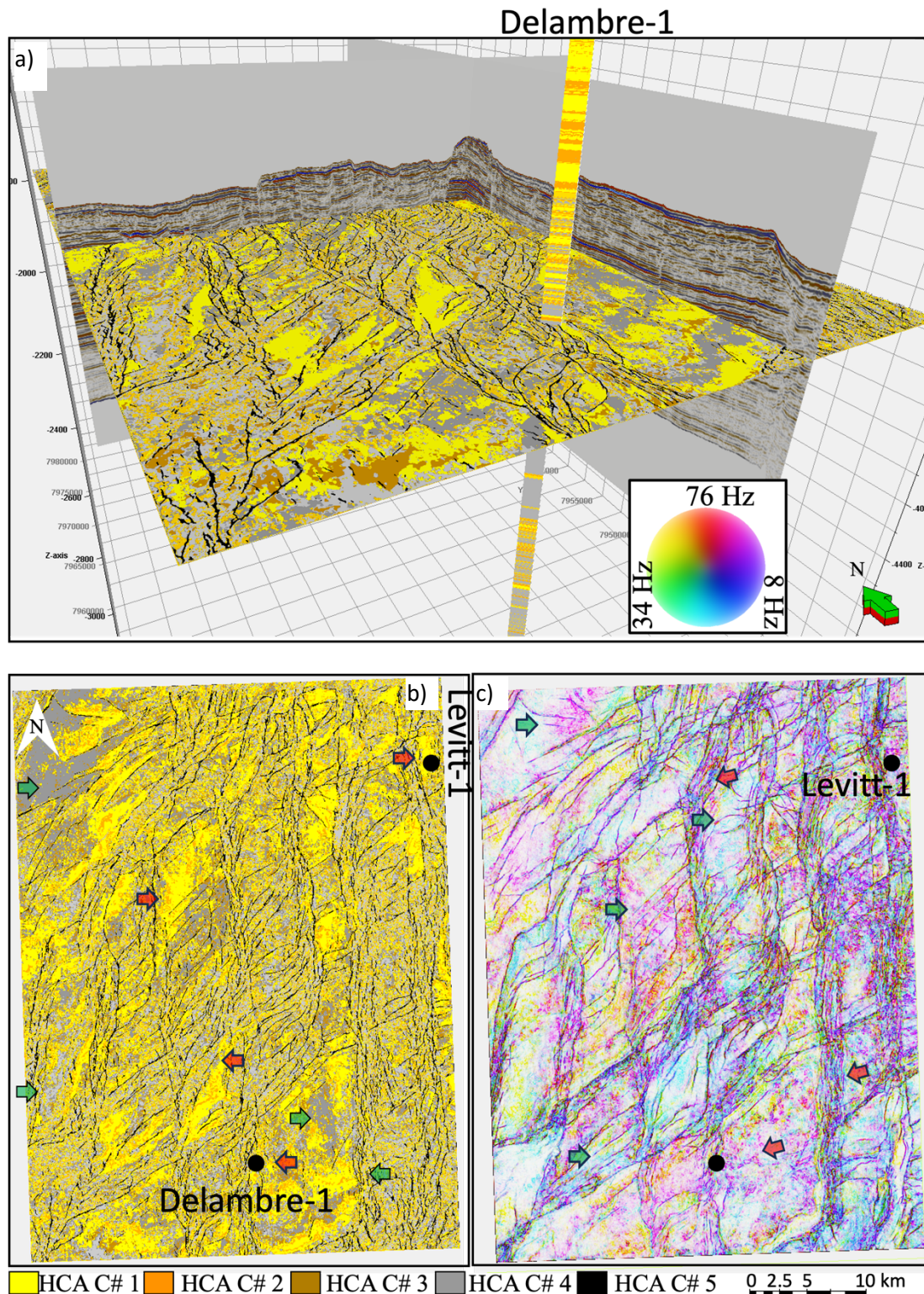


Figure 4.13. Panel a) A 3D perspective around the Delambre-1 borehole shows the lithology distribution from the calibrated HCA cluster. Panel b) RGB color mix used in Panel c). Panel c) shows multi-spectral similarity exhibiting structural and stratigraphic features. Panel c) A map view shows the channelized features where sandstone dominates the distribution. HCA Clusters improve the interpretation in contrast with multi-spectral similarity.

We use the only two boreholes available in the area to calibrate the HCA clusters; we assist our interpretation and associated channel-like clusters to a higher probability of being sand-filled. Notability and fault presence are imprinted in clusters 4 and 5. Levitt-1 borehole in Figure 4.14 shows their correlation with clusters to generate the sandstone volume.

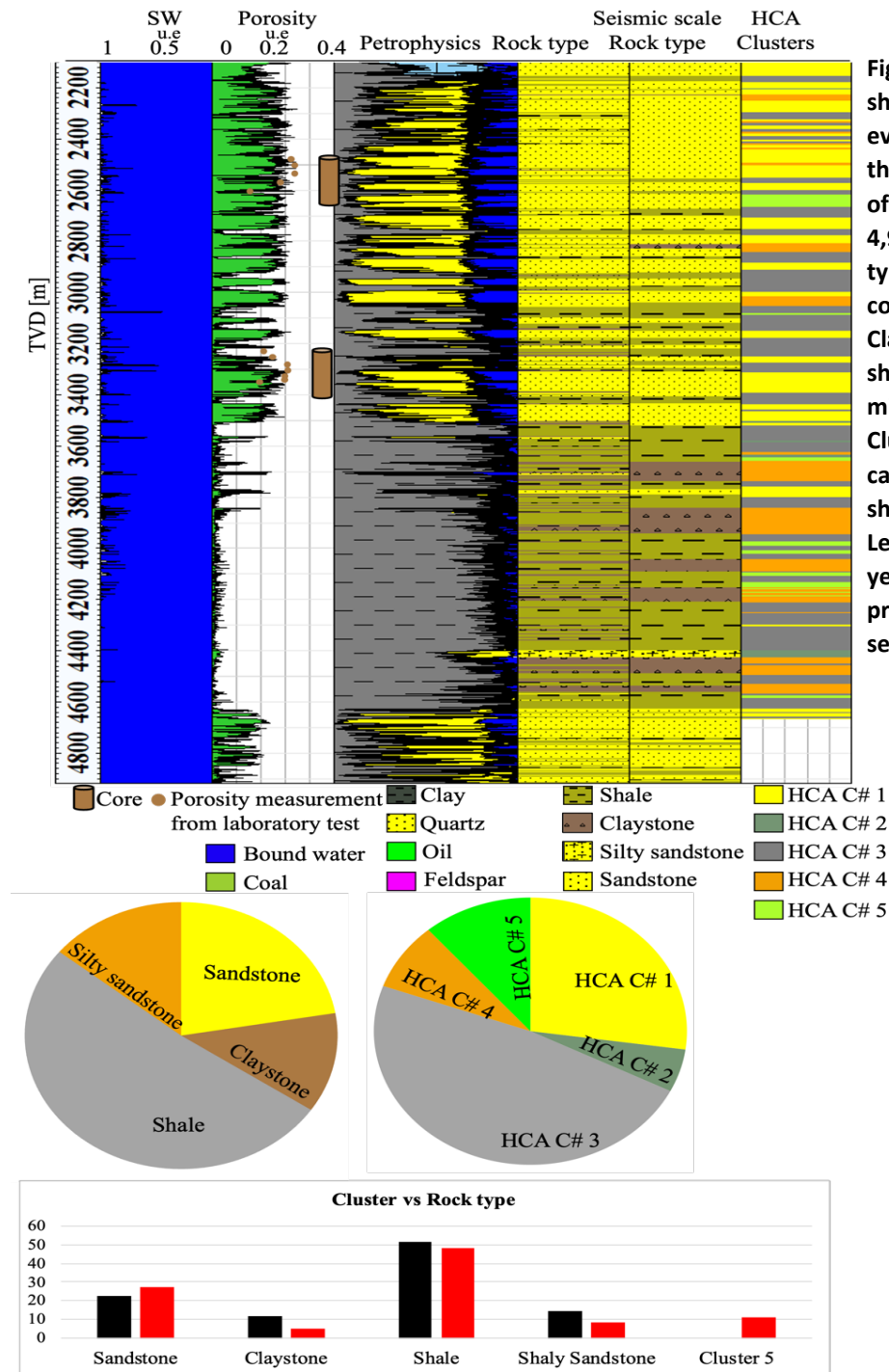


Figure 4.14. Levitt-1 data, showing the petrophysical evaluation and rock type from the mineral evaluation. The NTG of Levitt-1 from 2,200 m TVD to 4,900 m is 60%. Also, the rock types at seismic scale and its correlation with the HCA cluster. Claystones and silty sandstones shows the largest misclassification. Cluster 1 and Cluster 3 shows a good calibration with sandstones and shales. The high NTG in the Levitt-1 structure is promising yet the highly faulted nature presents a risk on the lateral the sealing.

Lastly, Figure 4.15 shows the fault probability overlapped on the HCA section, Fault probability is used as an automatic fault extraction input. The width of the fault plane (red arrows) is a consequence of seismic image resolution and bandwidth.

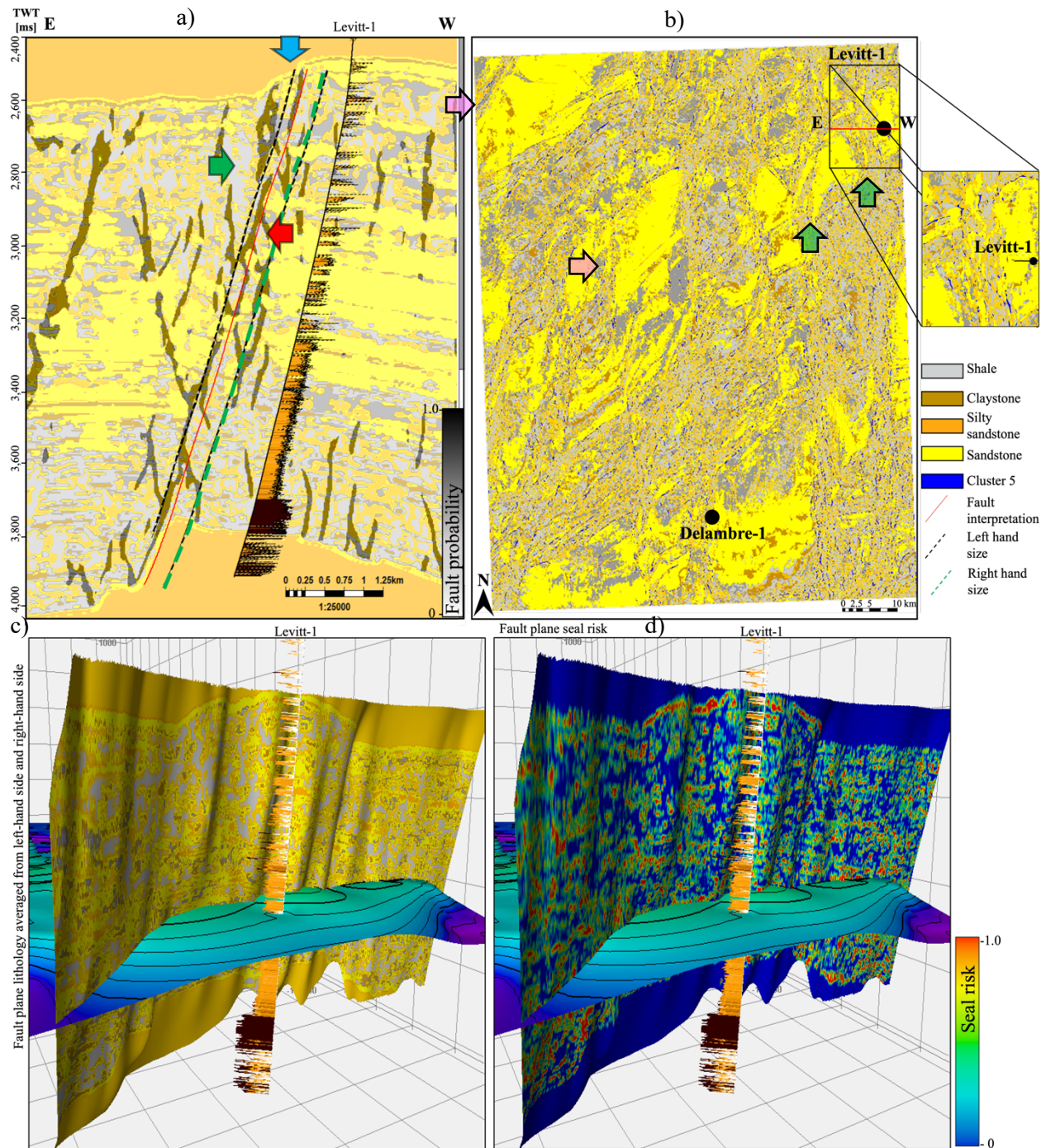


Figure 4.15. Canning seal risk analysis depicts a high risk in Canning seals. Panel a) displays HCA classification already transformed into rock types intersected with fault probability, delineating lithology computation within the fault zone. The dashed black line to the green line is the zone where the computation

of lithologies with the fault plane takes place, and the green and red arrows show the left and right-hand sides of the central fault (cyan arrow). The solid black line corresponds to the next panel. Panel b) illustrates sandstone distribution in the Lower Cretaceous via map view, revealing NE-trending clinoforms and geofoms characterized by hundreds of meters of height sandstones over tens of kilometers (magenta and green arrows) indicative of a continental deposit. Panel c) projects HCA lithologies, emphasizing more sandstone-related clusters than shale clusters (black arrows). Panel d) shows seal risk probability on the fault plane, indicating a high sealing risk in the Lower and Middle Jurassic deposits and traps due to sandstone juxtaposition.

DISCUSSION

The goal of automatically characterizing seismic facies is to abate time in picking surface and seismic facies interpretation. Defining seismic facies is important for subsurface energy as these are targets of drilling and further analysis. Before the widescale adoption of machine learning across energy exploration, one outstanding question remains: can it be trusted? Our calibration and seismic interpretation provide a sound answer and a good correlation with the structural and stratigraphic geology of the region (McCormack and McClay, 2018). Chopra and Marfurt (2014) suggest using spectral, geometric, textural, and instantaneous attributes to capture structural and stratigraphic features in seismic data. Our study incorporates a variety of seismic attributes to investigate the structural and stratigraphic elements in both datasets. Our HCA clustering lithology calibration and fault probability enhance seismic interpretation in a promising direction for seismic exploration and risk analysis. Machine learning studies are becoming routine, and the interpreter must validate or calibrate the outputs (Sacrey and Roden, 2014; Roden and Sacrey, 2016).

HCA, in contrast with GMM (Wallet and Hardisty, 2019) and SOM (Zhao et al., 2016; La Marca et al., 2022), does not require an initial assumption of a particular number of clusters (i.e., k-means, GMM), and it outputs discrete clusters that may correspond to meaningful taxonomies. Its primary disadvantage is that it is the computationally expensive time for large data sets, $O(n^2 \log(n))$.

Clustering analysis studies (Zhao et al., 2016; La Marca and Bedle, 2022; Ji et al., 2023) show the capabilities of machine learning to cluster seismic geomorphologies. Our study expands on this feature and introduces hierarchical clustering as an alternative tool. The geological features for the Macedon Formation observed in Figure 4.10 show basin-floor fans from a distributary channel, agreeing with Schenk et al. (2019). Coarser materials remain in the center of the fan, while silty sandstones are on the front lob, as the sedimentary model would expect (Posamentier and Kolla, 2003). The Barrow Formation in Figure 4.10 shows the deltaic system, expected in the latter stages of continental rifting, and is rich. The fault frame and structures herein agree with Alves's (2020) NW Australian Basin development description.

Borehole data in Figures 4.11 and 4.14 suggest using four clusters; even the minerals found at the cored sandstones suggest one predominant sandstone cluster and a transition to shale; the gap statistics method suggests five, nine, and seventeen clusters as the optimum cluster number. We select the use of five clusters to correlate them with the lithologies in the seismic cubes. The overlap of the seismic response between clusters is expected as the lithological contrast is slight. Borehole data needs to be carefully calibrated to avoid misinterpretation. While it provides factual rock data at borehole locations, its spatial uncertainty is considerable. Nonetheless, geology varies more vertically than horizontally, and the seismic facies allow the interpretation of Slatt and Weimer (2004). In the Stybarrow 3D, four boreholes help in reducing the uncertainty regarding the rock-type distribution and presence by providing information at the rate of 1 borehole every 14 km²; in the Canning 3D, the scenario is the opposite, one borehole every 1,763 km², increasing the uncertainty in the lithological control. Overall, borehole data suggest using four clusters; even the minerals found at the cored sandstones mainly suggest one predominant sandstone cluster and a transition to shale; the gap statistics method suggests five, nine, and seventeen clusters as the

optimum cluster number. We select the use of five clusters to correlate them with the lithologies in the seismic cubes.

In Figure 4.13, we showed the lithological distribution in the Canning 3D associated with the transitional to shallow marine deposits of the Upper Triassic. The reservoir architecture of the amalgamated channels is sandstone-rich with good porosities as described by Levitt-1 borehole. However, the complex fault system makes it challenging to have well-confined sandstone and to find a shale and sandstone juxtaposition across the multiple faults, making the accumulation of hydrocarbon risky in the area. Figure 4.15 around the Levitt-1 borehole data shows thick sandstone intervals with good porosity and, during its drilling, did not show any hydrocarbon. The lateral seal risk reflects the complex faulting history and the abundant sandstone presence. Comparatively with the fault plan example of the Stybarrow, the risk in the Canning 3D is higher than the risk in Stybarrow 3D. Magee and Jackson (2020) highlighted the Canning 3D complex normal faulting and structural inversion, and Gratell et al. (2022) pointed out the implication of the area's inversion tectonics and stratigraphy architecture.

In both datasets, HCA outperforms multi-spectral similarity in defining facies distribution. On the map view, the HCA better highlights the levees and channel edges characteristic in the northern portion of Northwestern Australia, as pointed out by Lech (2013). HCA exhibits excellent geological agreement, emphasizing its potential for improving seismic facies characterization in subsurface energy exploration.

CONCLUSION

We have presented a workflow based on unsupervised learning for automatic facies classification and showed the implementation of calibrated clusters in evaluating lateral seal risk. We defined the number of clusters as the depth of the linkage in the dendrogram; unlike most existing unsupervised works, HCA classifies without determining the cluster number. Another uniqueness of our study is the calibration of clusters to borehole lithology data. We calibrated seismic clusters to lithologies and used fault probability to impose lateral constraints and sharpen the classification; multi-spectral Sobel similarity reinforces the lateral discontinuities. In addition, we feed the seismic clusters transformed to lithology into a fault network to estimate the lateral risk, helping the prospect evaluation and assisting risk identification. Some limitations still need to be improved in our method. One limitation is that HCA is just one of many clustering techniques, and its effectiveness depends on the quality of data, the relevance of selected features, and the geological context. A second limitation is the computational time for large data sets, $O(n^2 \log(n))$, and lastly, once a decision is made to combine two clusters, the clusters can't be reversed.

Finally, we have demonstrated the method's effectiveness by applying it to siliciclastic systems in the Northern Carnarvon Basin, offshore Australia, and the two different paleoenvironments of deposits showcase the effectiveness of our method. Combining clustering methods with geological expertise enhances reservoir rock identification accuracy in subsurface exploration. In summary, HCA enables accurate delineation of lithologies, complex faulting, and key stratigraphic features. The structural and stratigraphic interpretations agree with previous geological studies of the area.

ACKNOWLEDGMENTS

We want to thank the Australian Governmental Department of Geoscience for providing both seismic surveys used in this study (HCA2000A 3D and Canning 3D) and borehole data, including wireline logs and associated reports required for this study. AASPI, MathWorks, and SLB are greatly acknowledged for providing the following software to the University of Oklahoma and the School of Geoscience: AASPI version 2023, Matlab 2023, Techlog 2019, and Petrel G&G software v.2022, respectively. Thanks to Dr. Brett Carpenter for the valuable discussion.

DATA AND MATERIAL AVAILABILITY

Datasets related to this article can be found at <https://portal.ga.gov.au/persona/gadds>, an open-source online data repository hosted by the Australian government and Geoscience Australia.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

CHAPTER 5: ASSESSING MACHINE LEARNING ACCURACY IN PREDICTING IN-SITU GAS SATURATION AND LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS TO UNDERSTAND ITS UNCERTAINTY: A CO₂ EXAMPLE FROM KEVIN DOME. *

*This chapter was submitted to the SEG journal, *The Leading Edge*, in the 2024. Subsurface Uncertainty special issue

ABSTRACT

As demands for cost-effective solutions in CO₂ projects grow, machine learning models, notably random forest, and recurrent neural networks, are increasingly favored in geoscience for precise and automated assessment of petrophysical properties in mature fields earmarked for CO₂ injection and sequestration. However, inherent uncertainties in subsurface conditions, coupled with challenges in interpreting situ gas saturation predictions made by these models, underscore the critical need for understanding contributing factors to uncertainty and evaluating confidence in pre-trained model parameters.

This study leverages Local Interpretable Model-Agnostic Explanations (LIME), a tool in Explainable Artificial Intelligence, to elucidate in-situ gas saturation predictions produced by random forest regression and recurrent neural networks models (RNN). By analyzing borehole data from the Kevin Dome, we investigate the impact of borehole and data density, relevant wireline logs, and algorithm selection on gas saturation estimation. Our analysis, supported by variational Bayesian Monte Carlo simulations and 3D geostatistical modeling, highlights the paramount importance of data quality over quantity and underscores the necessity for interpreter supervision, as well as the high sensitivity to model selection.

Key findings emphasize these factors' significant roles in enhancing gas saturation prediction accuracy for CO₂ sequestration projects utilizing machine learning models, where comprehension of uncertainty sources and confidence assessment remain critical needs.

INTRODUCTION

Uncertainty poses a fundamental challenge in geosciences, especially when evaluating subsurface conditions with limited data. This uncertainty stems from three interrelated issues: (1) the inherent spatial sparseness of subsurface data acquisition; (2) the high degree of subsurface heterogeneity; and (3) intrinsic variations across data collection methods. Two techniques—petrophysical evaluation and 3D geological modeling—play an integral role in energy resource characterization. However, both approaches harbor substantial uncertainties (Asquit & Krywoski, 2004; Yarin & Ghahramani, 2016; Kausik et al., 2021). The severity of such uncertainties grows from the confluence of low spatial data density, input data quality issues, and the scale of subsurface characterization. Recent advances in machine learning offer new pathways to enhance prediction accuracy and refine control points for 3D models, particularly for assessing in-situ water saturation (Adebayo et al., 2015; Wu et al., 2018).

Understanding the range of in-situ water saturation S_w values is pivotal for project developers, enabling them to evaluate potential risks such as leakage and caprock integrity at storage sites (Ajayi et al., 2019). A distribution of uncertainty in water saturation and 3D modeling aids in assessing the safe and secure storage capacity of CO₂ in the subsurface. The 3D modeling of S_w becomes instrumental in ensuring the long-term integrity of storage sites and the containment of injected CO₂ (Bachu, 2000; Goodman et al., 2011; Orlic et al., 2011).

Recent studies underscore the significance of geological modeling in carbon storage site selection (Osundinde et al., 2023). Zaluski (2018) provided a comprehensive 3D depositional environment analysis, including porosity and permeability estimations, while Clochard et al. (2018) performed quadri-joint inversion of multi-component data. Despite challenges in linking fluid responses to seismic attributes, geostatistical 3D modeling is a reliable tool for estimating S_w

distribution (Prycz and Deutsch, 2014; Wade and Lee, 2019). Neff et al. (2021) utilized triple combo wireline logs, complemented by random forest with sonic logs, for water saturation predictions, benefitting from extensive training data. Prado et al. (2019) highlighted machine learning uncertainties and strategies to reduce them, emphasizing the importance of interpretable models for random forest and neural networks (Molnar, 2022).

Our approach involves a comparative accuracy analysis between two machine learning methods for predicting in-situ water saturation (S_w): random forest as a representative shallow learning algorithm versus recurrent neural networks exemplifying deep learning. As a precursor, we improve the quality of input wireline logs and petrophysical evaluations. We then present feature importance values, accuracy plots, and error measurements for both shallow and deep models. We also employ LIME to explain individual model predictions, revealing new insights into using recurrent neural networks for reservoir characterization. We quantify uncertainty through three complementary lenses: 1) algorithm selection uncertainty from the random forest and RNN comparison; 2) data quantity versus quality uncertainty from our data enhancement process; and 3) intrinsic S_w uncertainty via a variational Monte Carlo simulation of S_w (Rączkowska et al., 2019; Acerbi, 2020).

DATASET AND BACKGROUND GEOLOGY

The Big Sky Carbon Sequestration Partnership (BSCSP) aimed to extract naturally occurring CO_2 at the Kevin Dome, Montana, from the Danielson-22-17 borehole (Adeniyi et al., 2023). This captured CO_2 was injected down dip at the Wallawein-22-1 borehole (Damiani et al., 2018). In 2014, the Danielson-22-17 borehole extracted 1 million metric tons of CO_2 from the Duperow Formation, but the injected volume was not recovered, leading to the detection of fresh water and a project interruption. Situated on the sweetgrass arch in north-central Montana, the

Kevin Dome spans 750 mi² and has a rich history of oil and gas production over a century (Hennip, 1973). The Sweetgrass Arch's structural relief, dating back to the Precambrian era, was further shaped by mountain-building forces in the late cretaceous-early tertiary (Lorenz, 1982). The Duperow Formation, a naturally fractured carbonate formation within the Kevin Dome, consists of dolostone mixed with limestone intervals and includes variable-thickness anhydrite layers. Impermeable carbonates, shales, and anhydrite rocks serve as seal rocks.

As a mature basin transitioning to CO₂ storage, over 1,500 boreholes have been drilled in the area. This study categorizes information into two tiers: Tier 1 with 212 boreholes featuring triple wireline logs, and Tier 2 with triple combo data and expert petrophysical interpretation (35 boreholes). The decision of separating the dataset in two tiers obeys the following logic, borehole legacy data have limited wireline data and to estimate water saturation there must be resistivity, porosity and lithology indication, hence Tier 1 collects boreholes in the area that have at least the following logs: gamma-ray, neutron porosity, bulk density, shallow and deep resistivity. On the other hand, while Archie's equation (Archie, 1952) to estimate S_w is a regression formula from an experiment, on carbonate rocks the S_w estimation is more complex, therefore, Tier 2 includes subject-matter expert information in addition to the Tier 1 information i.e. Anhydrite, calcite, clay, dolomite, and Quartz volume, total and effective porosity. The petrophysical evaluation used to test and train S_w algorithms is taken from Zaluski (2018). Figure 5.1 illustrates the distribution of borehole data and seismic information, while Table 5.1 summarizes wireline information in each tier. The water saturation prediction at the Danielson-22-17 borehole (producer) is the primary

prediction target, with the Wallawein-22-1 borehole (injector) falling within Tier 2.

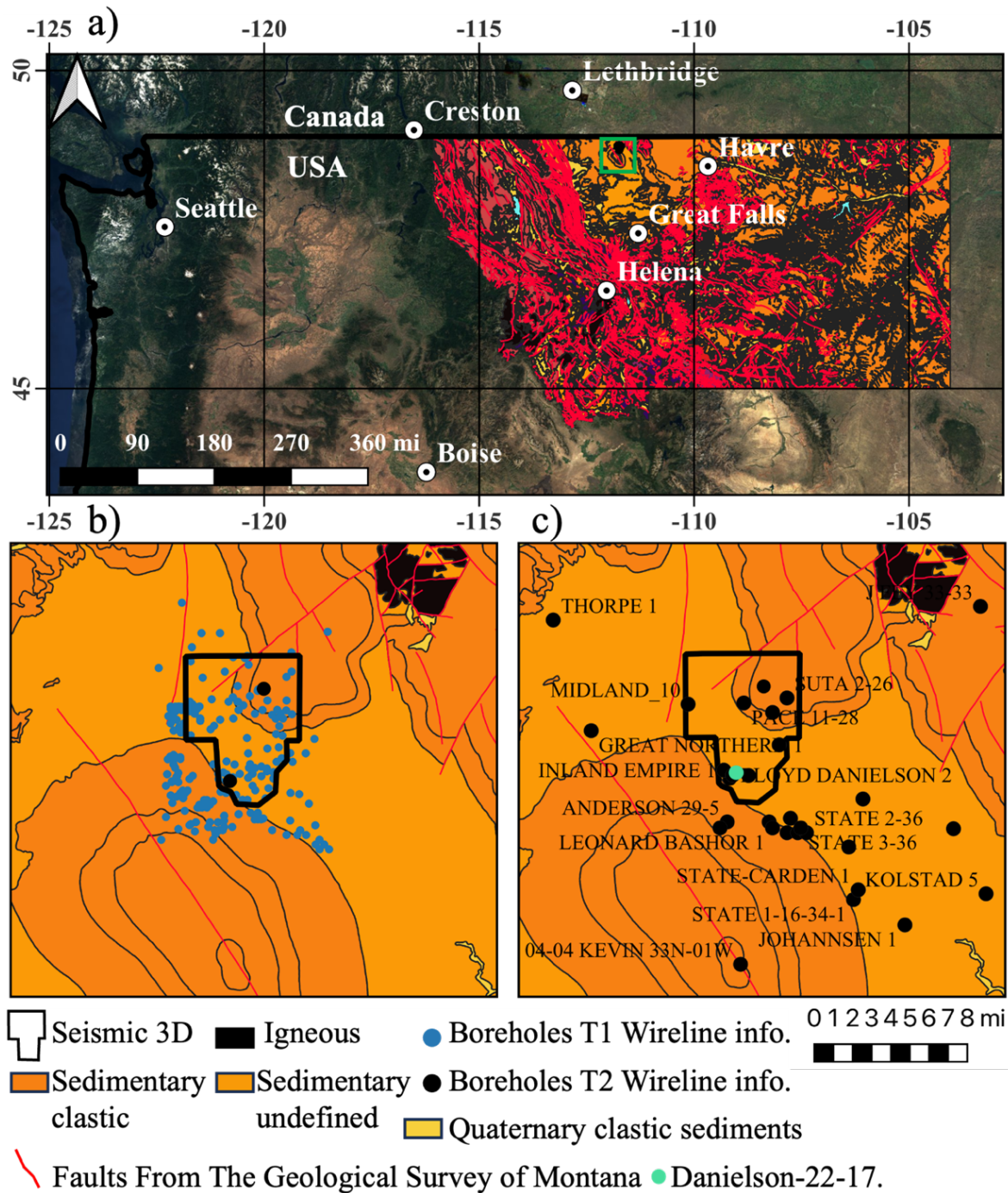


Figure 5.1. Location map and spatial distribution of available data. Panel a) Montana geological map and the documented faults from the Geological survey of Montana. Panel b) Borehole distribution around the CO₂ sequestration project, 212 boreholes with at least triple combo wireline logs. Panel c) Spatial distribution of 35 boreholes with enough information to estimate petrophysical properties (Total and effective porosity, vertical permeability and, water saturation), including water saturation, mineralogy volumes (Calcite, Shale, Dolomite, Anhydrite).

The reservoir intervals in the Duperow Formation are below seismic resolution, attempts to correlate seismic amplitude and seismic attributes will add extra uncertainty. While seismic data conditioning benefits seismic reservoir characterization we constrain its use to grid construction. Figure 5.2 shows an arbitrary seismic line connecting key boreholes and a depth-slice around 250 ft below sea level highlights the presence of faults and other smaller discontinuities. In Figure 5.2 we can see two reservoir zones highlighted by magenta (Bakken to Potlatch Formation zone) and green (Nisuk to Upper Duperow formation zone) circles. Seismic in the area is used to construct the 3D grid and Figure 2 shows the key interpreted horizons on the depth domain seismic with interpreted faults. It is clear both reservoir zones are seismically undetectable, and the seismic data is used to interpret structural and stratigraphical elements (Chopra and Marfurt, 2007; Barnes, 2016).

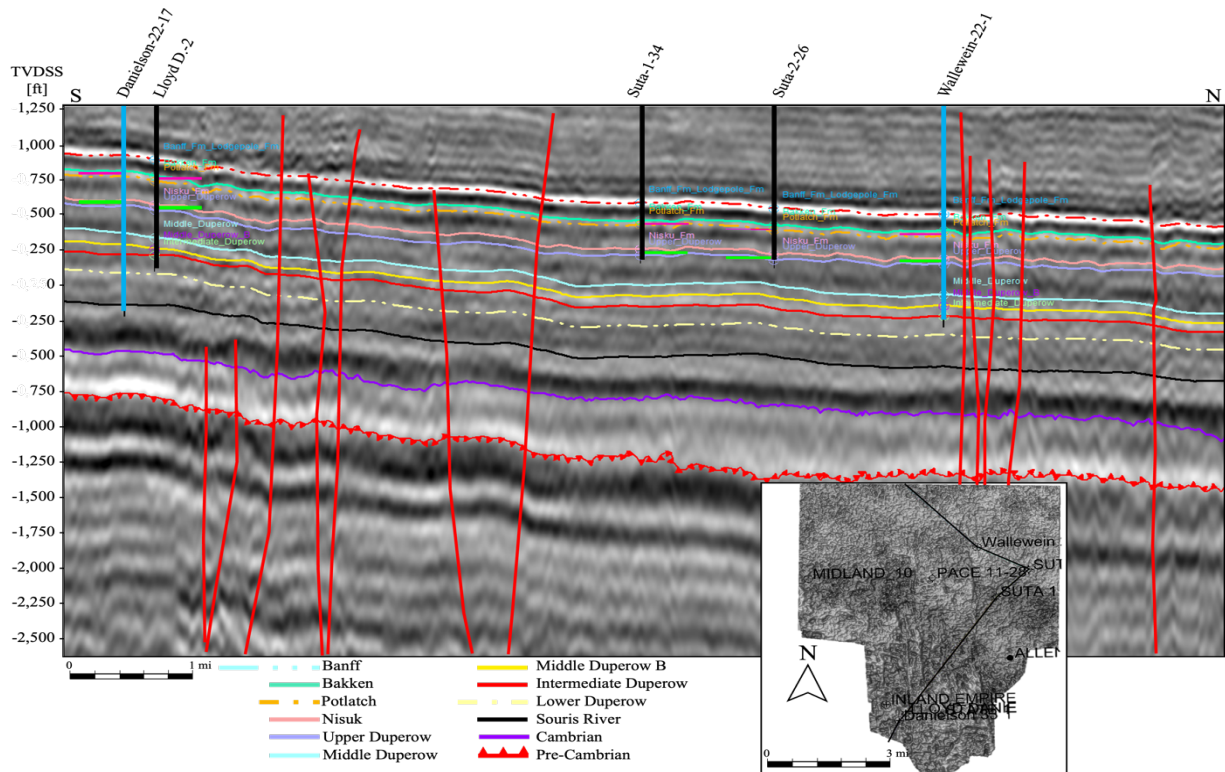


Figure 5.2. Arbitrary seismic section on a depth domain pre-stack migration including seismic interpretation of geological formations and faults. The high-dip angle faults display the structural complexity in the area and the risk associated with fluid injections. Two reservoir flags (green and red) show the impossibility in associate injection intervals to seismic amplitude.

METHODS

A crucial element in this research is the water saturation quantification (S_w). It is crucial to know the in-situ water saturation since pore space is filled with a combination of fluids in which water is present at all times, either as free or irreducible water (Tiab and Donaldson, 2015). Then the remaining fluids are the complement of the in-situ water, for example, oil saturation is $S_o = (1 - S_w)$. The calculation of water saturation we choose is dual water model because in carbonate rocks, the pore structure and fluid behavior can be quite complex due to factors such as vuggy porosity, fractures, and varying degrees of diagenesis. Archie's equation on the other hand is a widely used empirical relationship in petrophysics that relates water saturation (S_w) to the resistivity of the formation water (R_w), the porosity of the rock, and the resistivity of the uninvaded zone (R_t). Archie's equation (Archie, 1952) presented in 5.1 has three empirical coefficients, which contribute to the ontological uncertainties:

$$S_w^n = \frac{a R_w}{(\phi^m R_t)} \text{ Eq 5.1}$$

Where a is the tortuosity factor constant (often has a value of 1 unitless), n (the saturation exponent varying from 1.8 to 4 unitless, but normally 2 unitless) and m (cementation factor varying from 1.7 to 3 unitless, but normally 3 unitless). Dual water models for water saturation estimation complements the Waxman-Smiths-Thomas (WST) delving into the different porosity systems in the rock. The DW equation written in bound-water term and conductivity is presented in eq. 5.2:

$$\frac{1}{C_t} = \frac{1}{a} \phi_t^m S_{wt}^n \left[\left(\frac{S_{wt} - S_{wb}}{S_{wt}} \right) C_w + C_{bw} \frac{S_{wb}}{S_{wt}} \right]$$

Where:

$C_{bw} = \frac{\beta}{\alpha V_Q^H}$ and $S_{wb} = \frac{XBWA}{\phi_t} = \alpha V_Q^H Q v$ are the bound water conductivity (C_{bw}) and bound water saturation (S_{wb}). $\phi_t = \phi_e / (1 - S_{wb})$ is the total porosity and effective porosity. The V_Q^H is the clay water volume factor.

In carbonate rocks, Archie's equation may be applicable if the formation is relatively homogeneous and if the rock and fluid properties are well understood. However, in our zone of study the carbonate reservoirs are intensely fractured. Therefore, Dual water methods involve using multiple types of measurements, such as resistivity and neutron porosity, to estimate water saturation. These methods can be more robust in carbonate rocks because they provide complementary information about the formation properties. For example, resistivity measurements can be sensitive to the presence of connate water and hydrocarbons, while neutron porosity measurements are sensitive to the total porosity, including both water-filled and hydrocarbon-filled pores.

Our unique dual approach for predicting water saturation (S_w) while tackling uncertainty centers on employing both LIME and variational Bayesian Monte Carlo methods. We preprocess the input wireline log data by removing gaps and normalizing features following established best practices (McDonald, 2021). The training data consists of borehole petrophysical measurements from the Kevin Dome static characterization report from Zaluski (2018) which distribution is shown in Figure 1c. Since the S_w data exhibits data class imbalance, we address this statistical disproportion through data augmentation to improve model robustness. We train both a recurrent neural network and random forest regressor on this dataset. Based on superior accuracy, we select the random forest model trained on the Tier 2 data for subsequent analysis (see Supplementary Information 1 for data preprocessing and feature importance details). To interpret model

predictions, we conduct accuracy assessments alongside LIME analysis methods per guidance from Molnar (2022).

Seismic data aids only in constructing the 3D grid incorporating geological formations and large-scale faults. Sequential Gaussian simulation, maintaining constant variograms across zones, adjusts up-scaled water saturation inputs from the Tier 2 thirty-five boreholes. Model comparisons utilize spatial correlation, highlighting 3D differences and standard deviations as measures of uncertainty. Finally, we use the CNN S_w prediction as truth for a Variational Bayesian Monte Carlo exploration of prediction uncertainty. This streamlined methodology integrates machine learning, geological data, and statistical analyses to provide a comprehensive understanding of S_w prediction and uncertainty. Figure 5.3 summarizes the workflow and data integration points.

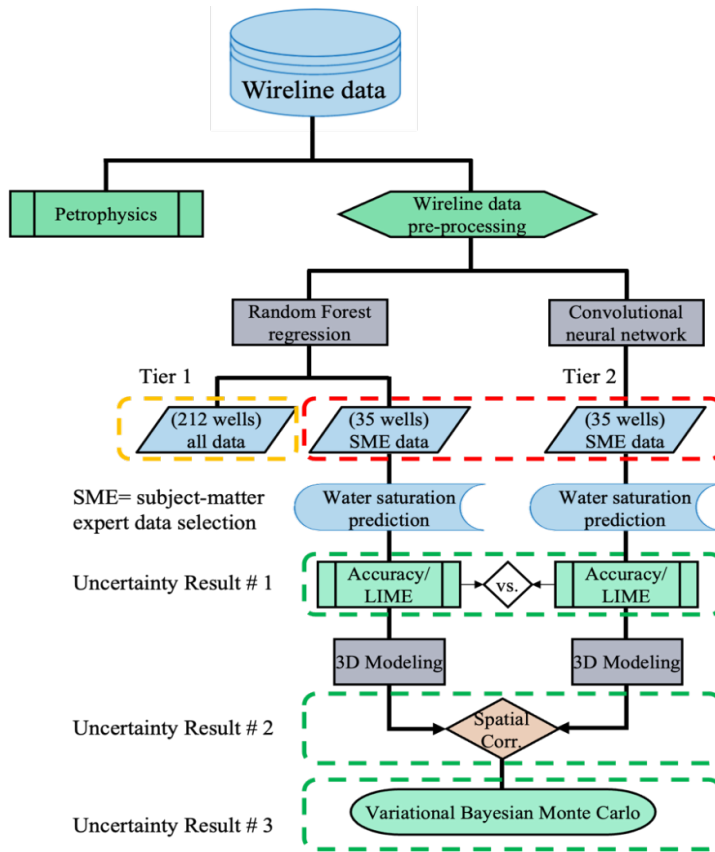


Figure 5.3. Our proposed Methodology, to include explainable machine learning in understanding prediction models. While Random Forest have plots to interpret the prediction and we keep a blind test borehole to see how the model performs with a blind borehole. CNN only evaluates the accuracy. However, it is important not to trust only the accuracy or loss function of the CNN prediction. Wireline data is pre-processed analyzed for missing data, augmented, and normalized. We consider subject-matter expert data, the inclusion of Sonic travel time, calcite volume, anhydrite volume and shale volume, total and effective porosity, in addition to gamma-ray, shallow and deep resistivity, density and neutron density.

The three pillars of our methodology are: 1) water saturation prediction using random forest and recurrent neural networks. 2) Local interpretable model-agnostic explanations and 3) Variational Bayesian Monte Carlo (VBMC). Herein is a brief description of our three methods, a detailed description on the methodology to estimate water saturation on carbonate rocks can be found in detail at Asquith and Krygowski, 2004; Alimoradi et al., 2011. For more detail about 3D reservoir modeling visit Deutsch, 2006; Pyrcz, and Deutsch, 2014.

Random forest, an extension of Classification and Regression Trees (CART) models, is an ensemble learning method. It involves constructing multiple decision trees through recursive partitioning of the input space. Random forest enhances diversity by randomly selecting subsets of input variables and data cases at each node of the tree. This randomization process contributes to the decorrelation of base learners, resulting in a robust and accurate predictive model. We split the processed data 80 % for train and 20% test, a maximum depth of the random forest of 10 units of depth and used 500 estimators. Grid search and cross-validation are used to tune the hyperparameter space.

Recurrent neural network is a neural network which maps from an input space of sequences to an output space of sequences in a stateful way. That is, the prediction of output y_t depends not only on the input x_t , but also on the hidden state of the system, h_t , which gets updated over time, as the sequence is processed. We use the Adam optimizer, rectified linear unit (ReLU) as an activation function and mean squared error over 1,000 epochs. Loss functions and a cross-plot of the predicted water saturation and the interpreted water saturation is presented as an accuracy test. The RNN model is used to make the prediction of water saturation on the Danielson-22-17 borehole.

Local Interpretable Model-Agnostic Explanations is an explainable AI method for complex machine learning models. LIME probes a model by introducing small perturbations to the input data and assessing how predictions change. For a specific data instance, LIME generates a new dataset, in our case a wireline dataset, by sampling perturbed variants of the input features alongside the corresponding model predictions. It then trains an interpretable surrogate model on this dataset, weighting the instances based on proximity to the original data point. Thus, the interpretable model provides a locally faithful approximation of the complex model's behavior. However, while LIME excels at local explanation, the simplified model may not furnish an optimal global representation.

Variational Bayesian Monte Carlo combines variational inference and Monte Carlo sampling to approximate full Bayesian analysis, aiming to characterize the posterior probability distribution of model parameters. This posterior captures uncertainties in both the estimates of model parameters (θ) and the resulting predictions. VBMC utilizes a flexible Gaussian mixture model to approximate the posterior distribution of parameters related to water saturation. In addition, VBMC incorporates prior distributions over parameters to represent inherent uncertainties in their true values. Variational inference identifies the most likely parameter sets given the data and specified priors, while Monte Carlo sampling generates multiple samples from these posterior distributions. These samples are then propagated through the water saturation prediction model, illustrating how parameter uncertainty influences predictions. The resulting distribution of water saturation predictions encapsulates overall prediction uncertainty.

Carbonate reservoir rocks in the area are highly fractured, the petrophysical evaluation of the most recent boreholes are presented in Figure 5.4 showing the petrophysical evaluation and rose diagrams of fractures interpreted from the FMI logs. In the plot, we show the S_w from

petrophysical evaluation which we considered ground truth and name it $S_{w\text{interpreted}}$. From the petrophysical inputs, the highest correlation is between $S_{w\text{interpreted}}$ and anhydrite volume. Although porosity and permeability coming from the magnetic resonance tool (NRM) and calibrated with core data, theoretically, will have a connection with S_w . Figure 5.4 suggest that the carbonate matrix hinders the correlation.

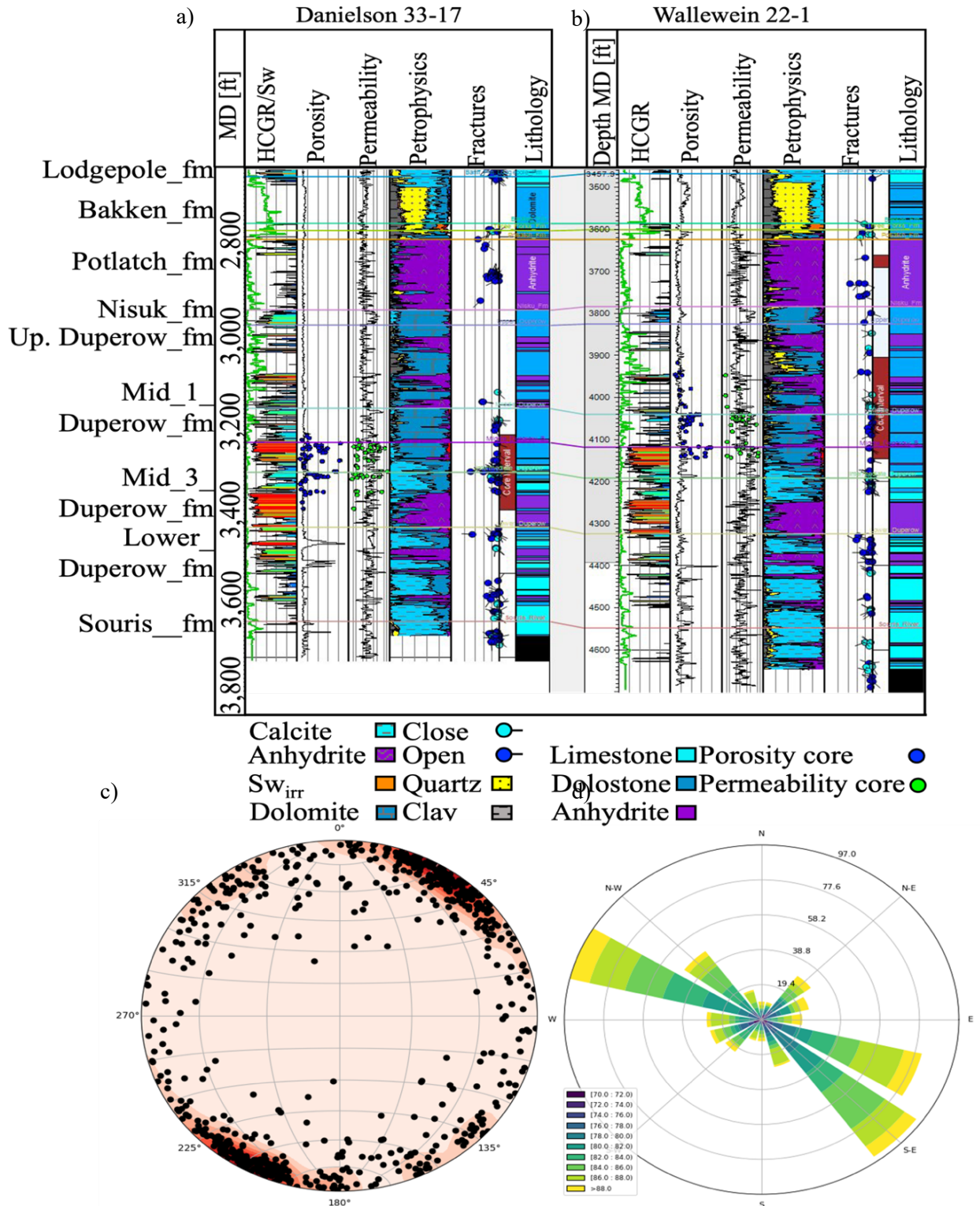


Figure 5.4. Panel a) Danielson-22-17 and Panel b) Wallawein-22-1 boreholes with petrophysical evaluation. The evaluation includes lithology and fluid effects. The template includes gamma-ray and water saturation, porosity, and permeability. Volume of Illite, anhydrite, calcite, dolomite, quartz, and irreducible water. The tadpoles coming from the FMI interpretation helps generate the stereonet and analyze fracture azimuth (SW-NE 40 °) presented in panel c) and panel d).

RESULTS

Our presentation of the results includes the exploration of S_w prediction using random forest on Tier 1 and Tier 2, and the accuracy differences. Followed by the comparison of random forest prediction of S_w on Tier 2 with the RNN method. We present LIME analysis as our first important result. LIME is done on random forest prediction of S_w on Tier 2 and the RNN prediction of S_w on Tier 2. We highlight the LIME explanation as a qualitative interpretation of uncertainty. Our second important result is a more quantitative analysis of the uncertainty and is the 3D sequential Gaussian simulation of the S_w predicted using (1) random forest and (2) RNN, accompanied by its spatial difference and the standard deviation of the differences. Finally, we present the VBMC as a novel approach to quantify the uncertainty in S_w prediction.

Our analysis of uncertainty starts with the method selection in estimating water saturation using machine learning. In Figure 5.5 we present the prediction of S_w at Danielson-22-17 borehole using Tier 2 dataset. Figure 5.5a shows the prediction of S_w using RF ($S_{w_{RF}}$ green line) and $S_{w_{interpreted}}$, black boxes indicate the intervals of interest (Bakken to Potlatch Formation zone) and (Nisuk to Upper Duperow formation zone). Figure 5.5b presents the S_w using RNN ($S_{w_{RNN}}$ black line) and $S_{w_{interpreted}}$ cyan boxes are equivalent to zone 1 and zone 2 of interest. While $S_{w_{RF}}$ and $S_{w_{RNN}}$ follow the in interest the S_w , $S_{w_{RF}}$ has a higher frequency and is less accurate. On the other hand, $S_{w_{RNN}}$ is a smother curve due to the convolutional layers follows the $S_{w_{interpreted}}$ more accurately than $S_{w_{RF}}$.

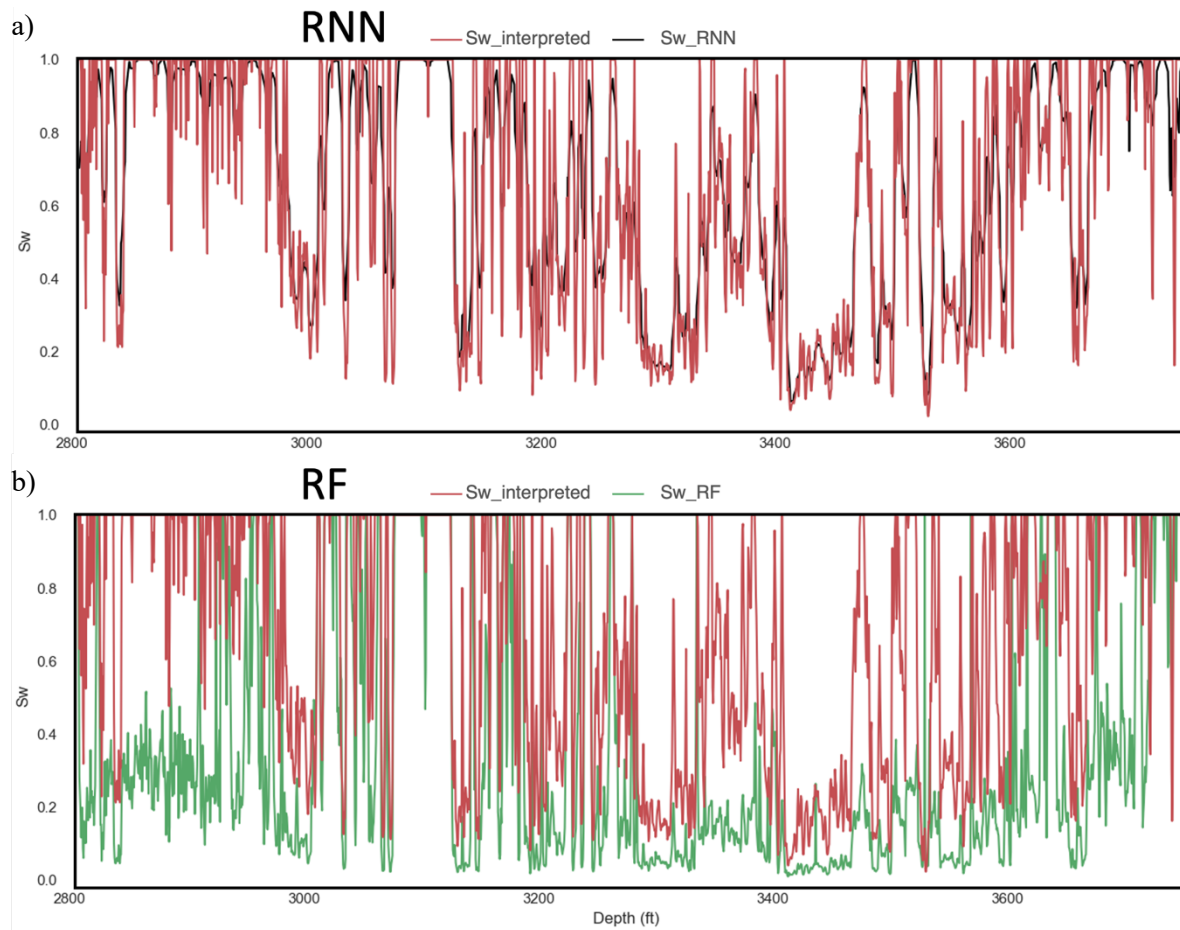


Figure 5.5. Sw predicted vs Sw interpreted using RF (Panel a) and RNN (Panel b).

Table 5.1 presents a summary of regression accuracies on the Danielson-22-17 borehole. We selected R-squared (R^2) as a regression measurement. The random forest regression using Tier 1 data provides an unambiguous score (0.51) and RNN provides the best accuracy.

Simulation Ranking	Algorithm	TIER	R^2
1	RNN	2	0.931
2	Random forest	2	0.716
3	Random forest	1	0.51

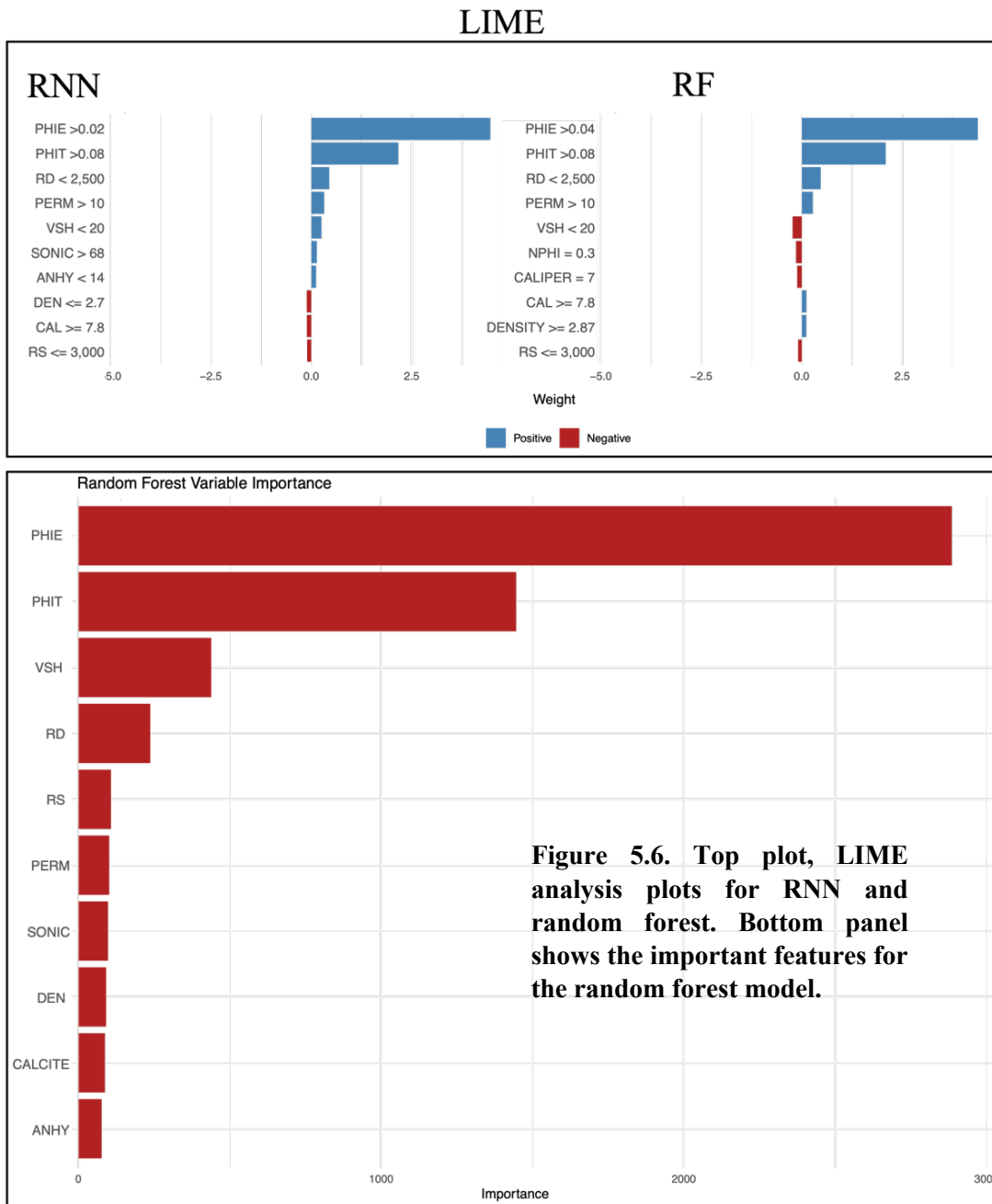
Table 5.1. Accuracy comparisons between the different quality/quantity datasets (Tier 1 represents more borehole data and less and raw wireline logs, Tier 2 represents less borehole data but more and better wireline logs). Simulation ranking # 3, random forest on Tier 1 has an ambiguous R^2 , since it is a statistical measure of the goodness of fit of a regression model, having a 0.5 is deceiving.

We use LIME to understand the contribution of features used to predict either with random forest or RNN. Figure 5.6 shows the model explanation for the prediction analysis and the feature

importance for the random forest prediction. It is worthwhile to highlight that CART methods and other regression algorithms have tools to interpret, yet the explanation and the interpretation might be different. The differences arise in LIME generating a new dataset of perturbations around the predicted water saturation, these perturbations are weighted as a function of its distance to the original model. This perturbation works either for random forest or RNN. Then, the trained machine learning model is used to predict the class of each instance in the new generated dataset. Finally, a simpler model easily interpretable e.g. a linear regression model, is fitted and used to explain the prediction of the classifier. In Figure 5.6 we can see several bar charts. The y-axis shows the features while the x-axis shows the relative contribution of each feature. Positive value (blue color) shows that the feature support or increase the value of the prediction, while the negative value (red color) has negative effect or decrease the prediction value.

Each observations have different explanation, for example effective and total porosity has positive effect toward the final predicted SW. The weight of the PHIE and PHIT increases when the values are larger than ($>$) 2% and 4 % in the RNN and 4 % and 8 % for the RF. On the other hand, bulk density has a negative effect toward the final prediction. This interpretation is that rocks with densities smaller than 2.7 g/cc have a negative effect towards the interpretation of Sw, in agreement with density values for sandstones or shales.

As we can see, water saturation performance for the total and effective porosity strongly affects the final score in both algorithms, followed by deep resistivity and permeability. The shale content has a positive effect on RNN and negative on random forest. Sonic travel time and neutron porosity (NPHI) have a positive contribution for RNN but for random forest, caliper and bulk density have a positive contribution to water saturation prediction.



Using the best two simulations (1 & 2) we ran two sequential Gaussian simulation (SGS) to show how the uncertainty escalates in 3D. This result serves as our second uncertainty estimation, using a conventional tool we show the differences between two machine learning models in a paramount petrophysical property. The grid has 300 ft by 300 ft horizontal size in eight zones of interest from Potlatch (top) to Cambrian (bottom). We focus in preserving the vertical

variability and selected a grid size of 6 ft. The first SGS had the Sw_RNN as input on the Danielson-22-17 borehole, the second SGS had the Sw_RF as input. Both times the variograms and hyper-parameters were preserved. The difference is the normalization of the data since both have different means and standard deviations. In 3D the differences expand and there is a 12 to -12 % difference between the Sw in model 1 and model 2. Figure 5.7 summarizes the differences in a section of interest, between Danielson-22-17 and Wallawein-22-1.

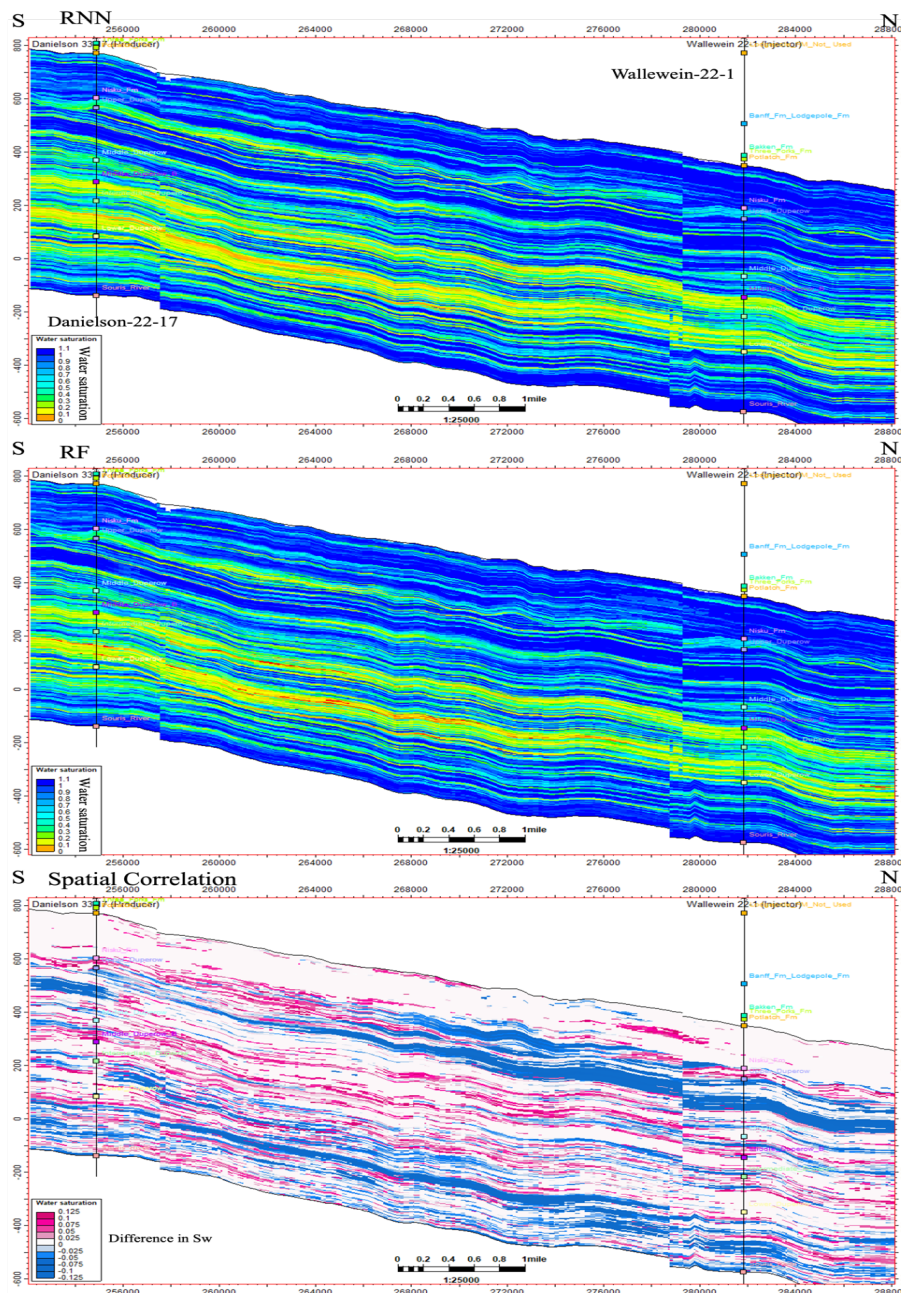


Figure 5.7. Geostatistical models in 3D using Sw_RNN and Sw_RF in top and middle panel. Panel a) shows the water saturation estimated using RNN 3D grid. Panel b) shows the water saturation estimated using random forest 3D grid, and Panel c) shows the spatial correlation performed at each grid cell, where the value of the cell for both models is correlated. The difference can be considered as a measure of uncertainty. Observe how inside the two zones of interest (black arrows) Bakken to Potlatch Formation and Nisuk to Upper Duperow Formation the differences are larger, casting doubt on the model.

DISCUSSION

Borehole data analysis is a crucial area of research where machine learning algorithms help interpreters to derive relationships or equations from data. Perhaps the most famous regression is the Archie's equation (1952) to estimate water saturation in sandstones with different porosities from the Gulf of Mexico. Nonetheless, having shown the level of uncertainty based on the algorithm selection a crucial petrophysical parameter such as S_w should be closely monitored. Although in the lights of CO_2 injection where cost needs to be considered machine learning models can be an initial tool to evaluate areas of opportunity with an understanding of the range of uncertainties. RNN algorithms are computationally more expensive, yet yields more accurate results than random forest, an observation Zhang et al., (2022) and McDonald (2021) agree. Our study categorized the quality of the data and included relevant wireline logs, a lesson learned from Neff et al., 2022. Neff et al., (2022) perform a water saturation prediction using 500 boreholes, with a reported an accuracy of 0.73, and our classification using 35 boreholes is similar in accuracy (0.71). The S_w _RNN prediction reported an R^2 above 0.9, Zhang in siliciclastics reservoirs observed similar values, we suspect the RNN is overfitted making the architecture and its weights useless for other lithologies, nonetheless our code and weights are open to test.

Overall, we find that LIME is a valuable tool to explain the feature contribution to the water prediction, including random forest and RNN. It is important to note the distinction between interpretable and explainable machine learning techniques. While the primary features identified as significant remain consistent across both random forest and RNN models, there are nuanced differences in their relative importance as revealed by the LIME plots. Notably, the contribution of shale volume emerged prominently in the importance plot, whereas its significance appeared relatively diminished in the S_w _RNN LIME plots, albeit accompanied by a quantifiable numerical

value indicating an improvement in Sw prediction with clay content below 20%. Furthermore, we emphasize the importance of integrating explainable AI as an intrinsic component of machine learning methodologies.

While it is uniquely to have total and effective porosity from a petrophysical evaluation and not have Water saturation, our study serves as a proof of concept, since we did not had control on the petrophysical evaluation, and it came from a commercial report submitted to the Department of Energy. Removing the petrophysical outputs, we observe that the most important features and the features that explain the most the Sw prediction are shallow and deep resistivity, sonic travel time, density, and neutron porosity. Caliper had a more significant role in the Sw_RF prediction and in either of the test gamma ray had a relevant contribution. It is likely that the carbonate matrix affected the contribution of the gamma ray log. Including the petrophysical outputs in both algorithms effective and total porosity had the largest contribution, with different values. Most likely the water saturation algorithm used to calculate the Sw interpreted includes a porosity, permeability, and a shallow/deep resistivity combination (Juhasz, 1979).

As a final measure of uncertainty, VBMC, depicted in Figure 5.8, provides insight into the variability of predictions and potential values of Sw. This figure illustrates the significant uncertainty spanning from theoretical (Target density) to predicted Sw (Stable iteration). Consistent with our findings, Grana et al. (2021) conducted uncertainty analysis on shale volume, porosity, and water saturation, corroborating the observed uncertainty. Similarly, Neff et al. (2022) show the uncertainty associated with the Sw prediction. A cumulative uncertainty in estimating a Sw should be considered when making decision and economical evaluation.

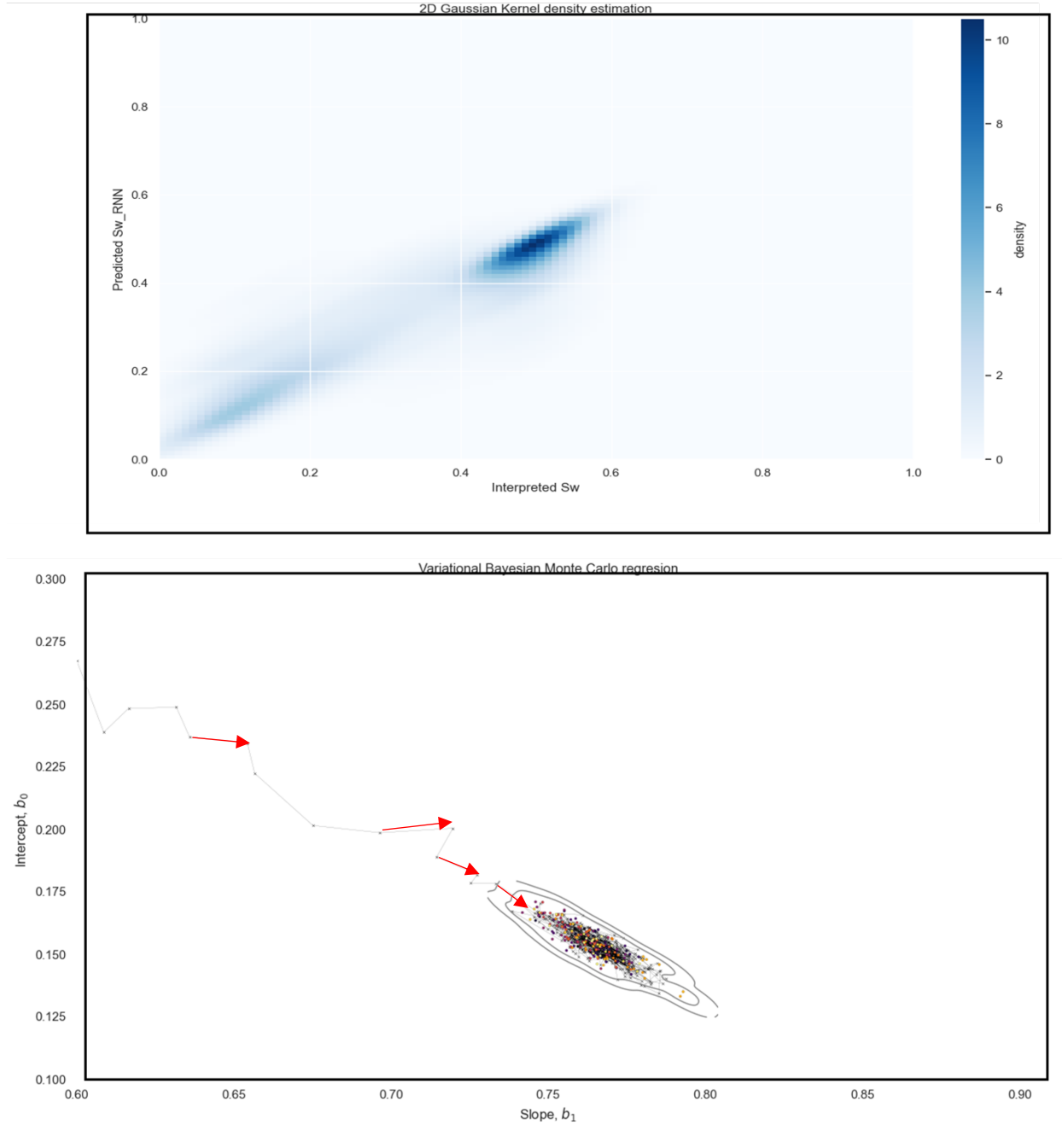


Figure 5. 8. Uncertainty plot of gas saturation. Panel a) shows the initial kernel density estimate. After assuming a prior model, the slope and intercept model parameters to the most likely linear regression, based only on the L2 norm minima of the error vector over all sample data. Noise is noise added to the synthetic data. Red and cyan arrow show the Bayesian uncertainty.

CONCLUSION

LIME provides means to evaluate neural networks since it is a model-agnostic tool. Water saturation is important for everything. Using subject matter expert data improves the classification.

In conclusion, our study addresses the critical challenge of uncertainty in predicting in-situ water saturation (S_w) in geoscience applications, particularly in the context of CO₂ injection and sequestration projects e.g. geothermal. By employing machine learning models such as random forest and recurrent neural networks, we have demonstrated the potential to enhance accuracy and automate the evaluation of petrophysical properties in mature fields. However, the inherent uncertainties in subsurface conditions, coupled with challenges in understanding model predictions, highlight the necessity of comprehending contributing factors to uncertainty and evaluating confidence in model parameters. We reckon recurrent neural networks have a better performance leveraged by the selection of relevant data.

Our analysis, supported by advanced techniques including Local Interpretable Model-Agnostic Explanations, variational Bayesian Monte Carlo simulations, and 3D geostatistical modeling, reveals several key insights. First, the quality of input data significantly impacts prediction accuracy, emphasizing the importance of data quality over quantity. Second, our study underscores the sensitivity of prediction outcomes to algorithm selection, with RNNs outperforming random forest models in accuracy albeit with higher computational costs, presenting challenges to its application of using seismic data to predict seismic scale reservoir saturations. Additionally, the interpretability provided by LIME sheds light on the varying importance of features across different models, highlighting the need for explainable AI techniques in enhancing model transparency and trustworthiness.

Furthermore, our findings underscore the importance of considering uncertainty in decision-making processes and economic evaluations associated with CO₂ storage and other reservoir storage projects. The variability in S_w predictions, as illustrated by our analysis and supported by previous studies, emphasizes the need for a comprehensive understanding of uncertainty and its implications.

Overall, our study contributes to the ongoing efforts in leveraging machine learning for improved subsurface characterization and highlights the importance of integrating interpretability and uncertainty quantification into model development and decision-making processes in geoscience applications.

ACKNOWLEDGMENTS

We thank the Department of Energy for providing the survey Kevin Dome 3D and the Big Sky Carbon Sequestration Partnership. Borehole data and the associated reports required for this study were downloaded thanks to Enverus. AASPI and SLB are greatly acknowledged for providing the software to the University of Oklahoma.

DATA AND MATERIAL AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

Advancing machine learning in geophysics requires careful validation of algorithm selection and data input. While geophysical and geological methods offer validation using geological analogs or well-established physical principles, machine learning (ML) algorithms are often perceived as "black box" methods. Therefore, creating and designing synthetic labels is crucial to ground-truth ML algorithms.

ML techniques have gained significant attention in geophysics due to their potential to explain data and work with large datasets for analysis, interpretation, and modeling. However, synthetic models are yet to catch up with the research on ML in seismic data. This thesis delves into the integration of ML algorithms in geophysical research to provide careful validation, ensuring reliable results. Unlike traditional geophysical and geological methods, ML algorithms lack the well-established physical principles, making validation a crucial step.

The versatility of generative adversarial networks (GANs) in geophysical modeling, as demonstrated in Chapter 2, showcases their ability to generate synthetic data with high fidelity for subsurface velocity modeling. By exploiting the capabilities of GANs, this work generated a pixel-pixel model of a reservoir image based on a geostatistical seismic inversion, testing its transferability across diverse reefs around the Ray reef. While reservoir modeling methodologies are well-established, the use of GANs presents an academic opportunity to study reservoir distributions with limited data, overcoming the challenges of collecting and integrating subsurface data.

Chapter 3 highlights the importance of synthetic model generation in ML validation of rock properties derived from seismic. The analysis demonstrates how the mechanisms, noise selection, and property values employed during synthetic model generation are as critical as the ML models

themselves. By simulating various geological scenarios, researchers can systematically evaluate the robustness and generalization capabilities of ML algorithms.

The quality of seismic attributes and inversion outcomes is contingent upon image quality and data quantity, as discussed in Chapters 2 and 3. High-resolution seismic images facilitate the extraction of detailed geological features, while inadequate data coverage may compromise inversion results. Addressing these challenges can enhance the reliability and interpretability of ML-based seismic analyses.

Chapter 4 addresses three limitations of machine learning clustering algorithms in seismic facies analysis: the variability of facies to fluid response, the vertical resolution of clustered data based on seismic attributes, and the need to calibrate the facies classifications from an unsupervised method with borehole data. While ML clustering algorithms offer valuable insights into seismic facies changes, they often struggle to detect subtle fluid effects within seismic data, highlighting the need for complementary approaches.

Chapter 5 demonstrates the impact of subject matter expert training on ML performance. Incorporating domain knowledge in ML model training significantly enhances classification and regression tasks in geosciences. The synergistic relationship between domain expertise and ML algorithms underscores the importance of expert-guided feature selection and model optimization. Random forest regression algorithms, in particular, offer robust performance in mitigating overfitting tendencies while making predictions, although their accuracy may lag behind deep learning models in certain geophysical applications.

In summary, the integration of ML techniques in geophysics requires meticulous validation to ensure robust and reliable results. Synthetic model generation, expert-guided training, and

algorithm selection are pivotal in enhancing the efficacy of ML applications in geophysical research. Despite the inherent challenges associated with ML validation, ongoing advancements in algorithm development and data-driven methodologies hold promise for addressing complex geophysical problems effectively. Nonetheless, regardless of the countless efforts of geoscientists in using machine learning with applications in seismic volumes for different objectives. Any product derived from a seismic cube is dependent on the quality of the image and machine learning efforts must be made on depth migrated cubes, mainly because depth migrated seismic cubes represent a truest position of seismic facies and faults and because the migration algorithm in the depth domain allows better image quality and resolution. In conclusion, by leveraging the synergies between domain expertise and computational capabilities, geophysicists can harness the full potential of machine learning for advancing the understanding and exploration of the Earth's subsurface.

APPENDIX

APPENDIX A: A short note on geostatistical inversion

The geostatistical seismic inversion in CGG Jason is based on Bayesian theory. It is a method developed by Total and CGG that combines geostatistical stochastic and a multivariate Bayesian approach by Buland and Omre (2003). This stochastic inversion process creates multiple possible solutions, which then can be averaged to create a probable solution. The deterministic inversion process produces one least-squares solution, which we try to make as likely as possible through the inputs. The geostatistical inversion in Jason method has three types of data that are used to create the result, these data are equally important: seismic amplitudes, well data and prior model parameters. The standard deviations of the prior model and the uncertainties of the seismic and well data will control where the importance of these elements is shifted i.e. small standard deviations in seismic combined with large standard deviation on the borehole shift the importance of the result towards the seismic data. If the result's match is increased to one type of data (e.g. if you weight that data source more), you will likely decrease the result's match to the other two types. These three data types can be viewed as probability distributions.

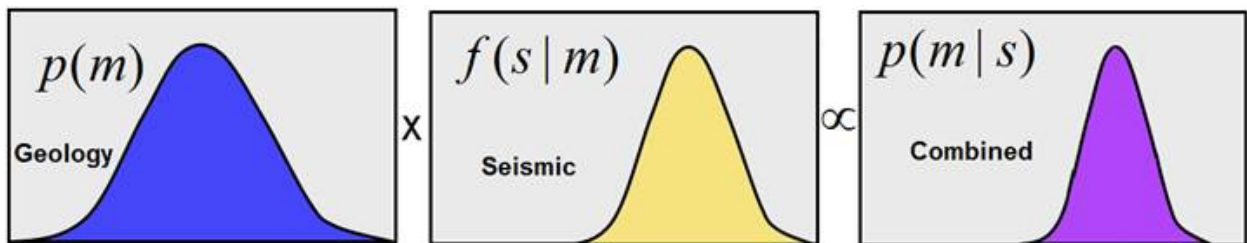


Figure A.1. The scheme of seismic amplitude distribution as a probability distribution combined with the geological parameters.

The workflow of geostatistical seismic inversion using in this thesis is summarized in these five steps:

1. Build a stratigraphic grid with horizons, well logs and layer-based kriging.
2. Well-tie and wavelet estimation of the partial angle stacks.
3. Compute the Bayesian posterior distribution by combining the model, the seismic data, and the well logs.
4. It creates multiple p-impedance and s-impedance realizations using Sequential Gaussian Simulation.
5. The program computes the mean and standard deviations from the multiple realizations.

In Figure A.2 we present the graphical summary in more detail for purposes of reproducibility.

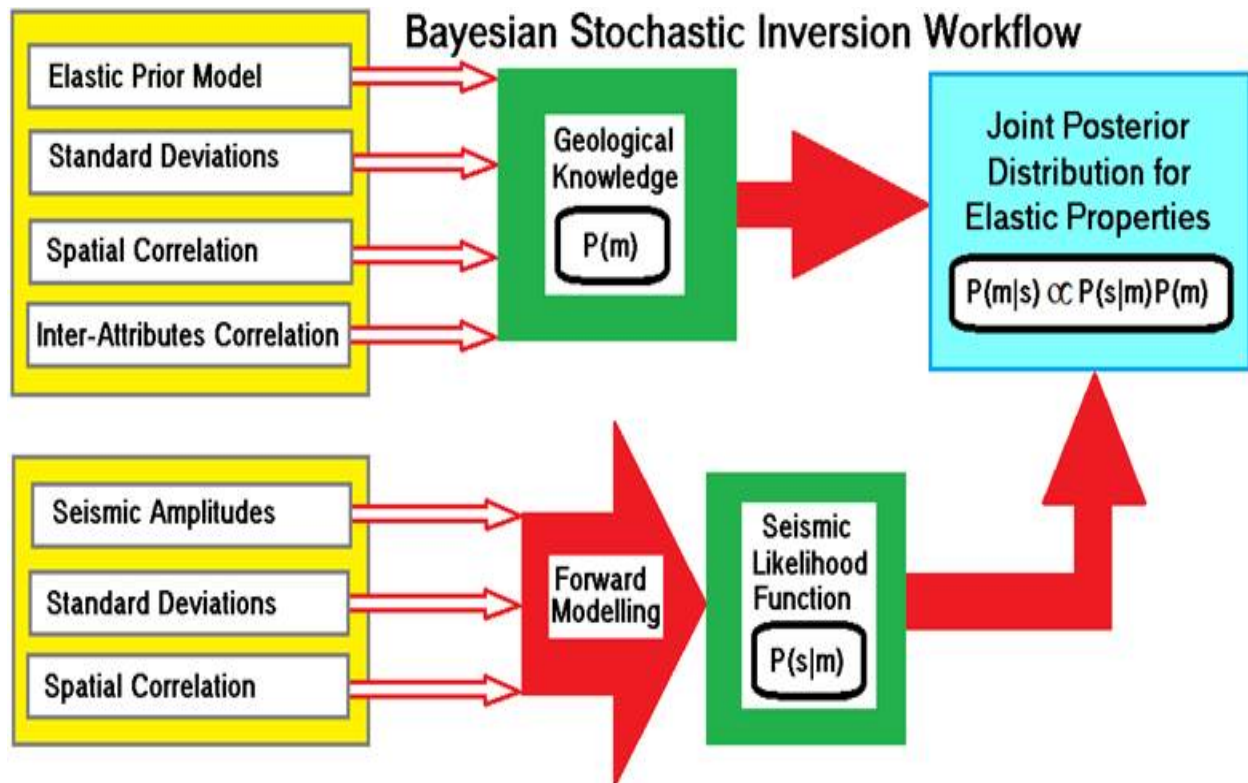


Figure A.2. Geostatistical seismic inversion workflow on Jason Geoscience workbench 10.2.

The importance of model building, and geosciences integration is shown in Figure A.3.

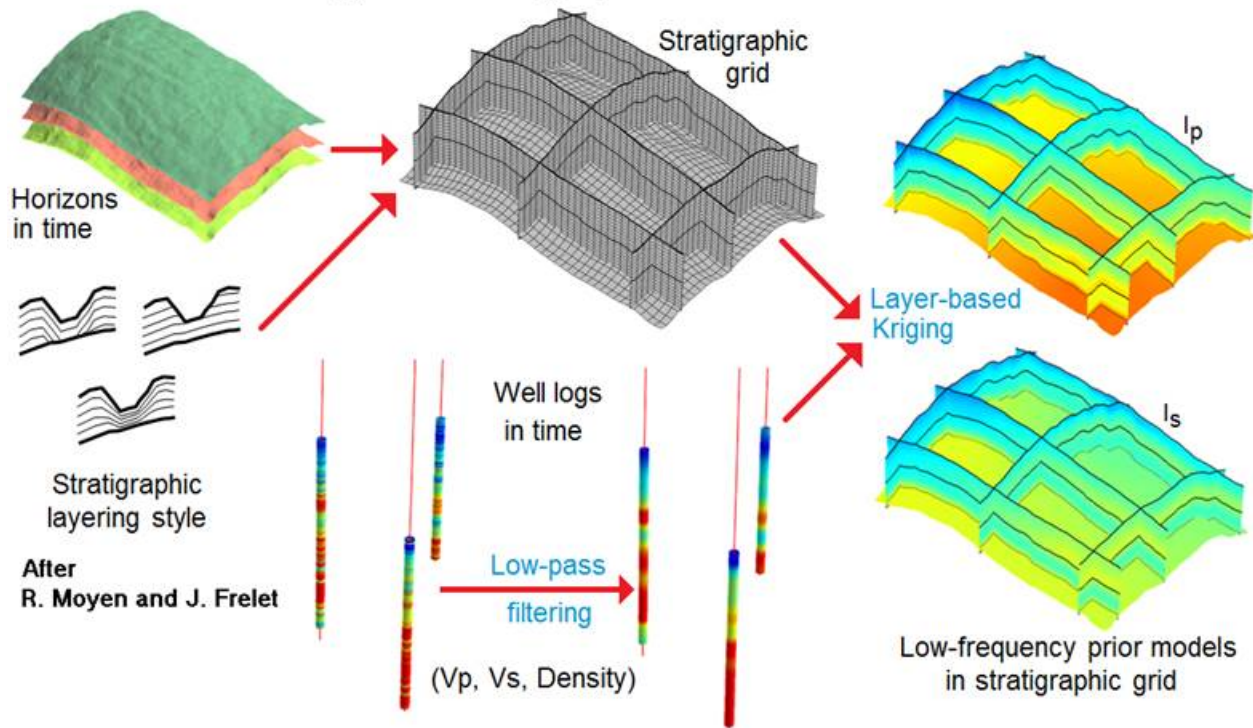


Figure A.3. Changing the reservoir shape and layer distribution impacts heavily the inversion calculations.

The Posterior Distribution, $P(\mathbf{m}|\mathbf{s}, \mathbf{wl})$ according to Buland and Omre, (2003) is calculated from the likelihood from the seismic data, $P(\mathbf{s}|\mathbf{m})$ where $\mathbf{s}$ is the observed seismic.

The likelihood from the well data, $P(\mathbf{wl}|\mathbf{m})$ and the likelihood from the prior model, $P(\mathbf{m})$: $P(P(\mathbf{m}|\mathbf{s}, \mathbf{wl})\mu P(\mathbf{s}|\mathbf{m})P(\mathbf{wl}|\mathbf{m})P(\mathbf{m}))$ assuming that all distributions are log-normal and that reflectivities can be linearly approximated by the 2-term Zoeppritz Fatti's (1987) equation.

This results in:

$$P(\mathbf{m}|\mathbf{s}) \propto \exp\{-\frac{1}{2}((\mathbf{s} - \mathbf{W}.\mathbf{m})^T \mathbf{S}^{-1}(\mathbf{s} - \mathbf{W}.\mathbf{m}) + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{S}_m^{-1}(\mathbf{m} - \mathbf{m}_0))\}$$

Where:

- $\mathbf{s}$ is the observed seismic,

- $W * m$ is the Wavelet convolution matrix
- S_m^{-1} is the Seismic covariance matrix,
- m is the Prior IP-IS model,
- S_T^{-1} is the Model covariance (e.g. the variograms).

Also, $P(m|s) \propto \exp\{-\frac{1}{2}((p - m)^T S_T^{-1}(p - m))\}$, where:

- p is the Posterior mean
- S_T^{-1} is the Posterior covariance.
- $S_T^{-1} = W^T S_s^{-1} W + S_m^{-1}$ and $p = S^T(W^T S_T^{-1} S + S_m^{-1} m)$

The Input Probability Distributions

You would need seismic angle stacks, with one wavelet and one seismic uncertainty per angle stack, in our case, we have seven partial stacks where we need to calculate these.

Seismic Likelihood:

$P(s | m) \propto \exp(-\frac{1}{2} S_q S_i (S_{i,q} - W_{i,q} m_i)^T S_{si,q}^{-1} (S_{i,q} - W_{i,q} m_i))$ where:

- $S_{i,q}$ is the input seismic trace i for the angle q . Use angle stacks with one wavelet per angle stack and one seismic uncertainty value (or map) per angle stack.
- $W_{i,q}$ is the matrix that groups the reflectivities with the wavelet convolution, to transform a model trace with irregular layers into a seismic trace with regular samples.
- $S_{si,q}$ is the matrix for seismic covariance.

AVA Approximation

Jason Geoscience Workbench uses a linear forward model linking elastic properties to reflectivity (Fatti et al, 2004).

$$r_{\theta} \approx \frac{1}{2\cos^2 \theta} \Delta \ln(I_P) - 4 \frac{\bar{I}_S^2}{\bar{I}_P^2} \sin^2 \theta \Delta \ln(I_S)$$

For a column i of layers in a stratigraphic grid, we can write: $r_{i,q} \gg A_{i,q} m_i$

- the m_i vector contains the logarithms of I_P and I_S
- the $A_{i,q}$ matrix computes the differences of $\ln(I_P)$ and $\ln(I_S)$ and multiplies them with
AVA (amplitude versus angle) coefficients

In practice, we group $A_{i,q}$ with the wavelet to form the $W_{i,q}$ matrix.

Well Likelihood

You would need well logs (vertical or deviated) and a well measurement error.

$$P(\mathbf{wl} | \mathbf{m}) \propto \exp(-\frac{1}{2} \mathbf{S}_j (\mathbf{m} - \mathbf{wl}_j)^T \mathbf{S}_j^{-1} (\mathbf{m} - \mathbf{wl}_j))$$

- $\mathbf{wl}_j$ is the log measurement for well j
- $\mathbf{S}_j$ is the matrix for the error covariance measurement

Prior Model

For this model, you need the means and standard deviations, normalized variograms for the inter-cell dependencies and $I_P \ll I_S$ correlation coefficients for the inter-attribute dependencies.

$$P(\mathbf{m}) \propto \exp(-\frac{1}{2} (\mathbf{m} - \mathbf{m}_m)^T \mathbf{S}_m^{-1} (\mathbf{m} - \mathbf{m}_m)),$$

- $\mathbf{m}_m$ is the prior means
- $\mathbf{S}_m$ is the matrix of prior covariances

The basic equation for the inversion process

It is not practical to invert the posterior covariance matrix directly due to its large size. We have three alternate options:

- Diagonalization of the covariance matrix by Fourier transform (Buland and Omre, 2003)
- Sampling of the full PDF with Monte Carlo-Markov Chain (MCMC) methods
- Trace-by-trace decomposition of the posterior using a Sequential Gaussian Simulation

Local Posterior at Trace 1, Trace 2,...,Trace N:

$P(m|s) \gg P(m_1|s_1)P(m_2|m_1, s_2) \dots P(m_N|m_{N-1}, \dots, m_1, s_N)$ that computes: Trace 1, Trace 2 ...

Trace N .

APPENDIX B: GTM methodology for unsupervised seismic clustering and the result of synthetic data.

Generative topographic maps (GTMs) are a type of unsupervised machine learning algorithm that have found application in seismic facies classification. Seismic facies classification is the process of categorizing subsurface geological units based on their seismic response. GTMs provide a probabilistic framework for analyzing and visualizing the distribution of seismic attributes to understand the underlying patterns and structures within the data. GTMs were introduced by Bishop et al. in 1998 as a generative model that combines elements of principal component analysis (PCA) and mixture models. The main advantage of GTMs is their ability to capture the nonlinear relationships and correlations among seismic attributes, which is crucial in facies classification where the relationships between seismic responses and geological units can be complex.

In the context of seismic facies classification, GTMs can be used to learn a low-dimensional representation of seismic attributes. This representation, often referred to as a latent space, encodes the essential features of the data. By projecting the seismic attributes onto the latent space, GTMs allow for a compact and informative visualization of the data distribution. This visualization can aid geoscientists in identifying clusters and patterns that correspond to different facies. The GTM algorithm involves two main steps: training and inference. During the training phase, the GTM learns the parameters that define the latent space and the probability distribution of the data. This is typically done using an expectation-maximization algorithm. In the inference phase, the trained GTM is used to map new seismic attribute data to the learned latent space and perform classification based on the proximity to the different clusters.

Figure B.1 shows the topology map on the latent space and the initial Gaussian noise distribution to iterate over and fit it to the topological map.

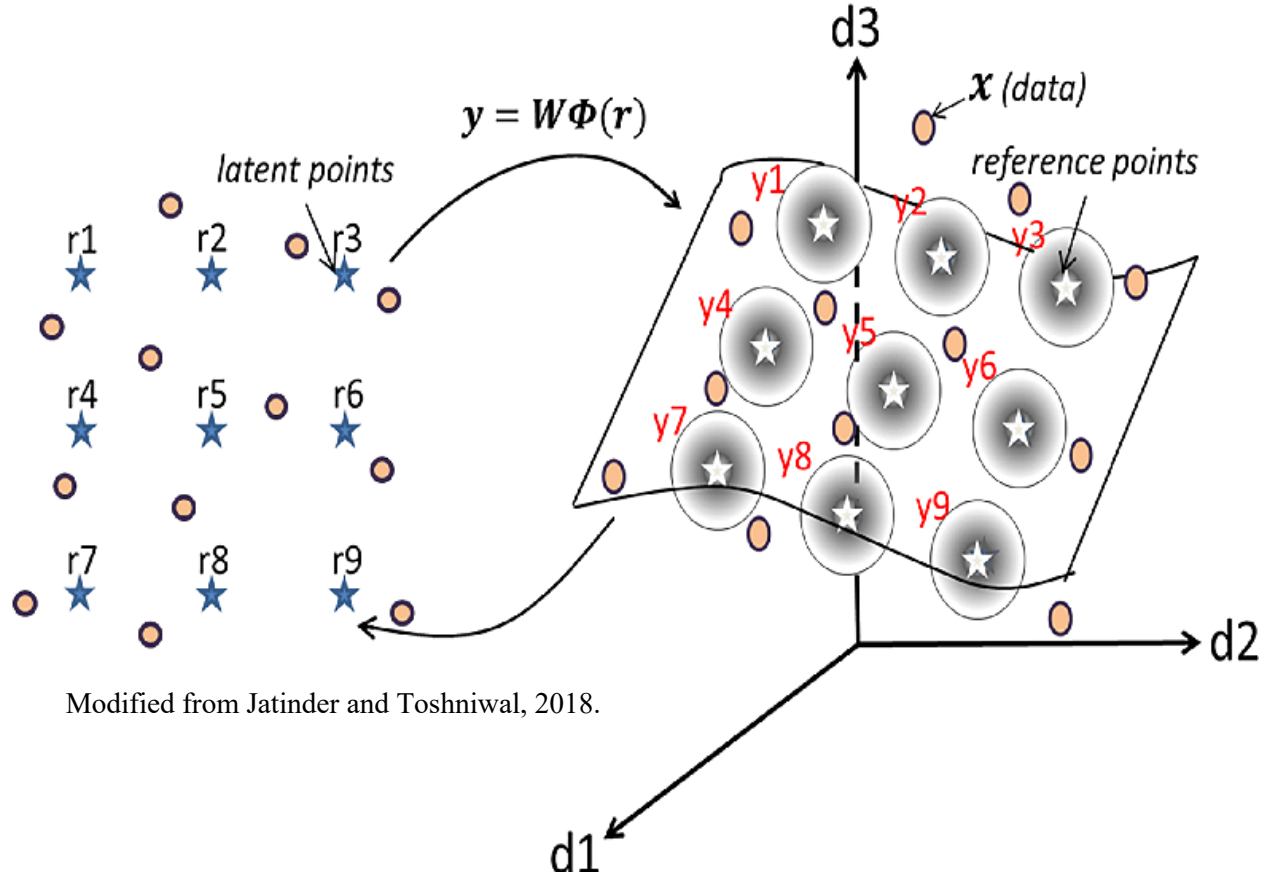


Figure B.1. An Example of Non-Linear Transformation from Data Space (3-Dimensions) to latent Space (2- Dimensions) by embedding a manifold in data space, where $r_1, r_2, \dots, r_n$ denotes the points in latent space, $y = W \cdot \phi(r)$ represents the mapping function.

Using the same list of seismic attributes for the SOM and GTM there are important differences for future work. GTM defines better the reservoir distribution at the cost of showing more noise in the classification. Figure B.2 shows the GTM the classification for the first reservoir at the location of the model with the highest hydrocarbon saturation.

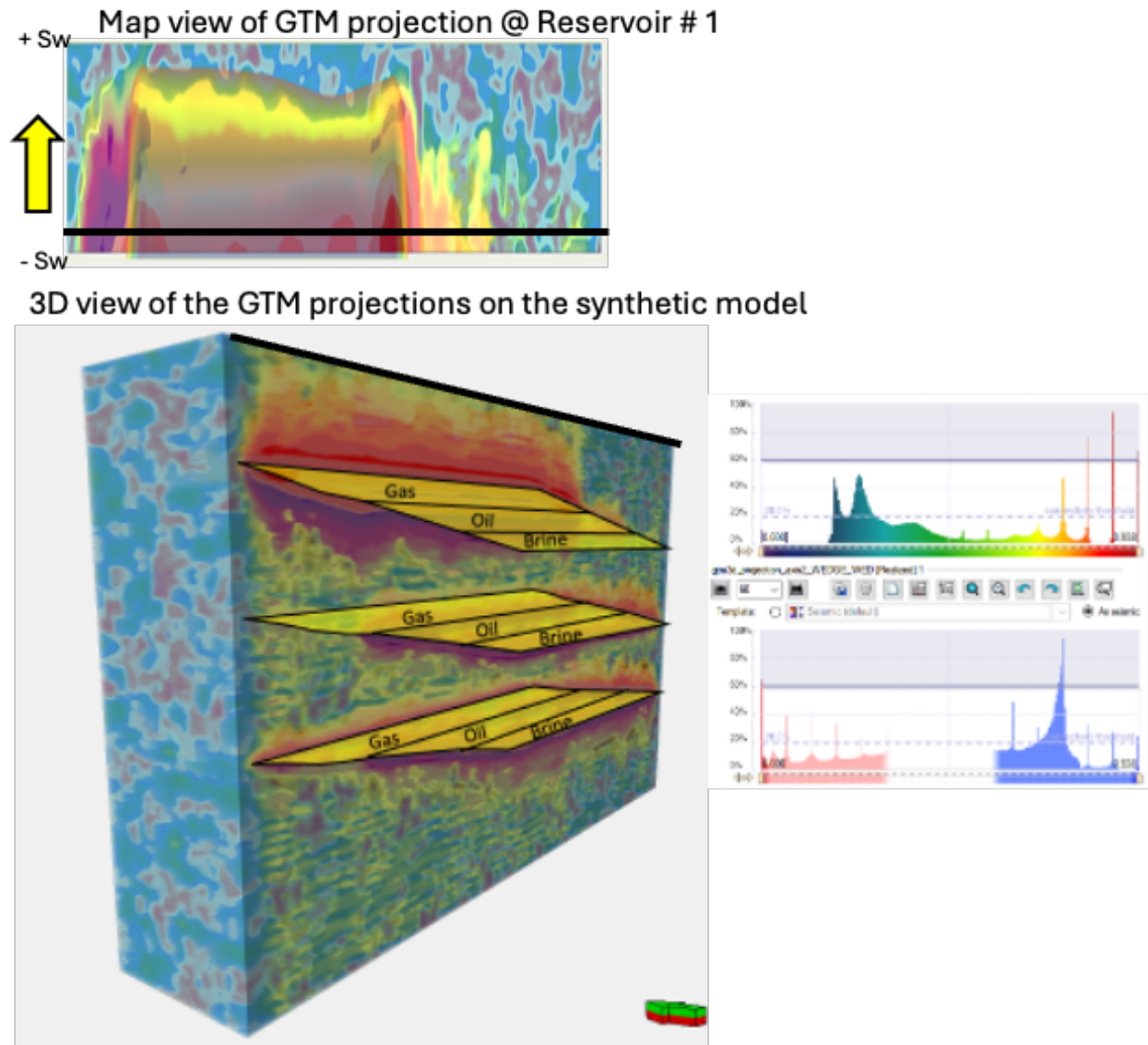


Figure B.2. Visualization of the GTM classification of the reservoirs. In this view the highest hydrocarbon saturation of the first reservoir is classified with a large contrast with the shale background.

As we move deeper into the model, the water saturation is increasing and the hydrocarbon saturation decreases. The GTM clustering does an acceptable clustering on the first reservoir but with larger differences in reservoir 2 and 3. This is shown in Figure B.3, a 3D visualization similar to B2.

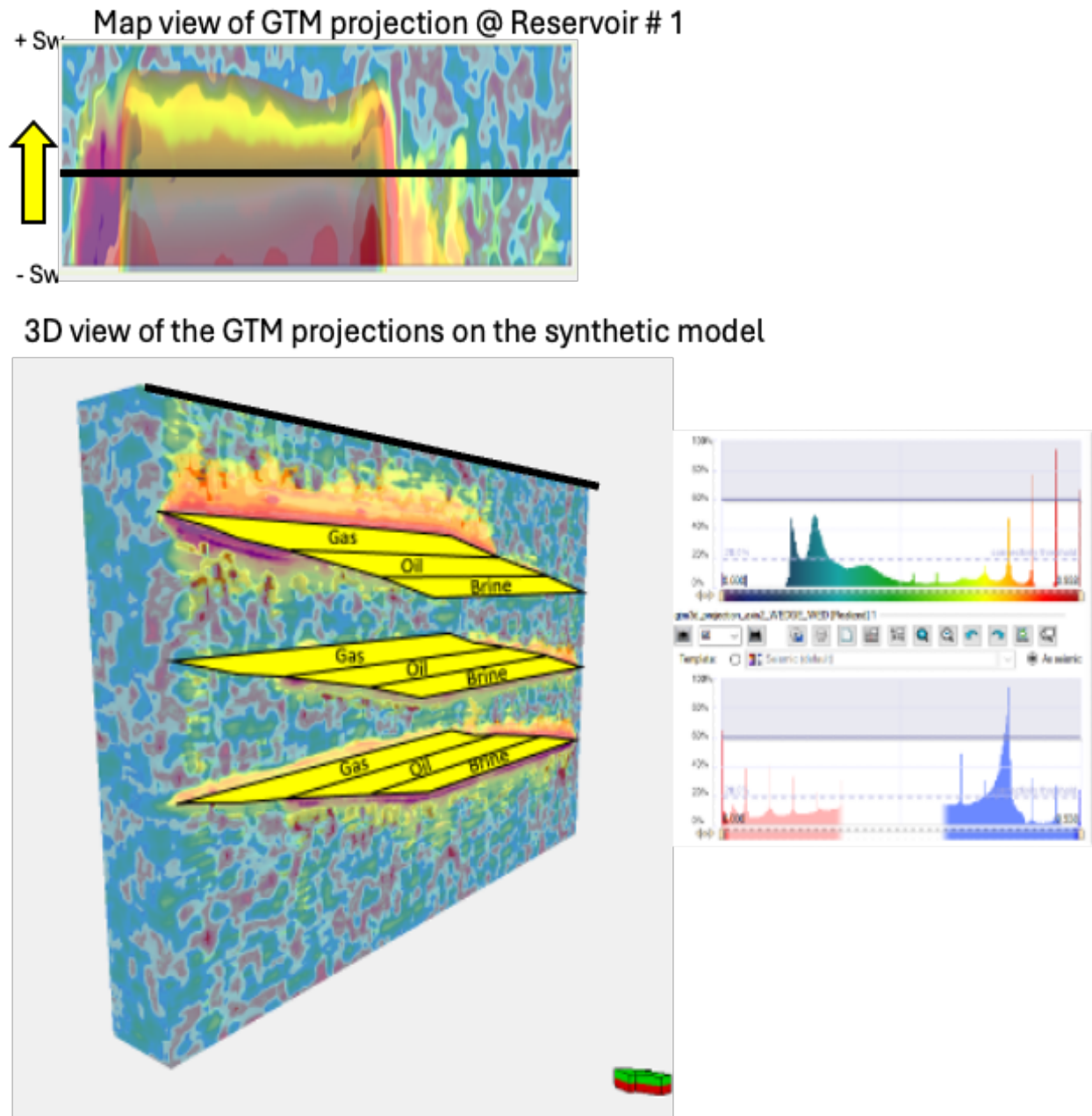
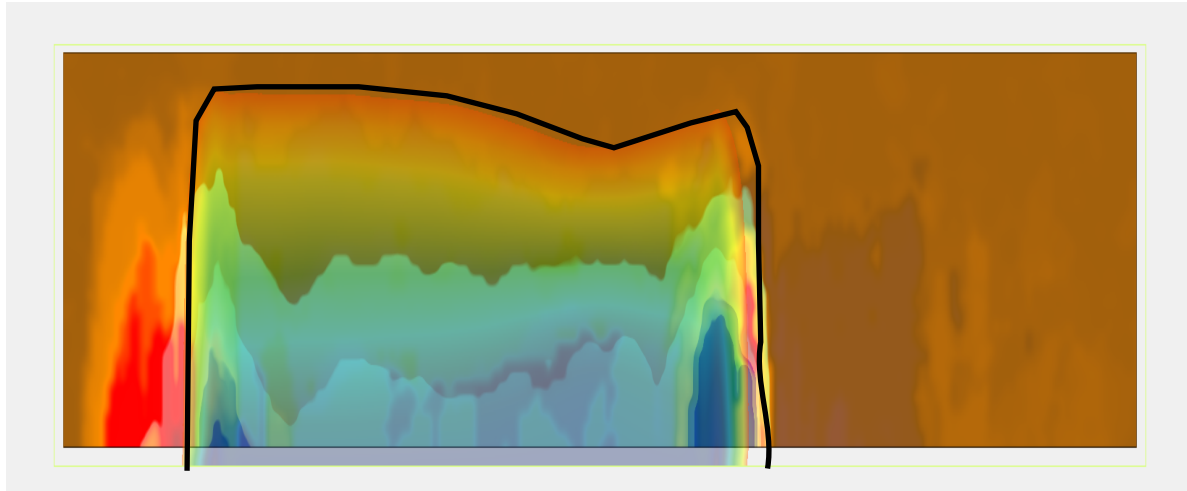


Figure B.3. A 3D perspective on the GTM clustering of the different reservoirs. The anomaly of the first reservoir is exaggerated vertically and beneath the wedge casting doubt on the true extension of the anomaly.

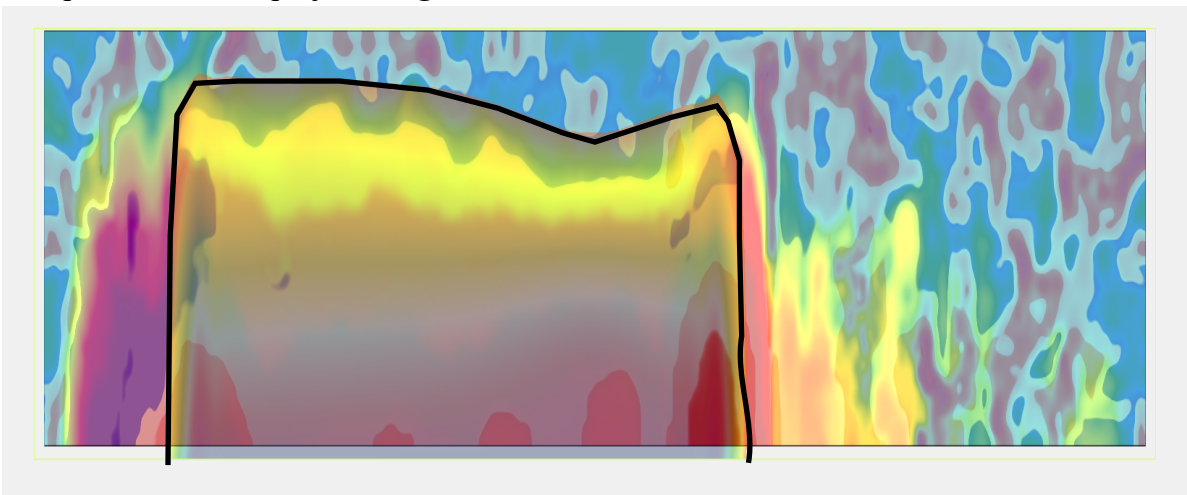
Finally, the comparison of top of shallowest reservoir of both unsupervised methodology shows the compromise between GTM larger accuracy at the cost of more added noise in comparison of

GTM clear delineation of the reservoir at a smaller detectability of the gas extension. Figure B.4, shows the map view of the comparison.

Map view of SOM projection @ Reservoir # 1



Map view of GTM projection @ Reservoir # 1



— Polygon representing the limit of the 80% hydrocarbon saturation.

Figure B.4. Map view comparing GTM and SOM clustering classification of the synthetic data.

APPENDIX C: When the levee breaks: Visualization of a deepwater crevasse splay in Late Cretaceous in Exmouth Basin Offshore Western Australia. *

Alexandro Vera-Arroyo and Heather Bedle, (2022), "When the levee breaks: Visualization of a deepwater crevasse splay in Late Cretaceous in Exmouth Basin Offshore Western Australia," *Interpretation* 10: C43-C51. <https://doi.org/10.1190/INT-2022-0019.1>

ABSTRACT

“The present is the key to the past” is a foundational geologic concept that helps us contextualize buried subsurface features in current geologic analogs. As seismic interpreters, the generation of the geologic model should be unbiased, yet as humans our unconscious biases are expected, and we sometimes overlook anomalous reflection patterns in our seismic data that do not fit the model. As a result, we often disservice ourselves when we overlook these characteristics, potentially ignoring additional geologic context. These anomalous geoforms or funny-looking things (FLTs) may provide further geologic context and aid us in solving the geologic model if included. Crevasse splay on a continental slope marine environment is described and analyzed using attributes, seismic inversion, and voxel-based classification. We discuss possible causes that may have triggered the break of the levee on the Exmouth Basin during the Early Eocene and why a crevasse splay on a steep slope is an FLT. A possible explanation is that the presence of preexisting faults beneath said feature is the likely culprit for a levee break that created the crevasse splay. Thus, in contrast with the Led Zeppelin song about why the levee breaks, it is equally important to understand the preexisting faults when analyzing sediment supply. In addition, this highlights the importance of integral stratigraphic sequence interpretation — from deep to shallow — to understand geology in a full context.

INTRODUCTION

The funny-looking thing (FLT) is in the Exmouth Basin, on the western shelf of Australia, and is part of the Carnarvon Basin and the Westralian Superbasin (Hocking et al., 1987). The Exmouth Basin is the result of a complex tectonostratigraphic evolution with different phases of rifting (Hocking et al., 1987), chiefly among them the top Permian, Late Jurassic extension (late Tithonian), and the Base Cretaceous with the final breakup between Australia and Greater India during the Valanginian (Tortopoglu, 2014).

The continental breakup triggered a sediment inversion, and the sediment supply increased as it evolved into a passive margin. The preexisting rift faults were reactivated during the late Miocene, linking the sediment distribution's lateral variations between the shelf and the deepwater areas. Moreover, the basin has been tectonically active due to the collision between the Australian and Eurasian plates. The increase of overburden due to the accelerated sediment supply combined with displacement of the Australian plate triggered intrabasin small-scale faults (Peel, 2014) and reactivation of preexisting faults during the Early Eocene. The area seismic survey with the available boreholes was located 80 km northwest of Exmouth, and the water depth ranged from 825 to 900 m (Figure C.1).

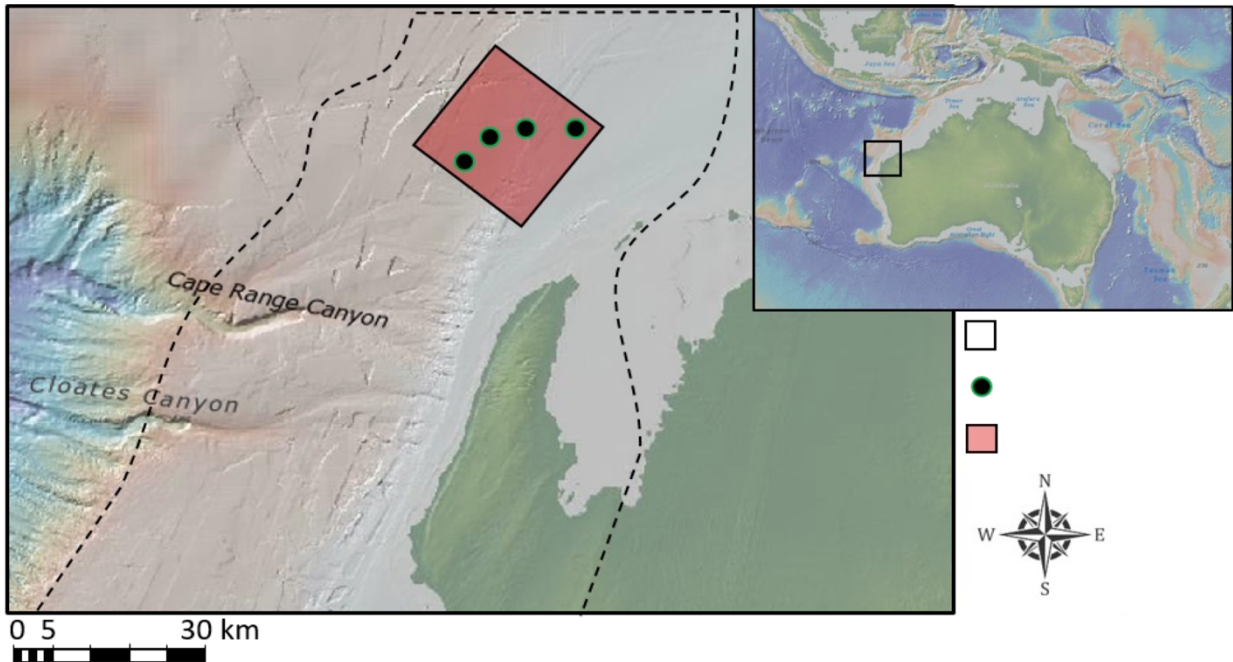


Figure C.1. The study area, denoted by the red polygon, represents the Exmouth Basin, the zone of interest, and the available borehole information.

In the area, four boreholes with a wide variety of geophysical logs are available to integrate into the study with the seismic data set. (The Stybarrow 2008 M4D MSS is a pre-stacked time-migrated seismic data with 13 ms sampling, acquired in 2007–2008 using streamers and conventional air guns.) It is of particular interest to the geochronology of faults and sediment features of Early Eocene since the small intrabasin and preexisting faults have favored the breach on the incision canyon walls (Prather, 2012; Kelman et al., 2013). A detailed structural definition of the area can be found in Exxon and Willcox (1978) and Tindale et al. (1998).

Appearance on seismic data

Figure C.2 shows two seismic sections: parallel to the canyon and sediment influx (Figure C.2a) and perpendicular to the canyon (Figure C.2b). In Figure C.2a, the reflectors associated with the levee walls are parallel reflectors dipping northeast–southwest and are associated with a rapid sand-shale transition. In Figure C.2b, we observe the terminations of the canyon walls and the

Silva, 2004). The p-impedance also is closely related to clay content and effective porosity. In Figure C.3, the base of the sequence highlights the levee walls, and well-1 shows the clay content with the porosity; it is observed in the impedance section (Figure C.3b) that high values of clay content correlate with high values of acoustic impedance.

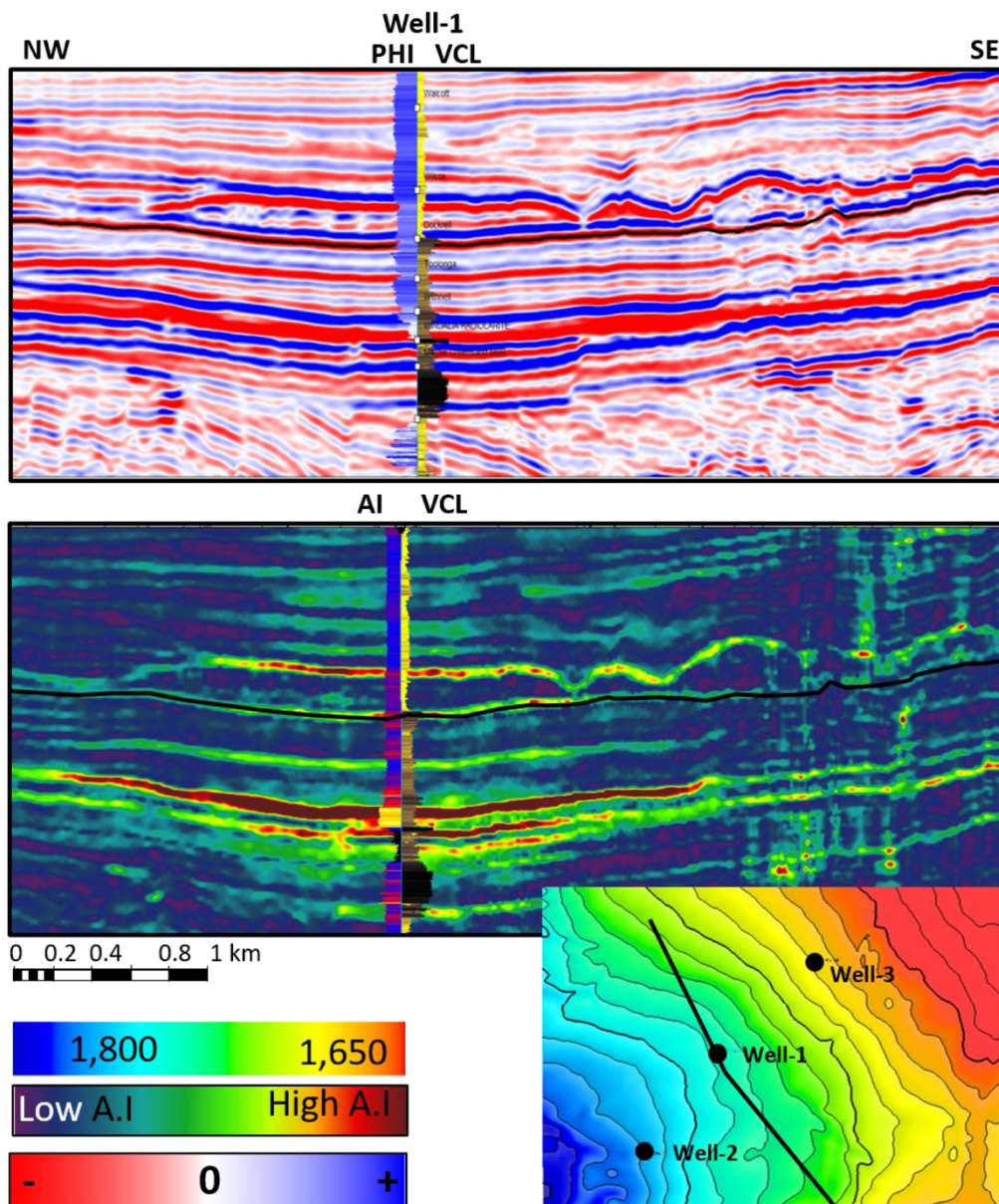


Figure C.3. (a) Seismic section parallel to the levee break and well-1 to calibrate the clay content and porosity. Inside the dashed black box, the chaotic reflectors of the FLT contrast with the sigmoid pattern northwest denoted by green arrows. (b) Acoustic impedance section showing the levee walls; the chaotic reflection patterns are transformed into intercalations of shales and sandstones.

Sediment input and modern levee breaks

The existence of modern analogs to steep-slope de- posits is limited in a literature search; the research is primarily focused on understanding crevasses splay on low- gradient systems, e.g., levee break examples from the Mississippi River (Nienhuis et al., 2018). Figure C.4a shows the continuous deposition of crevasses splay due to floods and the small elevation difference that allows the flood deposits. Figure 4b shows a modification of Sprague et al. (2005) that includes the proposed FLT and a sketch of both slope systems. Typically, the deep-water sediment deposition illustrations consider the base of depositional sequence homogeneous and isotropic. Nevertheless, the preexisting faults affect the deposits and the architecture of the sedimentary structures. In Figure C.4b, a dashed red line represents the fault that aligns with the crevasse splay and levee break.

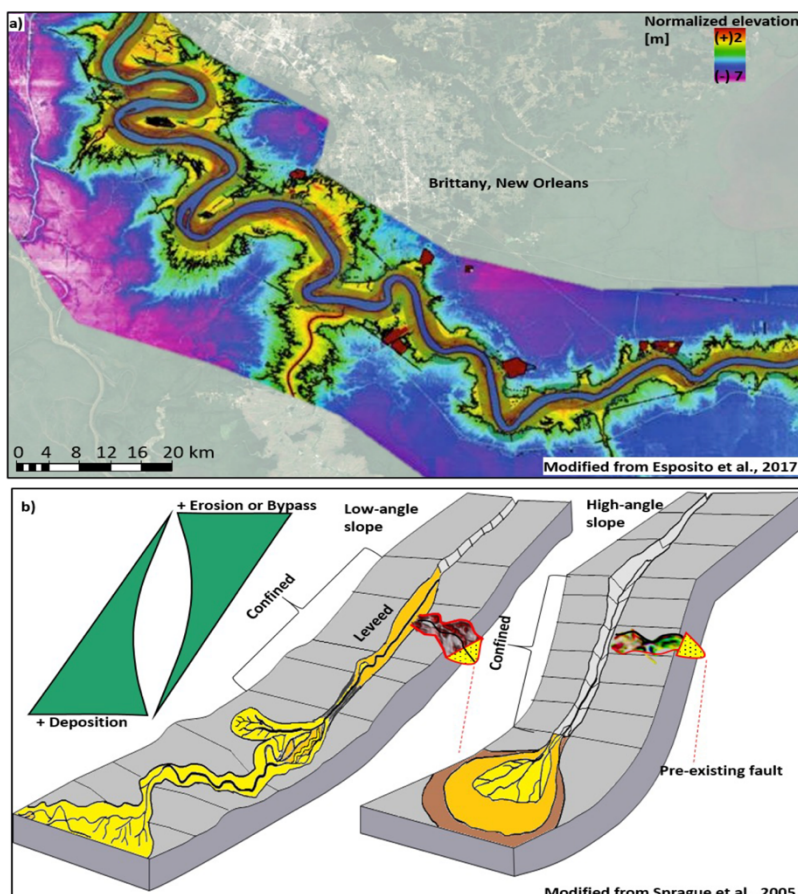


Figure C.4. (a) LIDAR elevation model of the Mississippi River embayment overlapping a satellite image showing the elevation difference that allows the sediments to spill and breach the levee (<10 m). The levee breaks continuously deposit mud sediments during floods. (b) Illustration modified from Sprague et al. (2005) showing the erosion and deposition ratio as a slope function. We include the existence of preexisting faulting as a factor that contributes to canyon breaches.

As in the illustration from Figure C.4b, most schematics of these systems in the literature assume an isotropic base, and the modification included for this example (the dashed red line) is the preexisting faults. A modern geologic analog with more similarities is the Arguello terrace offshore California (Figure C.5).

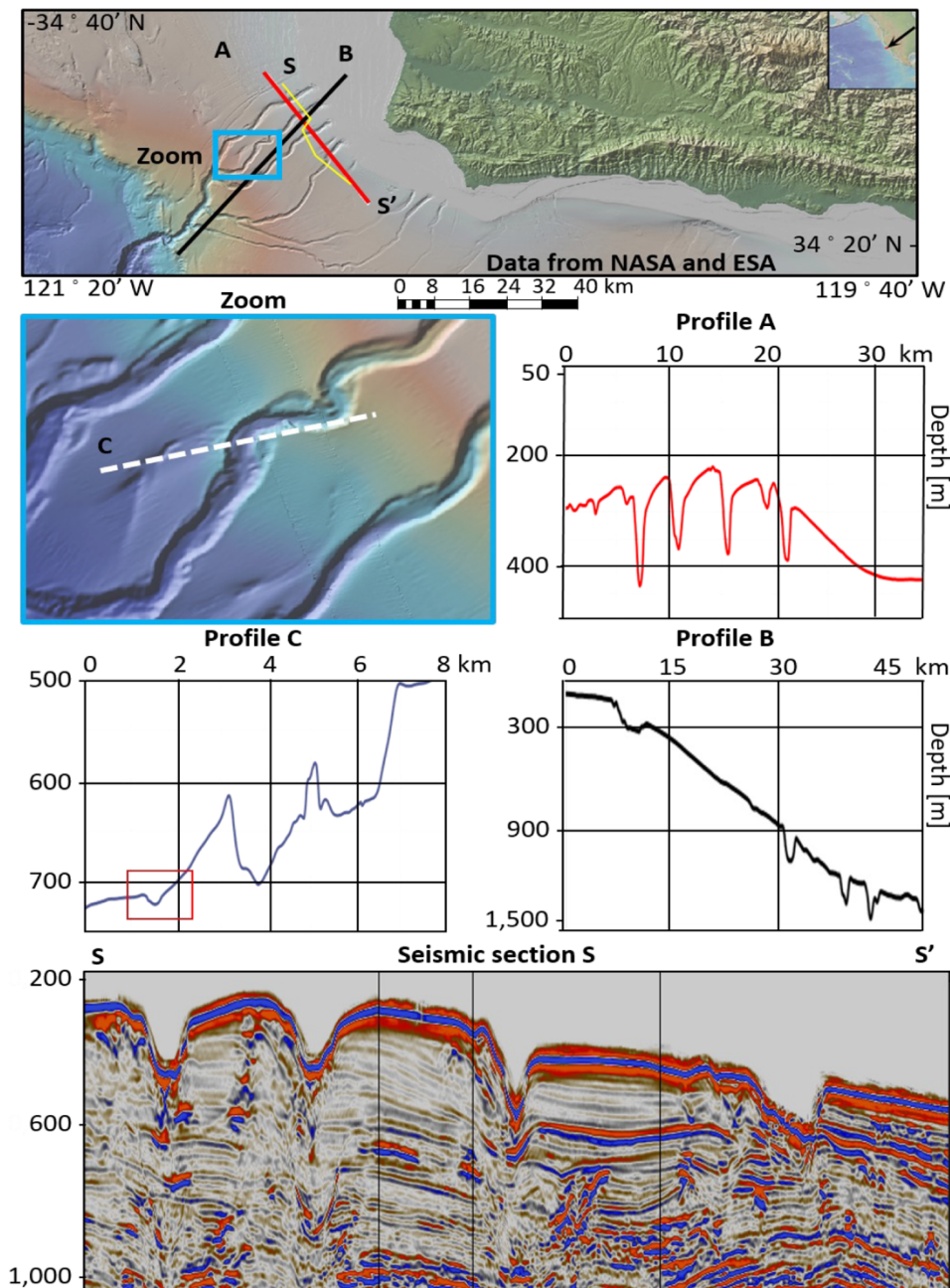


Figure C.5. The high-resolution bathymetry shows a high-angle slope around the Arguello terrace. The seismic section S-S' shows the seismic profile associated with the canyons. The red box on profile C shows the feeder of the crevasse splay for this system (the green box) and it also is on the profile. The black arrows show the splay's beginning, and the red arrows show a slump on the profile, most likely due to the seismic activity in the area. Letters on the map are related to the profiles.

METHODS

The seismic attributes used (acoustic impedance, broadband coherency, curvedness, flatness, envelope, and principal curvature) and the deep learning attribute (convolutional neural network fault predictor) are tools to visualize and better understand the geologic, particularly the tectonic and sedimentary, structures. Acoustic impedance results from a poststack seismic inversion (Russell et al., 1991), which can be briefly described as transforming seismic reflection data into an impedance profile by removing the wavelet effect. The rest of the seismic attributes are different filters that transform seismic data either by a constant time window or by geometric estimation of properties. We invite readers to check Chopra and Marfurt (2007a, 2007b) for detailed documentation on seismic attributes. On the other hand, the convolutional neural network (CNN) fault predictor is a novel technique (2019) that treats seismic sections as images and detects faults as discontinuities. The discontinuities are compared against preexisting synthetic data to assign a fault probability (Wu et al., 2019).

Seismic attributes play an essential role in seismic interpretation, and they are crucial to interpreting geometries thoroughly (Ementon et al., 2004). Calculating attributes alone is partially accurate and highly biased with borehole integration. Lithology and petrophysical properties integrated with the seismic information yield an understanding of the relationship between rocks present in the subsurface and seismic stratigraphy (Paumard et al., 2018). Visualization techniques such as flattening, and 3D interpretation increase the understanding of the geologic model. Flattening to a particular surface or stratal slice between two surfaces is a standard tool. For example, flattened seismic volume and stratal slices reveal the canyon and the channelized behavior of the center of the crevasse splay.

CNN fault detection and supervised fault extraction have accelerated the interpretation of the primary faults in the Mesozoic. The fault framework correlates with the levee's break in the Tertiary. The fault alignments are approximately perpendicular to the inferred dominant flow direction. A 3D voxel classification is a powerful tool that has helped us highlight and isolate the crevasse's splay, and the voxel representation in three dimensions isolates the splay improving the understanding of how the rift geodynamics impacted the architecture of a passive margin. Voxel classification is an interpreter's tool, combining acoustic impedance and seismic attributes into a cross plot where values are clustered based on 3D visualization. Spectral decomposition and flattening are visualization techniques that can be found in different interpretation packages. The frequencies of interest based on the analysis are 12–32–76 Hz. We have plotted these frequency bands as a horizon slice, and spectral coherency uses these frequencies as peak frequencies to estimate the voices. Tuning thickness is a function of frequency content; thus, spectral coherency and spectral root mean square are helpful tools to present the main feeders of the splay.

RESULTS

Attribute interpretation provides insights into the subsurface, but the information needs to be carefully integrated with borehole information to reduce uncertainty in the interpretation (Posamentier and Kolla, 2003; Paumard et al., 2018). Figure C.6a displays the modified Posamentier and Kolla (2003) illustration for deepwater deposits where the canyon stage suggests that sediments are well confined and only splayed on the base of the slope. Figure C.6b shows a constant time slice where the canyon system is well presented. The canyon's surface is transparent to show how the constant slice is mixing stratigraphic features. Figure C.6c shows the 3D visualization of the surface of interest; in Figure C.6, the dashed red line shows the location of the levee break. A stratal slice of the seismic amplitude clearly shows the main feeder of the crevasses

splay in Figure C.6d. Finally, the 3D visualization of the surface with the seismic section that runs perpendicular to the center of the levee break clearly shows the location of the FLT. Showing the existence of the levee break is half of the exciting feature. How did it get there?

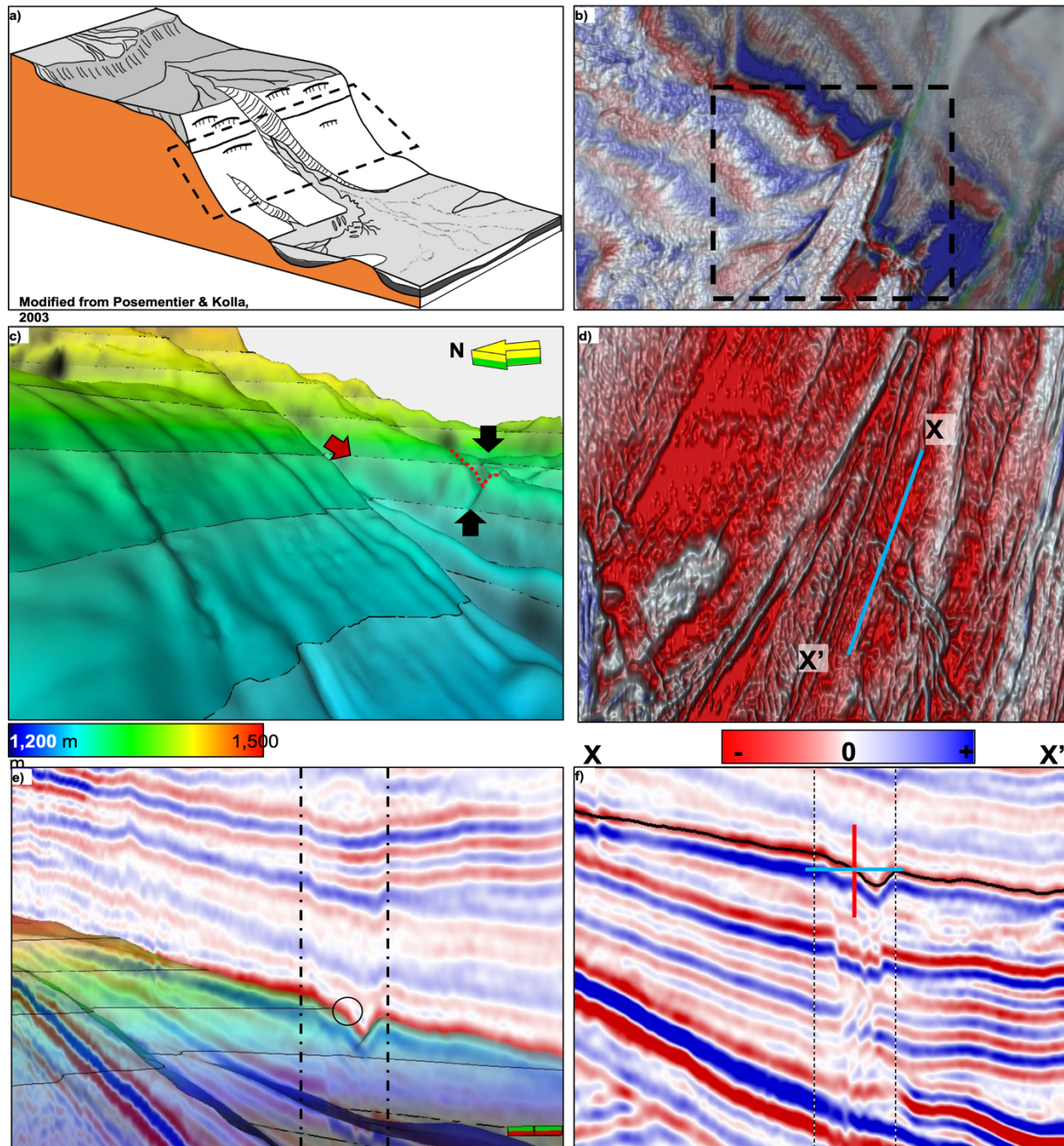


Figure C.6. (a) Schematic depiction of principle depositional elements in deepwater settings; the red square shows the study area of our FLT. (b) A constant time slice with the surface of interest denoted with opacity to show the area of improvement by using a stratal slice. (c) The surface of interest; the red arrow shows the canyon walls and points out the direction of the sediment supply. The black arrows and dashed red lines emphasize the levee breach. According to the sketches, at this location of the canyon, sediments

ought to be contained. (d) The amplitude stratal slice of the canyon reflector. The crevasses splay signature is enhanced, which invites flattening of the seismic volume and the follow-up attribute calculation. The blue line is the direction of the seismic section in the following panels. (e) Seismic section parallel to the canyon shown in a 3D view to enhance the canyon's slope and the levee's breach. A black circle shows the area of interest on the next panel. (f) The interesting interval and the black surface's chaotic reflectors. Cross confirmed by red, and blue lines represent the location of the black circle. The red line has a thickness of 85 m and the blue line represents 200 m.

Using CNN fault prediction, we estimate the complete 3D fault framework; most faults correspond to the Mesozoic. We will show the alignment of faults to the canyon system and the crevasses splay. The seismic attributes help us identify the canyon and the direction of sediment input. Considering the chrono-stratigraphic position, we can infer the distribution of sediment supply (northeast–southwest). In Figure A1.7, seismic volume is flattened, and we have calculated the spectral decomposition (12–32–76 Hz) that evidentiates tuning thickness of the crevasses splay using only the amplitude at different frequencies and the multispectral coherency corendered with seismic amplitude. Figure C.7a shows the seismic amplitude voices, and Figure C.7b shows multispectral coherency using the same frequency values.

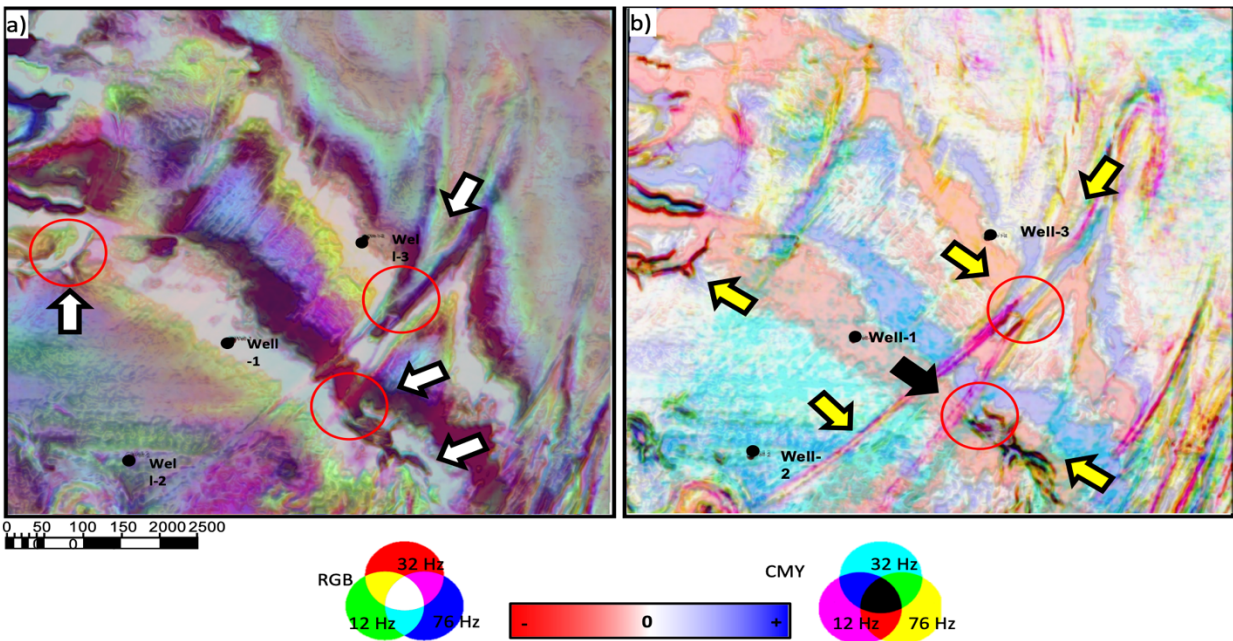


Figure C.7. (a) Multispectral decomposition of seismic amplitude showing braided channels associated with the splay. (b) Multispectral coherence corendered with amplitude exaggerating the break of the levee and the braided nature of the crevasses splay. The yellow and white arrows show features of interest and open different questions for interpretation.

To further discard other hypotheses such as avulsion channels or abandoned backfilling channels, and especially to present the link between preexisting faults in the Mesozoic with the stratigraphy and sediment deposits systems in the Tertiary, we use additional attributes. On the flattened amplitude data, structural most-positive and most-negative curvature delineate the boundaries of the synforms and antiforms interpreted as levee wells (Figure C.9a). The acoustic impedance is corendered with broadband coherency to show that the high acoustic layers associated with sand-prone layers are associated with the chaotic patterns (Figures C.9b). Those patterns are associated with low-contrast amplitude (Hurren et al., 2012). Flatness and curvedness attributes delineate the center of the feeder channels of the crevasses splay (Figure C.9c). A corendering of the multispectral Sobel similarity with envelope makes the breaching area of the levee break more evident (Figure C.9d).

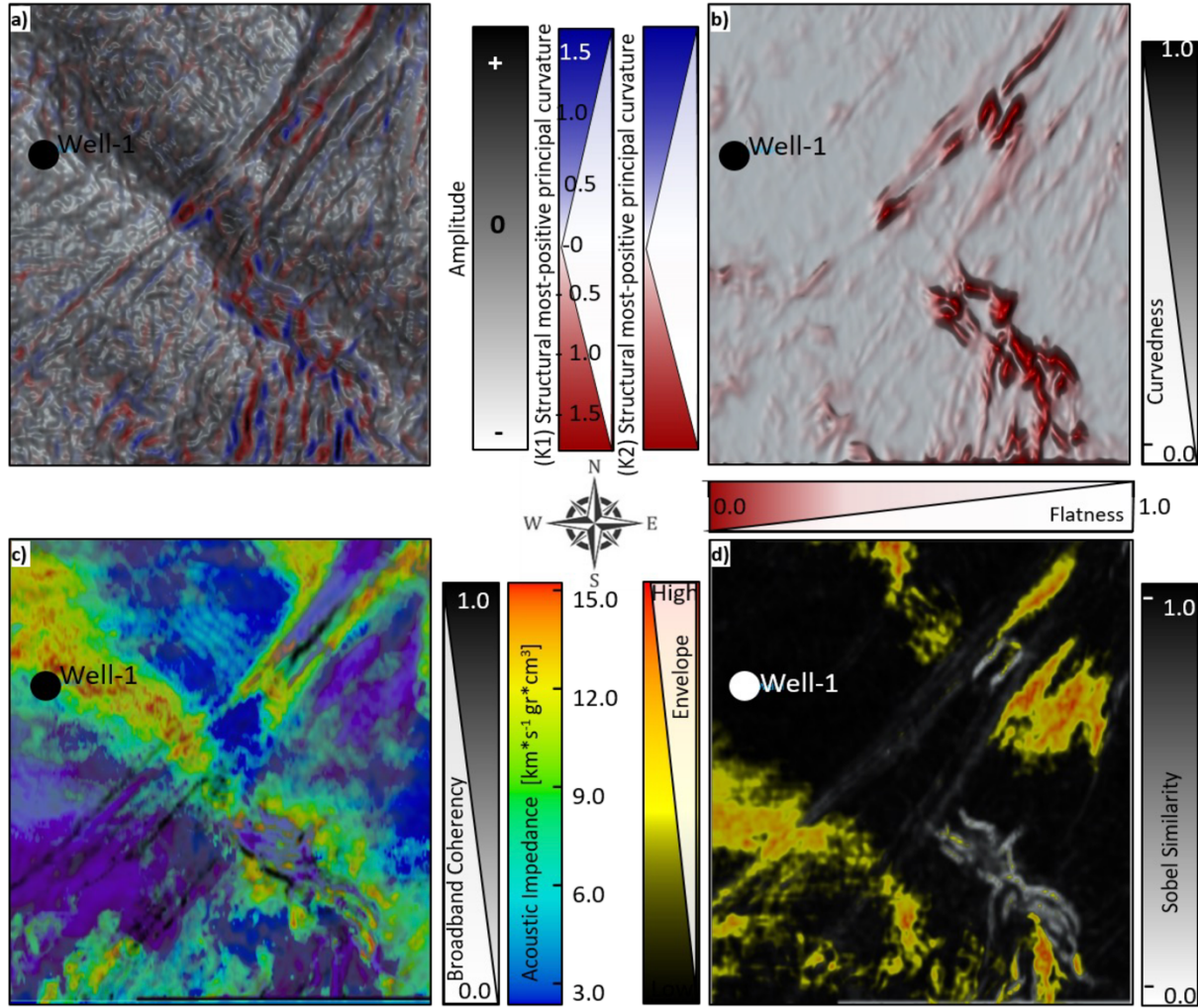


Figure A1.9. (a) Seismic amplitude corender with K1 and K2. (b) Horizon slice extraction of two attributes: flatness and curvedness; most of the extraction is white, but the red value of flatness shows the center of the feeder channels. (c) Acoustic impedance and coherency corendered to represent the crevasse splay, and high values of impedance indicate shales. (d) Envelope corendered with Sobel similarity to show the canyon walls and the edge of the crevasses splay. High-amplitude values are most likely associated with sandstones.

We have done an extensive vertical search using dip magnitude, dip azimuth, and multispectral energy-ratio similarity. The results of the analysis reveal that there are no faults present. We also have combined total aberrancy magnitude and curvedness to look for sub-seismic flexural folds (McArdle et al., 2014) in the Tertiary. In contrast, the fault intensity is remarkable on the rifting stage, yet the faults are limited to unconformity and do not continue. The result of the CNN fault detection algorithm confirms that the fault intensity in the Mesozoic is substantially more prominent than in the Tertiary. The result of the CNN fault prediction is shown in Figure

C.10, where the Aptian limits faults' continuity, separating the area's two tectonic styles. Despite this division, the existence of faults in older strata has affected the distribution of younger sediments.

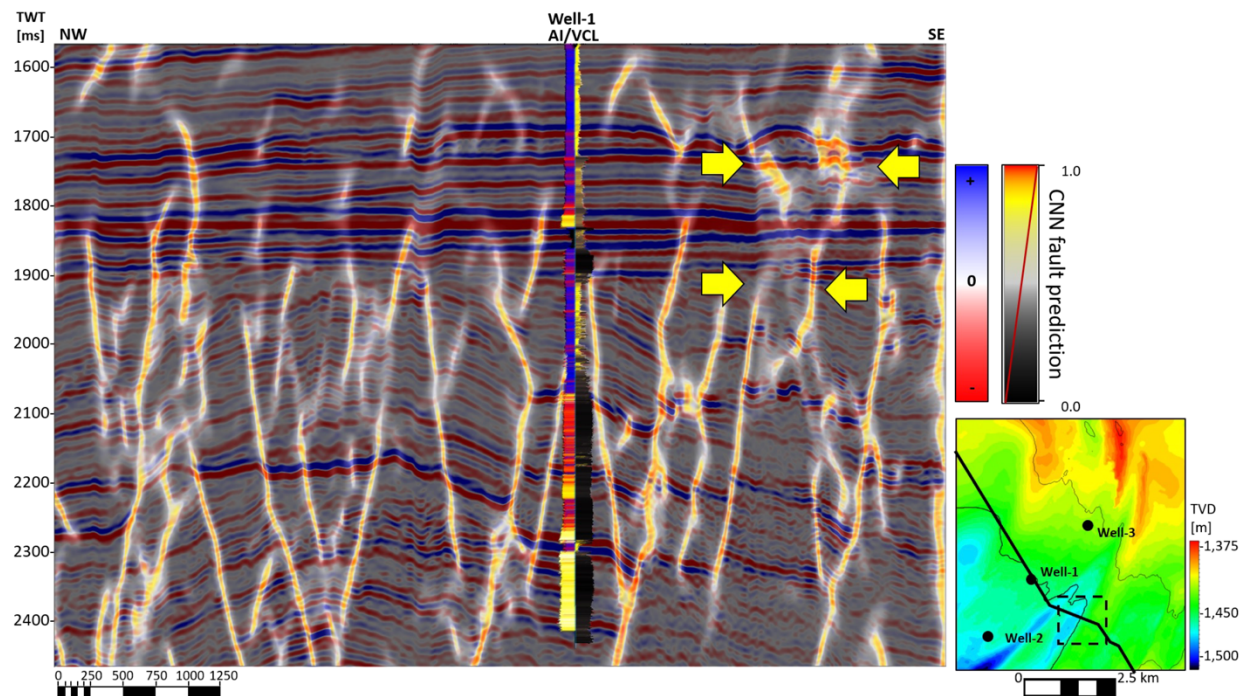


Figure C.10. CNN fault prediction showing the fault intensity of the rift stage. This fault system hardly continues the stability of the drift of the Tertiary.

Finally, we have combined Sobel similarity with envelope and acoustic impedance using a weighting function to extract the connected voxels of the geoforms. (Canyon and channels are a product of the levee.) We have combined the geobodies extracted from the Paleocene interval with the Mesozoic faults to show the effect of preexisting faults. The coincidence with the direction of the canyon is remarkable. In addition, the position of the levee breaks channels with a perpendicular fault (Figure C.11). The primary fault alignments are parallel to the dominant flow direction. The crevasse splay is generated due to a system of perpendicular faults below the breach of the wall levee. We have isolated in three dimensions the geologic features of interest to visualize the elements of interest from different angles. Some key characteristics are that the width of the

yet the faulted and inverted deposits of the Mesozoic contrast with the deposits in the Tertiary. Nonetheless, the slope of the channel does not suggest the possibility of an overbank deposit or a levee breach due to flooding.

CONCLUSION

Data integration is crucial to understanding the subsurface better. In this work, the attribute visualizations, seismic acoustic inversion, and borehole information show a possible crevasse splay on a submarine channel. Although the Early Eocene is an excellent reservoir rock in this portion of the Exmouth Basin, with porosities greater than 15%, the lack of evidence that connects reservoir rock to the documented source rock makes the Early Eocene play unattractive to drill. However, the identification of the crevasses splay should be relevant to the reservoir risk analysis. Drilling late Eocene thick shale sequence using high-density drilling muds and unexpectedly drilling through these high porosity sands could represent a drilling hazard.

The affinity of splay deposits with the basin floor is documented, yet the FLT displayed shows a crevasse splay on a slope deposit, considered a sedimentary bypass zone where the sediments are well confined. The Albian unconformity marks a change in structural and depositional style. However, the preexisting faults on the Mesozoic condition weakness zones have triggered the levee break. Our figures show the interplay between preexisting faults and deposit processes. The proposed attributes and visualization techniques help the understanding of the interplay between sedimentary processes and faults.

ACKNOWLEDGMENTS

We would like to take this opportunity to thank the Australian Geoscience department for providing the 3D data set, AASPI consortium, CGG, and Schlumberger for the software licenses offered to the University of Oklahoma and the attribute computation, and NASA and European Space Agency for the satellite images of the California coast and the bathymetry models used for the profiles.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be accessed via the following URL:
<https://portal.ga.gov.au/persona/gadds>.

APPENDIX D: Contrasting faulting styles of salt domes and volcanoes: Can unsupervised learning techniques differentiate fault styles? *

Alexandro Vera-Arroyo and Heather Bedle, (2023), "Contrasting faulting styles of salt domes and volcanoes: Can unsupervised learning techniques differentiate fault styles?," *Interpretation* 11: C1-C13. <https://doi.org/10.1190/INT-2022-0018.1>

ABSTRACT

An attractive feature seen on seismic data, also known as Funny looking thing or FLT henceforth has a wide range of interpretations, from noise patterns to amplitude anomalies. An example of an FLT is the similar faulting patterns between a volcanic intrusion and a salt intrusion from the point of view of a machine learning algorithm. Oftentimes seismic interpreters do not have a complete dataset, or geologic background to determine the genesis of observed features. This can be particularly perplexing when trying to determine if an intrusion is volcanic or halokenetic in origin as they exhibit similar geomorphologies. We examine the differences in these features in the Gulf of Mexico, which is a well-documented salt basin, and the Taranaki Basin in New Zealand, which is igneous prone. The analysis aims to discern geological features based on the geometries and attributes shared by seismic data and remote sensing tools. Seismic attributes and machine learning techniques highlight differences and similarities between the intrusions. We discuss machine learning techniques such as self-organized maps (SOM), an unsupervised machine learning technique, and cluster fault systems without regard to the geologic context. The attributes used in the SOM are fault probability, fault dip azimuth, fault dip magnitude, and thin-bed detector. Fault probability is performed through a combination of convolutional neural network fault prediction and a skeletonization process.

Once the faults are clustered using SOM, the visualization of fault architecture due to the existing mount (either volcano or salt dome) is done considering high fault probabilities ($>75\%$). The methodology consists of selecting the neurons from the SOM grid corresponding to the presence of faults and combining them with fault probability and a fault dip azimuth using a cross plot. The cross-plot product assists in the automatic extraction of the fault planes using; a) a voxel representation of the fault planes and b) fault patches representing the fault planes. Moreover, the visualization technique defined demonstrates that the cross-plot product yields better-defined fault planes. With the fault system characterized, we compare horizon slices using coherence, fault dip magnitude, and azimuth against remote sensing images with similar attributes. In conclusion, our methodology combines technologies to differentiate the genesis of intrusion –salt or igneous – using the fault presence and could be helpful in frontier exploration or planetary exploration.

INTRODUCTION

When examining seismic data and encountering a 'funny looking thing,' we typically address two questions: 1) what are we seeing? 2) how did it get there? With the blooming of machine learning (ML) interpretation, existing algorithms focus on accelerating seismic facies interpretation or fault detection. In this vein, we use ML fault detection to assess if we can improve on the complex question: is this a salt or igneous intrusion?

Are ML algorithms capable of detecting subtle patterns to differentiate between these two intrusions? Can ML help train seismic images and extrapolate the result to other remote sensing techniques? Two questions are relevant to which solutions we will aim to explore, combining legacy and open-source seismic data and two ML techniques. Naturally, analyzing reflection patterns and subtle fault trends is irrelevant in today's leading-edge seismic imaging capabilities, acquisitions, and processing techniques that produce high-resolution images of the subsurface. However, extrapolating our understanding of subtle fault patterns on horizon slices

using seismic interpretation and the experience in attribute interpretation can assist in remote sensing, particularly in the interpretation of planetary forms.

The analysis we present is the interpretation of uncertainty from the assessment of machine learning algorithms, focusing only on the z-perspective (depth/time or horizon slice) and comparing it with the map view perspective. Since satellite images sample the surface with different techniques yielding a set of different attributes, the comparison with seismic images can only be either constant time slice or horizon slice and attribute extraction. Zeng (2010) showed attribute extraction from formation is more accurate to explore sedimentary features. Moreover, focusing on the map view perspective of our methodology allows us to classify and help train labels to discern if an intrusion has either a salt or a volcanic origin.

Our methodology and the visualization technique proposed herein also seek to educate and plant the seed against experiential interpretation bias (Bond et al., 2008). In summary, the present FLT offers both efficient and exciting methodologies to rapidly distinguish the fault framework, depict the similarities, underline the difference between two relevant intrusions and compare them with satellite images with the same two geoforms (salt intrusion and volcano). Fault pattern interpretation from satellite images can expand our understanding of Mars surface for example. Moreover, an accelerated structural definition might lead to a better characterization of the stress field or tectonic regime, fluid connectivity, or identification of thermal sweet spots (Dos Santos et al., 2022).

The seismic datasets we are using to show the methodology are courtesy of the New Zealand geoscience agency (Kora and Nimitz) and the national archive of marine seismic surveys (NAMSS) of the United States of America (GoM A and GoM B for short names) shown on Figure D.1. On the other hand, the satellite images from the Landsat-8 image library and processed using

web-based engine services and the HiRISE from NASA/JPL-Caltech/UA Arizona. The locations where the satellite images are taken are Taranaki Mountains and the Zagros Mountains in Iran.

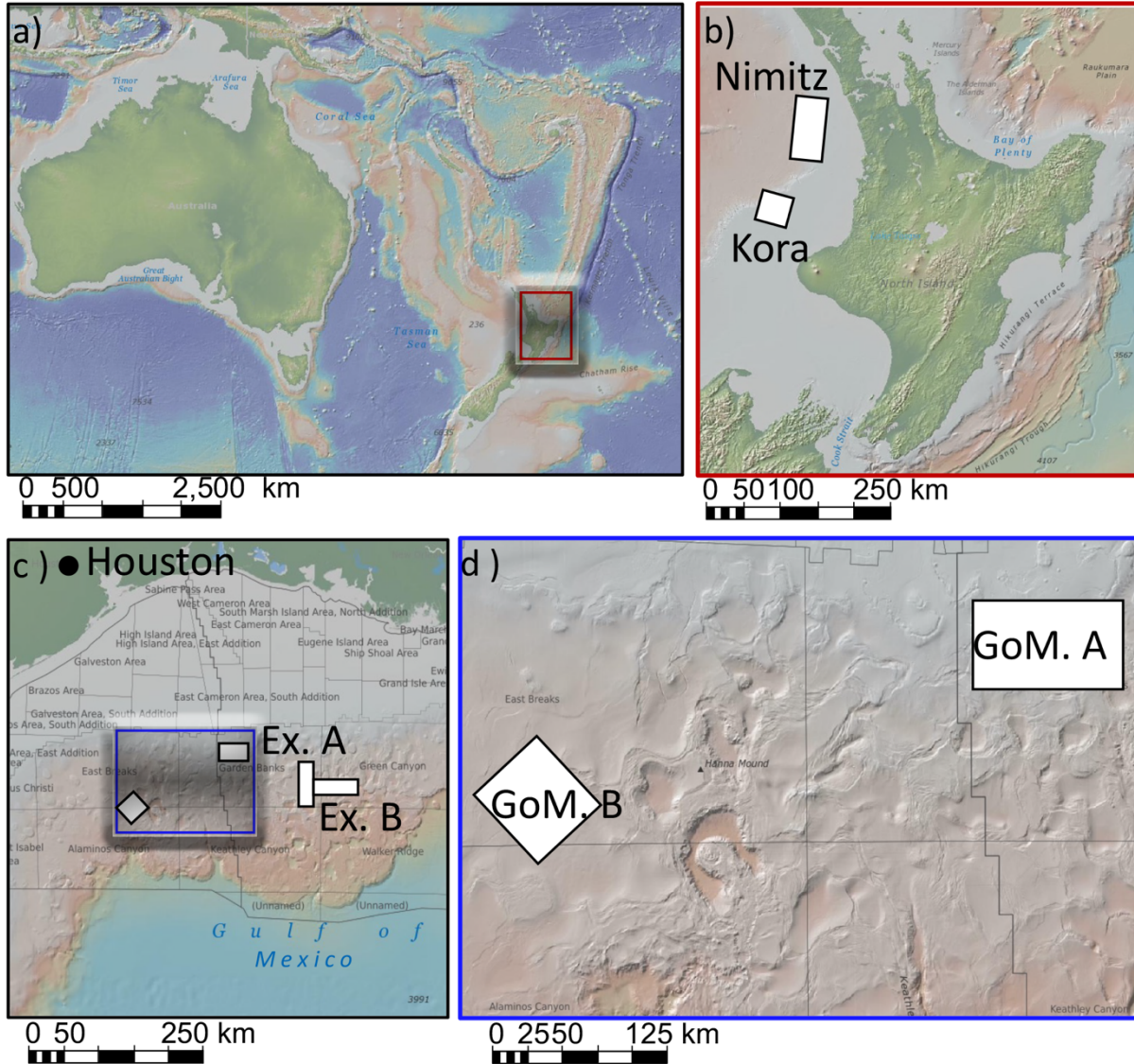


Figure D.1. Panel a) Zoom of the location of the datasets used to analyze volcanic intrusions. (d) Zoom of the Gulf of Mexico ocean bottom to show the location of the seismic dataset. Example A is survey B-38-94-LA-3D and Example B is survey B-69-94-LA-3D, both surveys are not used in the process but help as examples of salt intrusions geometries.

Volcano shapes are roughly classified as fissure, shield, cinder, strato, dome, and caldera volcanos (Jain, 2014). On the other hand, salt geoform shapes are complex and heterogeneous, varying from salt sheets to massive salt canopies (Hudec et al., 2013; Varela and Mohriak, 2013). The relevance of igneous and salt intrusions to hydrocarbon exploration is beyond comparison in

terms of hydrocarbon production. Moreover, both intrusions have been frequently studied due to their similarities, especially in seismic information (Alves, 2012).

The comparative advantage of seismic interpreters against machine learning is context interpretation. Seasoned interpreters have better chances in intrusion detection than facies classification algorithms, especially when analyzing legacy data. The interpretation of salt intrusions is meaningful to detect and interpret due to the challenges associated with said intrusions. For example, salt deposits on weak sub-horizontal dips in salt basins such as the Gulf of Mexico can create false flat spots –a well-known direct hydrocarbon indicator. On the other hand, volcanic intrusions are also essential to interpret. Volcanic cones to the untrained eye (or an unsupervised learning algorithm) worldwide have been easily mistaken for carbonate build-ups (Klarner et al., 2006) and remnant sandstone erosions (Stoppa, 2012).

Volcanic shows on the surface are well documented and evident in contrast, salt intrusions are less noticeable. Fortunately, halokinesis has shaped the ocean floor's bathymetry and allowed for hydrocarbon exploration on the shores of the US and Mexico. Similarly, in the Brazilian Equatorial Margin (BEM), cinder cones exist on the ocean floor and subsurface (Eide et al., 2017) and have modified submarine sediment distribution. In both cases, both intrusions have sparked technological developments and improved our understanding of deepwater sediment deposition and deformation. Since the discussion revolves primarily around the geometry of the faults produced by volcanoes or salt domes. We focus next on similarities and differences associated with the intrusions in map view and compare them with overlaid attributes.

Differences in fault appearance

The apparent similarity between the geometric features of salt domes and buried volcanoes is the normal radial faults generated at the crest. Indeed, the differences related to these faulting patterns

are subtle and chiefly related to halite dissolution from evaporitic bodies (Jackson et al., 2014). The general appearance and representation in the vertical seismic domain are funny and exciting to describe. Halokinetic processes cause multiple effects on the local tectonics, one of these effects are normal radial faults on top of the intrusion. However, as Coleman et al., (2018) discuss, these faults are complex, with differences in throw and heave on salt-filled and non-salt-filled faults.

On the other hand, crestal faults on top of the volcanic intrusions are more continuous. Additionally, sills near the shallower volcanic vents might resemble fracture-like shapes (Infante-Paez et al., 2017). Another exciting similarity is the evergreen tree shape of different volcanic and salt dome intrusions (Varela et al., 2013). Infante-Paez et al., (2017) suggested identifying the volcanic feeder and volcanic extrusions around the igneous intrusion to differentiate salt vs. igneous origins based on the shape of the intrusion. Volcanic feeders appear differently from salt feeders; as salt features have a variable feeder that is a product of the viscosity of the halite minerals. Sometimes, the salt feeder is a fault where the salt crept through a weaker zone, generating a welding fault that connects the autochthonous salt with the allochthonous salt (Varela et al., 2013).

The primary differences incorrectly classifying the salt vs. igneous origin are related to the visualization of the (1) internal reflections within the intrusion, (2) differential compaction on the surrounding strata, and (3) the top and base of the intrusion. While salt domes are vastly chaotic or reflection-free, parallel concordant reflections could also be present due to strata dragged during the salt emplacement (Haugen et al., 2009; Jones and Davison, 2014). Other interesting features to observe are the simulated bottom reflectors inside the salt dome (Figure D.2b), as pointed out by the yellow arrows.

On the other hand, volcanic bodies commonly have high amplitude discontinuous reflections or hummocky to chaotic reflections due to different lava flows. The presence of parallel high-amplitude reflectors is also possible due to dykes and sills (Figure D. 2a). We used the red arrows to show the parallel reflections and possible secondary magma flows. Igneous intrusions present processing artifacts to consider, e.g., the seismic multiples in Figure D.2a, highlighted by cyan arrows, which suggest that the top element is not a gas chimney as documented by Infante-Alvarez et al. (2017), but rather a high acoustic impedance intrusion (laccolith or sill), we can observe the multiple highlighted as well with the cyan arrows. We calibrated the polarity in this seismic volume using two criteria: 1) Ocean bottom polarity (positive) and 2) Borehole information from Lennon-Infante (2017). Moreover, the reflection coefficient of gas-bearing sand at shallow depths as 400 ms below the mud line is most likely a strong negative coefficient (Arevalo-Lopez et al., 2016, Wollner et al., 2017, Mavko 2020). These suppositions are valid if the seismic processing was performed correctly (Niranjan 2016). In Figure D.2a, the first cyan arrows show a strong positive reflection, instead of a bright spot.

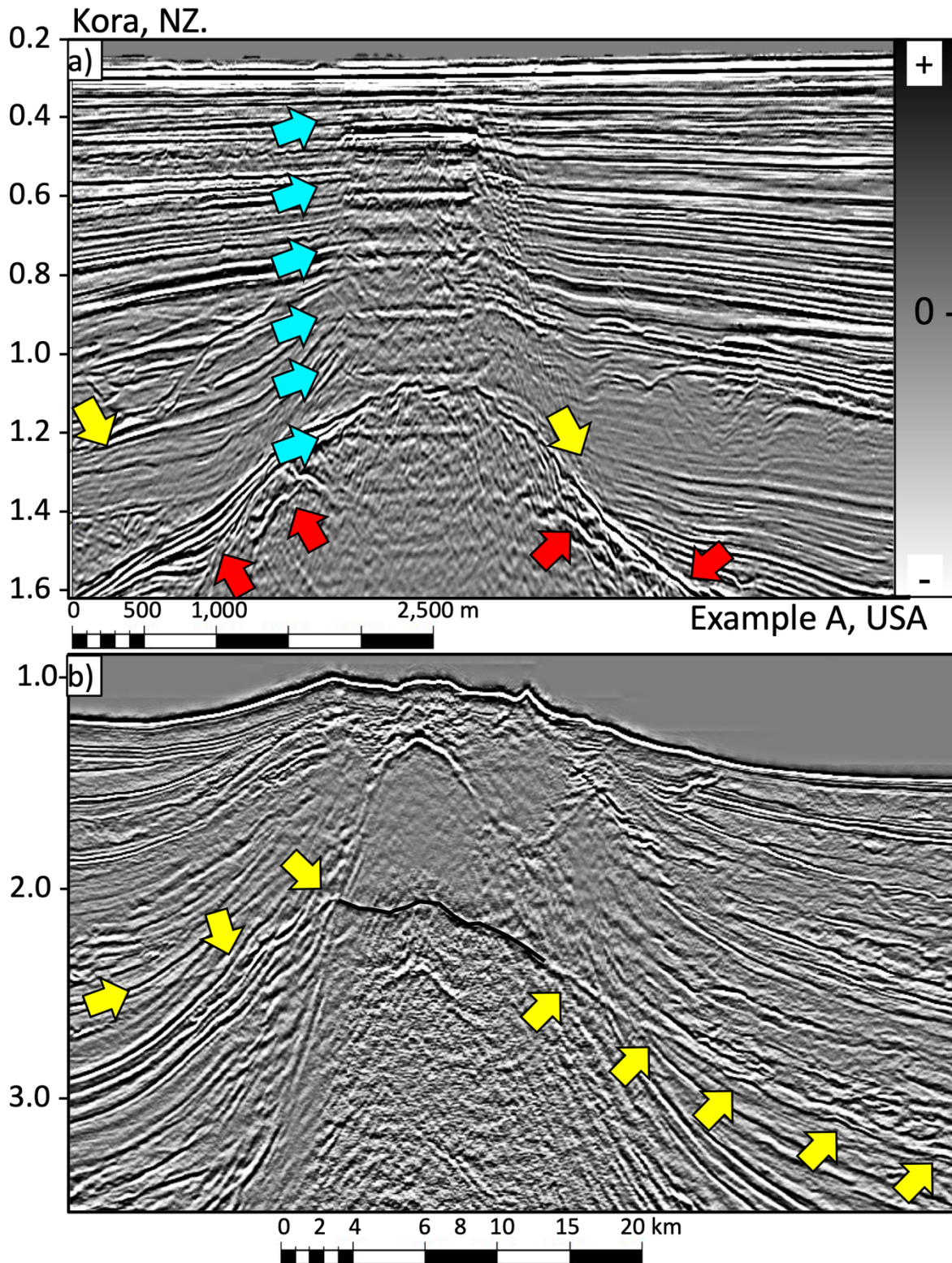


Figure D.2. Panel a) High amplitude chaotic reflections inside the Kora volcano, side chamber also present. Panel b) Null contrast with high amplitude on top of the salt dome, a parallel reflector crosscuts the salt dome and mimics the ocean floor. The effect is known as bottom simulating reflector or BSR. The BSR appears in porous rock, indicating the salt thickness delineating the salt dome feeder.

Another intriguing effect of looking out for is differential compaction, which is the topological change due to overburden. Differential compaction is commonly associated with syntectonic deposits (Chopra and Marfurt, 2007), a common effect in salt provinces. In back-arc basins, sediment input constantly creates continuous compaction causing autochthonous withdrawal, increasing space of accommodation and thus stress.

While syn-deposited sediments around salt domes might exhibit differential compaction (Figure D.3b), as shown by the yellow arrows, the parallel reflections above volcanic provinces show the opposite (Figure D.3a). For example, (Bischoff et al., 2017) described a listric system fault in Northern Taranaki Graben. Nevertheless, parallel reflectors or truncations around the Nimitz Volcano showed standard compaction (Figure D.2a) (Clairmont and Bedle, 2020). We seek to show the constant compaction on the Nimitz volcano with red arrows.

The inherent seismic characteristics between salt and volcanic intrusions are significantly different; for example, while volcanic intrusions have a variable reflection polarity such as dim positive, strong positive, or dim negative (Figure D.3a), the top of the salt on the Gulf of Mexico is commonly a positive reflection of SEG polarity (Figure D.3b).

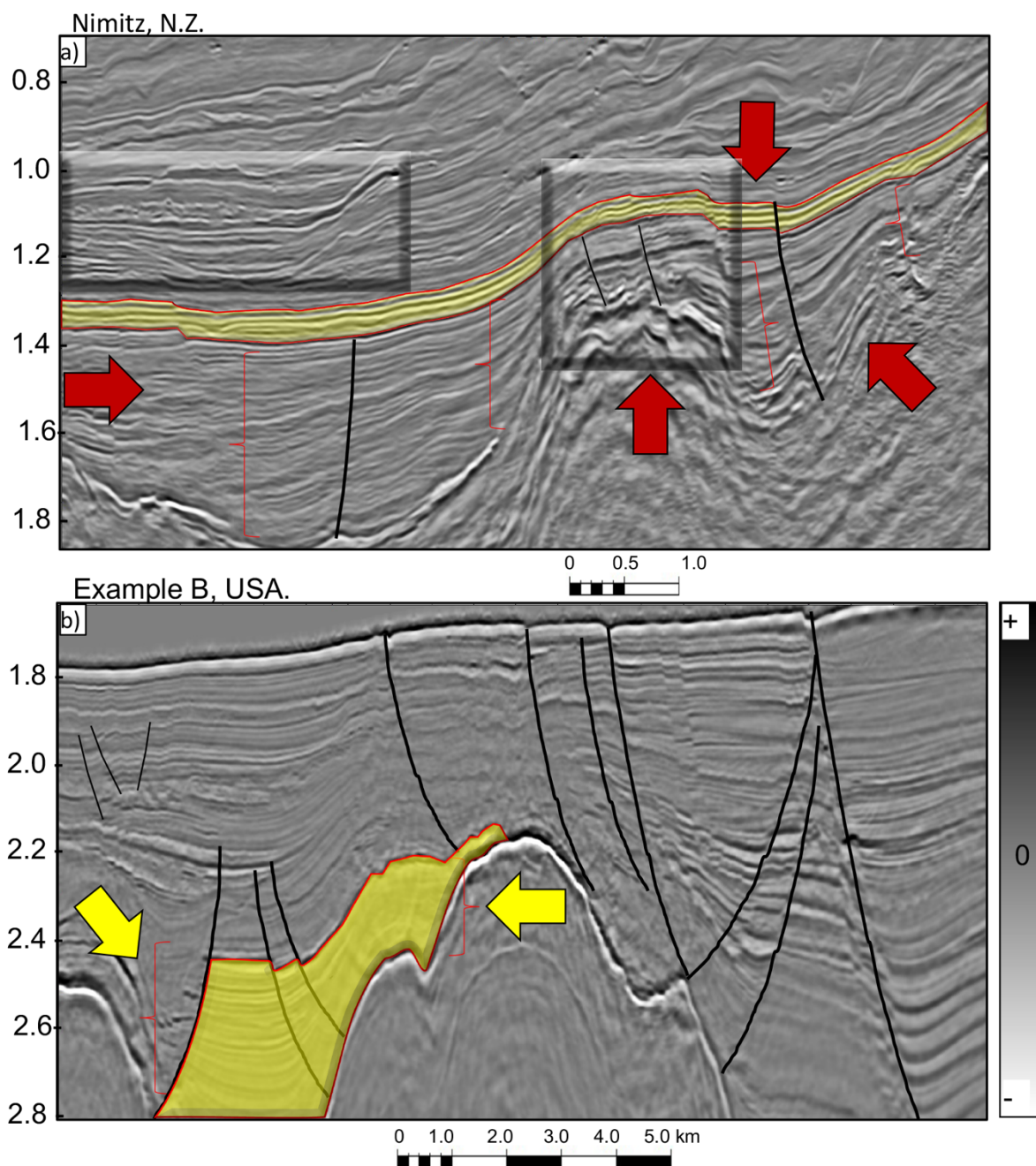


Figure D.3. Panel a) Salt dome is associated with a complex fault system; the ocean bottom reflects the nuance of faulting around the salt dome. Internal reflections inside the salt dome are null or the result of multiples in the data. The strata around the salt dome, particularly to the left, show an amplitude increase due to the differential compaction. Panel b, on the other hand, regardless of the larger accommodation space due to the half-graben present in the area, exhibits no differential compaction, probably due to the usual compaction trend.

The detailed appearance of seismic data

To contextualize the datasets used to contrast igneous and salt radial fault styles, we decided to show first a collection of salt shapes around the Gulf of Mexico (Figure D.4). The shapes vary from domes to sheets, canopies, and walls. In the datasets we used in the Gulf of Mexico, the top of salt is defined by high-amplitude positive reflection. Base and lateral extension are poorly defined due to the natural complexities of the salt dome and the type of imaging techniques used 30 years ago. However, the seismic section shows a plethora of exciting geoforms and faults associated with the salt intrusions, including fault planes with high dips and low amplitudes associated with the fault planes the crown faults on top of the salt diapir.

The kinematic evidence displayed in the sections can be observed chiefly in the immediate strata, where amplitude change, dip (magnitude and azimuth), and horizon continuity are the most notorious changes. Crown fault systems associated with salt intrusions might exhibit a change in polarity if fault planes are intruded by salt at their intersections (Jackson 2020). The conventional seismic processing workflow of the 1990s is still valid today and uses typical velocities for salt bodies around 4.5 km/s, corresponding to the halite velocity (Lu et al., 2007). Although dirty salt could account for variations in the velocity field, 4.5 km/s salt velocity is a practical approximation.

One of the possible consequences of using constant velocity is the presence of porous sediments within the salt that allows the presence of BSR, which mimics the ocean floor even across the salt dome (Figure D.4c). Otherwise, the internal reflections of the salt are chaotic or null in our region of study, except for Figure D.4b, where the base of salt could be vaguely delineated.

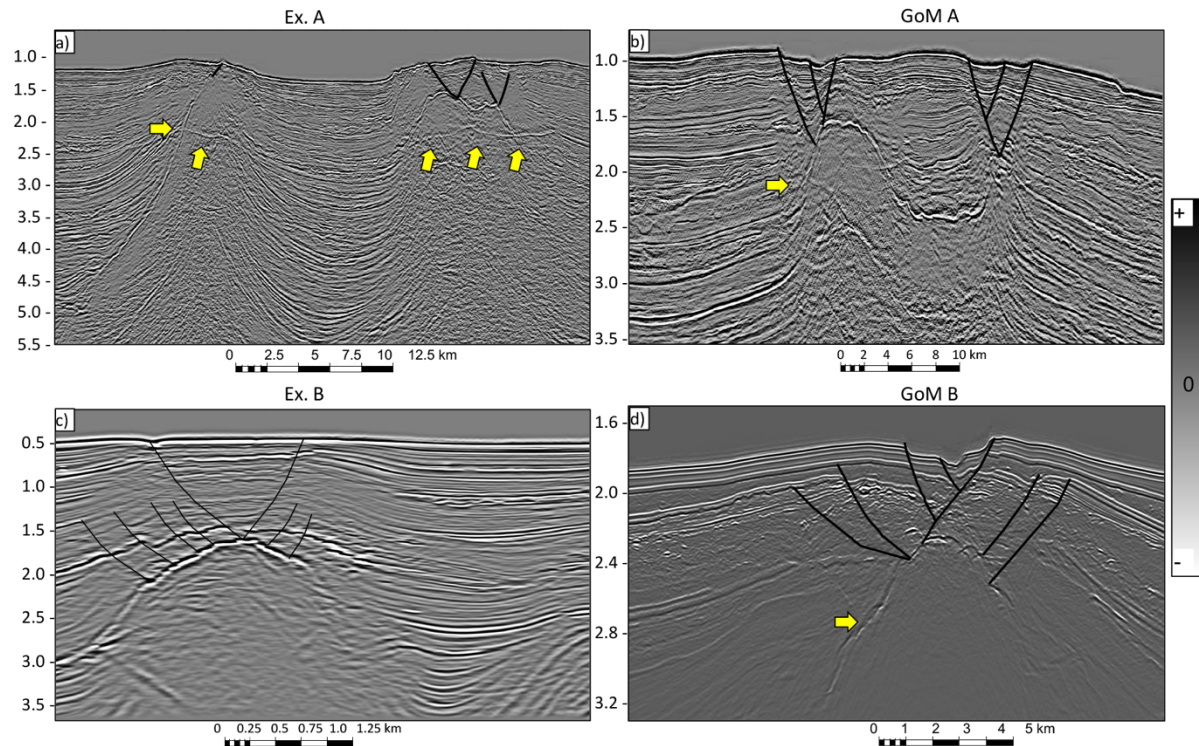


Figure D.4. a) High-low dip angles on top of the salt dome, faults are inferred, and exhibit a low amplitude contrast. The lower limit of the salt dome is unclear. b) A complex salt sheet with two different structural heights affected the strata above it with complex faults. Some of them continue to the ocean floor. The salt sheet thickness can be hypothesized from 0.5 to 0.8 seconds. c) two different salt shapes, a salt dome affects the ocean floor to the left, and it can be observed with a possible salt feeder. To the right, a salt canopy with a possible salt welded fault, the effect of the salt's arbitrary shape can also be observed on the strata deformed above. The ocean floor is almost undisturbed, except for the structural difference of the salt canopy. d) The most evident effect on the seafloor in GoM B the normal faults due to collapse (synthetic and antithetic faults) on the sediment above the salt dome. An evident seafloor deformation can show that the fault has similar dip magnitudes but opposite dip azimuth.

Figure D.5 shows an arbitrary line of two volcanoes on the available datasets to compare the salt with volcanic provinces. The Kora dataset (Figure D.5a) shows a sharp positive reflection in contrast with the surrounding strata. Internal configuration exhibits high amplitudes of chaotic reflections associated with internal magmatic chambers, silt intrusion, or volcanic flows. The Nimitz dataset (Figure D.5b) contains two volcanos with a positive polarity but random and chaotic reflection contrast.

The Nimitz volcano has a steeper side and an asymmetrical cone; these two characteristics define the Nimitz volcano as a stratovolcano (Clairmont et al., 2021). Internally chaotic, the volcano

boundary shows a very disturbing substrate; there is little evidence of fluid mud or molten lava supplied from these volcanos in contrast with Kora. (Figure D.5a). Of course, the absence of evidence is not evidence per se, and the feeders for the Nimitz volcano might be below seismic resolution. Also, the shape and size of the caldera are different between these two volcanoes. While the Kora volcano might have a smaller but steeper caldera (Bischoff et al., 2017), Nimitz resembles a shield-shaped volcano with a more expansive crater. A hypothesis can be evaluated by looking into erosion on the cone's flanks and detecting angular unconformities (Figure D.5c).

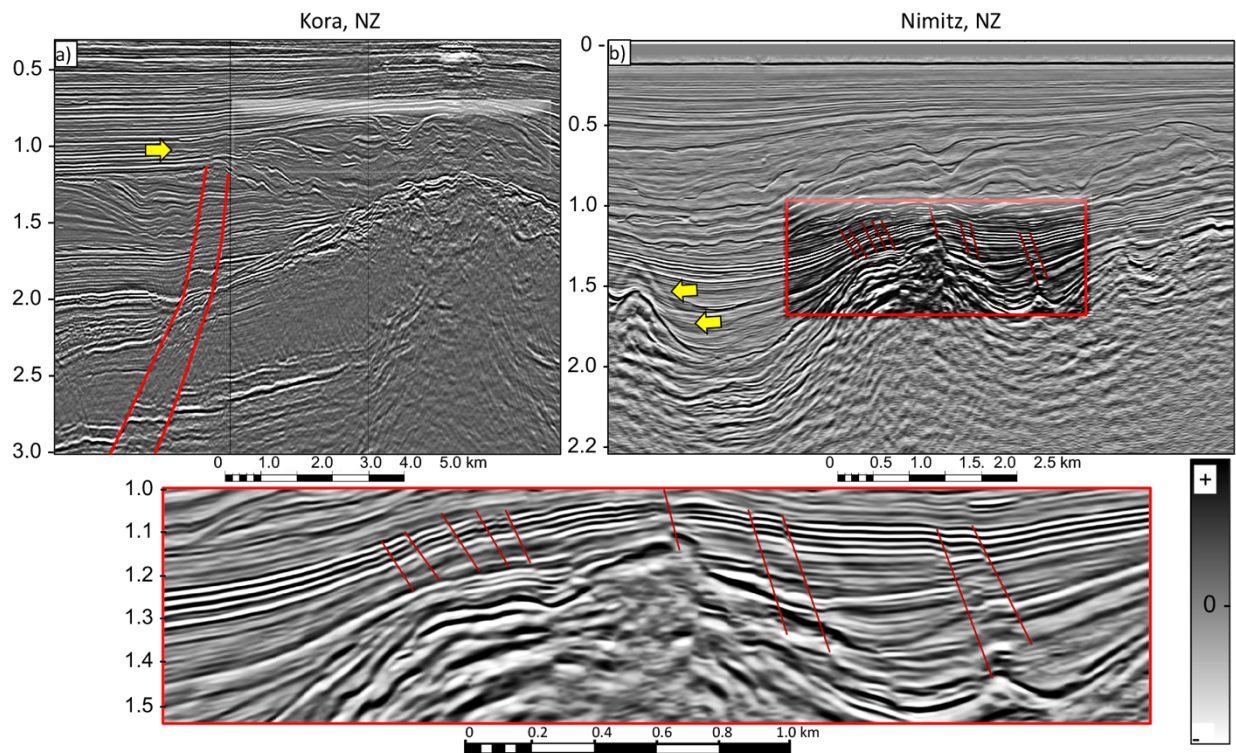


Figure D.5. The seismic appearance is associated with volcanic intrusions. Panel a) Kora arbitrary seismic line across the Kora intrusion. Yellow arrows highlight a few sedimentary structures. In particular, the thalweg – the central portion of the channel of most extensive channel present in the system correlates with a fault beneath it (channel centers are most likely to occur above a fault?). Panel b) Nimitz seismic section across the volcanic intrusion. Yellow arrows show different volcanic intrusions. The red rectangle is a zoom for the same seismic section to clearly show the faults above the volcanic intrusion.

As mentioned before, the objective is to evaluate whether machine learning algorithms can differentiate the semblance of fault patterns from a horizontal perspective - satellite view, time, depth, or horizon slice- among the foresaid geological features.

The radial faults "crown" faults typically occur in regional extensional deformation (Jackson and Hudec, 2017) or gravity spreading (Fort and Brun, 2012) framework. Crown faulting associated with salt diapirism has three stages: reactive, active, and passive (Hudec et al., 2013). Similar stages can describe the faults associated with volcanoes. Moving forward, we focus on the radial faulting aspect of these two different geological features in modern analogs and seismic reflection datasets with a horizontal resolution of 25 m, similar to the satellite images of the earth's surface ten years ago (Datasets acquired and processed more than 25 years ago).

Ideally, a patient Ph.D. student will pick all supra-salt or supra-volcano faults on all datasets. However, "*Ars longa vita brevis*" (i.e., skillfulness takes time and life is short), therefore the workflow we propose automatically selects the fault systems of both datasets. The proposed workflow includes four significant steps:

- 1) Data processing aims to condition the dataset to increase the discontinuities yielding a sharper contrast.
- 2) Attribute calculation and CNN fault prediction.
- 3) Self-organizing map using fault dip azimuth, fault dip magnitude, CNN fault probability, Thin-bed detector, and the seismic amplitude.
- 4) Automatic fault extraction using selected neurons representing faults in the Kohonen map in 2D and 3D.

In the first step, we used spectral balancing and structural-oriented filtering, or SOF henceforth is a filtering process that smooths along structural planes as a function of inline dip and crossline dip. SOF is an effective method of removing incoherent noise (Fehmers and Hocker, 2003), and different interpretation packages are available. The Calculation of SOF involves an anisotropic diffusion filter (Fehmers and Hocker, 2003) that enhances the capabilities of edge detection

methods such as the one used in the second step. Albeit the SOF modifies the relativeness of amplitude, we are only interested in emphasizing the fault edges.

The second step uses a recent machine learning application, the deep convolutional neural network, for fault prediction (Wu et al., 2019, Qi et al., 2020). First, we calculate the CNN fault prediction using the SOF amplitude data. Secondly, we skeletonize the fault prediction volume to transform it into a fault probability volume. The skeletonization is a sensible step that requires careful quality control to avoid tampering with the structural style and creating anomalous fault patterns; reader is invited to read (Qi et al., 2016, Li et al., 2022). The fault probability distribution is from 0 to 1, where 1 represents the highest probability.

Moreover, fault dip magnitude and azimuth are a byproduct of 3D fault skeletonization. The skeletonization enhances the resolution of the fault planes. It calculates fault dip and magnitude since part of the skeletonization process includes the 3D dip field and seismic coherency as a weighting function. It is noteworthy that there are multiple methodologies to generate a fault probability attribute, for example, Hale (2013), but we selected CNN fault prediction to input into the skeletonization process.

In the third step of the workflow, we used self-organizing maps or the Kohonen maps (Kohonen, 1990), including the attributes in Table D.1. Using self-organizing maps is ideal for automatic feature detection as they cluster mathematically independent attributes into geological interrelated weighting vectors from the inputs into a 2D space.

Similarly to Hussein (2021), the result transforms the five-dimensional space of the four attributes and the PSTM conditioned by the SoF process into a low-topology array of nodes (6x6). In contrast, we used four attributes. However, SOM is an unsupervised machine learning method since it does not require a label or training to perform the clustering. Let's consider the selection

of attributes (interpretational bias) and the selection of nodes or neurons that mean a fault plane (Figure D.6). We consider the process of fault clustering using SOM as a semi-supervised method.

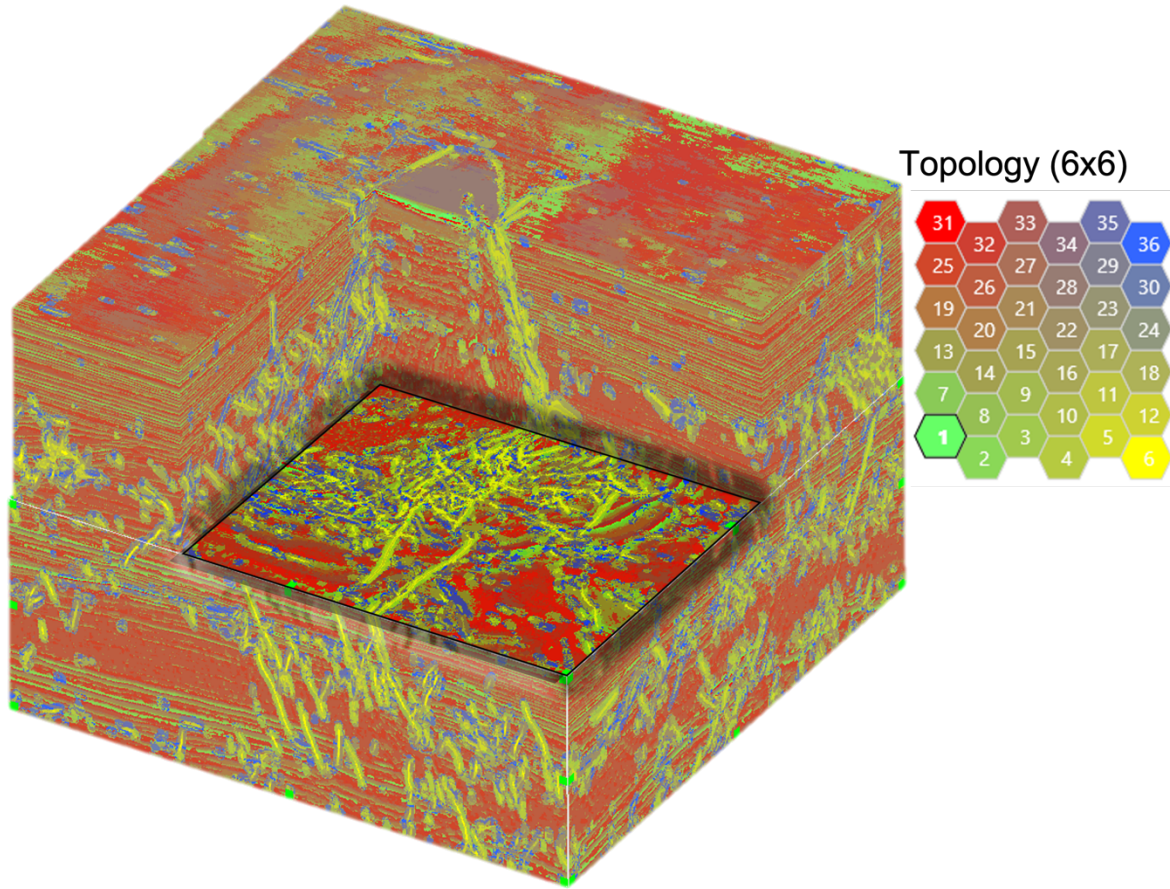


Figure D.6. 3D projection of a self-organizing map with a 6x6 topology showing clearly the nodes associated with faults. Mainly for these datasets, nodes, or neurons 11-12 and 17-18 have the highest probability associated with faults.

We should stress that the grid architecture among the four datasets is kept constant (6x6). However, the nodes representing a fault feature are independent of the datasets. The SOM grid of the Kora, the dataset is shown in Figure D.5 along with a seismic cube clustered with the same neuron grid. On the grid neurons 4,5,10,11, and 12 present faults in the Kora dataset, yet the same node numbers are related to parallel reflectors in the Gulf of Mexico datasets, both GoM A and GoM B. We changed the grid architecture, using 8x8 and 16x16 nodes, but fault system analysis using more than 36 nodes is unnecessary and time-consuming.

We should stress that the grid architecture among the four datasets is kept constant (6x6). However, the nodes representing a fault feature are independent of the datasets. The SOM grid of the Kora, the dataset is shown in Figure D.5 along with a seismic cube clustered with the same neuron grid. On the grid neurons 4,5,10,11, and 12 present faults in the Kora dataset, yet the same node numbers are related to parallel reflectors in the Gulf of Mexico datasets, both GoM A and GoM B. We changed the grid architecture, using 8x8 and 16x16 nodes, but fault system analysis using more than 36 nodes is unnecessary and time-consuming.

Finally, we compare two automatic fault extraction methods for the last step (step 4). The standard approach is the generation of fault geobodies using the transparency of the volume, typically CNN fault prediction or fault probability, even ant-tracking. Using the SOM volume and keeping opaque (visible) only the nodes associated with faults, we generated the fault geobodies. The second methodology extracts the fault patches from the selected nodes with fault dip magnitude and azimuth. The resulting geobody is rendered a new seismic attribute and inputted as a fault volume to extract fault patches or planes.

The second method yields better results, and it is geologically more meaningful. To examine why it is more meaningful, voxel definition or geobody capture is used, which overlooks any reference on the fault dip magnitude, fault dip azimuth, depth centroid, and surface area. The geobody shows fault presence and locations. On the other hand, fault planes calculate dip and azimuth that can be used to define whether a fault is critical-stressed or not i.e., when a fault is close to the friction angle, thus slowly failing, or if the fault is oriented parallel to the maximum horizontal stress.

RESULTS

As you can see from the map in Figure D.1, salt presence is evident in the high-resolution ocean bottom map. The first figure of our analysis, Figure D.7, shows the SOM fault results for the dataset associated with salt intrusion combined with Sobel similarity. Figure D.8, on the other hand, shows similar results for the fault systems corresponding to the datasets where volcanic intrusions are present, both from various locations in New Zealand, the Kora and the Nimitz surveys.

Figure D.9 and Figure D.10 contrast two different automatic fault extraction methods; the methods are (I) geobody extraction, where the faults are presented as voxels, and (II) fault patches, where faults are presented as planes with their accompanying feature intrusion. Following the same order. Figure D.9 is associated with the faults from the Gulf of Mexico, and Figure D.10 shows the automatic interpretation of the fault system product of volcanic intrusions. Given that it is possible to extract fault patches with multiple attributes (ant-tracking, similarity, fault probability), we compared the automatic extraction using the SOM nodes associated with faults. Comparing the pole dispersion presented in Figure D.11 shows that Figure D.11d is better defined than Figure D.11a. A statistical comparison between both pole plots indicates that the pole count in Figure D.11d is 70% smaller than in Figure D.11a. Finally, Figure D.12 directly compares the most representative and eye-catching examples of both intrusions against the modern geological equivalents on the surface.

SOM RESULTS

Let us begin analyzing the results in detail; Figures D.7a and D.7b show two images from the same dataset, with the difference that Figure D.7a is a time slice (horizontal cut) and Figure D.7b is a horizontal slice with the values extracted on the horizon. In both slices, radial faults are evident

and highlighted with yellow arrows. The cause of the radial faults or crown faults is a salt intrusion. In Figure D.7b, keeping opaque and visible the nodes associated with faults. Sobel similarity highlights the more evident faults on the SOM nodes than on the Sobel similarity attribute. For example, the fault to the southeast is highlighted with yellow arrows. Figure D.7c and Figure D.7d follow the same configuration for different salt intrusion and datasets. Figure D.7c shows the horizontal slice, and Figure D.7d represents the horizon slice with the same time (single horizon, actual value extraction) of the deposit affected by faults. Jackson 2020 proposed the mechanism for radial or crown faults due to salt emplacement. Comparing Figures D.7a and D.7b, the fault effect is present, yet the effect is exaggerated in Figures D.7b and D.7d. Selecting the exact value of the attributes as the extraction method to yield the most accurate seismic values on the horizon.

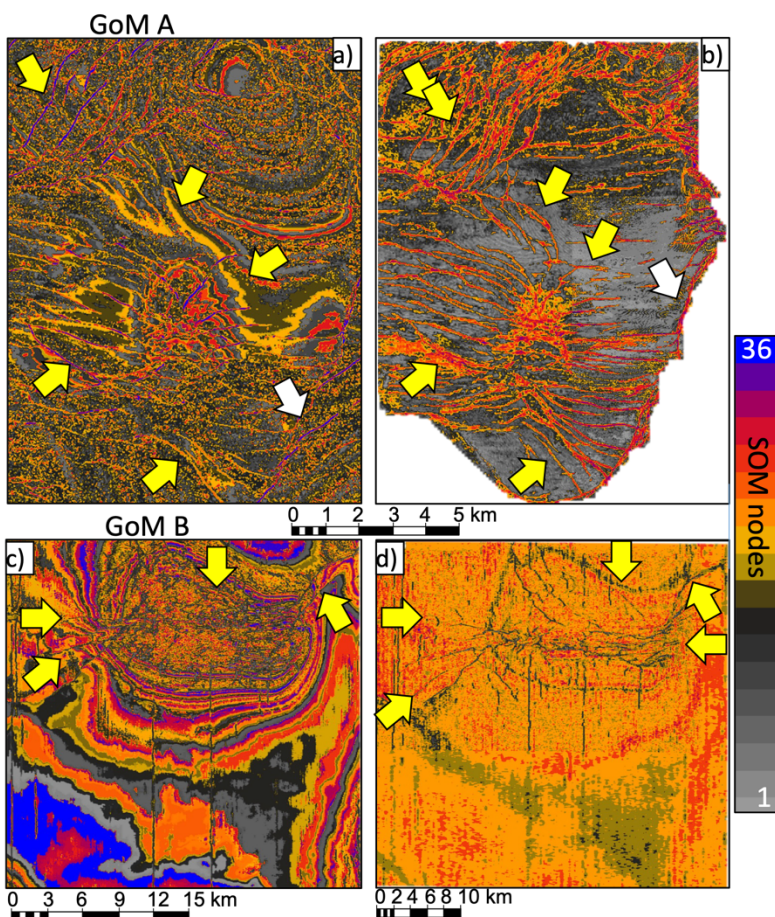


Figure D.7. Panel a) Time slice of the SOM fault classification associated with a salt dome in the Gulf of Mexico. (b) Horizon slice keeping the SOM nodes associated with faults opaque and corendering with Sobel similarity. On both slices, radial faults are evident. Panel (c) time slice of Self-Organizing Map on a different dataset on the Gulf of Mexico; in particular, the GoM B dataset confirms the salt intrusion on the seafloor. The ocean bottom collapsed and subserved deepwater sediment distribution. Panel (d) shows the horizon slice with better visualization of the radial faults, particularly the center of the fault collapsing.

The differentiation in the selection has two components, an esthetic component, i.e., the faults are more precise and better defined by the SOM. Faults are complex, despite characterizing them as planes, as the tortuosity associated with the faults is still a research subject. For example, Figure D.5 shows that the main fault has multiple ramifications. Thus, the horizontal slice captures at a specific time-or depth the effect of said fault on that specific time of deposit, and not so much the constant time slice. The second component is strategic to our objective since seeking to compare the horizon slice to present-day geology. Both perspectives (horizontal slice and satellite image of the subsurface) show the characteristics of the subsurface at the time of deposit. Naturally, the objective is to show the equivalent time of deposit and deformation and highlight the similitudes with salt domes currently visible in outcrops in Figure D.12.

Like Figure D.7, in Figure D.8, comparing the constant time slice against the horizon slice for the volcanic datasets, the horizons slice of Figure D.8, especially panels (b) and (d) exhibit an improvement in the fault detectability when comparing with panel (a) and panel (c). In contrast, with Nimitz dataset –panel b-, the Kora dataset has grid neurons between 4 and 6. While on the Nimitz dataset, grid neurons associated with faults around 28-32.

Using the yellow arrows in Figure D.8, fault trends are present in both datasets associated with the igneous intrusion. All yellow arrows highlight two elements; 1) The fault system's extension is associated with the intrusion and the radial pattern, while in contrast with the salt intrusion in Figure D.7, it is more discontinuous. The fact that both intrusions generate radial patterns opens different questions: how do the ductile and brittle rheology of the intrusions impact overburden. Does the geometry of the intrusions impact the fault pattern? It is crucial to solve the presented questions to extrapolate the results to planetary exploration tasks, for example, to identify landing sites for rovers and other exploratory probes.

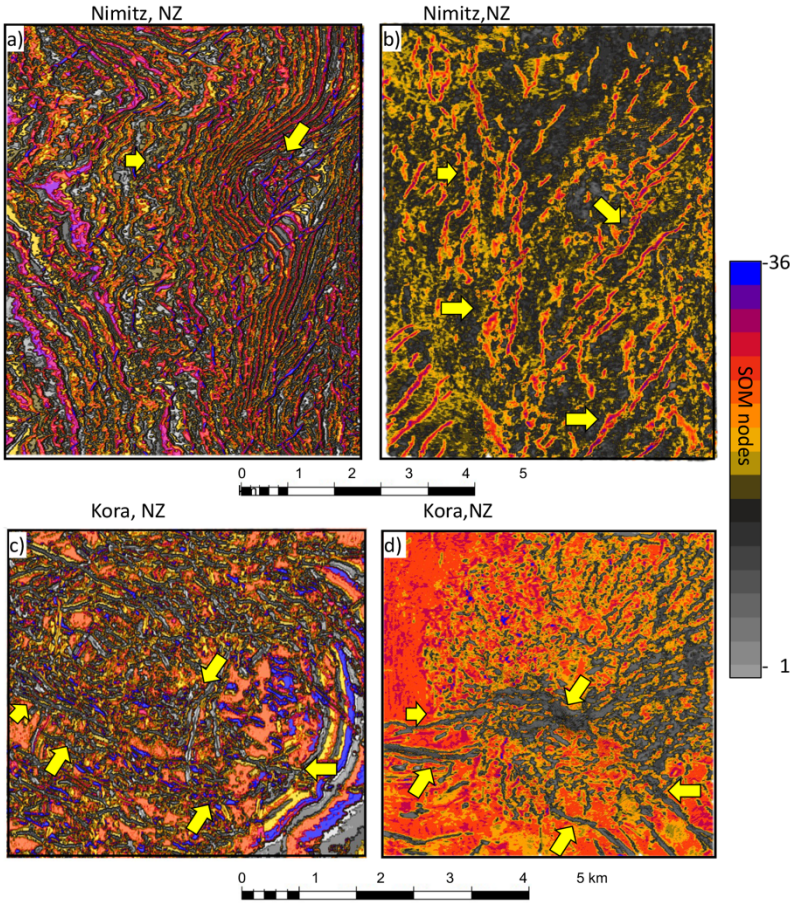


Figure D.8. a) Time slice of the SOM fault classification associated with a volcano in the New Zealand basin. b) Horizon slice with the same SOM fault classification. Radial faults are present in both slices, yet more evident on the horizon slice. Panel c) Time slice of Self-Organizing Map on the Kora volcano. Panel d) Horizon slice shows the radial pattern with a predominant fault on the southeast portion of the volcano cinder.

VOXEL AND FAULT PATCH INTERPRETATION

Figure D.9 contrasts two different methodologies to extract fault planes using the SOM of faults. The GoM B dataset helps evaluate the accuracy and speed of automatic fault detection. Figure D.9a shows the fault planes using geobody extraction on the Gulf of Mexico B dataset. Using two of the same seismic cubes from the Gulf of Mexico and Figures D.10a and D.10b do the same for the faults associated with the volcanic intrusion. Figure D.9a shows the geobody extraction using the SOM nodes identified as faults. The captured voxels, which visually represent faults from the SOM nodes with 26-way connectivity, are highlighted with yellow arrows. The straight planes in Figure D.9 captured as red voxels are missing or dead inlines (absent) on the seismic volume. As mentioned above, all SOM nodes are different among the datasets. The selected neurons are different in every SOM analysis, which means that neurons 4 to 6, which are

interpreted as a fault on dataset A from the Gulf of Mexico, are not necessarily a node that represents a fault on the Kora or Nimitz seismic information.

The voxel connection or geobody extraction parameters are kept constant among all fault extraction analyses to reduce variability; as seen in Figure D.9a, the red voxels represent different fault systems, not differentiable by azimuth direction or dip magnitude. Alternately, in Figure D.9b the fault patches are separated by azimuth fault dip magnitude. Figure D.9b only shows the faults dipping NE-SW. It is important to stress that the diapir walls have variable dips; on salt walls, the dip is steeper than the top of the salt diapir (horizontal). The node mixing might suggest that, as the SOF amplitude data is the pruning attribute, the discontinuity level of the faults is like the salt walls.

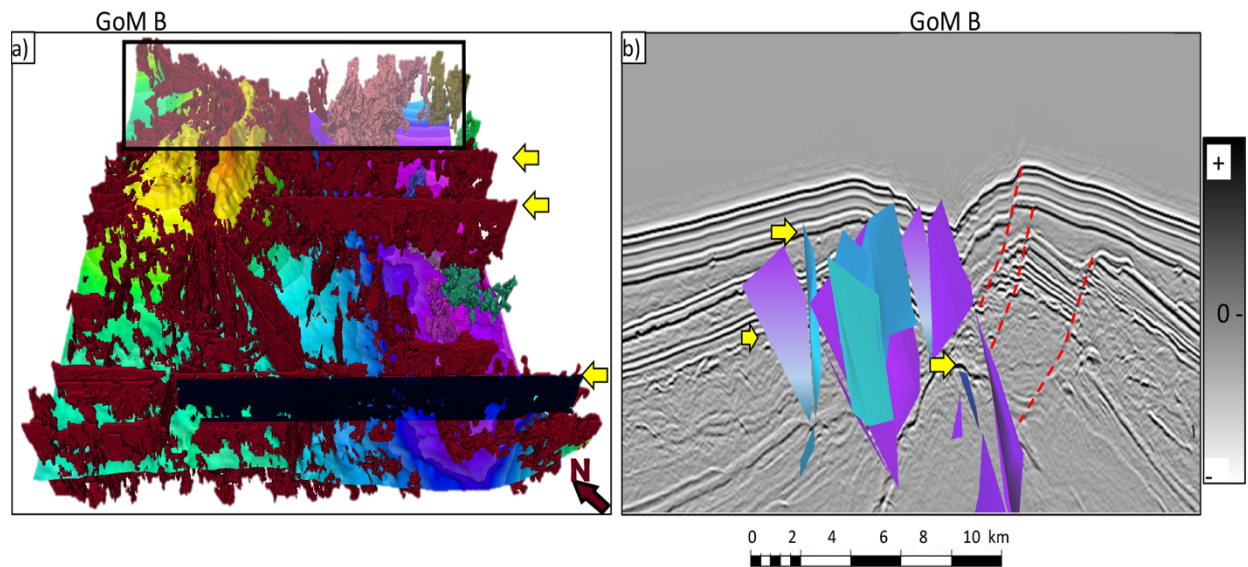


Figure D.9. The same dataset, GoM B presented in two different perspectives. Panel a) Voxel extraction of the fault system associated with selected SOM neurons. Panel b on the right shows the versatility of automatic fault extraction; using the selected SOM neurons, the fault planes are captured for the entire system and area of study. In Panel b) Fault planes are filtered. The filter removes low fault dip angle fault planes and faults with azimuth parallel to NE-SW. the high fault dip magnitude angle and similar fault dip azimuth are preserved.

To address how the fault planes, look for volcanic intrusions, Figure D.10 maps the fault planes with voxels and fault patches using the Nimitz (Figure D.10a) and Kora (Figure D.10b)

datasets, focusing on the voxel geobody extraction. The extraction quality varies due to geologic information (channels and canyons, sigmoid deposits) contaminating the fault definition.

However, mixing the SOM fault detection map with fault prediction and dip azimuth better separates the faults on the automatic fault extraction. The red inset square in Figure D.10a highlights the crustal faults as being a different system on top of one of the volcanic intrusions. Equally, the dashed cyan square shows the faults on top of the well-defined volcanic intrusion. Yellow arrows show well-defined planes that have the same genesis as the faults below but have a different color id. Since each geobody has a color code, it discerns each fault plane. However, the approach overlooks the geological genesis, that faults have the exact origin in most cases.

In Figure D.9b and Figure D.10b, we showed the automatic fault patch using a combination of attributes to create fault patches on both intrusions, Figure D.9b associated to salt intrusion and Figure D.10b associated with the volcanic intrusion. The input for extraction is the nodes from the SOM that represent a fault selected by the interpreter, which stresses the importance of context interpretation. The combination of transparencies increases the confidence in the selected fault planes when plotting the fault system and visualizing the classification resulting in a valuable tool for transfer learning. After the extraction of the fault planes, the stereographic projection is generated from the generated fault planes. In general, automatic fault planes are more valuable as fault planes than geobodies. In particular, the fault planes using the SOM fault attribute yield the most accurate fault planes, even more when combined using a cross-plot with control on the fault probability and dip or magnitude as the color axis.

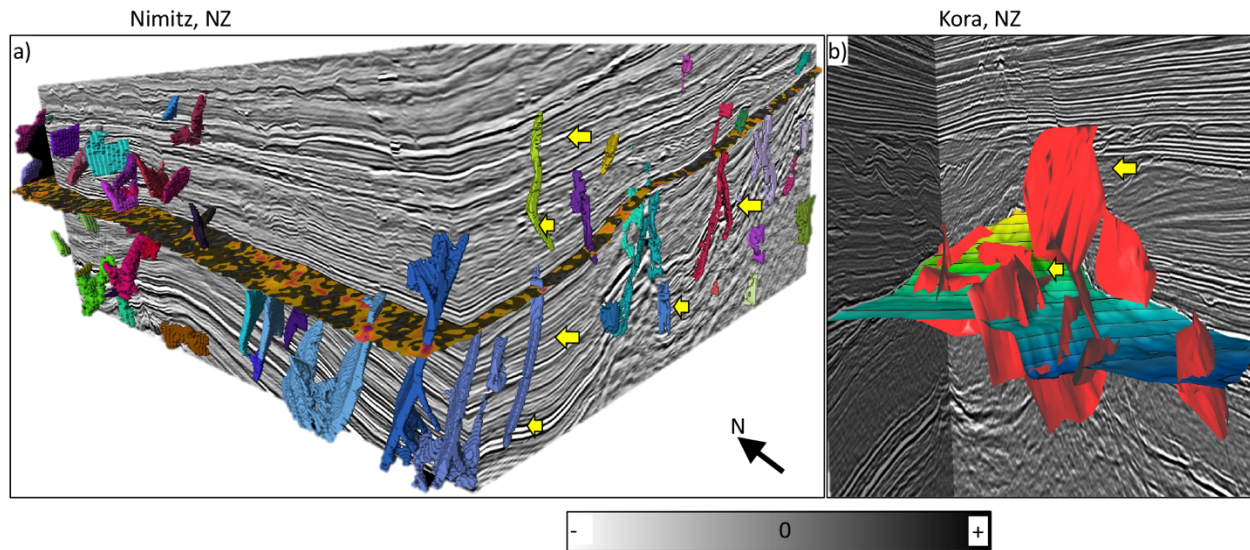


Figure D.10. a) Voxel extraction of the fault system associated with selected SOM neurons in a volcanic environment and the horizon slice with the SOM extraction b) demonstrates the versatility of automatic fault extraction. We capture the fault planes for the entire system using the selected SOM neurons, and using automatic fault extraction, we can extract specific values for each fault plane. For example, the top of the volcano with faults with a dip higher than 70 degrees is kept visible.

MODERN ENVIRONMENT COMPARISON

We selected two horizon slices to compare with similar modern environments from satellite imagery, one from the collapsing fault nucleated by a salt dome in the Gulf of Mexico (Figure D.9b) and the other from the Kora volcano (Figure D.10b). We compare the Landsat 8 imagery overlaying surface attributes, dip azimuth, elevation, and hill shade for the two locations of interest. The Mt. Taranaki Volcano (Figure D.13a) is located a hundred kilometers southeast of the Kora volcano and the salt dome in the Zagros Mountains in Bushier Province, Iran (Figure D.12b). A salt diapir that rapidly changes the region's topography has been described in detail due to the importance of halokinesis in the oil industry. The first horizon slice, the Kora horizon slice, has erosional channels and possible lava flows that exaggerate the similarities between the radial fault product of volcanic and salt intrusions.

The comparison of the Kora and Mt. Taranaki volcano shows well-defined cone craters and central vents with auxiliary cinders. The resolution of satellite image and seismic horizons slice are close

to 25 to 30 meters. Simultaneously, Figure D.11a and Figure D.11b show the modern environment for salt intrusion. The glacier salt or salt mountain from the Zagros mountains (Iran) and salt diapirs from the Gulf of Mexico reveal oblong shapes with their maximum direction parallel to the fault plane.

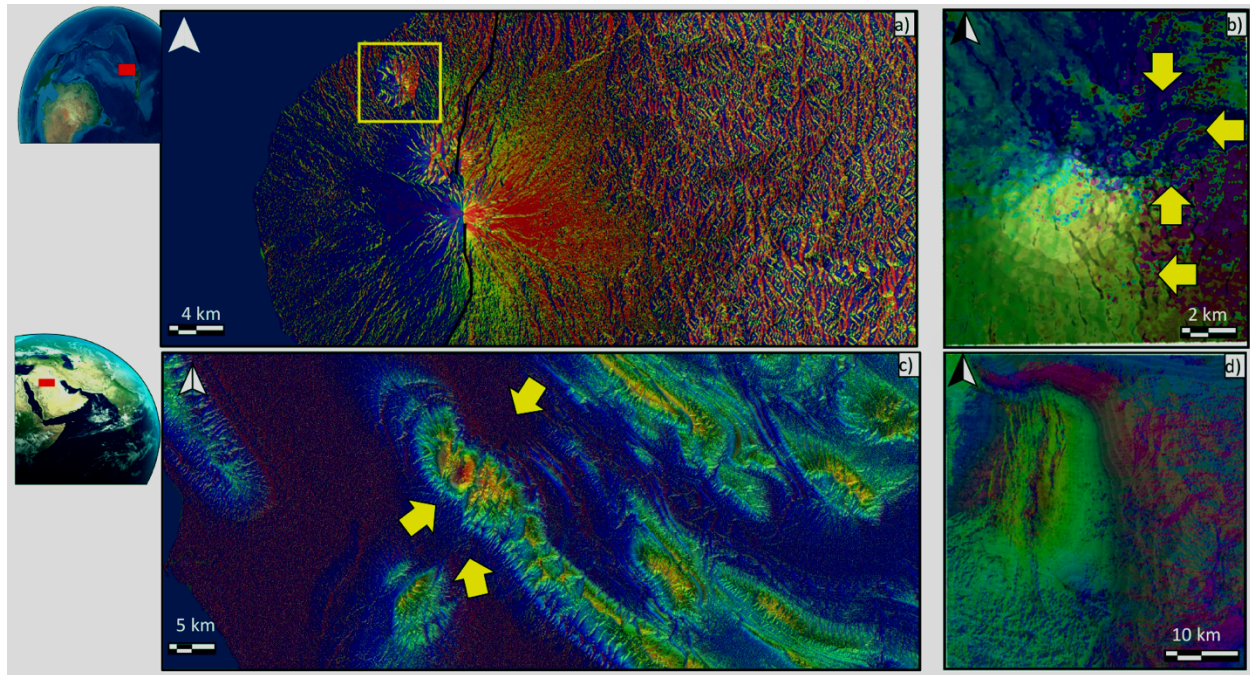


Figure D.11. a) Mount Taranaki overlaid with dip azimuth and hillshade and elevation, an image from the Landsat 8 and its sensors. A hyperbola separates the mount as it does in the Kora volcano. The radial faults present contrast with the drainage patterns very subtle. b) Kora volcano has a fault dividing the caldera next to the central cinder cone. The radial fault patterns are like the Taranaki Mount and the salt intrusion with collapsing blocks near the center of the intrusion. c) Zorghos salt intrusion and the fault trend mirror the salt intrusion on the Gulf of Mexico with a central portion collapsed by salt, an aperture, and a system of antithetic and synthetic faults. d) The salt dome triggers the seafloor collapse, which is dynamic; while the overburden increases around the salt diapir, it is constantly uplifted.

In Figure D.11a and Figure D.11b, the volcano perspective has three attributes overlain, slope ratio, dip azimuth, and dip magnitude. To exaggerate the similarities between the seismic attributes, we can use conventionally. Figure D.11a shows the synthetic faults collapsed on the center of the extrusion with antithetic highlighted with the yellow arrows. As the effect of the intrusion dampens, the coupled system (synthetic-antithetic fault) is mixed with the water drainage of the volcano. The kinematic changes are less evident yet, still present around the dome. It is

curious how the water drainage system can be easily confused with the radial fault planes; still, our work with SOM and attributes shows the difference in continuity and planarity. Also, the faults remain far away from the intrusions in contrast with drainage channels that disappear entirely at the slope change. With the salt comparison, it is evident that the faults extend predominantly perpendicular to the structure of the salt extrusion. In both figures, the salt collapse, and salt flows that creeps out of the main feeder toward the outsides create an elliptical shape. The salt flow follows the direction of the fault, explaining the preferred directionality of the salt distribution, while in contrast, the volcano intrusions are more circular.

All collaboration from seismic interpretation showing a pattern recognition on attributes common to both exploration techniques are valuable for remote sensing, not for geological structures where we can quickly go and study the area, but for planetary exploration or frontier plays with difficult access. We should capitalize on the experience using seismic data to interpret the nature of subsurface intrusions with only an educated guess from the integration of sparse data if we can calibrate the different examples from the open-access datasets to generate labels to train artificial intelligence algorithms using the same image perspective as satellite images or robotic probes that would use in space, for example, in planetary geology exploration.

ALTERNATIVE INTERPRETATIONS

The fault clustering around complex tectonic environments such as salt basins and volcanic arcs can further encompass multiple interpretations. Drainage channels around intrusions are perhaps the most robust alternative hypothesis, primarily due to erosion in the case of volcanoes and changes in sea level for salt domes that controlled the fluid direction. A steep slope will generate enough potential energy to promote flat (straight) channels radiating outside the intrusion. Another strong hypothesis applicable to volcanoes is lava flowing out of the cone. Even some

secondary cones can have lava flowing out in the radial pattern. Clustering channels and other sedimentary features using SOM is well documented and can easily capture the sedimentary features (La Marca et al., 2020).

DISCUSSION

Classifying the differences by hand is an unpleasant task despite the subtle differences mentioned above. The selected SOM cluster the faults that best define the fault systems and classify the origin to us. Context interpretation matters, and we, as interpreters still have control over what is classified as a fault when selecting the nodes. Our analysis with the correct geological context helps reconstruct the geological setting. In general, CNN is a supervised learning method, and SOM is a semi-supervised method; thus, we still have control as interpreters.

SOM nodes or neurons are arbitrary and cannot be correlated with each other. As seen in Figures D.6, Figure D.7, and Figure D.8, the accuracy of fault detection is the function of seismic quality. For example, the missing inline on the Gulf of Mexico is falsely detected as a part of the same fault system. However, the SOM clustering is reduced when using other edge attributes such as semblance of coherency. An alternative available to increase the accuracy of SOM for fault detection is to constrain the vertical analysis to a targeted window, but rarely are faults vertically limited by a constant time or depth window. Zeng & Hentz (2004) showed the benefits of horizon slices, benefits are also present when contrasting constant time slices against horizon slice.

The resemblance of SOM nodes with the drainage systems is remarkable in Figures D.11a, and D.11b and perhaps is our only interpretable feature to distinguish from volcanoes and salt diapirs. For example, in Figure D.11a, the Taranaki volcano has drainage marks and faults entangled, and semblance attributes only exaggerate the limits of said features. Looking closer at the yellow arrows, familiar pattern is observed; drainage systems have a straight pattern on steep

elevation, and at the slope break, the direction of the channel changes. On the other hand, faults radially extend away from the kept direction. The use of hillshade and color relief maps in Figure D.12, a satellite image from Mars exhibits an exciting example of discerning an intrusion from a map view using attributes and fault patterns. Can you guess the intrusion in Figure D.12 based on our discussion?



NASA/JPL-Caltech/Arizona State University

Figure D.12. Mars intrusion, local Mars time 15:23 acquired on 22 November 2006. Flight altitude 285.8 km. Solar latitude of reference 139 °, northern summer. Each pixel represents 57.2 cm, and the north is up. The geo (mars) morphology of the features inside the black box highlighted by the hillshade invites interpretation. In the blue box, the hillshade attribute helps to exaggerate the presence of a channelized system, or canyons maybe? Inside the yellow rectangle which intrusion is more likely? The pockmarks show the presence of craters from an asteroid impact. An alternative interpretation could be gas escaping the subsurface.

Conclusions

Nowadays, interpreters have many interpretation tools at the tip of their hands. Fault clustering using semi-supervised learning allows us to create 3D models of the tectonic settings in a fraction

of what it would take for an advanced interpreter. SOM detects faults successfully around the intrusions, but we cannot single out the origin of the intrusion using the information clustered. Hence pattern recognition and transfer learning are needed to complement the labeling and training process. For example, the ambiguity between two intrusions using only attributes extracted on the horizons representing the Kora Volcano top and salt diapir intrusions from the GoM with two similar modern geological environments. The modern geological environments selected here are the Mt. Taranaki Volcano in New Zealand and the Zagros salt glacier in Iran. Both modern geological environments are corendered with elevation, aspect ratio, and dip azimuth from Landsat-8 satellite data courtesy of USGS, NASA.

However, despite similarities, subtly different geometries can be discerned using our interpretation skills, and a certain degree of experience and bias. For example, the two main differences between these two fault style patterns to distinguish between salt or igneous intrusion are:

- 1) In both intrusion systems, the faults are well-defined straight planes. Nonetheless, the hydrological drainage patterns associated with the intrusions are different; for example, in volcanic intrusions, the drainage pathways can be confused with faults; however, when the slope ends, the channels meander and sometimes cause a slope floor fan. However, drainage patterns in volcanic intrusions are chaotic and different from fault planes. The SOM maps show straight lines defined as faults; channels are represented with other nodes.
- 2) Volcanic intrusions tend to be well-defined and round-shaped; even volcanic eruptions are discordant with their surroundings, taking advantage of pre-existing pathways. The dense nature of salt allows the intrusions to creep into irregular shape spaces. Salt tends to occupy the available space spilling out from the feeder. The SOM maps show that the elongated shape of the salt corresponds to the dip azimuth of the main fault.

The comparison is helpful for us as interpreters; however, the ML methods cannot differentiate between intrusion type (salt or volcanic) or fault style (normal, inverse, lateral). Nonetheless, these methods help underline edges and discontinuities present in the seismic data to provide the geological context.

Fault interpretation requires a broader window to capture the trends as accurately as possible. In some cases, where the signal-to-noise ratio (S/N) is below the threshold, the SOM indistinctively groups fault systems into the same neuron (Figure 9). Nevertheless, in datasets with higher S/N (Figure 10a), several SOM neurons independently represent both fault systems. In conclusion, drainage channel radial faults and stratigraphic features are tangled when the signal-to-noise ratio is low. While the channels have a similar dip azimuth with the fault, the dip magnitude has a different value, thus making the extraction more accurate and the classification easier with another dimension to differentiate similar structures. The horizon slice of the top of the intrusions improves the visualization.

On the other hand, automatic fault patch extraction shows multiple advantages versus voxel visualization, which is commonly used (Wu et al., 2019). A crucial flaw of geobody extraction is that voxels are visually aesthetic but computationally inefficient. The voxel representation has six faces (top, middle, bottom, left, right, center). Further computations require an unequivocal position (x, y, z); otherwise, it creates a debate on accuracy and further input from the user. Furthermore, automatic fault patch interpretation assigns a confidence value that allows the connection of the highest values to build the fault plane.

In conclusion, SOM combining CNN fault probability and fault geometry characteristics (fault dip magnitude and fault dip azimuth) improves the fault detectability but fails in determining the origin of faults since both CNN and SOM methods detect the intrusion as part of the fault

system, making the separation ambiguous. However, these features generate complex geoforms. The complexities of these geometries can hinder fault extraction in general, and our methodology could assist in labeling distinctive features. Primarily, fault dip magnitude and fault dip azimuth of the faults are directly related to the geological context that can satisfy constraints, for example, synthetic from antithetic faults. While these ML methods do not singularly solve our original investigation, if ML can be used to discriminate volcanic or salt origins based on fault geometry, these methods do aid, with interpreter insight, in gaining further confidence into the seismic FLT's origin.

ACKNOWLEDGMENTS

We would like to take the time and opportunity to thank New Zealand and the United States governments via the Department of Interior and The United States geological survey (USGS) for providing the 3D datasets, AASPI consortium, CGG, Geophysical Insight, and Schlumberger for the software licenses offered to the University of Oklahoma.

APPENDIX E: Elastic properties of an idealized sandstone

The elastic parameters used to estimate the rock matrix and what constitutes the hypothetical sandstone are presented in Table E.4.1, which presents the minerals and their elastic moduli used in the rock physics model. A small percentage of calcite and cement are included using the constant cement model and mixed using rock physics mixing models (Dvorkin & Nur, 1996).

Mineral	Mineral ratio [%]	Shear modulus [GPa]	Bulk's modulus [GPa]	Density [kg/m ³]
Quartz	70	45	36.6	2,650
Clay	15	9	25	2,580
Feldspar	10	25.9	75.9	2,630
Calcite	3	32	76.8	2,710
Cement	2	0.036	0.064	2,000

Table E.1. Minerals and percentages that constitute the model sandstone and the elastic moduli of each mineral are used to calculate dry and saturated velocities as a function of porosity.

The second step is defining the rock physics template (RPT), which yields the creation of the elastic moduli using and following the modified Hashin-Shtrikman bounds. The modified Hashin-Shtrikman bounds on the effective elastic properties are independent of grain texture or fabric for a two-phase material (Figure 4.3a and Figure 4.3b). The third step is the fluid substitution to include the effect of pressure, temperature, and different fluid types. The model uses the VpVs ratio and Acoustic impedance template or RPT (Figure E.1).

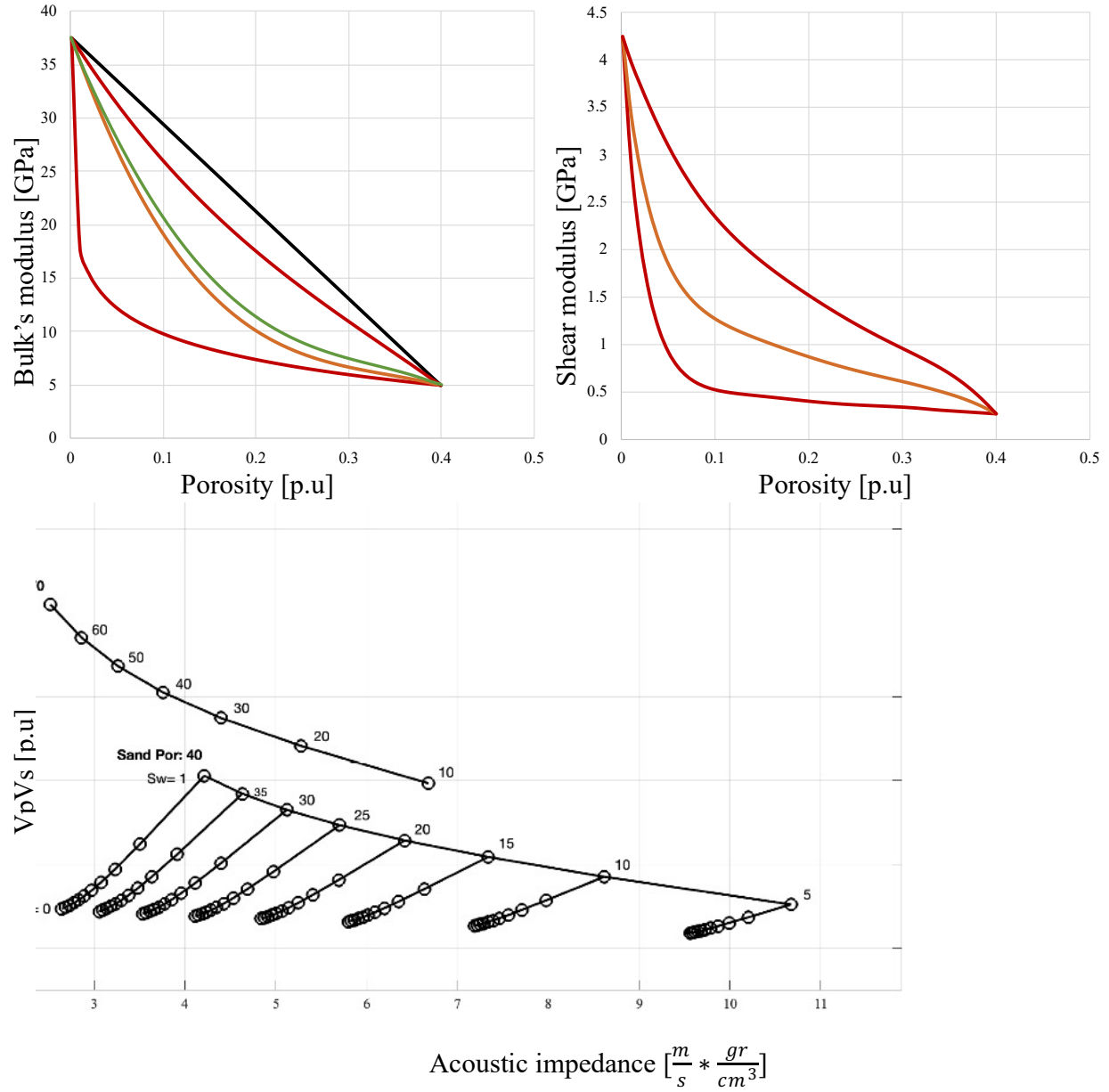


Figure E.1. a) Bulk's modulus versus porosity distribution for the minerals mixed in Table 2. The selected distribution is the modified Hashin-Shtrikman model. b) Shear modulus distribution against porosity. c) The result of fluid substitution for the given different fluid cases.

Having a continuous porosity distribution with physical bases of what constitutes a sandstone, and the effect of different water saturations, two facies (shale/sand) are selected to build a pseudo-borehole. Three sandstone intervals are selected from the VpVs ratio and acoustic impedance distribution. The model considers that each interval has 600 depth units, and each

reservoir has three different fluid contacts with a thickness of 200 depth units. The pseudo-borehole and the elastic logs are displayed in Figure 4.4.

APPENDIX F: Brief documentation on the clustering selection methods

SOM and other unsupervised clustering techniques require an initial guess for the clusters, and the selection of the initial cluster is not trivial. Heuristic methods are used to select the optimum number of clusters to validate the clustering. These techniques are heuristic methods used to validate grid selection. Figure A4 shows three clustering validation techniques.

- **The elbow method** calculates the sum of squares at each number of clusters, where the first derivation of the sum's square indicates the optimal number of clusters. This method is inexact but helpful.

- **The gap statistics**, which compares the total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data. The estimate of the optimal clusters will be a value that maximizes the gap statistic (i.e., that yields the most significant gap statistic). This means that the clustering structure is far from the random uniform distribution of points.

- **The Silhouette method** is a visualization that can help determine the optimal number of clusters. The average silhouette method computes the average silhouette of observations for different values of k . The optimal number of clusters k is the one that maximizes the average silhouette over a range of possible values for k . Notably, these methods are based on statistical methods; in our methodology, there is a range for the optimum number of clusters (4-10) based on context interpretation.

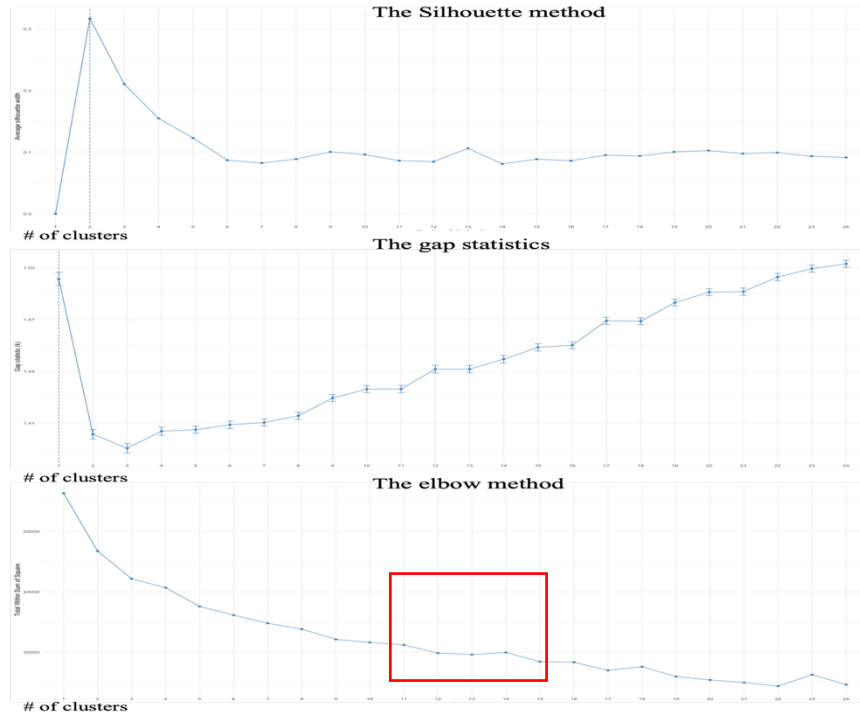


Figure F.1. Comparison of three clustering evaluation techniques. These three techniques yield a different suggestion for the number of clusters in the dataset. The red square is where the derivative of the elbow plot shows a constant. The elbow plot suggests 12 to 14 clusters, and the rest of the plots suggest 2 to 3 clusters.

REFERENCES

- Acquaviva-Luebbert, L. (2016). Quantitative Analysis of Dissimilarities Between Different Methods of Seismic Inversion to Facies Realizations \[Master of Science thesis, Stanford\].
- Acerbi, L. (2020). Variational Bayesian Monte Carlo with Noisy Likelihoods. In *Advances in Neural Information Processing Systems* 33:8211-8222 (paper + supplement on arXiv, NeurIPS Proceedings).
- Adebayo, A. R., Abdulraheem, A., & Olatunji, S. O. (2015). Artificial Intelligence Based Estimation of Water Saturation in Complex Reservoir Systems. *Journal of Porous Media*, 18, 893-906. <https://doi.org/10.1615/JPorMedia.v18.i9.60>
- Adegbite, J. O., Belhaj, H., & Bera, A. (2021). Investigations on the Relationship among the Porosity, Permeability and Pore Throat Size of Transition Zone Samples in Carbonate Reservoirs Using Multiple Regression Analysis, Artificial Neural Network and Adaptive Neuro-Fuzzy Interface System. *Petroleum*, 6, 321--332. <https://doi.org/10.1016/j.ptlrs.2021.05.005>
- Adeniyi, E. O., Tyne, R. L., Barry, P. H., Darrah, T. H., Hubbard, M. S., Myers, M. L., Shaw, C. A., Bowen, D. W., & Calavan, C. W. (2023). Origin of CO₂ in Upper Devonian Duperow Formation and the Bakken Petroleum System at Kevin Dome, Northwest Montana. *Chemical Geology*, 639, 121733. <https://doi.org/10.1016/j.chemgeo.2023.121733>
- Ahmed, T. H. (1946). *Reservoir Engineering Handbook* (3rd ed.). Elsevier/Gulf Professional.
- Ajayi, T., Gomes, J. S., & Bera, A. (2019). A review of CO₂ storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. *Petroleum Science*, 16(5), 1028--1063. <https://doi.org/10.1007/s12182-019-0340-8>
- Akhundi, H., Ghafoori, M., & Lashkaripour, G. R. (2014). Prediction of Shear Wave Velocity Using Artificial Neural Network Technique, Multiple Regression and Petrophysical Data: A Case Study in Asmari Reservoir (SW Iran). *Open Journal of Geology*, 4, 303--313. <https://doi.org/10.4236/ojg.2014.47023>
- Aki, K., & Richards, P. G. (2002). *Quantitative Seismology* (2nd ed.). Univ. Science Books.
- Al-Jawad, S. N., & Saleh, A. H. (2020). Flow Units and Rock Type for Reservoir Characterization in Carbonate Reservoir: Case Study, South of Iraq. *Journal of Petroleum Exploration and Production Technology*, 10, 1--20. <https://doi.org/10.1007/s13202-019-0736-4>
- Alimoradi, A., Moradzadeh, A., & Bakhtiari, M. R. (2011). Methods of water saturation estimation: Historical perspective. *Journal of Petroleum & Gas Engineering*, 2(3), 45--53.
- AlRegib, G., & AlRegib, A. (2005). Integration of seismic attributes using self-organizing maps. *Geophysics*, 70(2), P1--P9.
- Alves, T. M. (2012). Scale-relationships and geometry of normal faults reactivated during gravitational gliding of Albian rafts (Espírito Santo Basin, SE Brazil). *Earth and Planetary Science Letters*, 331--332, 80--96.

- Amini, H. (2018). Comparison of Xu-White, Simplified Xu-White (Keys & Xu) and Nur's Critical Porosity in Shaley Sands. *European Association of Geoscientists & Engineers*, 2018, 1--5. <https://doi.org/10.3997/2214-4609.201801016>
- Archie, G. E. (1952). Classification of carbonate reservoir rocks and petrophysical considerations. *AAPG Bulletin*, 36(2), 218--298.
- Arévalo-López, H. S., & Dvorkin, J. P. (2016). Porosity, mineralogy, and pore fluid from simultaneous impedance inversion. *The Leading Edge*, 35(5), 423--429. <https://doi.org/10.1190/tle35050423.1>
- Arévalo-López, H. S., & Dvorkin, J. P. (2017). Simultaneous impedance inversion and interpretation for an offshore turbiditic reservoir. *Interpretation*, 5(3), SL9--SL23. <https://doi.org/10.1190/INT-2016-0192.1>
- Arevalo Lopez, H., Wollner, U., & Dvorkin, J. (2018). Effect of Vs Predictors on AVO Response. *Geophysics*, 83, 1--49. <https://doi.org/10.1190/geo2017-0259.1>
- Armstrong, D. K., & Carter, T. R. (2010). *The Subsurface Paleozoic Stratigraphy of Southern Ontario*. Ontario Geological Survey, Mines and Minerals Division.
- Asquith, G., & Krygowski, D. (2006). *Basic Well Log Analysis* (Second Edition, 2nd ed., Vol. 1, No. 16). AAPG.
- Avseth, P. (2010). Explorational Rock Physics -- The Link Between Geological Processes and Geophysical Observables. In K. Bjorlykke (Ed.), *Petroleum Geoscience* (pp. 403--426). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-02332-3_18>
- Avseth, P., & Lehocski, I. (2016). Combining burial history and rock-physics modeling to constrain AVO analysis during exploration. *The Leading Edge*, 35(6), Article 6. <https://doi.org/10.1190/tle35060528.1>
- Avseth, P., & Lehocski, I. (2021). 3D Subsurface Modeling of Multi-Scenario Rock Property and AVO Feasibility Cubes\---An Integrated Workflow. *Frontiers in Earth Science*, 9, 160. <https://doi.org/10.3389/feart.2021.642363>
- Avseth, P., & Lehocski, I. (2021). 3D Subsurface Modeling of Multi-Scenario Rock Property and AVO Feasibility Cubes---An Integrated Workflow. *Frontiers in Earth Science*, 9, 160. <https://doi.org/10.3389/feart.2021.642363>
- Avseth, P., Mukerji, T., & Mavko, G. (2005). *Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk*. Cambridge: Cambridge University Press.
- Ba, J., Carcione, J. M., Du, Q., Zhao, H., & Muller, T. (2014). *Seismic Exploration of Hydrocarbons in Heterogeneous Reservoirs: New Theories, Methods, and Applications*. Elsevier Ltd. eBook ISBN: 9780124202054.
- Backus, G. (1962). Long-wave elastic anisotropy produced by horizontal layering. *Journal of Geophysical Research*, 67(11), 4427.

Bachu, S. (2000). Sequestration of CO₂ in geological media: Criteria and approach for site selection in response to climate change. *Energy Conversion and Management*, 41(9), 953–970. [https://doi.org/10.1016/S0196-8904\(99\)00149-1](https://doi.org/10.1016/S0196-8904(99)00149-1).

Ballinas, M. R., Bedle, H., & Devegowda, D. (2023). Supervised machine learning for discriminating fluid saturation and presence in subsurface reservoirs. *Journal of Applied Geophysics*, 217, 105192. <https://doi.org/10.1016/j.jappgeo.2023.105192>

Barajas-Olalde, C., Haffinger, P., Gisolf, D., Zhang, M., Droujinina, A., Doulgeris, P., Khatibi, S., Jin, L., Burnison, S. A., Hamling, J. A., & Gorecki, C. D. (2019). Simultaneous Time-Lapse WEB-AVO Inversion for Seismic Reservoir Monitoring: Application to CO₂ Enhanced Oil Recovery at the Bell Creek Oil Field. In *SEG Technical Program Expanded Abstracts 2019* (pp. 564–568). Society of Exploration Geophysicists. <https://doi.org/10.1190/segam2019-3216895.1>

Barnes, A. E. (2007). Redundant and useless seismic attributes. *Geophysics*, 72(3), 33–38.

Barnes, A. E. (2016). *Handbook of Poststack Seismic Attributes*. Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560803324>

Barnes, A. (2006). Too many seismic attributes? *CSEG RECORDER*, 31(03), Article 03. <https://www.csegrecorder.com/articles/view/too-many-seismic-attributes>

Barnes, A., & Laughlin, K. J. (2002). Investigation of methods for unsupervised classification of seismic data. 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221–2224. doi: 10.1190/1.1817152.

Batzle, M., & Wang, Z. (1992). Seismic properties of pore fluids. *Geophysics*, 57(11), Article 11.

Bendat, J. S., & Piersol, A. G. (2010). *Random Data: Analysis and Measurement*.

Bischoff, A. P., Nicol, A., & Beggs, M. (2017). Stratigraphy of architectural elements in a buried volcanic system and implications for hydrocarbon exploration. *Interpretation*, 5, SK141–SK159.

Bishop, M. G. (1999). Total Petroleum Systems of the Northwest Shelf, Australia: The Dingo-Mungaroo/Barrow and the Locker-Mungaroo/Barrow (USGS Numbered Series No. 99-50- E), Total Petroleum Systems of the Northwest Shelf, Australia: The Dingo-Mungaroo/Barrow and the Locker-Mungaroo/Barrow, Open-File Report. Geological Survey (U.S.). <https://doi.org/10.3133/ofr9950E>

Bosch, M., Mukerji, T., & Gonzalez, E. F. (2014). Seismic, rock physics, spatial models, and their integration in reservoir geophysics. In *Encyclopedia of Exploration Geophysics* (pp. M1-1–M1-56). Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560803027.entry7>

Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, 134(4), 477–492. <https://doi.org/10.1037/0033-2909.134.4.477>

Bradshaw, J., Sayers, J., Bradshaw, M. T., Kneale, R., Ford, C., Spencer, Lisk, M. (1998). Palaeogeography and its impact on the petroleum systems of the North West Shelf, Australia. In Purcell, P. G., & Purcell, R. R. (Eds.), *The Sedimentary Basins of Western Australia 2:*

Proceedings of the Petroleum Exploration Society of Australia Symposium, Perth, WA, 1998 (pp. 95–121).

Buist, C., Bedle, H., Rine, M., & Pigott, J. (2021). Enhancing Paleoreef Reservoir Characterization through Machine Learning and Multi-Attribute Seismic Analysis: Silurian Reef Examples from the Michigan Basin. *Geosciences*, 11, 142. <https://doi.org/10.3390/geosciences11030142>

Caetano, H. (2012). Integration of Seismic Information in Reservoir Models: Global Stochastic Inversion. *IntechOpen*. <https://doi.org/10.5772/52308>

Carcione, J., Gei, D., Picotti, S., Misnan, M. S., Amir Rashidi, M. R., Bakar, Z., Harith, Z., Bahri, N., & Hashim, N. (2020). Porosity and Permeability of the Overburden from Wireline Logs: A Case Study from Offshore Malaysia. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 6(3). <https://doi.org/10.1007/s40948-020-00172-y>

Castagna, J. P. (1993). AVO Analysis - Tutorial and Review. In M. M. Backus (Ed.), *Offset-Dependent Reflectivity - Theory and Practice of AVO Analysis* (pp. 3–36). Society of Exploration Geophysicists.

Castagna, J. P., & Backus, M. M. (1993). *Offset-Dependent Reflectivity - Theory and Practice of AVO Analysis*. Investigations in Geophysics. Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560802624>

Castagna, J. P., Batzle, M. L., & Eastwood, R. L. (1985). Relationships between compressional-wave and shear-wave velocities in elastic silicate rocks. *Geophysics*, 50, 571–581.

Cavcic, M. (2022, April 22). Australian giant seeking approval for subsea equipment removal. *Offshore Energy*. <https://www.offshore-energy.biz/australian-giant-seeking-approval-for-subsea-equipment-removal/>

Chenin, J., & Bedle, H. (2020). Multi-attribute machine learning analysis for weak BSR detection in the Pegasus Basin, Offshore New Zealand. *Marine Geophysical Research*, 41(4), 21. <https://doi.org/10.1007/s11001-020-09421-x>

Chenin, J., & Bedle, H. (2022). Unsupervised Machine Learning, Multi-Attribute Analysis for Identifying Low Saturation Gas Reservoirs within the Deepwater Gulf of Mexico, and Offshore Australia. *Geosciences*, 12(3), Article 3. <https://doi.org/10.3390/geosciences12030132>

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, PMLR, 1597–1607.

Chi, X. G., & Han, D. H. (2009). Lithology and Fluid Differentiation Using a Rock Physics Template. *Geophysics*, 28, 60–65. <https://doi.org/10.1190/1.3064147>

Chopra, S., & Castagna, J. P. (2014). *AVO*. Investigations in Geophysics. Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560803201>

- Chopra, S., & Marfurt, K. J. (2005). Seismic attributes—A historical perspective. *GEOPHYSICS*, 70(5), Article 5. <https://doi.org/10.1190/1.2098670>
- Chopra, S., & Marfurt, K. J. (2007). Seismic attributes for prospect identification and reservoir characterization. Society of Exploration Geophysicists.
- Chopra, S., & Marfurt, K. J. (2016). Understanding the seismic disorder attribute and its applications. *The Leading Edge*, 35(8), 695–702. <https://doi.org/10.1190/tle35080695.1>
- Chopra, S., Castagna, J., & Portniaguine, O. (2006). Seismic resolution and thin-bed reflectivity inversion. *CSEG Recorder*, 31(1), 19–25.
- Chopra, S., & Marfurt, K. J. (2007a). Curvature attribute applications to 3D surface seismic data. *The Leading Edge*, 26, 404–414. doi: 10.1190/1.2723201.
- Chopra, S., & Marfurt, K. J. (2007b). Seismic attributes for prospect identification and reservoir characterization. Society of Exploration Geophysicists.
- Chopra, S., Marfurt, K. J., & Kumar, R. (2014). Seismic discontinuity attributes and Sobel filtering. 84th Annual International Meeting, SEG, Expanded Abstracts, 1624–1628. doi: 10.1190/segam2014-0465.1.
- Clairmont, R., & Bedle, H. (2020). Mass transport deposit or complex channel system? A case study of the 3D Nimitz seismic survey, Taranaki Basin, New Zealand. *Interpretation*, 8, SR59–SR64.
- Clairmont, R., Kolawole, F., Omale, A. P., & Bedle, H. (2021). Controls of pre-existing structures on clinoform architecture and the associated progradational system elements. *Basin Research*, 33, 875–902.
- Clochard, V., DeVault, B. C., Bowen, D., Delépine, N., & Wangkawong, K. (2018). Quadri-joint inversion: Method and application to the Big Sky 9C 3D data set in northern Montana. *Interpretation*, 6(4), SN101–SN118. <https://doi.org/10.1190/INT-2018-0035.1>
- Coleman, A., Jackson, C., Duffy, O., & Nikolinako, M. (2018). How where and when do radial faults grow near salt diapirs? *Geology*, 46(7), 655–658. doi: <https://doi.org/10.1130/G40338.1>
- Coléou, T., Poupon, M., & Azbel, K. (2003). Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge*, 22(10), Article 10. <https://doi.org/10.1190/1.1623635>
- Côte, G., MacBeth, C., & Amini, H. (2019). North Sea Field Application of 4D Bayesian Inversion to Pressure and Saturation Changes. In 81st EAGE Conference and Exhibition 2019.
- Côte, G., Dramsch, J., Amini, H., & MacBeth, C. (2020). Deep Neural Network Application for 4D Seismic Inversion to Changes in Pressure and Saturation: Optimizing the Use of Synthetic Training Datasets. *Geophysical Prospecting*, 68, 2164–2185. <https://doi.org/10.1111/1365-2478.12982>

- Damiani, D., & Rodosta, T. (2018). Kevin Dome Project final report (Final 2018–11; p. 2). National Energy Technology Laboratory. <https://www.netl.doe.gov/sites/default/files/2018-11/Kevin-Dome-Project.PDF>
- Daley, T. M., Schoenberg, M. A., Rutqvist, J., & Nihei, K. T. (2006). Fractured Reservoirs: An Analysis of Coupled Elastodynamic and Permeability Changes from Pore-Pressure Variation. *Geophysics*, 71, O33–O41. <https://doi.org/10.1190/1.2231108>
- Dandekar, A. Y. (2013). *Petroleum Reservoir Rock and Fluid Properties* (2nd ed.). CRC Press.
- Deutsch, C. V. (2006). A sequential indicator simulation program for categorical variables with point and block data: Block SIS. *Computers & Geosciences*, 32(10), 1669–1681. <https://doi.org/10.1016/j.cageo.2006.03.005>
- De Matos, M. C., & Marfurt, K. (2014). Complex wavelet transform spectral broadening. In *SEG Technical Program Expanded Abstracts 2014* (pp. 1465–1469). Society.
- De Matos, M. C., Osorio, P. L., & Johann, P. R. (2007). Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *GEOPHYSICS*, 72(1), P9–P21. <https://doi.org/10.1190/1.2392789>
- Dos Santos, A. C., Bonifácio, J. F., Monteiro, L. G. P., Rocha-Júnior, E. R. V., de Guerra, L. S., Heilbron, M., Magini, C., Jeck, I. K., & Bruno, H. (2022). Chapter 12 - Brazilian Equatorial Margin: Magmatic Genesis and Evolution. In A. C. dos Santos & P. C. Hackspacher (Eds.), *Meso-Cenozoic Brazilian Offshore Magmatism* (pp. 433–472). Academic Press.
- Downie, R. C., Grapenthin, R., Boylan, T., (2019). Shale gouge ratio theory—An advanced workflow for analysing lateral sealing. *Petroleum Geoscience*, 25(4), 413–422.
- Duffaut, K., Alsos, T., Rognø, H., Al-Najjar, N. F., & Landrø, M. (2000). Shear-wave elastic impedance. *The Leading Edge*, 19(11), 1222–1229. <https://doi.org/10.1190/1.1438510>
- Duncan, G., Hurren, C., Hill, R., Stanley, M., Woodward, J., & Lumley, D. (2013). The Stybarrow Field—A 4D Case Study. *ASEG Extended Abstracts*, 2013(1), 1–5. <https://doi.org/10.1071/ASEG2013ab276>
- Dvorkin, J., & Nur, A. (1996). Elasticity of high-porosity sandstones: Theory for two North Sea data sets. *GEOPHYSICS*, 61(5), Article 5. <https://doi.org/10.1190/1.1444059>
- Efnik, M. S., Taib, S. H., & Islami, N. (2016). Using time-lapse 4D seismic to monitor saturation changes in carbonate reservoirs. *Journal of the Geological Society of India*, 88, 736–742. <https://doi.org/10.1007/s12594-016-0541-3>
- Eide, C., Schofield, N., Lecomte, I., Buckley, S., & Howell, J. (2017). Seismic interpretation of sill complexes in sedimentary basins: Implications for the sub-sill imaging problem. *Journal of the Geological Society*, 175. <https://doi.org/10.1144/jgs2017-096>

Ementon, N., Hill, R., Flynn, M., Motta, B., & Sinclair, S. (2004). Stybarrow oil field — From seismic to production, the integrated story so far. Asia Pacific Oil and Gas Conference and Exhibition. <https://doi.org/10.2118/88574-MS>

Esposito, C. R., Shen, Z., Tornqvist, T. E., Marshak, J., & White, C. (2017). Efficient retention of mud drives land building on the Mississippi Delta plain. *Earth Surface Dynamics*, 5, 387–397. <https://doi.org/10.5194/esurf-5-387-2017>

Everitt, B., & Dunn, G. (2001). *Applied multivariate data analysis* (Vol. 2). London: Arnold.

Exon, N. F., & Willcox, J. B. (1978). Geology and petroleum potential of Exmouth Plateau area off Western Australia. *AAPG Bulletin*, 62, 40–72. <https://doi.org/10.1306/C1EA47F2-16C9-11D7-8645000102C1865D>

Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P. J., & Levitt, P. R. (1994). Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique. *GEOPHYSICS*, 59(9), Article 9. <https://doi.org/10.1190/1.1443695>.

Fehmers, G. C., & Hocker, C. (2003). Fast structural interpretation with structure-oriented filtering. *Geophysics*, 68, 1286–1293.

Gartrell, A., Keep, M., van der Riet, C., Paterniti, L., Ban, S., & Lang, S. (2022). Hyperextension and polyphase rifting: Impact on inversion tectonics and stratigraphic architecture of the North West Shelf, Australia. *Marine and Petroleum Geology*, 139, 105594.

Gegenhuber, N. (2015). Application of Gassmann's Equation for Laboratory Data from Carbonates from Austria. *Austrian Journal of Earth Sciences*, 108, 239–244. <https://doi.org/10.17738/ajes.2015.0024>

Gilks, W., Richardson, S., & Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice* (Vol. 1). Taylor & Francis.

Glinsky, M. E., Asher, B., Hill, R., Flynn, M., Stanley, M., Gunning, J., Thompson, T., Kalifa, J., Mallat, S., White, C., & Renard, D. (2005). Integration of uncertain subsurface information into multiple reservoir simulation models. *The Leading Edge*, 24(10), 990–999. <https://doi.org/10.1190/1.2112372>

Goodfellow, I. (2016). NIPS 2016 Tutorial: Generative Adversarial Networks. *Communications of the ACM*, 63(4).

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2016). Generative Adversarial Networks. *Communications of the ACM*, 63(4).

Goodway, B., Chen, T., & Downton, J. (1997). Improved AVO fluid detection and lithology discrimination using Lamé petrophysical parameters, lambda, mu and Fluid stack, from P and S inversions. In *SEG Technical Program Expanded Abstracts 1997* (pp. 183–186). Society of Exploration Geophysicists. <https://doi.org/10.1190/1.1885795>

Grammer, G. M. III, W. B. H., & Barnes, D. A. (2018). Paleozoic Stratigraphy and Resources of the Michigan Basin. Geological Society of America.

Grana, D., & Mukerji, T., Doyen, P. (2021). Seismic Reservoir Modeling: Theory, Examples, and Algorithms. United Kingdom: Wiley.

Grana, D., & E. Della Rossa (2010). Probabilistic petrophysical properties estimation integrating statistical rock physics with seismic inversion. *Geophysics*, 75(3), O21–O37. <https://doi.org/10.1190/1.3386676>.

Guo, Q., Ba, J., Luo, C., & Pang, M. (2021). Seismic rock physics inversion with varying pore aspect ratio in tight sandstone reservoirs. *Journal of Petroleum Science and Engineering*, 207, 109131. <https://doi.org/10.1016/j.petrol.2021.109131>

Ha, T. N., Lubo-Robles, D., Marfurt, K. J., & Wallet, B. C. (2021). An in-depth analysis of logarithmic data transformation and per-class normalization in machine learning: Application to unsupervised classification of a turbidite system in the Canterbury Basin, New Zealand, and supervised classification of salt in the Eugene Island minibasin, Gulf of Mexico. *Interpretation*, 9(3), Article 3. <https://doi.org/10.1190/INT-2021-0008.1>

Hale, D. (2013). Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. *Geophysics*, 78(2), 33–43. <https://doi.org/10.1190/GEO2012-033.1>

Hall, F., & Wang, Y. (2009). Elastic wave modelling by an integrated finite difference method. *Geophysical Journal International*, 177(1), 104–114. <https://doi.org/10.1111/j.1365-246X.2008.04065.x>

Harmon, R. S., Thompson, P., Schwarcz, H. P., & Ford, D. C. (1978). Late Pleistocene Paleoclimates of North America as Inferred from Stable Isotope Studies of Speleothems. *Quaternary Research*, 9, 54–70. [https://doi.org/10.1016/0033-5894\(78\)90082-0](https://doi.org/10.1016/0033-5894(78)90082-0)

Harrison, W. (2020). Assessment of Enhanced Oil Recovery Using Carbon Dioxide in Michigan Basin Silurian Pinnacle Reefs. U.S. Department of Energy Report, 2, DOE --42589—Michigan Basin, 1773069. <https://doi.org/10.2172/1773069>

Haugen, J., Mispel, J., & Arntsen, B. (2009). Seismic Imaging Below "Dirty" Salt: 71st EAGE Conference and Exhibition Incorporating SPE EUROPEC 2009.

Hayes, M. H. (1996). Statistical Digital Signal Processing and Modeling. Wiley.

Haynie, J. M. (2009). Characterization and modeling of petrofacies and pore-volume distribution within a gas-storage reservoir, Ray Reef Field, Southern Michigan Basin, Michigan (M.S. Geology thesis). Associate Professor Matthew J. Pranter (Advisor).

- Helland-Hansen, D., Magnus, I., Edvardsen, A., & Hansen, E. (1997). Seismic inversion for reservoir characterization and well planning in the Snorre field. *The Leading Edge*, 16(03), Article 03.
- Hennip, R. D., (1973) History of the crude oil industry in Montana. Graduate Student Theses, Dissertations, & Professional Papers. 5846. <https://scholarworks.umontana.edu/etd/5846>.
- Hermana, M., Lubis, L. A., Ghosh, D. P., & Sum, C. W. (2016). New Rock Physics Template for Better Hydrocarbon Prediction. *OnePetro*. <https://doi.org/10.4043/26538-MS>
- Hill, R., O'Halloran, G., Napalowski, R., Wanigratne, B., Elliott, A., & Jackson, M. (2008). Development of the Stybarrow field, Western Australia. *SPE*, 115373(9), Article 9.
- Hocking, R. M., Moors, H. T., & Van de Graaff, W. J. E. (1987). Geology of the Carnarvon Basin, Western Australia: Geological Survey of Western Australia Bulletin.
- Hong Bing, L. I., JiaJia, Z., ShengJuan, C. a. I., & HaoJie, P. a. N. (2019). 3D rock physics template for reservoirs with complex pore structure. *Chinese Journal of Geophysics*, 62, 2711–2723. <https://doi.org/10.6038/cjg2019K0672>
- Hongliu Zeng, & Tucker F. Hentz. (2004). High-frequency sequence stratigraphy from seismic sedimentology: Applied to Miocene, Vermilion Block 50, Tiger Shoal area, offshore Louisiana. *AAPG Bulletin*, 88, 153-174. <https://doi.org/10.1306/10060303018>.
- Hudec, M., Jackson, M., & Peel, F. (2013). Influence of deep Louann structure on the evolution of the northern Gulf of Mexico. *AAPG Bulletin*, 97, 1711–1735.
- Hussein, M., Stewart, R., & Wu, J. (2021). Which seismic attributes are best for subtle fault detection? *Interpretation*, 9(2), T299-T314.
- Hussein, M., Stewart, R.R, Sacrey D., Johnston, D. H., & Wu J. (2021). b Unsupervised machine learning for time-lapse seismic studies and reservoir monitoring. *Interpretation*, 9(3), T791–T807. <https://doi.org/10.1190/INT-2020-0176.1>
- Infante-Paez, L., & Marfurt, K. J. (2017). Seismic expression and geomorphology of igneous bodies: A Taranaki Basin, New Zealand, case study. *Interpretation*, 5(3), SK121–SK140. <https://doi.org/10.1190/INT-2016-0244.1>
- Iqbal, R., Ali, I., & Saira. (2017). Petro-physical investigation and reservoir characterization through well logging. *Journal of Applied and Emerging Sciences*, 7(1), 1–11. <https://doi.org/10.36785/jaes.71210>
- Jackson, C., Jackson, M., Hudec, M., & Rodriguez, C. (2014). Internal structure, kinematics, and growth of a salt wall: Insights from 3-D seismic data. *Geology*, 42, 307–310.

- Jackson, M. P. A., & Hudec, M. R. (2017). *Salt Tectonics: Principles and Practice*. Cambridge University Press.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
- Jain, S. (2014). Volcanoes. In: *Fundamentals of Physical Geology*. Springer Geology. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1539-4_16
- Jodry, R. L. (1969). Growth and dolomitization of Silurian reefs, St. Clair County, Michigan. *American Association of Petroleum Geologists Bulletin*, 53, 957-981.
- Jones, I. F., & Davison, I. (2014). Seismic imaging in and around salt bodies. *Interpretation*, 2, SL1–SL20.
- Jolly, B., Anyiam, O., & Omeru, T. (2017). Structural controls on channel-related seismic facies distribution in the toe-thrust of deepwater Niger Delta. *Journal of African Earth Sciences*, 125, 151–165. <https://doi.org/10.1016/j.jafrearsci.2016.11.012>
- Juhasz, I. (1979). The Central Role of Qv and Formation-Water Salinity in the Evaluation of Shaly Formations. *The Log Analyst*, 20(4), 3–13. SPWLA1979-vXXn4a1
- Kallweit, R., & Wood, L. C. (1982). The Limits of Resolution of Zero-Phase Wavelets. <https://doi.org/10.1190/1.1441367>
- Kausik, R., Prado, A., Venkataramanan, L., Datir, H., & Johansen, Y. (2021). Dual neural network architecture for determining permeability and associated uncertainty. *SPWLA 61th Annual Logging Symposium*.
- Keller, E. F. (2008). *Organisms, Machines, and Thunderstorms: A History of Self-Organization, Part One. Historical Studies in the Natural Sciences*, 38(1), 45–75. <https://doi.org/10.1525/hsns.2008.38.1.45>
- Kelman, A., Nicoll, R., Kennard, J., Mory, A., Mantle, D., Poidevin, S., Bernardel, G., Rollet, N., & Edwards, D. (2013). Northern Carnarvon Basin biozonation and stratigraphy. *Geoscience Australia*.
- Knott, C. G. (1899). *Reflection and Refraction of Elastic Waves with Seismological Applications*. *Philosophical Magazine*, 48, 64–97. Biodiversity Heritage Library
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480. doi: 10.1109/5.58325.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>

Kohonen, T. (1990). The self-organizing map. In *Proceedings of the IEEE*, 1464-1480.

Kuwahara, M., Hachimura, K., Eiho, S., & Kinoshita, M. (1976). *Digital processing of biomedical images*. Plenum Press, 187–203.

Kumar, D. (2006). A Tutorial on Gassmann Fluid Substitution: Formulation, Algorithm and Matlab Code. *Geohorizons*, 9.

La Marca, K., Bedle, H., Stright, L., Marfurt, K. J., & Langenkamp, T. (2021). Sensitivity analysis of seismic attributes parametrization for interpretation of a multi-story deepwater channel system: Tres Pasos Formation, Magallanes Basin Chile. In *First International Meeting for Applied Geoscience & Energy, Expanded Abstracts* (pp. 1191–1195). <https://doi.org/10.1190/segam2021-3584026.1>.

La Marca, K. (2020). Seismic attribute optimization with unsupervised machine learning techniques for deepwater seismic facies interpretation: Users vs machines. MSc thesis, The University of Oklahoma. <https://shareok.org/handle/11244/324893>

La Marca, K., & Bedle, H. (2022). Deepwater seismic facies and architectural element interpretation aided with unsupervised machine learning techniques: Taranaki Basin, New Zealand. *Marine and Petroleum Geology*, 136, 105427. <https://doi.org/10.1016/j.marpetgeo.2021.105427>.

La Marca, K., Lubo-Robles, D., Marfurt, K. J., & Wallet, B. C. (2021). An in-depth analysis of logarithmic data transformation and per-class normalization in machine learning: Application to unsupervised classification of a turbidite system in the Canterbury Basin, New Zealand, and supervised classification of salt in the Eugene Island minibasin, Gulf of Mexico. *Interpretation*, 9(3), Article 3. <https://doi.org/10.1190/INT-2021-0008.1>

Lech, M. (2013). New observations of the post-Triassic succession in the central Beagle Sub-basin, Northern Carnarvon Basin, Northwest Shelf, Australia. In M. Keep & S. J. Moss (Eds.), *The Sedimentary Basins of Western Australia 4. Proceedings of the Petroleum Exploration Society of Australia Symposium, Perth, 2013* (pp. 32).

Li, J., Wu, X., Ye, Y., Yang, C., Hu, Z., Sun, X., & Zhao, T. (2023). Unsupervised contrastive learning for seismic facies characterization. *GEOPHYSICS*, 88(1), WA81–WA89. <https://doi.org/10.1190/geo2022-0148.1>

- Li, Y., Zhong, B., Xu, X., & Liang, Z. (2022). Fault detection by skeleton extraction based on orientation field consistency. *PLoS ONE*, 17(7), e0271615. <https://doi.org/10.1371/journal.pone.0271615>
- Locke, A. (2004). WA-255-P(2), Stybarrow-1/1CH Well Completion Report - Interpretive Volume. BHP Billiton Petroleum Report.
- Lorenz, J. C. (1982). Lithospheric flexure and the history of the sweetgrass arch Northwestern Montana. *Geological Studies of the Cordilleran Thrust Belt*, Vol 1. Rocky Mountain Association of Geologists.
- Lubo-Robles, D., Bedle, H., Marfurt, K. J., & Pranter, M. J. (2023). Evaluation of principal component analysis for seismic attribute selection and self-organizing maps for seismic facies discrimination in the presence of gas hydrates. *Marine and Petroleum Geology*, 150, 106097. <https://doi.org/10.1016/j.marpetgeo.2023.106097>
- Lu, R., Willis, M., Campman, X., Ajo-Franklin, J., & Toksöz, M. (2007). Redatumming through a salt canopy – Another salt flank imaging strategy. *SEG Technical Program Expanded Abstracts*, 26.
- Luo, Y. (1991). Calculation of wave paths for bandlimited seismic waves. 61st Annual International Meeting, 1509–1512.
- Luo, Y., al-Dossary, S., & Marhoon, M. (2002). Edge-preserving smoothing and applications. *The Leading Edge*, 21, 136–158.
- Ma, X.-Q. (2001). Global joint inversion for the estimation of acoustic and shear impedances from AVO derived P- and S-wave reflectivity. *First Break*, 19(10), Article 10.
- Ma, Z., Jin, S., Chen, J., & Wang, H. (2002). Quantitative estimation of seismic imaging resolution. 72nd Ann. Internat. Mtg, 2281–2284.
- Maas, M. V. R., Bedle, H., & de Matos, M. C. (2023). Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data. *Marine and Petroleum Geology*, p. 151, 106199. <https://doi.org/10.1016/j.marpetgeo.2023.106199>
- Maurya, S. P., & Sarkar, P. (2016). Comparison of post stack seismic inversion methods: a case study from Blackfoot Field, Canada. *International Journal of Scientific and Engineering Research*, 7(8), 1091-101.

- Marfurt, K. J. (2018). Seismic Attributes as the Framework for Data Integration Throughout the Oilfield Life Cycle. Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560803522>
- Mavko, G., Mukerji, T., & Dvorkin, J. (2020). The Rock Physics Handbook (3rd ed.). Cambridge: Cambridge University Press. doi:10.1017/9781108333016
- McArdle, N. J., D. Iacopini, M. A. KunleDare, and G. S. Pa- ton, (2014), The use of geologic expression workflows for basin scale reconnaissance: A case study from the Ex- mouth Subbasin, North Carnarvon Basin, northwestern Australia: Interpretation, 2, no. 1, SA163–SA177, doi: 10.1190/INT-2013-0112.1.
- McCormack, K. D., & McClay, K. R. (2020). Orthorhombic faulting in the Beagle Sub-basin, North West Shelf, Australia. Geological Society, London, Special Publications, 476(1), 205-230.
- McDonald, A. (2021). Data Quality Considerations for Petrophysical Machine Learning Models. SPWLA 62nd Annual Logging Symposium 2021.
- Mesoella, K. J., J. D. Robinson, L. M. McCormick, & A. R. Ormiston, (1974), Cyclic Deposition of Silurian Carbonates and Evaporites in Michigan Basin: American Association of Petroleum Geologists Bulletin, v. 58, p. 34-62
- Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). christophm.github.io/interpretable-ml-book/
- Nadhim, Z. (2019). Depositional facies and sequence stratigraphic model for the Silurian reefs, Michigan Basin, USA.
- Nanda, N. C. (2016). Seismic data interpretation and evaluation for hydrocarbon exploration and production (p. 24). Springer.
- Neff, P., Steineder, D., Stummer, B., & Clemens, T. (2021). Estimation of initial hydrocarbon saturation applying machine learning under petrophysical uncertainty. SPE Reservoir Evaluation & Engineering, 24(02), 325–340. <https://doi.org/10.2118/203384-PA>
- Nienhuis, J. H., Törnqvist, T. E., & Esposito, C. R. (2018). Crevasse splays versus avulsions: A recipe for land building with levee breaches. Geophysical Research Letters, 45, 4058–4067. <https://doi.org/10.1029/2018GL077933>
- Niranjan, N. C. (2016). Chapter 2 Seismic reflection principles: Basics. In Seismic data interpretation and evaluation for hydrocarbon exploration and production: A practitioner's guide (pp. 19–35). Springer.

- Ødegaard, E., & Avseth, P. (2003). Interpretation of Elastic Inversion Results Using Rock Physics Templates. European Association of Geoscientists & Engineers. <https://doi.org/10.3997/2214-4609-pdb.6.E17>
- Oppenheim, A.V., & Schaffer, R. W. (2010). Discrete-time signal processing. Pearson.
- Ostrander, W. J. (1984). Plane-wave Reflection Coefficients for Gas Sands at Nonnormal Angles of Incidence. *Geophysics*, 49, 1637–1648. <https://doi.org/10.1190/1.1441571>
- Orlic, B., Heege, J. ter, & Wassing, B. (2011). Assessing the integrity of fault- and top seals at CO₂ storage sites. *Energy Procedia*, 4, 4798–4805. <https://doi.org/10.1016/j.egypro.2011.02.445>
- Osunrinde, T. O., Yelisetti, S., & Sanchez, V. (2023). The effect of depositional environment on sustaining a carbon storage site. *The Leading Edge*, 42(11), 737–747. <https://doi.org/10.1190/tle42110737.1>
- Pan, G. S., Young, Chi. Y., & Castagna, J. P. (1994). An integrated target-oriented prestack elastic waveform inversion: Sensitivity, calibration, and application. *GEOPHYSICS*, 59(9), 1392–1404. <https://doi.org/10.1190/1.1443697>
- Palaz, I., & Marfurt, K. J. (1997). Carbonate seismology. *Geophysical Developments Series*, Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560802099>
- Paumard, V., Bourget, J., Payenberg, T., Ainsworth, R. B., George, A. D., Lang, S., Posamentier, H. W., & Peyrot, D. (2018). Controls on shelf-margin architecture and sediment partitioning during a syn-rift to post-rift transition: Insights from the Barrow Group (Northern Carnarvon Basin, North West Shelf, Australia). *Earth-Science Reviews*, 177, 643–677. doi: 10.1016/j.earscirev.2017.11.026
- Peel, F. J. (2014). The engines of gravity-driven movement on passive margins: Quantifying the relative contribution of spreading vs. gravity sliding mechanisms. *Tectonophysics*, 633, 126–142. doi: 10.1016/j.tecto.2014.06.023
- Perico, E., Bedle, H., Buist, B., & Damasceno, A. C. (2023). Fault characterization in a postsalt reservoir interval, Jubarte Field (Campos Basin), using seismic attributes and machine learning. *Interpretation*, 11(2), T199–T214. <https://doi.org/10.1190/INT-2022-0061.1>
- Phillips, M., & Fomel, S. (2017). Plane-wave Sobel attribute for discontinuity enhancement in seismic images. *Geophysics*, 82(6), WB63–WB69. doi: 10.1190/geo2017-0233.1
- Posamentier, H. W., & Kolla, V. (2003). Seismic geomorphology and stratigraphy of depositional elements in deep-water settings. *Journal of Sedimentary Research*, 73, 367–388. doi: 10.1306/111302730367
- Prasad, M. (2003). Velocity-permeability relations within hydraulic units. *Geophysics*, 68, 108–117. <https://doi.org/10.1190/1.1543198>

Prather, B. E. (2012). Controls on reservoir distribution, architecture, and stratigraphic trapping in slope settings. In *Phanerozoic regional geology of the world* (pp. 552–579). Elsevier.

Pyrzcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. New York: Oxford University Press.

Qi, J., Fangyu Li, Lyu B., Olorunsola, O., Marfurt, K., & Zhang, B. (2016). Seismic fault enhancement and skeletonization. *SEG Technical Program Expanded Abstracts*, 1966-1970.

Qi, J., Lyu, B., Al Ali, A., Machado, G., Hu, Y., & Marfurt, K. (2019). Image processing of seismic attributes for automatic fault extraction. *GEOPHYSICS*, 84(1), Article 1. <https://doi.org/10.1190/geo2018-0369.1>

Qi, J., Machado, G., & Marfurt, K. (2017). A workflow to skeletonize faults and stratigraphic features. *GEOPHYSICS*, 82(4), O57–O70. <https://doi.org/10.1190/geo2016-0641.1>

Qi, J., Zhang, B., Lyu, B., & Marfurt, K. (2020). Seismic attribute selection for machine-learning based facies analysis. *Geophysics*, 82(2), 017-035. doi: 10.1190/geo2019-0223.1

Qi, J., Lin, T., Zhao, T., Li, F., & Marfurt, K. (2016). Semi supervised multi-attribute seismic facies analysis. *Interpretation*, 4(1), SB91–SB106. doi:10.1190/INT-2015-0098.1.

Quakenbush, M., Shang, B., & Tuttle, C. (2006). Poisson impedance. *The Leading Edge*, 25(2), 128–138. <https://doi.org/10.1190/1.2172301>

Rine, M. J., J. D. Garrett, & S. E. Kaczmarek. (2017). A New Facies Architecture Model For the Silurian Niagaran Pinnacle Reef Complexes Of The Michigan Basin: Characterization and Modeling of Carbonates–Mountjoy Symposium 1. doi:10.2110/sepmssp.109.02.

Rine, M. J., McLaughlin, P. I., Bancroft, A. M., Harrison III, W. B., Kuglitsch, J., Caruthers, A. H., ... & Emsbo, P. (2020). Linked Silurian carbon cycle perturbations, bursts of pinnacle reef growth, extreme sea-level oscillations, and evaporite deposition (Michigan Basin, USA). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 554, 109806.

Rine, Matthew. (2015). Depositional Facies and Sequence Stratigraphy of Niagaran-Lower Salina Reef Complex Reservoirs of the Guelph Formation, Michigan Basin. 10.13140/RG.2.2.20337.79203.

Roden, R., & Chen, C. W. (2017). Interpretation of DHI characteristics with machine learning. *First Break*, 35(5), Article 5. <https://doi.org/10.3997/1365-2397.35.5.88069>

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

Roden, R., & Chen, C. W. (2017). Interpretation of DHI characteristics with machine learning. *First Break*, 35(5), Article 5. <https://doi.org/10.3997/1365-2397.35.5.88069>

Roden, R., Smith, T. A., Santogrossi, P., Sacrey, D., & Jones, G. (2017). Seismic interpretation below tuning with multiattribute analysis. *The Leading Edge*, 36(4), 330–339. <https://doi.org/10.1190/tle36040330.1>

Ross A, W. (1962). *Ashby-Principles of Self-Organization.pdf*. Symposium on Self-Organization, 8. <http://cast.b-ap.net/wp-content/uploads/sites/8/2011/09/ashby-Principles of Self-Organization.pdf>

Ross, S. M. (2009). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press.

Rączkowska, A., Możejko, M., Zambonelli, J., & Szczurek, E. (2019). ARA: Accurate, reliable and active histopathological image classification framework with Bayesian deep learning. *Scientific Reports*, 9(1), 14347. <https://doi.org/10.1038/s41598-019-50587-1>

Roy, A., Dowdell, B. L., & Marfurt, K. J. (2013). Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma. *Interpretation*, 1(2), SB109–SB124. <https://doi.org/10.1190/INT-2013-0023.1>

Roy, A., Matos, M., & Marfurt, K. (2010). Automatic Seismic Facies Classification with Kohonen Self Organizing Maps -a Tutorial.

Russell, Hampson, & Bankhead. (2006). An Inversion Primer. *CSEG RECORDER Magazine*, 31(Special). <https://csegrecorder.com/articles/view/an-inversion-primer>

Sabeti, H., & Javaherian, A. (2009). Seismic facies analysis based on k-means clustering algorithm using 3D seismic attributes. 1st EAGE International Petroleum Conference and Exhibition. <https://doi.org/10.3997/2214-4609.20145876>

Sam-Marcus, J., Enaworu, E., Oluwatosin, J., Rotimi, & Seteyeobot, I. (2018). A proposed solution to the determination of water saturation: Using a modelled equation. *Journal of Petroleum Exploration and Production Technology*, 8, 1009-1015. <https://doi.org/10.1007/s13202-018-0453-4>

Sams, M., Begg, P., & Manapov, T. (2017). Seismic inversion of a carbonate buildup: A case study. *Interpretation*, 5(4).

Schenk, O., Dempsey, C., Benson, R., Cheng, M., Tewari, S., Karvelas, A., & Bancalà, G. (2019). New insights into the Exmouth Sub-basin from basin and petroleum system modeling. *ASEG Extended Abstracts 2019*, 1-5. <https://doi.org/10.1080/22020586.2019.12072963>

Schuber, E. (2022). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, 25(1), 36-42. <https://doi.org/10.1145/3606274.3606278>

Schon, J. H. (2004). Physical properties of rocks, Volume 18: Fundamentals and principles of petrophysics. Elsevier.

Sears, S. O., & Lucia, F. J. (1979). Reef-growth model for Silurian pinnacle reefs, northern Michigan reef trend. Geological Society of America Bulletin, 7, 299. <https://doi.org/10.1130/0091-7613>

Sheriff, R. E., & Geldart, L. P. (1995). Exploration seismology (2nd ed.). Cambridge University Press.

Silver, C., Bedle, H., & Rine, M. (2021). Applications of prestack seismic inversion on a Silurian pinnacle reef gas storage reservoir for enhanced geologic interpretations. First International Meeting for Applied Geoscience & Energy Expanded Abstracts, 322-326. <https://doi.org/10.1190/segam2021-3594319.1>

Slatt, R. M. (2013). Geologic controls on reservoir quality. In Developments in Petroleum Science (Vol. 61, pp. 229-281). Elsevier. <https://doi.org/10.1016/B978-0-444-56365-1.00006-7>

Slatt, R., & Weimer, P. (2004). Petroleum systems of deepwater settings. Society of Exploration Geophysicists.

Song, S., Mukerji, T., & Hou, J. (2021). Geological facies modeling based on progressive growing of generative adversarial networks (GANs). Computational Geosciences, 25, 1251-1273.

Sprague, A. R. G., Garfield, T. R., Goulding, F. J., Beaubouef, R. T., Sullivan, M. D., Rossen, C., Champion, K. M., Sickafoose, D. K., Abreu, V., Schellpeper, M. E., & Jensen, G. N. (2005). Integrated slope channel depositional models: The key to successful prediction of reservoirs presence and quality in offshore West Africa. Conference on CIPM.

Stoppa, F. (2012). Updates in volcanology: A comprehensive approach to volcanological problems. BoD - Books on Demand.

Taner, M. T. (2001). Seismic attributes. CSEG Recorder, 26(7). <https://csegrecorder.com/articles/view/seismic-attributes>

Taner, M. T., Koehler, F., & Sheriff, R. E. (1979). Complex seismic trace analysis. Geophysics, 44(6), 1041-1063. <https://doi.org/10.1190/1.1440994>

Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898717921>

Tahmasebi, P., Javadpour, F., & Sahimi, M. (2017). Data mining and machine learning for identifying sweet spots in shale reservoirs. Expert Systems with Applications, 88, 435-447. <https://doi.org/10.1016/j.eswa.2017.07.015>

Tiab, D., & Donaldson, E. C. (2012). Petrophysics: Theory and practice of measuring reservoir rock and fluid transport properties. Elsevier Science.

Tindale, K., Newell, N., Keall, J., & Smith, N. (1998). Structural evolution and charge history of the Exmouth sub-basin, northern Carnarvon Basin, Western Australia. The Sedimentary Basins of Western Australia, 2, 37.

Toelle, B., & Ganshin, Y. V. (2018). Porosity characterization in a Silurian reef, northern Michigan Basin, using azimuthal seismic data and potential impacts for enhanced oil recovery. The Geological Society of America, Special Paper 531, 157-173.

Torres-Verdín, C., Victoria, M., Merletti, G., & Pendrel, J. (1999). Trace-based and geostatistical inversion of 3-D seismic data for thin-sand delineation: An application in San Jorge Basin, Argentina. *Geophysics*, 18, 1070-1077. <https://doi.org/10.1190/1.1438434>

Tortopoglu, B. (2014). The structural evolution of the Northern Carnarvon basin, Northwest Australia [Doctoral dissertation, Colorado School of Mines].

Trout, J. L. (2012). Faunal distribution and relative abundance in a Silurian (Wenlock) pinnacle reef complex-Ray Reef, Macomb County, Michigan.

Van Engelen, J. E., and H. H. Hoos, (2020), A survey on semi-supervised learning: Machine Learning, 109, 373–440, doi: 10.1007/s10994-019-05855-6.

Varela, C., and W. Mohriak, (2013), Halokinetic rotating faults, salt intrusions, and seismic pitfalls in the petroleum exploration of divergent margins: AAPG Bulletin, 97, 1421–1446.

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T., & Lähdesmäki, H. (2015). Self-organization and missing values in SOM and GTM. *Neurocomputing*, 147, 60–70. <https://doi.org/10.1016/j.neucom.2014.02.061>

Veeken, P., and M. Da Silva, (2004), Seismic inversion methods and some of their constraints: First Break, 22, 47–70, doi: 10.3997/1365-2397.2004011.

Vera-Arroyo, Blanco-Dufu, P. A, and Bedle, H. (2023). Joint PP and PS Elastic Seismic Inversion to Investigate Seal Integrity in CO2 Projects. Carbon Capture, Utilization, and Storage (CCUS) expanded abstract conference, Houston, USA.

Verma, S., & Scipione, M. (2020). The early Paleozoic structures and its influence on the Permian strata, Midland Basin: Insights from multi-attribute seismic analysis. *Journal of Natural Gas Science and Engineering*, 82, 103521. <https://doi.org/10.1016/j.jngse.2020.103521>

Wallet, B. C., & Hardisty, R. (2019). Unsupervised seismic facies using Gaussian mixture models. *Interpretation*, 7(3), SE93–SE111. doi:10.1190/INT-2018-0119.1.

Wang, T., Trugman, D., & Lin, Y. (2021). SeismoGen: Seismic waveform synthesis using GAN with application to seismic data augmentation. *SeismoGen*, 1. <https://doi.org/10.1029/2020JB020077>

West, B. P., May, S. R., Eastwood, J. E., & Rossen, C. (2002). Interactive seismic facies classification using textural attributes and neural networks. *The Leading Edge*, 21, 1042–1049. doi:10.1190/1.1518444

White, R. E. (1991). Properties of instantaneous seismic attributes. *The Leading Edge*, 10(07), 26-32.

Wilkinson, R. (2015, July 1). BHP, Woodside move to decommission Stybarrow field. *Oil & Gas Journal*. <https://www.ogj.com/drilling-production/production-operations/article/17246384/bhp-woodside-move-to-decommission-stybarrow-field>

- Wollner, U., Yang, Y., & Dvorkin, J. P. (2017). Rock-physics diagnostics of an offshore gas field. *Geophysics*, 82(4), MR121MR132. <https://doi.org/10.1190/geo2016-0390.1>
- Wollner, U., Arevalo Lopez, H., & Dvorkin, J. (2017). Seismic-scale petrophysical interpretation and gas-volume estimation from simultaneous impedance inversion. *The Leading Edge*, 36, 910–915. <https://doi.org/10.1190/tle36110910.1>
- Wrona, T., Pan, I., Gawthorpe, R. L., & Fossen, H. (2018). Seismic facies analysis using machine learning. *Geophysics*, 83(5), O83–O95. doi: 10.1190/geo2017-0595.1
- Wu, P., Jain, V., Kulkarni, M. S., & Aria, A. (2018). Machine learning-based method for automated well-log processing and interpretation. Paper presented at the 2018 SEG International Exposition and Annual Meeting, Anaheim, California, USA, October 2018. doi: <https://doi.org/10.1190/segam2018-2996973.1>
- Wu, X., Luming, L., Yunzhi, S., Sergey., F. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84(3), 0016-8033. <http://doi.org/10.1190/geo2018-0646.1>
- Wold, J. L., & Grammer, G. M. (2008). Rock-Based 3-D Reservoir Characterization of a Silurian (Niagaran) Reef---Ray Gas Storage Field, Macomb County, Michigan. In *Paleozoic Stratigraphy and Resources of the Michigan Basin* (pp. xxx–xxx). Geological Society of America. [https://doi.org/10.1130/2017.2531\(06\)](https://doi.org/10.1130/2017.2531(06)).
- Xu, S., Chen, G., Zhu, Y., Zhang, J., Payne, M., Deffenbaugh, M. Dunsmuir, J. H. (2007). Carbonate Rock Physics: Analytical Models and Validations Using Computational Approaches and Lab/Log Measurements. International Petroleum Technology Conference. doi:10.2523/iptc-11308-ms.
- Yajun Tian, Gao, J., & Wang, D. (2022). Improving seismic resolution based on enhanced multi-channel variational mode decomposition. *Journal of Applied Geophysics*, 199, 104592. <https://doi.org/10.1016/j.jappgeo.2022.104592>
- Yarin, G., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. International Conference on Machine Learning.
- Yenwongfai, H. D., Mondol, N. H., Faleide, J. I., Lecomte, I., & Leutscher, J. (2017). Prestack inversion and multi-attribute analysis for porosity, shale volume, and sand probability in the Havert Formation of the Goliat Field, SW Barents Sea. *Interpretation*, 5, 1-54.
- Yeste, L. M., Varela, A. N., Viseras, C., Mcdougall, N. D., & García-García, F. (2020). Reservoir architecture and heterogeneity distribution in floodplain sandstones: Key features in outcrop, core and wireline logs. *Sedimentology*, 67, 3355–3388. doi: 10.1111/sed.12747
- Yielding, G., Freeman, B., & Needham, D. T. (1997). Quantitative fault seal prediction. *AAPG Bulletin*, 81(6), 897-917.
- Zaluski, W. (2018). Kevin Dome Static Model Development Final Report (p. 76). Schlumberger.

- Zeng, H. (2010). Stratal slicing: Benefits and challenges. *The Leading Edge*, 29(9), 1040-1049. <https://doi.org/10.1190/1.348764>
- Zhang, B., Zhao, T., Jin, X., & Marfurt, K. J. (2015). Brittleness evaluation of resource plays by integrating petrophysical and seismic data analysis. *Interpretation*, 3(2), T81–T92. <https://doi.org/10.1190/INT-2014-0144.1>
- Zhang, H., Chen, T., Liu, Y., Zhang, Y., & Liu, J. (2021). Automatic seismic facies interpretation using supervised deep learning. *Geophysics*, 86(1), IM15–IM33. <https://doi.org/10.1190/geo2019-0425.1>
- Zhang, J., Liu, Z., Zhang, G., Yan, B., Ni, X., & Xie, T. (2022). Simultaneous prediction of multiple physical parameters using gated recurrent neural network: Porosity, water saturation, shale content. *Frontiers in Earth Science*, 10, 984589. <https://doi.org/10.3389/feart.2022.984589>
- Zhang, K., Lin, N., Tian, G., Yang, J., Wang, D., & Jin, Z. (2022). Unsupervised-learning based self-organizing neural network using multi-component seismic data: Application to Xujiahe tight-sand gas reservoir in China. *Journal of Petroleum Science and Engineering*, 209, 109964. <https://doi.org/10.1016/j.petrol.2021.109964>
- Zhang, R., Sen, M. K., Phan, S., & Srinivasan, S. (2012). Stochastic and deterministic seismic inversion methods for thin-bed resolution. *Journal of Geophysics and Engineering*, 9(5), 611–618. <https://doi.org/10.1088/1742-2132/9/5/611>
- Zhao, T., Jayaram, V., Roy, A., & Marfurt, K. J. (2015a). A comparison of classification techniques for seismic facies recognition. *Interpretation*, 3(4), SAE29–SAE58. <https://doi.org/10.1190/INT-2015-0044.1>
- Zhao, T., Jayaram, V., Roy, A., & Marfurt, K. J. (2015b). A comparison of classification techniques for seismic facies recognition. *Interpretation*, 3(4), SAE29–SAE58. <https://doi.org/10.1190/INT-2015-0044.1>
- Zhao, T., Li, F., & Marfurt, K. J. (2017). Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. *Interpretation*, 5(2), T163–T171. <https://doi.org/10.1190/INT-2016-0132.1>
- Zhao, T., Zhang, J., Li, F., & Marfurt, K. J. (2016). Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps. *Interpretation*, 4(1), SB79–SB89. <https://doi.org/10.1190/INT-2015-0094.1>
- Zheng, Q. (1999). Carbonate diagenesis and porosity evolution in the Guelph Formation, Southwestern Ontario [Doctoral thesis, University of Waterloo]. Waterloo, Ontario, Canada.
- Zhu, D., Cui, J., Li, Y., Wan, Z., & Li, L. (2022). Adaptive Gaussian mixture model and convolution autoencoder clustering for unsupervised seismic waveform analysis. *Interpretation*, 10(1), T181–T193. <https://doi.org/10.1190/INT-2021-0087.1>