UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

OPTIMIZATION OF DEEPWATER CHANNEL SEISMIC RESERVOIR CHARACTERIZATION USING SEISMIC ATTRIBUTES AND MACHINE LEARNING

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

Karelia La Marca Molina

Norman, Oklahoma

2023

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340051

OPTIMIZATION OF DEEPWATER CHANNEL SEISMIC RESERVOIR CHARACTERIZATION USING SEISMIC ATTRIBUTES AND MACHINE LEARNING

A DISSERTATION APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Matthew Pranter

Dr. Deepak Devegowda

Dr. Sarah George

© Copyright by Karelia La Marca 2023

All Rights Reserved.

I have no special talent. I am only passionately curious.

Albert Einstein

ACKNOWLEDGEMENTS

Today, as I reach this significant milestone, a profound realization has crystallized in my mind: A Ph.D. is a deeply humbling experience. The more you delve into it, the more you appreciate the vastness of what you don't know. It's a continuous journey of learning, emphasizing the importance of the path over the destination—a path that cannot be walked alone. The beauty of our accomplishments reflects the support, love, and kindness that surrounds us. Therefore, in the following paragraphs, I want to express my heartfelt gratitude to the individuals and institutions whose invaluable support, guidance, and encouragement made the completion of my dissertation possible.

First and foremost, to my exceptional advisor, **Dr. Heather Bedle**. Thanks for the numerous roles you have played in my life: advisor, friend during joyful moments, a comforting presence in challenging times, and a source of inspiration as a woman in STEM. I cannot thank enough for being there when I needed it the most, for offering words of encouragement, or a comforting hug. Thank you for your genuine support, understanding, and being an integral part of my life.

To my committee members, **Dr. Deepak Devegowda**, **Dr. Matthew Pranter**, **Dr. Sarah George**, and formerly Dr. Mike Soreghan, thank you for your contribution to this dissertation and helping me grow professionally and personally. I admire you all.

I extend my special appreciation to two pillars in my journey, **Dr. Marfurt** and **Dr. Slatt**, who have been simply inspirational. You are the kind of people someone wants to be when they grow old.

I'm equally grateful to my mentors (Adam, Carrie, Alejandro, David and Katarina- to name some) and my OU professors; you have illuminated my professional path with your wisdom and guidance.

To my family—especially **mom** Zule, **dads** Enrique and Wime, Mi honey Andre (**sister**), **brothers** Enzo and Erik, and my guardian angel: **Grandma** Cloris—your unwavering support, understanding, and encouragement have been my foundation. **This achievement belongs to all of you :**)

My heartfelt thanks go to fellow students, including the **AASPI group**, who provided ideas, and offered assistance when needed. To my friends—**Steff**, **Julian**, **Pame**, **Vane**, **Aintzane**, **Emily**, **Lau**, **Delcio**, **Yolmir**, **Nany**, **Dayis**, **Esteban, Maty, Joe, Diana, Ducha, and my ULA**, **BLN and SLB** friends**.**.— you have been my support system, my vitamin people! I cherish the countless memories we've created and look forward to many more adventures together. I know this list goes longer and apologize if I have not name y'all, but you know who you are and what you mean to me. Gracias!

I'm thankful to my collaborators (and friends)—**Rafael Pires de Lima** and **Mario Ballinas**—for their dedication and input, especially **Dr. Lisa Stright** and **Teresa Langenkamp** for providing the foundation of my projects and always giving input with kindness. I also extend my gratitude to **AASPI**, School of geosciences, the **SEG** Foundation and **COLSA** for the financial support.

A special mention goes to our administrative staff: **Rebecca Fay** for making things possible, **Leah** for her words of encouragement, and Ashley and Ginger for their unwavering assistance.

Finally, I want to acknowledge the countless unnamed faces who have crossed my path, leaving an indelible mark on my journey. Each interaction, no matter how brief, has contributed to the person I've become. I can't wait for what's coming next.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **V**

LIST OF FIGURES ..... **VIII**

LIST OF TABLES ..... **XVII**

ABSTRACT ..... **XIX**

CHAPTER I: INTRODUCTION **1**

CHAPTER 2: SENSITIVITY ANALYSIS OF SEISMIC ATTRIBUTES PARAMETRIZATION TO REDUCE MISINTERPRETATIONS: APPLICATIONS TO DEEPWATER CHANNEL COMPLEXES* **3**

ABSTRACT 3
MOTIVATION AND OBJECTIVES 5
INTRODUCTION 6
GEOLOGICAL MODEL DESCRIPTION 8
METHODS 13
RESULTS 22
DISCUSSION 31
CONCLUSIONS 38
DATA AVAILABILITY 40
ACKNOWLEDGMENTS 40

CHAPTER 3: UNCERTAINTY ASSESSMENT IN UNSUPERVISED MACHINE LEARNING METHODS FOR DEEPWATER CHANNEL SEISMIC FACIES USING OUTCROP-DERIVED 3D MODELS AND SYNTHETIC SEISMIC DATA **41**

ABSTRACT 41
INTRODUCTION 42
GEOLOGIC MODEL 45
METHODS 48
RESULTS 66
DISCUSSION 74
CONCLUSION 81
ACKNOWLEDGMENTS 83

CHAPTER 4: UNDERSTANDING UNCERTAINTY IN DEEPWATER CHANNEL SEISMIC FACIES CLASSIFICATION APPLYING RANDOM FOREST ON OUTCROP-CONSTRAINED 3D MODELS AND SYNTHETIC SEISMIC DATA 84

ABSTRACT 84
INTRODUCTION 85
METHODOLOGY 89
RESULTS AND DISCUSSION 97
CONCLUSIONS 110
ACKNOWLEDGMENTS 112
DATA AND MATERIALS AVAILABILITY 112

SPECIAL SECTION: WORKING WITH IMBALANCED DATA IN MACHINE LEARNING ALGORITHMS: BRIDGING THE GAP BETWEEN REAL AND OPTIMAL IN SEISMIC FACIES INTERPRETATION * 113

UNSUPERVISED METHODS AND IMBALANCED DATA 115
SUPERVISED METHODS AND IMBALANCED DATASETS 117

CHAPTER 5: CONCLUSIONS 120

APPENDIX 122

APPENDIX A: SPECS FOR EACH SYNTHETIC SEISMIC VOLUME 122
APPENDIX B: DETAILED EXPLANATION OF CHANNEL ARCHITECTURAL FACIES (SOURCE: JACKSON ET AL., 2019) 122
REFERENCES 123

LIST OF FIGURES

Figure 1.1: Stratigraphic hierarchy: from a single element to complex sets. A vertical slice within the Pipeline 3D dataset offshore New Zealand shows the seismic appearance of different architectural hierarchies in deepwater channels, with corresponding cartoons below. The smallest architecture (4^{th} to 5^{th} order) is the channel element (box 1 in green). The second hierarchy (6^{th} order) occurs when the channel elements stack together, forming a channel complex (box 2 in blue). The higher-order hierarchy (7^{th} order and higher) occurs with the amalgamation of channel complexes, developing a channel complex set (box 3 in magenta). The color legend indicates the distinct facies that commonly occur within each element of each architecture. Measuring the sizes of each architectural element indicated on the right as well as their hierarchy provides key insight into the underlying depositional processes as well as a prediction of the more common lithologies. Hierarchies mentioned follow Pickering and Cantalejo (2015) classification.

Figure 1.2. A visual guide showing the steps to convert outcrop measurements to a synthetic model. (A) Location and exposure of the outcropping deepwater channels at Laguna Figueroa. Paleoflow is from North to South (obliquely and to the right into the outcrop at this location) (B) Conceptual diagram of the Upper and Lower Figueroa outcrops showing channel elements, complexes and complex sets. The red line indicates the outcrop profile. Left of the line is into the outcrop face, and the right has been eroded away. (C) geocellular model using the constraints from (B) augmented by facies and corresponding rock properties, including acoustic impedances from Shallow Offshore West African modeled rock properties (Stright et al., 2014). (D) The Ormsby wavelet and a representative vertical slice through the 3D synthetic seismic data volume generated from the model shown in (C). Courtesy of Teresa Langenkamp and Lisa Stright.

Figure 1.3. Explanation of stratigraphic event mixing or vertical smearing of stratigraphic features. (A) Vertical resolution of channel elements related to different peak frequencies (modified after Nielson, 2011). (B) Vertical slice through the 3D Pipeline 3D offshore New Zealand seismic survey showing a channel complex and a Horizon used for interpretation at different stratigraphic levels indicated by yellow arrows. The dominant frequency at this level is 40 Hz giving a dominant period of 25 ms. (C) Stratal slices at approximately 25 ms intervals through the coherence volume computed using a ±20 ms analysis window show "stratigraphic" mixing by the seismic wavelet. Note that the relatively straight channel form at 2050 ms can be seen at 2080 ms and other stratigraphic levels (1,2,3 from deeper to shallower) where other channel forms (green arrow) appear causing interference. The cause of this mixing could be 1) mixing of reflectivity by the 25ms dominant period seismic wavelet, 2) mixing of discontinuities through the 40 ms coherence computation, 3) shifting of the basal channel element thalweg due to compensational style as you move up or 4) differential compaction over deeper discontinuities between the floodplain and the channel element fill.

Figure 1.4. Workflow of the study. Four different parameters were evaluated in the sensitivity analysis: 1) the effect of frequency content, 2) the impact on the choice of the seismic attribute, 3) the analysis window effect, and 4) the sensitivity to band-limited random noise addition. The best cases were selected to be shown in each case, and analysis was performed by comparing them with the original 3D geological model derived from the Laguna Figueroa Deepwater outcrop. This resulted in 4 cases for analysis, ultimately leading to a workflow and documentation of best practices in channel architecture interpretation.

Figure 1.5. A representative analysis window used in attribute calculations. Signals are sampled at discrete points, not continuous recordings. Therefore, each seismic trace will contain as many samples as the sample rate allows. For the example shown, if we consider the dominant period (distance between two peaks), our analysis window will contain 20 samples, equivalent to ~ 20 msec. Other examples of analysis windows are depicted, including the smallest possible equivalent to the sample increment, in this case, 1 msec.

Figure 1.6. A representative vertical slice (224) and time slices at *t*=-191ms) through the (A) 3D model that shows (B) 15 Hz data, (C) 60 Hz data, (D) 90 Hz data, and (E) 180 Hz data. Notice the improvement in the channel architecture's detail with the increase in the dominant frequency and corresponding spectral bandwidth, being able to interpret complexes from figure B on.

Figure 1.7. Results on the attribute sensitivity analysis to thickness and extent of architectural elements using a default analysis window. We present results on the 30 Hz dataset, analogous to vintage seismic data (to the left), and the 60Hz dataset, representing modern seismic data (to the right). Each attribute is presented for both frequencies in a representative inline (224) and time slices (-280 ms and –191ms) as follows: (A) RMS (Root Mean Square) amplitude for 30 Hz (B) coherence (Sobel filter) for 30Hz, (C) Instantaneous frequency for 30 Hz, (D) RMS amplitude for 60 Hz (E) coherence (Sobel filter) for 60Hz, (F) Instantaneous frequency for 60 Hz.

Figure 1.8. RMS amplitude sensitivity analysis to evaluate the effect of the window size *W* on the vertical smearing of the different channel elements in the model. (A) 15 Hz with *W*=2ms (B)50 Hz with *W*=50ms (C) 30Hz with *W*=2ms (D) 50 Hz with *W*=50ms (E) 180 Hz with *W*=2 ms (F) 180 Hz with *W*=50 ms. High-frequency data with a small analysis window provides the most suitable representation of the true model. In RMS amplitude, the most accurate facies depiction is given by

high frequencies and small window combination, imaging channel element base, and inner and outer levee facies.

Figure 1.9. Sobel filter coherence attribute sensitivity analysis to evaluate the effect of the window size *W* on the vertical smearing and displacement of the different architectural elements in the model. (A) 15 Hz with *W*=2ms (B)30 Hz with *W*=2ms (C) 60Hz with *W*=2ms (D) 60 Hz with *W*=50ms. Notice how at a higher dominant frequency (60Hz) and smaller window of analysis, there is better detection of the channel element edges and channel element fill. In contrast, at lower frequencies and or larger windows of analysis, there is interference from deeper channel elements, as in the case presented on Figure 3.

Figure 1.10. Sensitivity analysis of the effect of random noise in channel architecture interpretation. Results shown refer to the 60 Hz dominant frequency seismic volume (which would be analog to real datasets). (A) 5% band-limited random noise added evaluated in RMS, Coherence, and Instantaneous frequency attributes. Interpretation of prominent features like channel complexes is only possible. (B) 5% band-limited random noise added applied to RMS, coherence, and instantaneous frequency attributes. Notice how the increase in noise is detrimental in the channel architecture interpretation, especially when using coherence. This may be due to the sensitivity that small windows have on high noise content.

Figure 1.11. Generalized workflow for a geoscientist to avoid pitfalls in interpretation by getting optimized results according to their dataset.

Figure 2.1: A visual guide showing the steps used to convert outcrop measurements to a synthetic model. (A) Location of Laguna Figueroa outcrop in the Patagonia-Chile. (B)Upper and Lower

Figueroa channel complexes identified in the outcrop, including boundaries and inner channel elements. The red line indicates the outcrop profile. (C) Creation of a 3D geocellular model using the constraints from (B) augmented by facies associations and corresponding rock properties, including acoustic impedances from GOM (Gulf of Mexico) and Nigerian analogs. (D) The Ormsby wavelet and a representative vertical slice through the 3D synthetic seismic data volume generated from the model shown in (C). Courtesy of Teresa Langenkamp and Lisa Stright.

Figure 2.2 Workflow of the study.

Figure 2.3. (A) Representative vertical slice of the original amplitude volume and seismic attributes used as input: (B) Peak Magnitude, (C) Envelope, (D) Root Mean Square amplitude (E) Spectral component 20 Hz, (F) Spectral component 40 Hz, (G) Spectral component 55Hz. The spectral components were used as a spectral decomposition volume.

Figure 2.4. Representation of (A) Self-Organizing Maps and (B) Generative Topographic Maps (adapted from Roy et al., 2013). Given an initial set of datapoints, they will be inputted in a lower dimensional latent space. In both cases the data is later projected in a high dimensional latent space where a manifold space best will fit the arranged data. For (A) the BMU or winning neuron is selected via weighting and each BMU will be representative of the group (cluster) the datapoints that get closer to them will form. For (B) there are a series of basis function centers whose linear combination will allow the projection of the grid points in the non-Euclidean manifold plane as m_k vectors. Adaptation to the manifold plane occurs as a product of Gaussian Pdf that occur around each vector. The main difference between the SOM and the GTM regarding their result presentation is that while SOM assigns each data point to exactly one place on the map GTM calculates a distribution for each data point on the map.

Figure 2.5. Geobodies per facies in the 3D model that serves as basis for the uncertainty quantification analysis.

Figure 2.6. SOM and GTM results per cluster arrangement and heat map per case. (A) SOM of 256 clusters, (B) GTM of 2566 clusters.

Figure 2.7. SOM and GTM results per cluster arrangement and heat map per case. (A) SOM of 36 clusters, (B) GTM of 36 clusters, (C) SOM of 81clusters, (D) GTM of 81 clusters, (E) SOM of 256 clusters, (F) GTM of 256 clusters.

Figure 2.8. Geobody comparison per case and facies compared to the original/ expected geobodies.

Figure 2.9. Visual representation of uncertainty in unsupervised ML methods. (A) Shows a section of the original 3D model, which has been filtered y the channel facies for visualization purposes. (B)Presents the same section in the results of GTM36 once they have been concatenated and resampled. Notice how some voxels represent the correct facies but others do not. This demonstrates that the use of ML methods will add uncertainty in the reservoir understanding (dimensions, facies, and volumetrics).

Figure 3.1. Workflow used in the uncertainty assessment of deepwater facies classification using a random forest classification algorithm, applied to an outcrop-derived 3D synthetic seismic model (Langenkamp, 2021). The choice of seismic attributes used in this model is discussed in La Marca et al (2023). * Notice that the original model has been divided into two: upper portion to perform the RF model, and the lower portion (green) to validate or apply in this unseen piece of model. In

this study, it's important to note that the terms 'test' and 'validation' are not interchangeable. Validation is employed here to understand uncertainty.

Figure 3.2. First row: 3D Model and synthetic amplitude volume, second and third row: calculated seismic attributes used in the study. Peak magnitude, envelope and Root Mean Square amplitude (RMS) are often used to highlight changes in lithology, whereas spectral magnitudes are used to interpret architectural elements of different thicknesses.

Figure 3.3. (A) Scatter plot displaying the correlation between input seismic attributes used as training data. Datapoints color code represent the facies defined (Yellow: channel axis, Orange: channel off-axis, Brown: channel margin, Red: Mass Transport deposits (MTD), and Gray (background shale) (B) Box plots showing the uniqueness of each attribute in representing each facies. Notice how shale can always be distinct from the channel facies overall.

Figure 3.4. Histogram of samples count per facies (label) in the training dataset. Although the shale has greater representation than the other four facies, this imbalance is representative of actual deepwater channel facies distribution.

Figure 3.5. A histogram that shows seismic attribute (feature) importance. Notice how peak magnitude considerably outperforms in the analysis. This might be attributed to this attribute sensitivity to changes in tuning which benefits different architectural facies recognition.

Figure 3.6. Correlation matrices (left) and classification report (right) to evaluate RF classifier performance. Confusion matrix compare original (on the Y-axis) versus predicted (on the X-axis) classes in a n x n configuration. The main diagonal of the matrix reveals the correctly predicted classes (facies), while off-diagonal elements represent misclassifications into other classes. (A)RF

results applying test #1 hyparameters indicated in Table 2, (B) Optimized random forest model results, applying test #6 hyperparameter values in table 2 Notice how accuracy increases substantially with the hyperparameter optimization.

Figure 3.7. Comparison of (A) original model (B) unoptimized RF- test#1 in table 2-, and (C) RF classification with optimized hyperparameters – test #6 in table 2-. Notice how the prediction is highly improved in C, over B as compared to the truth model. If we didn't optimize hyperparameters we would be overestimating the channel axis facies and underestimating thinner facies such as the channel margin and MTD.

Figure 3.8. A) confusion matrix and (B) classification report for the validation model. Here we see a decrease of ~ 20% in accuracy when applying the optimized RF model to unseen data. See Figure 6 for reference. Support refers to the number of voxels per facies.

Figure 3.9. Comparison of (A) original vs (B) predicted RF validation. In this chart, facies have been isolated via opacity to compare areal distribution and proportions visually. Channel complexes' location and orientation are predicted well. Chanel axis, shale. and MTD facies show predominance over the other channel facies, which has been misclassified into either of these predominant channel facies or shale.

Figure 3.10. Comparison of true vs predicted facies in the validation model in (A) Representative crossline, (B) representative time slice. Notice how channel position and orientation has been correctly predicted to be able to differentiate between shale (non- reservoir) facies and channel (reservoir) facies. It is observed a predominance or likelihood of prediction of class 1 (channel

axis) and class 5 (MTD) over other channel facies. Channel axis facies are more evident were channels stack vertically.

Figure 4.1. (A) samples per facies histogram in a training dataset showing an imbalance between predominant shale facies and minoritarian channel facies. (B) Feature relation vis scatterplot showing that gray (shales) predominates and in some cases its range is wide so it overlaps with other classes.

Figure 4.2. Elbow plot indicates that the optimal number of clusters is 3, and GTM results reveal MTDs in a light green geobody (GTM green), Shale in red geobody (Purple/blue in GTM), and channel facies combined in dark green geobody (orange in GTM). This demonstrates that not every cluster has to represent a single facies. The elbow plot is a valuable tool for estimating the optimal number of clusters. The amount of clusters may be related to the quality and resolution of the data and distintive patterns found.

Figure 4.3. Workflow to optimize machine learning results for facies interpretation in the face of imbalanced datasets.

LIST OF TABLES

Table 1.1. Tuning thicknesses for shallow and deep elements in each synthetic seismic volume. Modified from Langenkamp (2021).

Table 1.2. Selected seismic attributes to perform the sensitivity analysis showing a representative image, feature measurement, and use in channel element interpretation.

Table 1.3. Common pitfalls in seismic interpretation and how to avoid them.

Table 2.1. Summary of seismic attributes explored to use as input. Their category, principle, references, and common use in channel architecture interpretation are explained.

Table 2.2. Self-Organizing Maps cases' parameterization specifications

Table 2.3. Generative Topographic Map's cases' parameterization specifications

Table 2.4. Metrics to evaluate performance of the ML evaluated. References used: Towards data science and Kuhn and Johnson (2013)

Table 2.5. Clusters interpreted per case study and facies.

Table 2.6. Results of percentage error per case and facies

Table 2.7. Results of confusion matrix per facies for the SOM36, GTM36, SOM81, GTM81. Precision, recall, F1 and accuracy metrics are shown to the right of each case studied.

Table 3.1. Metrics to evaluate performance of the ML evaluated. Modified after Kuhn and Johnson (2013).

Table 3.2. Summary of the best results obtained per hyperparameter optimization test. Each test comprised a grid search with variations of the tested hyperparameter. Test #1 was selected randomly as a start point, and each subsequent test evaluated a different parameter (in bold). Test #6 ultimately shows the best hyperparameter configuration, which is the result of a last grid search using the results/ tendency of Tests 2-6. Test #6 shows the parameters employed in the RF model.

.

ABSTRACT

Accurate subsurface reservoir mapping is essential for resource exploration. In uncalibrated basins, seismic data, often limited by resolution, frequency, quality etc., algorithms become the primary information source due to the unavailability of well logs and core data.

Seismic attributes, while integral for understanding subsurface structures, visually limit interpreters to working with only three of them at once. Conversely, machine learning, though capable of handling numerous attributes, is often seen as inscrutable "black boxes," complicating the interpretation of their predictions and uncertainties.

To address these challenges, a comprehensive approach was undertaken, involving a detailed 3D model from Chilean Patagonia's Tres Pasos Formation with synthetic seismic data. The synthetic data served as a benchmark for conducting sensitivity analysis on seismic attributes, offering insights for parameter and workflow optimization. The study also evaluated the uncertainty in unsupervised and supervised machine learning for deepwater facies prediction through qualitative and quantitative assessments.

Study key findings include: 1) High-frequency data and smaller analysis windows provide clearer channel images, while low-frequency data and larger windows create composite appearances, particularly in small stratigraphic features. 2) GTM and SOM exhibited similar performance, with error rates around 2% for predominant facies but significantly higher for individual channel-related facies. This suggests that unbalanced data results in higher errors for minor facies and that a reduction in clusters or a simplified model may better represent reservoir versus non-reservoir facies. 3) Resolution and data distribution significantly impact predictability, leading to non-uniqueness in cluster generation, which applies to supervised models as well. Strengthening the argument that understanding the limitations of seismic data is crucial. 4)

Uncertainty in seismic facies prediction is influenced by factors such as training attribute selection, original facies proportions (e.g., imbalanced data, variable errors, and data quality). While optimized random forests achieved an 80% accuracy rate, validation accuracy was lower, emphasizing the need to address uncertainties and their role in interpretation.

Overall the utilization of ground truth seismic data derived from outcrops offers valuable insights into the strengths and challenges of machine learning in subsurface applications, where accurate predictions are critical for decision-making and safety in the energy sector.

CHAPTER I: INTRODUCTION

The accurate interpretation of seismic facies plays a pivotal role in understanding subsurface reservoirs, making informed decisions, and optimizing resource extraction strategies. In the realm of deepwater exploration, where the complexity of geological features often poses significant challenges, achieving precise facies interpretation becomes even more critical. This dissertation delves into the intricate world of seismic facies interpretation in deepwater environments, addressing key challenges and uncertainties through three main chapters.

In the second chapter, we introduce the objective of the study and embark on a journey to test the sensitivity of seismic attributes in identifying deepwater channel facies. We explore how various seismic attributes respond to subtle geological variations, aiming to enhance the accuracy of interpretation. Additionally, this chapter sheds light on the potential pitfalls in attribute interpretation. By pinpointing these pitfalls, we aim to develop a deeper understanding of the limitations and challenges faced when attempting to discern complex geological features from seismic attribute data.

Chapter 3 delves into the realm of unsupervised machine learning techniques for deepwater seismic facies interpretation. Here, we contrast the clusters obtained through these techniques with ground truth models, emphasizing the uncertainties inherent in such approaches. This chapter serves as a critical examination of the effectiveness and limitations of these state-of-the-art methods, enabling us to better grasp the nuances of seismic facies interpretation in deepwater environments.

The fourth chapter explores the application of a robust supervised machine learning method, Random Forest, in the context of seismic facies interpretation. We investigate the challenges posed by imbalanced data and delve into strategies for optimizing results. By addressing these uncertainties and offering potential solutions (shown in a section after Chapter 4), this chapter aims to provide a more accurate and reliable framework for seismic facies classification using unsupervised ML in deepwater settings.

The ability to accurately interpret seismic facies in deepwater environments carries immense implications for decision-making across industries. Beyond the hydrocarbon relevance, this dissertation highlights the relevance of precise facies interpretation in geothermal energy exploration and carbon capture, utilization, and storage (CCUS) initiatives. Accurate seismic facies interpretation not only informs volumetric assessments but also influences critical decisions related to reservoir management, well placement, and risk mitigation.

All in all, this dissertation navigates the multifaceted landscape of deepwater seismic facies interpretation, addressing sensitivity, uncertainty, and optimization. By addressing these challenges, we contribute to the advancement of knowledge and practices in subsurface exploration, with a broader impact that extends to industries beyond hydrocarbon extraction. The journey through these chapters exemplifies the necessity for accurate seismic facies interpretation as a cornerstone for informed decision-making and sustainable resource management.

CHAPTER 2: SENSITIVITY ANALYSIS OF SEISMIC ATTRIBUTES PARAMETRIZATION TO REDUCE MISINTERPRETATIONS: APPLICATIONS TO DEEPWATER CHANNEL COMPLEXES*

*This chapter is published in the journal, *Marine and Petroleum Geology*, Vol. 153, (July 2023); p. T585-T598:

La Marca, K., H. Bedle., L. Stright., and K.J. Marfurt, 2023, Sensitivity analysis of seismic attributes parametrization to reduce misinterpretations: Applications to deepwater channel complexes: Marine and Petroleum Geology, **153**, 106309, Elsevier. https://doi.org/10.1016/j.marpetgeo.2023.106309

*This chapter was presented at the 2021 AAPG & SEG (IMAGE) Annual Convention & Exhibition:

La Marca, K., K. Marfurt., H. Bedle., L. Stright, and T. Langenkamp, 2021, Sensitivity analysis of seismic attributes parametrization for interpretation of a multi-story deepwater channel system: Tres Pasos Formation, Magallanes Basin Chile: *First International meeting for Applied Geosciences and energy expanded abstracts*: 1191-1195. https://doi.org/10.1190/segam2021-3584026.1

Abstract

Geoscientists apply algorithms such as seismic attributes to better interpret depositional systems that enhance various aspects of the seismic data. However, they are limited by the original seismic amplitude or frequency content, data quality, and algorithm parameters considered.

Additionally, our capacity to interpret depositional system architecture is limited by seismic resolution, which results in potential misinterpretations associated with the correct position of stratigraphic features. This is particularly important as mapping reservoir architecture (geobody size, shape, and stacking patterns) in the subsurface is critical for exploring and producing hydrocarbons, CO_2 storage, and geothermal resource development since it can define connectivity or compartmentalization of flow zones.

To address these concerns, we investigated five synthetic seismic volumes from low to high-frequency bandwidths of 15 Hz, 30 Hz, 60 Hz, 90 Hz, and 180 Hz based on an architectural model

of an outcropping deepwater channelized slope system in the Magallanes Basin, Chile. We analyzed 1) how seismic bandwidth affects the resolution of stacked stratigraphic features (i.e., deepwater channel elements and Mass Transport Deposits (MTDs)) and their subsequent seismic interpretation, and 2) the effect of different seismic attributes commonly employed in channel interpretation on our data to understand the "mixing" or "vertical smearing" of stratigraphic features by comparing the seismic with the true geological model 3) we explored how the attributes' parametrization affects the imaging of differently sized features by modifying the analysis window in each case from +/-2ms to +/- 50 ms. Finally, 4) we evaluated the effect of different noise levels in the sensitivity analysis.

Results show that the "mixing" of events occurs mainly as a result of 1) the seismic bandwidth, 2) the algorithm used for each seismic attribute calculation, 3) the attribute vertical analysis window, and 4) the signal-to-noise ratio of the data. Broadband, higher frequency data, and small analysis windows provide clearer images of the stacked channels. In contrast, low-frequency data and larger analysis windows result in more mixing or "composite" appearances, affecting interpretations and net-to-gross estimates, especially in small-size stratigraphic features such as individual channel elements and Mass Transport Deposits (MTDs). Our observations warn of potential misinterpretations in applying default attributes to actual seismic data, especially in geometrical attributes and window-dependent ones. Recognizing these misinterpretations is paramount for reconstructing deepwater architecture (this study), sedimentary and structural studies for drilling locations, reserves estimation, and overall uncertainty assessment.

Keywords: deepwater; seismic facies; architectural elements; seismic geomorphology; interpretation; seismic attributes; channel complex, analysis window.

Motivation and objectives

Seismic exploration of deepwater channels is challenging due to the physical properties inherent to the seismic, the variability in fill and stacking of reservoir geobodies, and the uncertainty that can occur due to the lack of hard data (core or well data). Reservoir architecture controls the distribution of fluids in the subsurface and the connectivity (or compartmentalization) of the reservoir that impacts recovery or injectivity. Therefore, geoscientists seek to understand how sensitive the interpretation of reservoir architecture is to different quality and types of seismic data, as well as different attributes and parameters commonly used to identify the architecture better and to make appropriate well plan decisions, volumetric and recovery/storage estimates.

The primary questions we aimed to address with the study are:

- How does the seismic data's frequency content affect the imaging of deepwater architecture?
- What is the effect of each seismic attribute on the architecture interpretation and its true position, both vertically and horizontally?
- What is the effect of the seismic attribute analysis window size on the vertical smearing of architecture?
- What is the effect of seismic noise on our sensitivity analysis?

The study's importance is identifying common pitfalls in seismic interpretation using synthetic seismic data created from an outcrop-derived architectural model of a seismic-scale deepwater channel system. Studies with synthetic data like this allow geoscientists to understand uncertainty in interpreting channel architecture from seismic data.

Introduction

When it comes to reservoir characterization using seismic reflection data, even if we employ all the tools available to interpret, locate, and measure the reservoir that will contain (oil, gas, and water) or allow for the storage (geothermal, CO_2) of economic resources, uncertainty prevails. This interpretational ambiguity occurs due to changes in various physical parameters in response to the media and the seismic records' inherent acquisition and processing characteristics.

The imaging and interpretation of different-sized stratigraphic features in the subsurface using seismic reflection data are often compromised due to limits in seismic resolution, which in addition to tuning effects, can influence volumetric interpretations and gross rock volume calculations (Pemberton et al., 2018). Therefore, geoscientists need to understand the common pitfalls associated with seismic interpretation: the impact of the frequency content on the imaging of reservoir architecture, the choice of parameters and attributes' influence on the interpretation of architectural elements, and the detrimental impact noise can have on the overall picture. To address this, we performed a sensitivity analysis that evaluated four parameters: frequency content, the effect of the seismic attribute, the impact of the window of analysis, and noise level combined with five 3D synthetic datasets.

The synthetic data that was derived from an outcrop analog in Magallanes Basin, Chile (Ruetten, 2021), and employed realistic acoustic impedances. The models that used a series of zero-phase Ormsby wavelets and 1D convolution (Langenkamp et al., 2021) allowed us to better understand how seismic bandwidth and seismic attribute parametrizations affect the resolution of stacked stratigraphic features in a seismic-scale channel system, providing insights that could be beneficial to the industry for drilling decisions, whether it is for hydrocarbon, geothermal, or CO_2 storage purposes.

In order to extract the most value and information from the seismic data, seismic interpreters often derive seismic attributes from the data to reveal additional stratigraphic or structural features. These attributes provide a means to enhance vertical and lateral changes in reflectivity, thickness, continuity, and orientation of seismic features. From the exhaustive list of seismic attributes existent, we focus on amplitude-derived, instantaneous, and geometric attributes for offering promising results in channel architecture definition (La Marca, 2020). All coherence algorithms (that belong to the geometric attributes' class) use a vertical and lateral analysis window, whether they are based on cross-correlation, semblance/variance, eigenstructure analysis, or the gradient structure tensor. For good quality data, Marfurt et al. (1998) found it best to analyze stratigraphic features using a temporal analysis window as narrow as possible, determined by the highest frequency in the data or the 3rd frequency corner in the Ormsby wavelets. For poor-quality data, a larger window approximating the dominant period of the data provides improved results with minimal stratigraphic mixing.

Pemberton et al. (2018) and Langenkamp (2021) provided insights into the effect of amplitude and frequency on architectural element imaging and interpretation and facies classification. Nonetheless, the impact of seismic attributes, the parameters, and the noise content were not evaluated. Hence, this is a one-of-a-kind study that focuses on assessing the complexities of attribute parameterization using synthetic data based on a known geologic model.

With the true model known, attribute parameterization and its effects on stratigraphic interpretation can be quantified, particularly highlighting the parameters that impact the apparent stratigraphic mixing or smearing of events such as the windows of analysis.

We first describe the aspects related to the architectural (outcrop) model and the characteristics of the synthetic datasets used. Then, our workflow is explained, providing details

on the four parameters evaluated. Results are presented and focused on the cases derived from parameter combinations/sensitivity analysis. In the end, we provide a table and workflow that allows geoscientists to identify potential pitfalls in interpretation and address them according to their individual datasets, as best practices in interpretation should be documented and available to the geoscientific community to help reduce uncertainty in reservoir characterization.

Geological model description

The geological model that is the basis of this study is derived from a sandstone-rich deepwater channel system along a progradational slope system (Hubbard et al., 2010). These deepwater slope deposits from the Late Cretaceous (70-80 Mya) Tres Pasos Formation are exposed on approximately 3 km long, 200m thick outcrops near Laguna Figueroa in the Magallanes Basin, Southern Chile (Macauley and Hubbard, 2013; and Hubbard et al., 2014). The high quality of the stacked channel systems has been used to construct a seismic-scale 3D architectural model of the deepwater channel system (Pemberton et al., 2018; Jackson et al., 2019; Langenkamp et al., 2020; Ruetten, 2021). According to Fildani et al. (2013) the outcrop is analogous to many slope channel systems globally in stratigraphy and depositional setting, which makes it an excellent benchmark for any study that aims to address problems associated with channel interpretation.

The models are the result of several studies from Macauley and Hubbard (2013), Fletcher (2013), and Southern et al. (2017) combining measured sections, hierarchical stratigraphic interpretations, paleoflow measurements and thousands of GPS data points that calibrated a drone-derived photomosaic. For these models, the fundamental architectural component are channel elements, defined as distinct, mappable channelized sedimentary bodies (Figure 1.1). Multiple stacked, related channel elements form a channel complex, and two or more complexes form a

channel complex set (McHargue et al., 2011; Macauley and Hubbard, 2013, Meirovitz et al., 2020; Figure 1.1). The outcrops at Laguna Figueroa contain two complex sets, simply referred to as the Upper and Lower Figueroa. The upper complex set consists of eight channel elements and are grouped into four distinct channel complexes. The lower complex set contains twelve channel elements grouped into three channel complexes. Elements are modeled with a standardized width of 400 m and thickness of 25 m. Three additional architectural components are present in the outcrop: mudstone drapes at the base of channel elements, mass transport deposits (MTDs) at the base of channel complexes, and inner-levee thin-bed deposits encasing the channelized elements (Macauley and Hubbard, 2013; Hubbard et al., 2014). The geological models consist of five facies: 1) channel element axis in yellow, 2) channel element off-axis in orange, 3) channel element margin in brown, 4) homogeneous shale in gray, and 5) background shale (inner and outer levee facies) in white (Figure 1.2C).

Jackson et al. (2019) developed the first fine-scale geocellular model combining channel planforms and vertical stacking for the lower outcrop section (lower channel system) but did not include hierarchical groupings in the architecture. Pemberton et al., (2018) generated forward seismic models using Jackson et al. (2018)'s model and analyzed seismic interpretation of architecture as a function of seismic resolution. Nielson (2018) analyzed the tuning effects of single channel elements. Ruetten (2021) updated Jackson's initial model with new interpretations and added an upper channel system separated from the lower system by a debris flow, and studied how stacking patterns impact reservoir connectivity and fluid flow. Finally, Langenkamp (2021) analyzed the influence of stacked channel element architecture on facies classification using Ruetten's model. This work utilizes the geocellular model of Ruetten (2021) and synthetic seismic models from Langenkamp (2021).

The five synthetic seismic models used in this study were built using a series of zero-phase Ormsby wavelets of 15 Hz, 30 Hz, 60 Hz, 90 Hz, and 180 Hz and 1D convolution (Chile Slope Systems research consortium; Langenkamp et al., 202; Figure 1.2C) with a reflectivity model. More aspects of each model are found in Langenkamp (2021). Facies-based rock properties (Figure 1.2B), adopted from Stright et al. (2014), show that amplitude peaks represent an increase in acoustic impedance (Figure 1.2D). In contrast, troughs depict a decrease in acoustic impedance. The synthetic volumes have a vertical window of 500 ms. For analysis purposes, we cropped the volume from 120ms to 380ms to avoid dead/blank zones in the reflectivity and focus on the target channel systems.

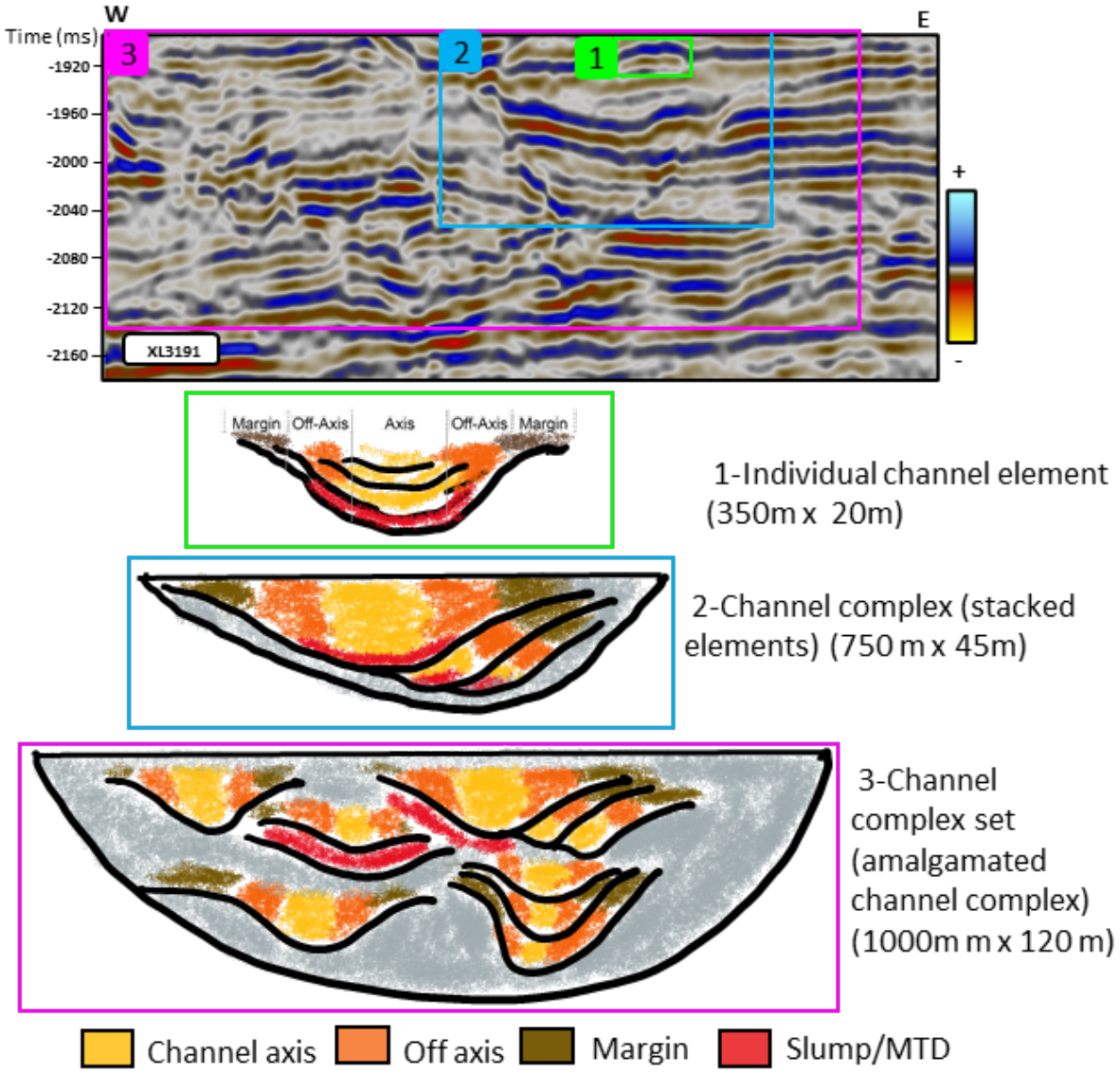


Figure 1.1. Stratigraphic hierarchy: from a single element to complex sets. A vertical slice within the Pipeline 3D dataset offshore New Zealand shows the seismic appearance of different architectural hierarchies in deepwater channels, with corresponding cartoons below. The smallest architecture (4^{th} to 5^{th} order) is the channel element (box 1 in green). The second hierarchy (6^{th} order) occurs when the channel elements stack together, forming a channel complex (box 2 in blue). The higher-order hierarchy (7^{th} order and higher) occurs with the amalgamation of channel complexes, developing a channel complex set (box 3 in magenta). The color legend indicates the distinct facies that commonly occur within each element of each architecture. Measuring the sizes of each architectural element indicated on the right as well as their hierarchy, provides key insight into the underlying depositional processes as well as a prediction of the more common lithologies. Hierarchies mentioned follow Pickering and Cantalejo (2015) classification.

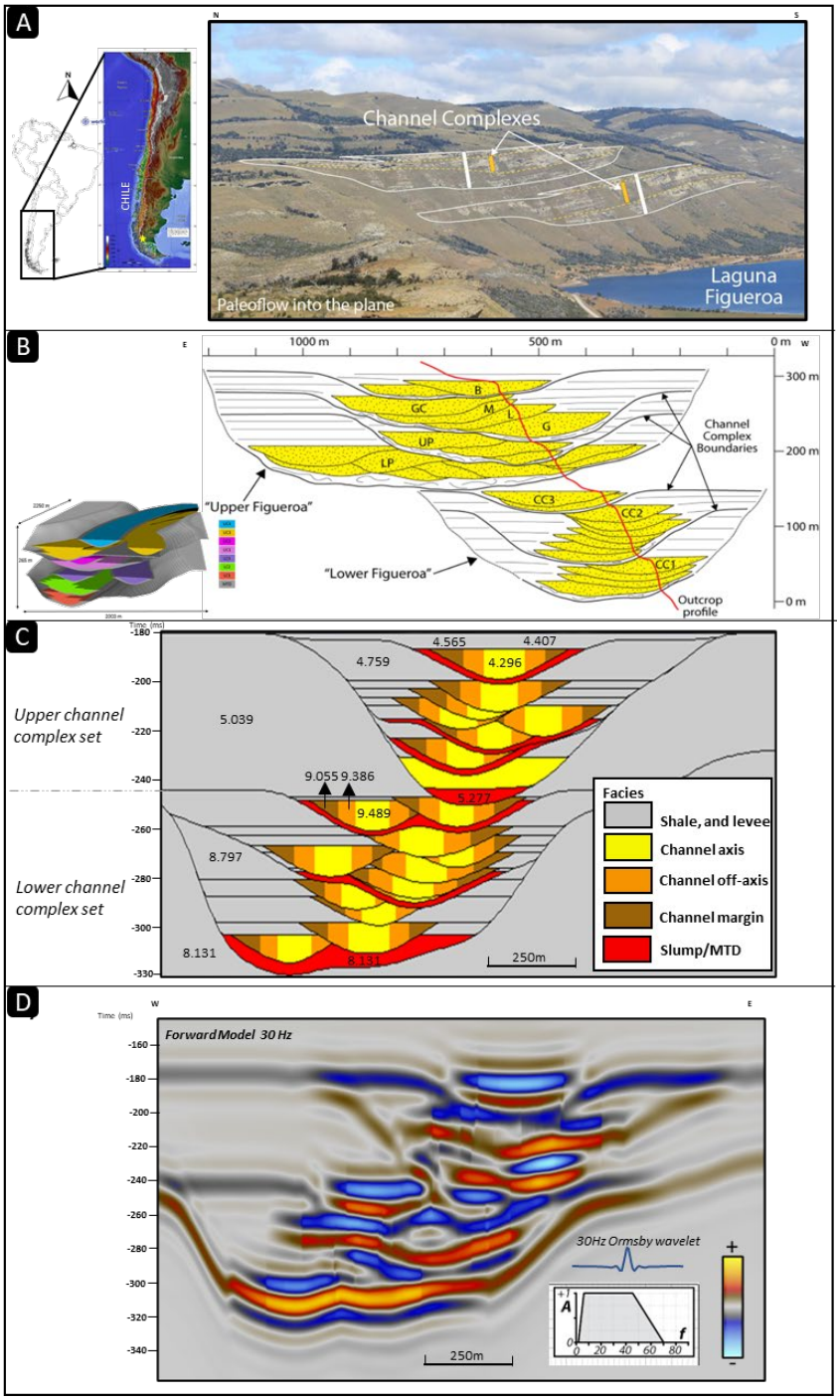


Figure 1.2: A visual guide showing the steps to convert outcrop measurements to a synthetic model. (A) Location and exposure of the outcropping deepwater channels at Laguna Figueroa. Paleoflow is from North to South (obliquely and to the right into the outcrop at this location) (B) Conceptual diagram of the Upper and Lower Figueroa outcrops showing channel elements, complexes, and complex sets. The red line indicates the outcrop profile. Left of the line is into the outcrop face, and the right has been eroded away. (C) geocellular model using the constraints from (B) augmented by facies and corresponding rock properties, including acoustic impedances from Shallow Offshore West African modeled rock properties (Stright et al., 2014). (D) The Ormsby wavelet and a representative vertical slice through the 3D synthetic seismic data volume generated from the model shown in (C). Courtesy of Teresa Langenkamp and Lisa Stright.

Methods

To address the questions posed in the study and focus on analyzing the effect of bandwidth on vertical resolution, we introduce the term "stratigraphic event mixing." This vertical smearing phenomenon is explained in Figure 1.3, using a real example where the interference of channels from other stratigraphic levels is evident. As depicted in Figure 1.4, first, we performed an exploratory data analysis to define a vertical window of interest constrained to the objective of the study: 120 ms and 380 ms from our five volumes of synthetic seismic data. After cropping the 15 Hz, 30 Hz, 60 Hz, 90 Hz, and 180 Hz dominant frequency volumes, we performed a sensitivity analysis on four parameters: 1) frequency content, 2) seismic attribute effect, 3) the effect of the analysis window, and 4) the influence of band-limited random noise.

Next, we calculated a series of seismic attributes from amplitude accentuating, geometric, and instantaneous attributes commonly employed in seismic interpretation of channel systems. Due to the number of cases to evaluate, we decided to explore and present the most representative seismic attributes for each case in detail (the most commonly used and that provided better results). Next, for each scenario, we defined a suite of 3-trace by 3-trace analysis windows with various vertical lengths from 2 ms to 50 ms. For visualization and interpretation purposes, we used co-rendering techniques. Finally, we explored the impact of the addition of low and high levels of noise.

The final analysis of the results was performed by combining the aforementioned parameters into the following cases: 1) the response on the same attribute and analysis window in the different bandwidth volumes, 2) the impact of changing the analysis window for different seismic

attributes, 3) the effect of changing the window of analysis size for the same seismic attribute, and 4) the effect of noise. All cases were contrasted with the initial actual data/model.

More details and considerations taken in each parameter evaluation are presented below.

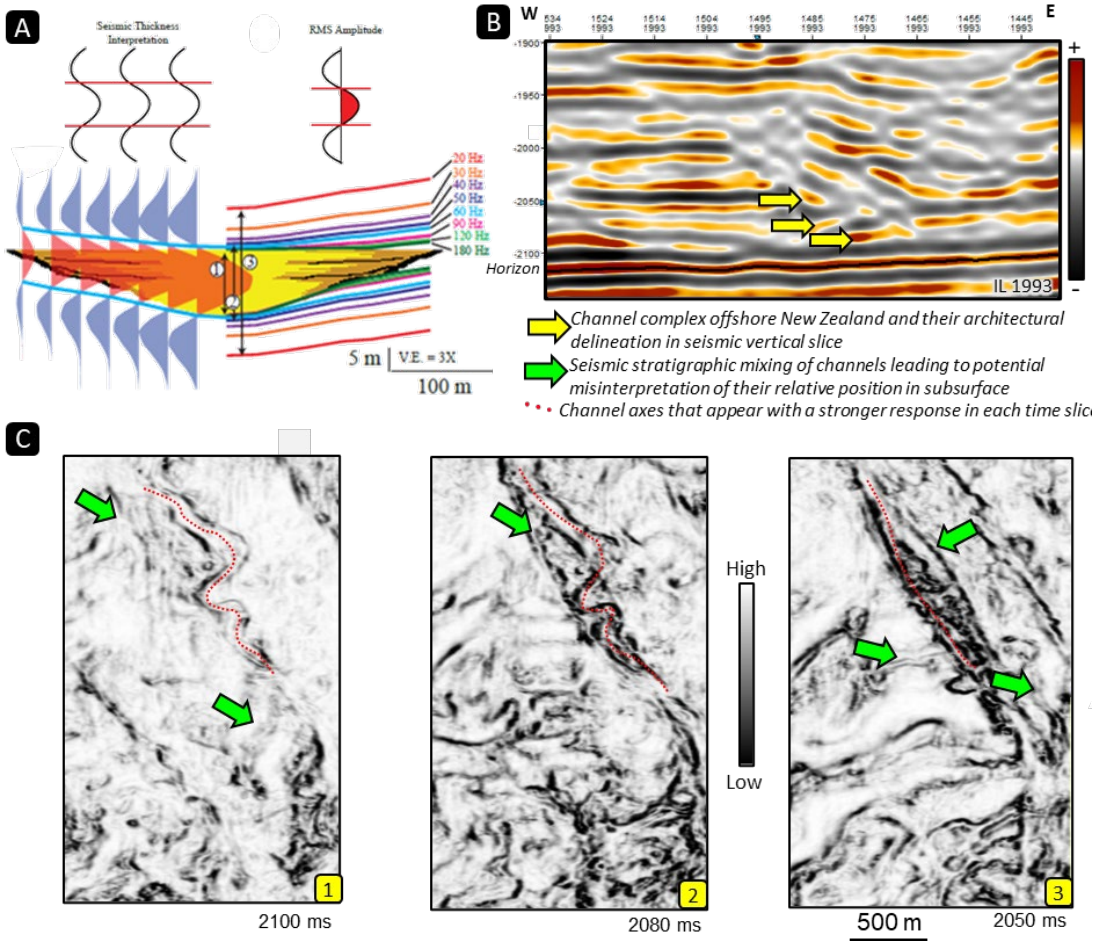


Figure 1.3. Explanation of stratigraphic event mixing or vertical smearing of stratigraphic features. (A) Vertical resolution of channel elements related to different peak frequencies (modified after Nielson, 2011). (B) Vertical slice through the 3D Pipeline 3D offshore New Zealand seismic survey showing a channel complex and a Horizon used for interpretation at different stratigraphic levels indicated by yellow arrows. The dominant frequency at this level is 40 Hz giving a dominant period of 25 ms. (C) Stratal slices at approximately 25 ms intervals through the coherence volume computed using a ±20 ms analysis window show "stratigraphic" mixing by the seismic wavelet. Note that the relatively straight channel form at 2050 ms can be seen at 2080 ms and other stratigraphic levels (1,2,3 from deeper to shallower) where other channel forms (green arrow) appear causing interference. The cause of this mixing could be 1) mixing of reflectivity by the 25ms dominant period seismic wavelet, 2) mixing of discontinuities through the 40 ms coherence computation, 3) shifting of the basal channel element thalweg due to compensational style as you move up or 4) differential compaction over deeper discontinuities between the floodplain and the channel element fill.

4.1. Parameter 1: frequency content effect

Tuning thickness is the temporal resolving power of seismic data. Some authors use resolution and tuning thickness terms interchangeably, although tuning starts right below the vertical resolution. Table 1.1 presents the tuning thickness for each element in the synthetic volumes studied here. Knowing the resolvability in seismic is paramount to understanding the effect of other parameters considered in the sensitivity analysis.

Resolution is the ability to resolve by seismic interpretation methods two features that are close together. By definition, the vertical resolution of seismic data is ¼ of the wavelength (λ), where the λ is determined by dividing the average velocity by the dominant frequency.

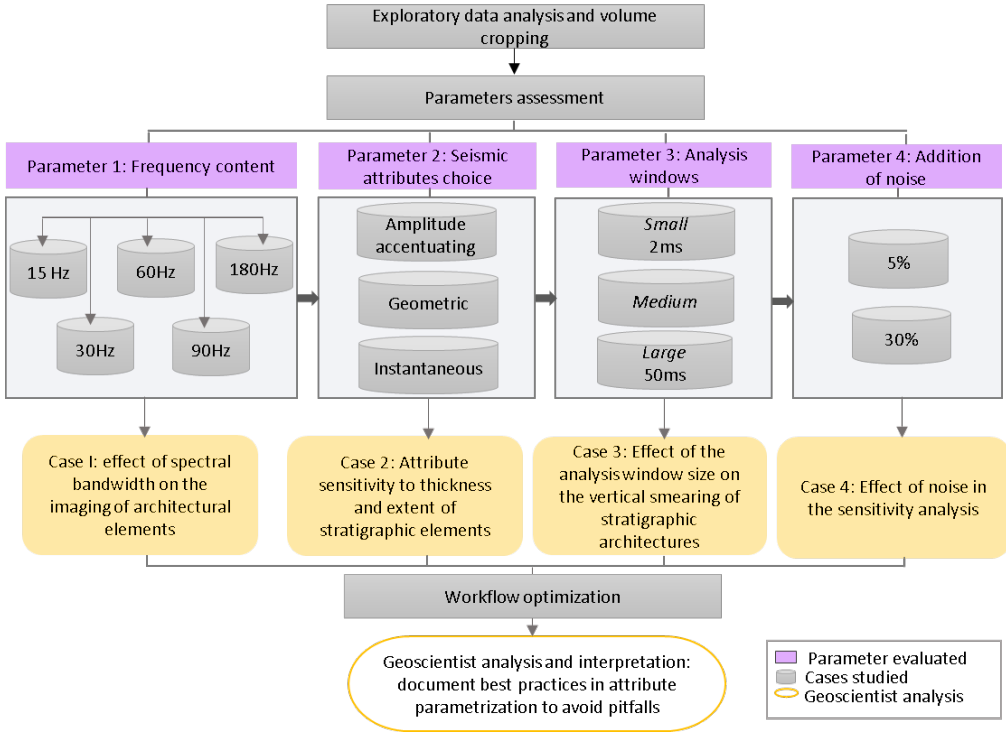


Figure 1.4. Workflow of the study. Four different parameters were evaluated in the sensitivity analysis: 1) the effect of frequency content, 2) the impact on the choice of the seismic attribute, 3) the analysis window effect, and 4) the sensitivity to band-limited random noise addition. The best cases were selected to be shown in each case, and analysis was performed by comparing them with the original 3D geological model derived from the Laguna Figueroa Deepwater outcrop. This resulted in 4 cases for analysis, ultimately leading to a workflow and documentation of best practices in channel architecture interpretation.

The frequency content was the first parameter evaluated for the impact on the interpretation and resolvability of the different architectural elements. The five dominant frequency synthetic seismic volumes were analyzed, and we compared the results with those of Pemberton et al. (2018), and Langenkamp (2021).

	Channel elements	**MTDs**
Dominant Frequency (Hz)	**Tuning Thickness (m)**	**Tuning Thickness (m)**
15	48.9	53.2
30	24.4	26.6
60	12.2	13.3
90	8.1	8.9
180	4.1	4.4

Table 1.1. Tuning thicknesses for shallow and deep elements in each synthetic seismic volume. Modified from Langenkamp (2021).

4.2. Parameter 2: Choice of seismic attributes

A seismic attribute is a computation made from algorithms applied to seismic data to get a more interpretable output. These responses relate to rock physical properties (La Marca, 2020) in rocks and fluids in the subsurface. However, there are tens if not hundreds of seismic attributes (Barnes, 2016), and time constraints do not allow for testing them all. Some of the latest studies (Posamentier and Kolla, 2003; Chopra and Marfurt, 2007; Hossain, 2020; and La Marca and Bedle, 2022) have proven the successful application of amplitude accentuating, geometrical and instantaneous attributes applied to PSTM data to interpret and characterize channel elements and complexes in both fluvial and deepwater settings. Generally, we need the

combination of a geometrical attribute that allows defining edges and at least one attribute that provides insights into stratigraphy (La Marca et al., 2019) to characterize channel features in seismic data. Therefore, we focused on testing attributes that belonged to those three attribute categories and chose the most prominent of each class.

Table 1.2 summarizes the most representative seismic attributes selected for each class: root mean square amplitude (RMS), Sobel filter coherency, and instantaneous frequency.

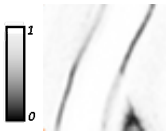
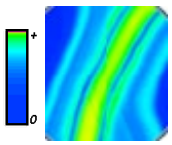
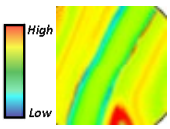
Attribute	Appearance	Attribute Category	Measurement	Use in architecture interpretation
Coherence		Geometric	Direct measure of waveform similarity or how similar waveforms or traces are in a volume- used to emphasize continuous events or edges.	Delineates edges of channel elements
RMS amplitude		Amplitude/Energy	Measures the square root of the average energy within a vertical window.	Provides measure of channel element vs inner levee Provides statistical measures of channel element fill between two picked horizons
Instantaneous frequency		Spectral	A simple approximation of the mean frequency of the seismic wavelet.	Channel element thickness

Table 1.2. Selected seismic attributes to perform the sensitivity analysis showing a representative image, feature measurement, and use in channel element interpretation.

4.2.1. RMS amplitude

The RMS amplitude is an amplitude accentuating attribute often used for stratigraphic and lithologic variations enhancement. It is defined by the standard deviation, σ(t), of the data, d(t), within a running analysis window, subsequently measuring the reflectivity within that

window (Meek, 2015). For a window that ranges from -T=-KΔt to +T=+KΔt about a sample j, the RMS amplitude is:

$$d_{RMS}(j\Delta t) = \sigma(j\Delta t) = \left(\frac{1}{2K+1}\sum_{K=K}^{+K}\{d[(j+k)\Delta t)]\}^2\right)^{1/2} \quad . \text{Eq (1)}$$

In this study, we evaluated how well RMS showed stratigraphic variations.

4.2.2. Coherence (Sobel Filter)

Seismic coherency is a measure of how similar traces are among their neighbors, which is a response to lateral changes in the seismic record caused by variations in structure, stratigraphy, lithology, porosity, and the presence of hydrocarbons (Marfurt et al., 1998), and it is determined computing amplitude derivatives along structural dip. Sobel filter (Luo et al., 1996) is one of the many coherence methods, which for seismic data normalizes coherence data to produce results between 0 and 1, where 0 is the lowest coherence, and 1 is the highest coherence. It has proven to be effective in delineating channel element edges (La Marca and Bedle, 2021; Hossain, 2020; and Herron, 2011); therefore, we aimed to test the definition of the channel elements' edges.

4.2.3. Instantaneous frequency

The instantaneous frequency is computed sample by sample, is the time derivative of the instantaneous phase φ(*t*):

$$F(t)=\mathrm{d}[\varphi(t)]/\mathrm{d}t. \quad \text{Eq (2)}$$

and provides a simple estimate of the mean frequency of an isolated seismic event.

Subrahmanyam and Rao (2008) find that the instantaneous frequency attribute can indicate bed thickness and provide lithology insights. Chopra and Marfurt (2007) emphasize

their usefulness in identifying abnormal attenuation and thin-bed tuning. Our model has thin features, such as channel elements (from axis, off-axis, to margins), that we aimed to test by using this instantaneous attribute.

4.3. Parameter 3: Analysis window effect

According to Lin et al. (2014), the scale of the window height, H, is a function of the dominant or peak frequency f_{peak}:

$$H=1/(2f_{peak}) \quad . \quad \text{Eq (3)}$$

Applying this concept to our dataset spectrum, we would need to use a window of 2ms for the highest frequency volume (180Hz) and ~33 ms for the lower frequency volume (15Hz).

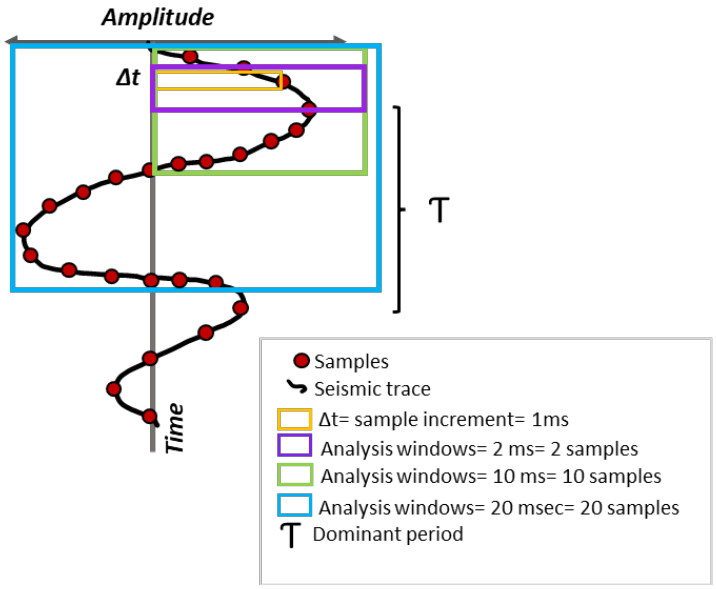


Figure 1.5. A representative analysis window used in attribute calculations. Signals are sampled at discrete points, not continuous recordings. Therefore, each seismic trace will contain as many samples as the sample rate allows. For the example shown, if we consider the dominant period (distance between two peaks), our analysis window will contain 20 samples, equivalent to ~ 20 msec. Other examples of analysis windows are depicted, including the smallest possible equivalent to the sample increment, in this case, 1 msec.

To understand the concept of analysis window, we have drawn a cartoon (Figure 1.5) that shows a seismic trace with its respective samples. It illustrates how a small, medium or large analysis window would look and what a dominant period is. Usually, the number of samples considered in a default analysis window parameter setting is around 11 samples (default window 60ms) for a 6ms sample rate dataset; in our dataset, the sample rate is 1ms. In real data, shallower, higher frequency data often shows smaller periods than the deeper strata/intervals where lobes of the traces are wider. Due to this change in frequency, some authors, like Lin et al. (2014), recommend using an adaptive window. However, this is not often possible or available in the software packages commonly used by geoscientists. Nonetheless, the geoscientist/interpreter can control the analysis window in the attribute settings. Therefore, we want to provide insights into the difference between a default analysis window and a shorter or larger one (which reduces or increases the number of samples, respectively). We compared the results from each different analysis window to the original, true data (the geological model), which provides a good sensitivity analysis for all kinds of datasets and any of the attribute families studied here.

4.4. Parameter 4: Addition of band-limited random noise

When seismic data is recorded, we find two components: signal and noise. The latter comprises all the unwanted recorded energy that contaminates seismic data, and it can be random or coherent (Kumar and Ahmed, 2021). Random noise is generated by activities in the environment where seismic acquisition work is being carried out, and this noise appears in a seismic record as spikes (Enwenode, 2014). Seismic noise levels depend on the type of acquisition—land or marine—and the intrinsic conditions unique to every site, such as climate, the burial of sensor, and wind (Tanimoto et al., 2015). Although there are many types of noise,

like Gaussian, Pink, Brownian, violet, and blue, in this study, our fourth parameter incorporated in the sensitivity analysis is the band-limited random noise.

Signal-to-noise ratio (SNR) is a measure used to compare the degree of signal to the level of background noise, in which case a ratio larger than 1:1 suggests more signal than noise. So, the lower the SNR, the noisier our dataset. As Chen et al. (2019) mentioned, this will lower the quality of the seismic, affecting subsequent analyses such as imaging and inversion.

We incorporated different noise levels, from a low noise added of 5% to a high 30%. Higher noise levels were not presented due to the incapacity to extract meaningful interpretations from the data. Nowadays, noise can be added to synthetic datasets thanks to available software like the one used in this study.

Results

To address the study questions and link the parameters taken into account in the sensitivity analysis, we present the results summarized in four cases:

5.1. Case 1: The effect of the spectral bandwidth on the imaging of architectural elements

After evaluating the five synthetic models of 15 Hz, 30 Hz, 60 Hz, 90 Hz, and 180 Hz dominant frequencies, we observed that the level of detail of the different sized architectural elements increases with frequency; therefore, broadband, higher frequency provides better resolution.

Figure 1.6 shows the effect of each frequency in imaging each geological element. Figure 1.6B corresponds to the lowest frequency, and we can distinguish channel complexes and mass transport deposits (MTDs). However, smaller features like the 25m thick channel elements and MTDs are mixed in thicker, unresolvable reflectors.

Figures 1.6C and 1.6D show frequencies commonly encountered in the subsurface (30Hz and 60Hz for vintage and recent data, respectively). Here, the individual complexes are well-defined. Nonetheless, the amalgamation or stacking of elements presents a single response and minimal acoustic impedance contrast.

Higher frequencies shown in Figures 1.6E and 1.6F present the best responses compared to the original model. Figure 1.6E can even differentiate some of the stacked packages; overall, the individual and stacked channel elements can be better resolved. It is noteworthy how the highest frequency (180 Hz) illustrated in Figure 1.6F starts to lose the definition of the inner reflectors. Another observation is that, in all cases, the vertical channel axis (center) is resolved better than the channel's margins, both as a function of thickness (channel elements are thickest

in the center at the axis) and acoustic impedance contrast (rock properties of channel element margin is more similar to inner levee than the channel element axis).

In general, channel complexes sets are visible at all frequencies analyzed, whereas individual complexes start to be resolvable from 30 Hz and higher. However, when elements have vertical stacking, channels do not show contrast in acoustic impedance due to repeated material/ similar composition and properties, therefore the attribute response is also affected by this phenomenon.

5.2. Case 2: Attribute sensitivity to thickness and extent of stratigraphic elements

After performing the seismic attribute sensitivity analysis, we noticed that the thickness and extent of stratigraphic events imaging are inherently linked to the frequency content of the seismic, analysis window, and seismic attribute used. At peak frequencies commonly encountered in the subsurface (around 30- 60Hz), the number of complexes was underestimated, and the size, shape, and type of architectural bodies (channels elements vs. margins, elements vs. complexes) were difficult to differentiate (Figure 1.7).

In most cases, channel complex sets were able to be interpreted. However, the amalgamation of smaller-scale channel elements results in an incorrect estimation of thicknesses. Nielson (2016) documented the same phenomenon.

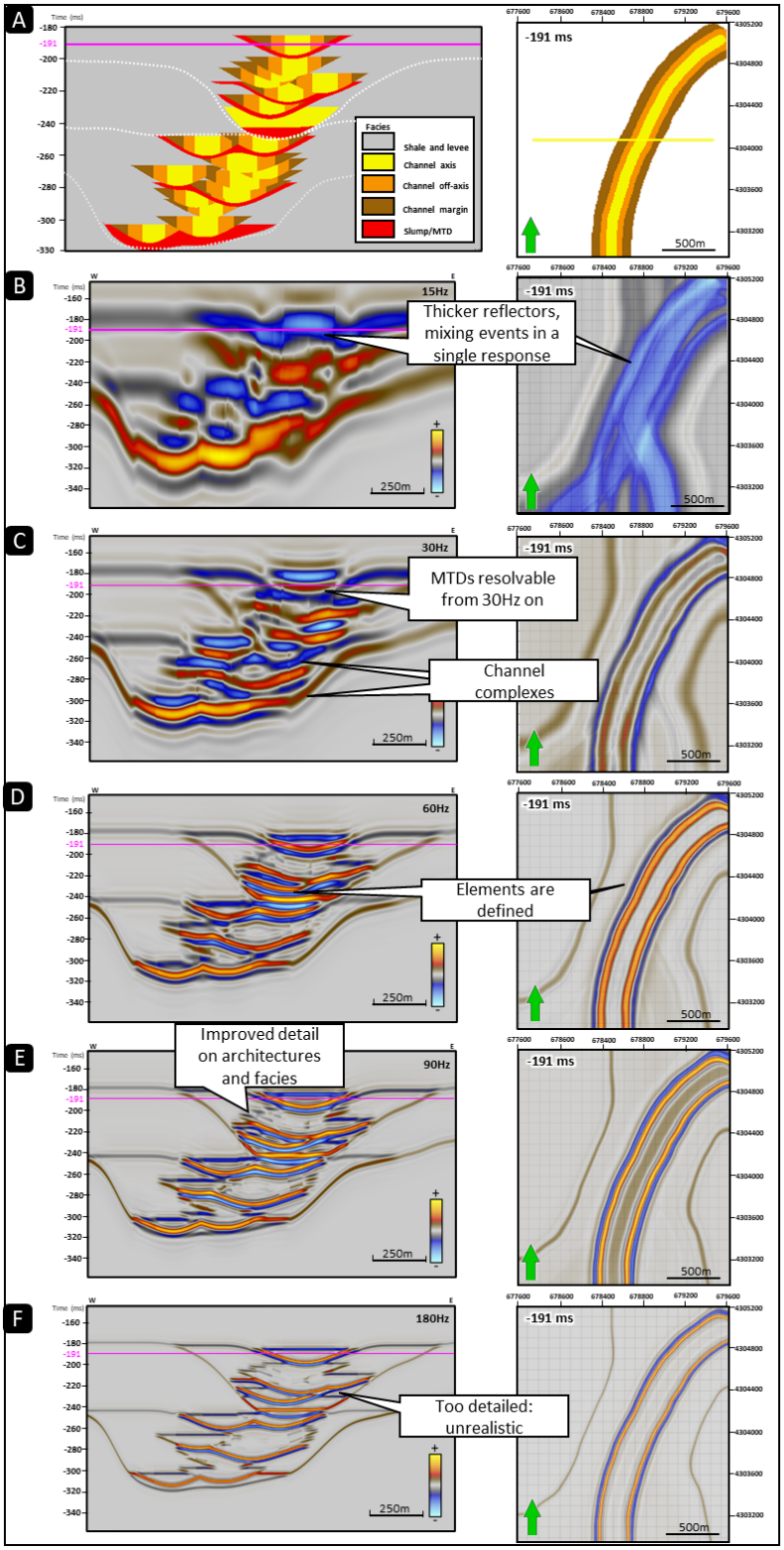


Figure 1.6. A representative vertical slice (224) and time slices at *t*=-191ms) through the (A) 3D model that shows (B) 15 Hz data, (C) 60 Hz data, (D) 90 Hz data, and (E) 180 Hz data. Notice the improvement in the channel architecture's detail with the increase in the dominant frequency and corresponding spectral bandwidth, being able to interpret complexes from figure B on.

For example, in cases where similar facies were in contact, the acoustic impedance similarities did not allow for the differentiation of individual channel elements. Also, in tests performed with a large analysis window, the interpretation of smaller features was not possible (Figure 1.8B, 1.8D).

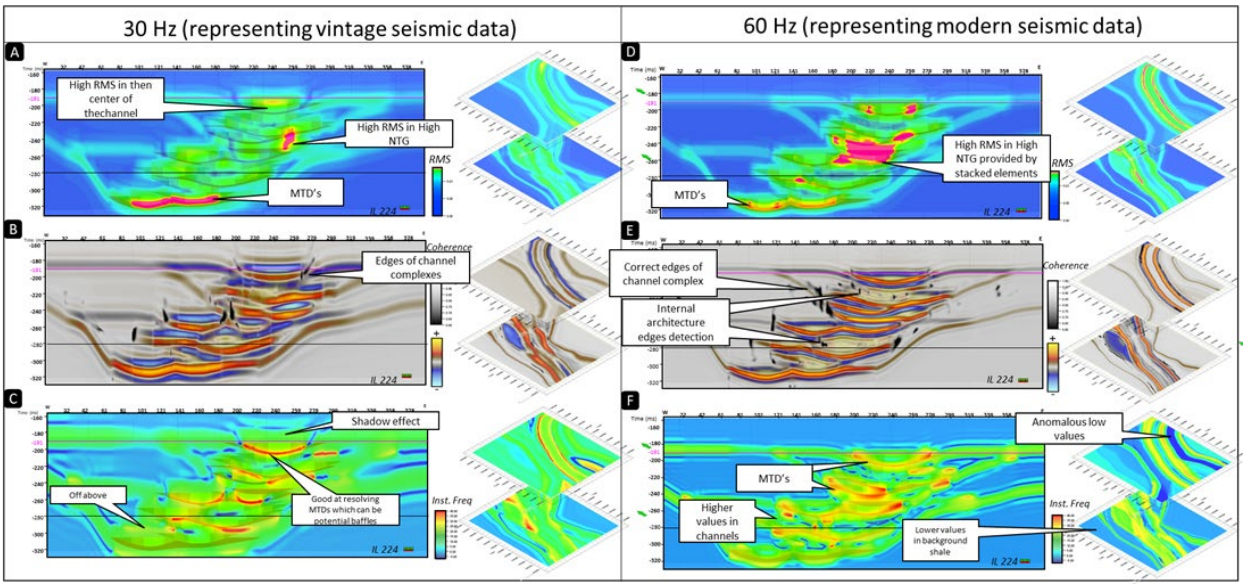


Figure 1.7. Results on the attribute sensitivity analysis to thickness and extent of architectural elements using a default analysis window. We present results on the 30 Hz dataset, analogous to vintage seismic data (to the left), and the 60Hz dataset, representing modern seismic data (to the right). Each attribute is presented for both frequencies in a representative inline (224) and time slices (-280 ms and –191ms) as follows: (A) RMS (Root Mean Square) amplitude for 30 Hz (B) coherence (Sobel filter) for 30Hz, (C) Instantaneous frequency for 30 Hz, (D) RMS amplitude for 60 Hz (E) coherence (Sobel filter) for 60Hz, (F) Instantaneous frequency for 60 Hz.

5.3. Case 3: Effect of the analysis window size on the vertical smearing of stratigraphic architecture

To demonstrate the effect of the window size on the vertical smearing of the different architectural elements in the model, we show an RMS amplitude sensitivity analysis (Figure 1.8). We perceived that RMS amplitude offers a detailed picture of the various facies in the 180 Hz volume, including imaging of the MTD associated with the channel complexes and the

individual channel elements. However, for larger analysis windows and lower frequency volumes, for example, 50 ms window combined with a 15Hz dominant period seismic volume, such details are lost, resulting in an accentuated vertical stratigraphic smearing effect. Therefore, the application of amplitude-derived seismic attributes results more effective in higher frequency content datasets and using a small analysis window than in another configuration. It is noteworthy that a higher value of RMS is presented where an amalgamation of events occurs, which potentially leads to the interpretation of higher NTG (Figures 1.7 and 1.8).

Since the Sobel filter coherence seismic attribute aids in detecting discontinuities, such as geological structures and edges, especially in time slices, we included this attribute in our analysis to identify the channel complexes and the edges of the channel elements (Figure 1.9).

We observe stratigraphic mixing occurring in the vertical slices when we utilize a larger analysis window. It is also noticeable that higher dominant frequency data in combination with a small analysis window allows for a correct placing of feature edges (Figure 1.9C), which is supported when compared to the true model. In contrast, the position of the channel element edges deviates from the original/ true location or becomes distorted when the analysis window size increases (Figure 1.9D) or the frequency of the data is small (Figure 9A) regardless of the analysis windows used.

5.4. Case 4: Effect of noise in the sensitivity analysis

Figure 1.10 shows the most prominent results of the sensitivity analysis on the noise effect. We compared, in this case, how instantaneous frequency, coherence, and RMS amplitude seismic attributes respond to variations in band-limited random noise from 5% to 30%.

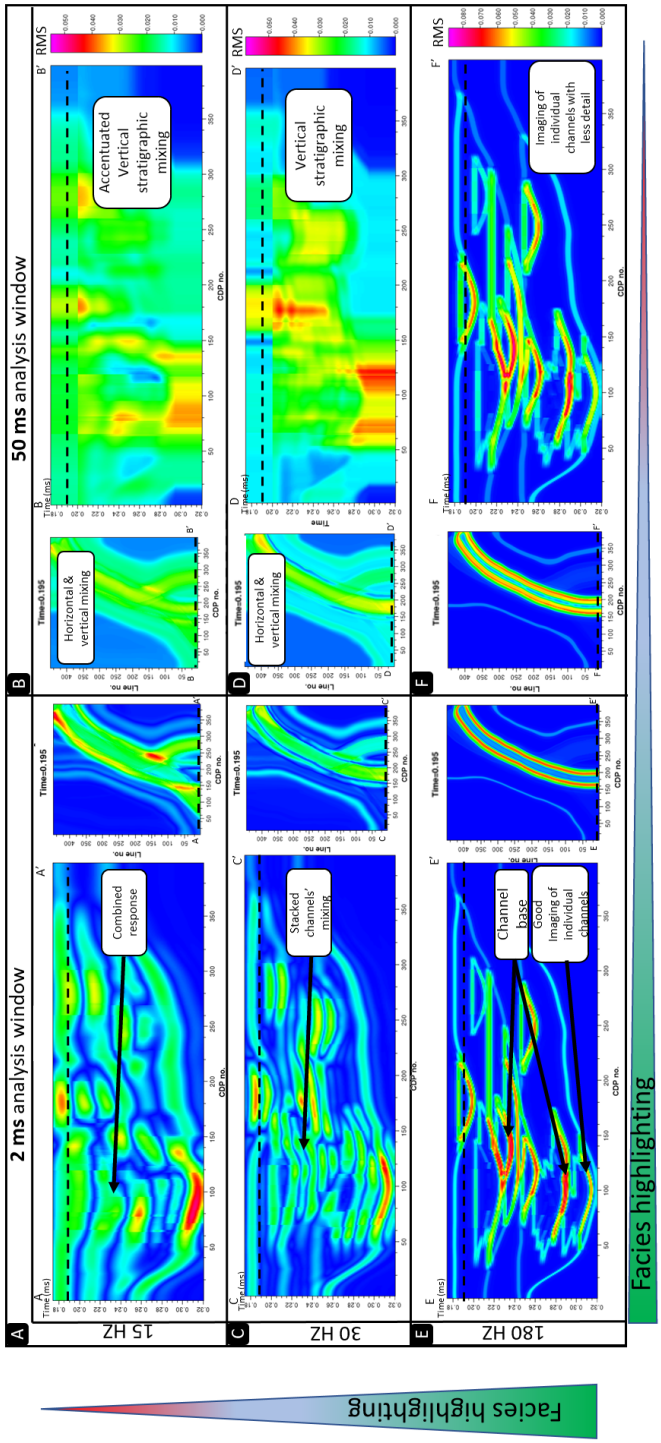


Figure 1.8. RMS amplitude sensitivity analysis to evaluate the effect of the window size *W* on the vertical smearing of the different channel elements in the model. (A) 15 Hz with *W*=2ms (B)50 Hz with *W*=50ms (C) 30Hz with *W*=2ms (D) 50 Hz with *W*=50ms (E) 180 Hz with *W*=2 ms (F) 180 Hz with *W*=50 ms. High-frequency data with a small analysis window provides the most suitable representation of the true model. In RMS amplitude, the most accurate facies depiction is given by high frequencies and small window combination, imaging channel element base, and inner and outer levee facies.

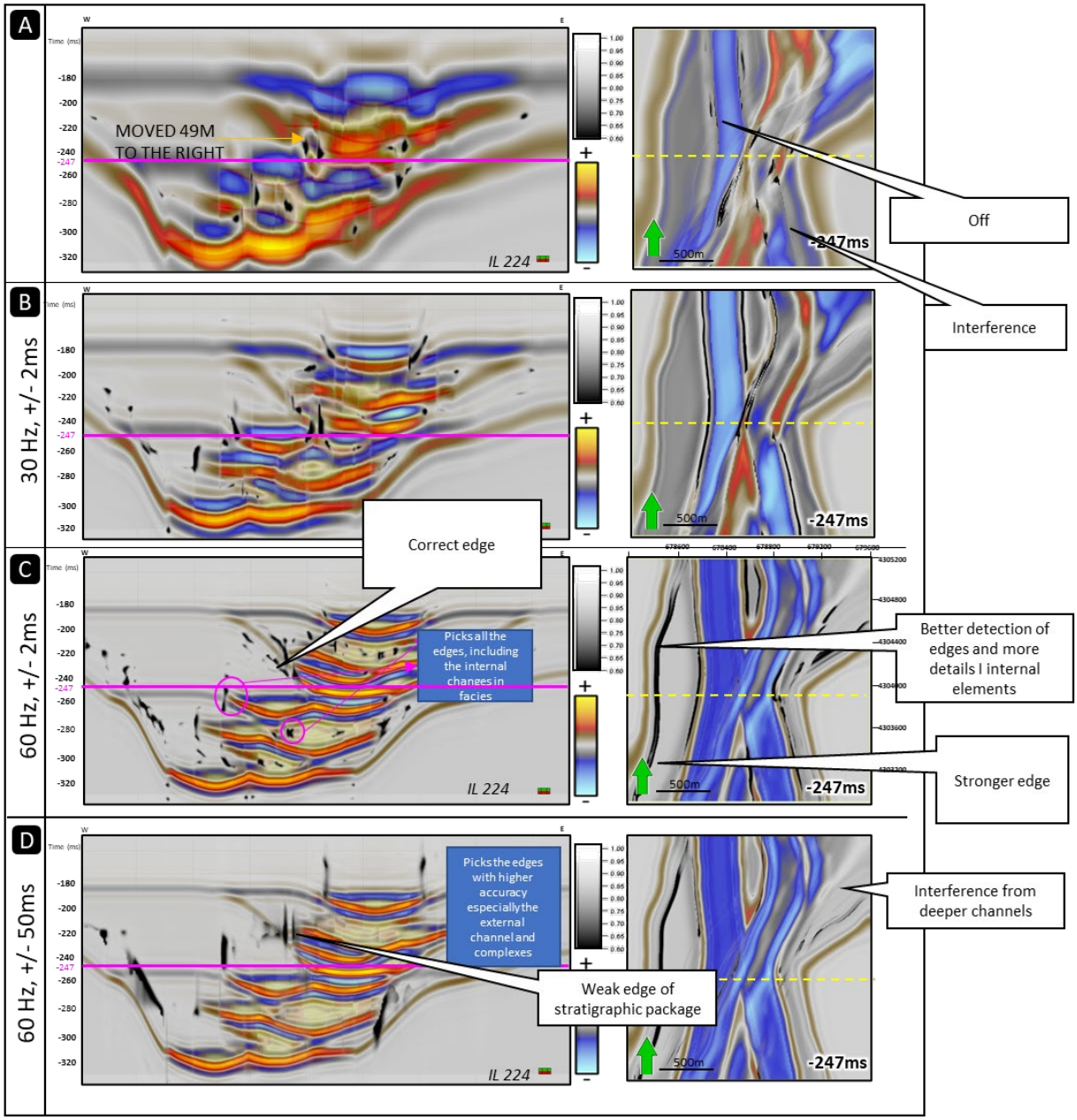


Figure 1.9. Sobel filter coherence attribute sensitivity analysis to evaluate the effect of the window size *W* on the vertical smearing and displacement of the different architectural elements in the model. (A) 15 Hz with *W*=2ms (B)30 Hz with *W*=2ms (C) 60Hz with *W*=2ms (D) 60 Hz with *W*=50ms. Notice how at a higher dominant frequency (60Hz) and smaller window of analysis, there is better detection of the channel element edges and channel element fill. In contrast, at lower frequencies and or larger windows of analysis, there is interference from deeper channel elements, as in the case presented on Figure 1.3.

The most evident finding is that with the increase in noise content, interpretation becomes more challenging since the amount of noise hinders channel element edge and facies detection. A

particular result found is a low coherence value predominance when evaluating the 30% noise case (Figure 1.10B).

5.5. Pitfalls in attribute interpretation and how to minimize them

The results mentioned above warn of potential pitfalls when interpreting seismic data. For example, not understanding the frequency content and quality of the seismic dataset could result in the inappropriate use of seismic attribute parameters and subsequent misinterpretations. To summarize all these results, we present a list of common pitfalls (Table 1.3) in attribute interpretation and how the interpreter can attempt to avoid them.

Common misinterpretation	How to address it	Comments
Not inspecting the frequency content and dominant frequency of the data	Inspect the seismic volume and understand dominant frequency and variations with depth	Higher frequency, broader bandwidth data provide higher vertical and lateral resolution
Not understanding the quality of the data and resolution	Calculate the seismic resolution	Imaging of different scales of architecture improves with a higher signal-noise ratio
Using meaningless attributes	Have a clear geological goal and calculate attributes effective for similar targets in literature	Avoid attributes that make pretty pictures but are not directly related to the target objective.
Using default parameters	Select parameters adequate to your study. In windows-based attributes, smaller windows provide less vertical mixing whereas larger windows are less sensitive to noise. Because data quality and resolution change with depth, parameters appropriate for the shallow part may be suboptimum for the deeper part of your survey.	Defaults are provided for the most common cases encountered by the most common user (e.g., oil and gas exploration). If your data are unique in either acquisition or objective, test a wide variety of parameters and choose the ones that best delineate your target.
Interpreting data that has a low S/N ratio, or high noise	Inspect your seismic volume and categorize the noise type and level when possible. Then apply a noise reduction or removal algorithm	Noise can hinder the interpretation of seismic facies and calculating attributes won't help. If data are noisy, we recommend just using amplitude-related attributes.

Table 1.3. Common pitfalls in seismic interpretation and how to avoid them.

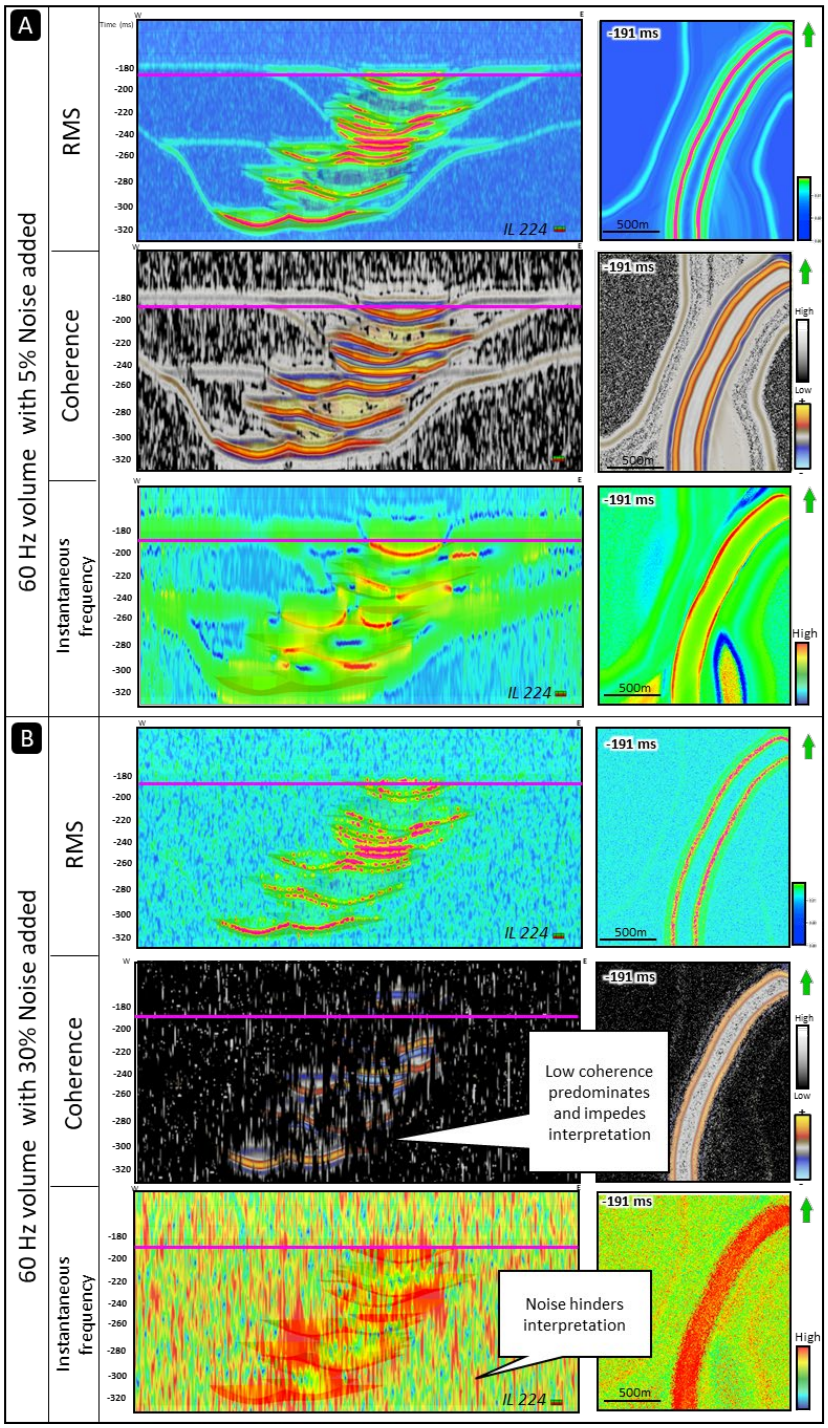


Figure 1.10. Sensitivity analysis of the effect of random noise in channel architecture interpretation. Results shown refer to the 60 Hz dominant frequency seismic volume (which would be analog to real datasets). (A) 5% band-limited random noise added evaluated in RMS, Coherence, and Instantaneous frequency attributes. Interpretation of prominent features like channel complexes is only possible. (B) 5% band-limited random noise added applied to RMS, coherence, and instantaneous frequency attributes. Notice how the increase in noise is detrimental in the channel architecture interpretation, especially when using coherence. This may be due to the sensitivity that small windows have on high noise content.

Discussion

6.1. The importance of mapping channel architectures

Channels systems, whether fluvial or deepwater in origin, exhibit petrophysical characteristics that make them excellent reservoirs for oil, gas, or other resources (Slatt et al., 2009). However, channel system mapping in the subsurface becomes challenging, especially when only seismic data is available.

In seismic reflection data, the amplitude, frequency, and noise content will condition the quality of the seismic image, affecting the interpretation of different architecture and facies embedded in the reflector configuration and overall seismic response. It is, therefore, paramount to understand how size-dependent architecture is displayed and imaged under different conditions and how the application of seismic attributes could help or hinder the interpretation of such architecture.

In this study, we used a unique approach by employing 3D synthetic seismic datasets as a benchmark to perform a sensitivity analysis to understand how different-scales of architecture appear as a function of frequency content, noise level, type of seismic attribute, and parametrization, especially the analysis window selected to calculate attributes.

The analysis of all the scenarios resulted in the following observations: broadband higher frequency data (e.g., 90 and 180 Hz) combined with a short analysis window (e.g., 2 ms, 20 ms) minimized the stratigraphic mixing (Figure 1.6). In contrast, lower frequency data that were analyzed using a large vertical analysis window (Figure 1.8) resulted in poor imaging of the channel architecture, vertically mixing stratigraphic architecture at different hierarchical levels. This affects the temporal evaluation of features that show an overlap of individual architecture

(i.e., channel elements) in the system in consecutive time slices, or in other words, a vertical offset from the known position (Figure 1.9) of sedimentary units (also shown by Pemberton et al., 2018).

This observation warns of potential interpretation pitfalls in applying such seismic attributes to actual seismic data. Stratigraphic mixing can hinder the correct temporal and spatial representation of individual channel elements and boundaries of channel complexes, leaving the internal architectures and potential fluid flow barriers imprecisely imaged (Coleman, 2000), hence, incorrectly mapping and estimating the volume of the reservoir units of interest, which could result in important economic loses.

6.2. The effect of frequency content in the imaging of architectural elements

Resolution has always been one of the main concerns for seismic interpreters since important features like small channels, or DHIs/ bright spots can be overlooked in seismic data. Also, hazards and baffles (Cardona, 2020; Meirovitz et al., 2020; Ruetten, 2021), like the MTDs, may not be imaged in the seismic picture analyzed. In this study, MTDs are only five meters thick and usually mantle the base of the channel complexes and complex sets.

To improve the seismic data's resolution, as demonstrated in this study by the enhanced definition of each element in higher frequency datasets, frequency content should be increased. This suggests that ideally, modifying the initial design of the seismic data and obtaining higher frequency data will allow us to get the resolution required to image submarine channel complexity to the detail commonly observed in outcrop (Coleman et al., 2000). This type of data is, however, costly.

Although the cost of high-quality data is currently high, with the advancement in technology, we presume that access to higher-quality data at reduced prices will become available soon. An example of this kind of data is the use of OBN (ocean bottom nodes) – high-resolution 3D seismic acquired with the addition of P-Cable, an offshore seismic data acquisition system that provides highly detailed ultrahigh-resolution images of the seafloor and subsurface geology (McGregor et al., 2022) – and high-resolution (1-6 kHz) chirp data that offer a better image of the stratigraphy and structure of rocks in the shallow subsurface.

It is likely that soon most of the seismic data acquired will be high resolution, without necessarily needing to be high frequency, and that we will find ways to improve our algorithms to treat the noise associated with acquisition and processing or overall improve the data quality from the early stages of seismic acquisition. But, until then, we need to understand the limitations of various kinds of seismic data and become aware of potential pitfalls in interpretation.

6.3. Selecting seismic attributes and parameters that are ideal for mapping channel architecture

In this study, we evaluated the effect of seismic attributes on the channel system architecture interpretation and their true position in the outcrop model. Our first insight is that there is no one-size-fits-all kind of seismic data.

Therefore, we suggest that the first step when interpreting seismic reflection data should be to 1) define the geological goal, 2) become familiar with the acquisition and processing of information, 3) make an initial inspection of the data, 4) determine if there is some noise or

artifact that the interpreter needs to correct or be aware of. Finally, 5) select seismic attributes based on established purpose/geological goal.

One of the most critical steps is defining a clear geological goal. The interpreter should reflect on what attribute would better suit their purpose. For example, we would use amplitude-derived or instantaneous attributes in studying bright spots. For channel systems, instantaneous or frequency attributes can help highlight the differences between the channel elements (Fedorova, 2016), whereas geometrical attributes (such as Sobel filter coherence) aid in delineating the external shape of the architecture (La Marca, 2020).

Because there are tens to hundreds of attributes, testing many of them can be time-consuming. Imagine that in this study, with just three attributes shown and the case combinations, we had a total of 90 volumes to evaluate. Therefore, it is recommended to work only with attributes whose principles the interpreter understands or, if using multiple attributes, rely on experimental designs like Box Behnken (Ferreira et al., 2007) that will help to synthesize a large amount of data.

Attribute computation varies from software to software, and some attributes are computed trace by trace, whereas others are sample by sample. Understanding this initially would help in setting the correct parameters in each case. For attributes that are window based, like wavelet or average frequency, and the ones studied here, the variation of the analysis window affects the imaging of architectural elements.

We used instantaneous frequency since it can indicate the edges of thin low-impedance thin features. Additionally, it is an excellent bed thickness indicator, where higher frequencies indicate sharp interfaces or thinly bedded strata and lower frequencies indicate thickly bedded sandstone-rich strata (Subrahmanyam and Rao, 2008). Interestingly, in the study, the true

thickness of isolated channel elements (approximately 25m of thick-bedded amalgamated sandstone at the channel element axis) was more interpretable from the dataset of 60 Hz dominant frequency and above and only partially distinguishable in the shallow portion of the 30Hz dataset. This finding was also noted by Nielson (2018) and Langenkamp (2021), although neither study included seismic attribute analysis.

RMS amplitude measures reflectivity within a time window (Meek, 2015). It computes the square root of the sum of squared amplitudes divided by the number of samples within the windows used (Equation 1). Therefore, the number of samples and windows used affects the strength of reflectivity we get with the algorithm used. In contrast, some attributes like Sobel filter coherence do not necessarily improve channel architecture imaging by selecting a minimal analysis window. In this case, since the wavelength increases with increasing velocity, which increases with depth, we agree with Lin et al. (2014) that coherence attributes should use a shorter analysis window in the shallow section and a larger vertical analysis window in the deeper section. Some software has this already integrated as an adaptive window, which we consider to be one route to improve attribute results in the future.

6.4. Impact of noise on the interpretation of channel features

Different types of noise could be found in our seismic data: coherent noise, a series of unwanted signals that appear when the source is applied (Alderton and Elias, 2021), and incoherent noise, which would appear whether we shoot or not. For this sensitivity analysis, we only explored the effect of band-limited random noise, which belongs to the latter category, on imaging the different architecture in a deepwater channel system.

When using instantaneous attributes such as instantaneous frequency, it was noticed that at low noise levels (e.g., 5%), smaller features like MTDs were still visible. However, with the increase in noise, it was extremely difficult to interpret individual geologic features in the seismic volume.

As Herron (2011) stated, the output's quality of interpretability relies on the input's noise content. In fact, a highly noisy dataset will likely contain very little reliable information. Moreover, we noticed that the effect of stratigraphic mixing is emphasized in the coherence attribute, where anomalous low or high values of the attribute were unexpectedly found and that the interpretation of small or large size features in the seismic was very hard to impossible to perform. These findings stress the importance of noise removal using adequate techniques related to the type of noise. This must be done in the early stages of seismic interpretation to avoid misinterpretations (Figure 1.11).

In terms of the analysis window selected in the presence of noise, using a very large data-analysis window (in three dimensions) will include plentiful data and likely produce output with a marked structural overprint. On the contrary, a window that is too small will barely include data and produce an outcome that is more a manifestation of noise in the data rather than geological content (Herron, 2011).

Our final thoughts are that although studies that use synthetic data and perform sensitivity analyses provide a tool to address misinterpretations encountered in the seismic interpretation of potential reservoirs, especially in the exploration stage, it is necessary to take into account that each seismic dataset and geologic setting are unique to the exploration area and that every aspect should be considered carefully before making impactful decisions.

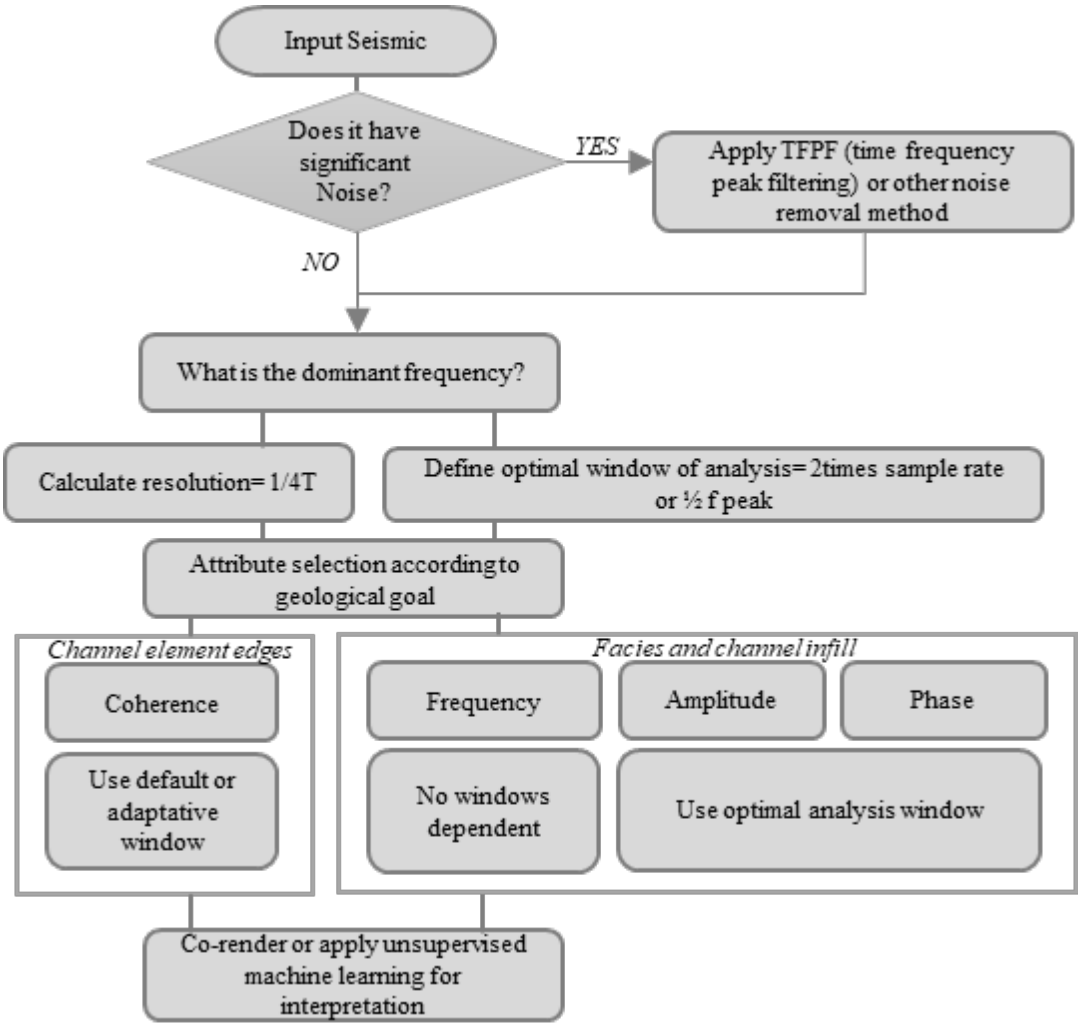


Figure 1.11. Generalized workflow for a geoscientist to avoid pitfalls in interpretation by getting optimized results according to their dataset.

Conclusions

The sensitivity analysis of 3D synthetic seismic volumes derived from a model of an outcropping deepwater slope-system and assigned reservoir properties was performed. The study comprised the combination of four parameters: 1) frequency content, 2) attribute selection, 3) windows of analysis and 4) noise content to identify their effects on the imaging of different scales of deepwater architecture.

The results are summarized in four cases that allow the depiction of common pitfalls in channel interpretation: 1) the effect of the spectral bandwidth on the imaging of the different scales of architectural elements, 2) the seismic attribute sensitivity to the thickness and extent of element and complexes, 3) the effect of the analysis window size on the vertical smearing of channel elements and MTDs, and 4) the effect of noise in the sensitivity analysis.

In this study, we introduced the term "stratigraphic mixing" to define the combined picture resulting from the inability to resolve a single channel element. In this sense, broadband, higher bandwidth (e.g., 90 Hz) data combined with a short analysis window (e.g., 2ms) minimizes stratigraphic mixing. In contrast, lower bandwidth data, in addition to a large analysis window, results in poor imaging of the channel element and channel complexes that exhibit a "composite" appearance, vertically mixing geological features at different stratigraphic levels.

The importance of this analysis resides in that stratigraphic mixing affects the temporal evaluation of features that show an overlap of individual architecture at different scales in the system in consecutive time slices or a vertical offset from the known position of sedimentary units, which may result in important economic losses by misplacing an exploration well (e.g., actual target not in place) or overestimating reserves due to an incorrect NTG estimation and subsequent volumetric calculation (e.g., baffles like MTD are not imaged).

In order to determine the sensitivity of the different architectural elements to seismic attributes, we explored three classes of algorithms: amplitude-based, geometrical, and instantaneous, and showed the results of the most prominent attributes: RMS amplitude, instantaneous frequency, and Sobel filter coherence.

Attributes that are window-dependent, such as RMS amplitude, show an improved imaging of the actual thickness of the channel architecture when calculated using a short analysis window (e.g., 2ms) over a higher frequency dataset (from 30Hz and above). With larger analysis windows and or a small frequency dataset, there is significant vertical stratigraphic mixing. High RMS values and a composite effect were found in stacked channel element configurations indistinctly of the parameters used, which would likely result in an overestimation of NTG.

Conversely, for edge detection attributes like coherence, a small analysis window does not provide a better depiction of channel element/complex/complex set edges. Instead, we observed displacement with respect to the actual position and composite pictures of them, especially in vertical sections, which makes their interpretation cumbersome. Therefore, we suggest using an adaptative window with depth or a default analysis window (half of the peak frequency is a good approximation).

When evaluating the last parameter, which corresponds to the effect of band-limited noise content, it was observed that the mapping of channel elements is hindered by adding noise to the data. When the noise level increases, as expected, the interpretation of features is hindered by the impact of the noise. We suggest applying algorithms that will allow us to eliminate or reduce the noise before calculating any seismic attribute.

Our observations warn of potential interpretation pitfalls in applying default attributes to real seismic data, especially when using geometrical attributes and others that are windows dependent. We offered a simplified workflow for geoscientists to understand and address these concerns depending on their available data.

The importance of using synthetic data to reduce uncertainty is proved. This data allows essaying multiple scenarios to provide tools that serve geoscientists that face different kinds of datasets around the world and help reduce uncertainty by applying best practices in seismic interpretation, especially in channel deposit settings.

Best practices in interpretation should be documented more often to better address uncertainty and optimize reservoir characterization.

Data Availability

Dataset presented at this article can be found at https://data.nzpam.govt.nz/GOLD/system/mainframe.asp, an open-source online data repository hosted at New Zealand and Petroleum Minerals. Synthetic data and model need to be requested to authors.

Acknowledgments

We thank SLB and AASPI (University of Oklahoma) for the software used for attribute and noise computation and visualization; and the Chile Slope Systems Consortium (and sponsors) for providing the synthetic data and 3D model. Thanks to NZPM for providing the Pipeline 3D dataset. Additionally, thanks to the AASPI group at OU and SEG organization for their financial support. Special appreciation to Thang Ha, David Lubo-Robles, Emily Jackson, Teresa Langenkamp, and Diana Salazar for their technical input for this study.

CHAPTER 3: UNCERTAINTY ASSESSMENT IN UNSUPERVISED MACHINE LEARNING METHODS FOR DEEPWATER CHANNEL SEISMIC FACIES USING OUTCROP-DERIVED 3D MODELS AND SYNTHETIC SEISMIC DATA *

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

La Marca, K., H. Bedle., L. Stright., and K.J. Marfurt, 2023, Uncertainty assessment in unsupervised machine learning method for deepwater channel seismic facies using outcrop-derived 3D models and synthetic seismic data. INT2023-0094

*This chapter was presented as a talk at the 2022 IMAGE conference (August 2022):

La Marca, K., H. Bedle., L. Stright., R. Pires de Lima. and KJ. Marfurt, 2022, Quantifying uncertainty in unsupervised machine learning methods for seismic facies using outcrop-derived 3D models and synthetic seismic data. In *SEG International Exposition and Annual Meeting*. SEG. https://doi.org/10.1190/image2022-3746992.1

*Part of this chapter was presented as a talk at the 2023 *Seismic reflection centennial* conference (April 2023):

La Marca, K., H. Bedle., L. Stright., and KJ. Marfurt, 2022, Determining generative topographic map's suitability as an unsupervised machine learning method to interpret channel-related seismic facies: Seismic *reflection centennial* conference. SEG.

*Part of this chapter was presented as a talk at the 2023 *AAPG midcontinent sectional* (October 2023):

La Marca, K., H. Bedle., L. Stright., and KJ. Marfurt, 2022, Uncertainty assessment in unsupervised machine learning methods for deepwater channel seismic facies: *AAPG midcontinent sectional.* AAPG.

Abstract

Unsupervised machine learning (ML) techniques have been widely applied to analyze seismic reflection data, including the identification of seismic facies and structural features. However, interpreting the resulting clusters often relies on geoscientists' expertise, necessitating a robustness assessment of these methods. To evaluate their reliability, synthetic data generated from an actual outcrop model were employed to demonstrate how two unsupervised methods, Self-Organizing Maps (SOM) and Generative Topographic Maps (GTM), cluster deepwater channel-related seismic facies and then measure the associated error. Six seismic attributes,

comprising RMS amplitude, instantaneous envelope, peak magnitude, and spectral decomposition frequencies at 20Hz, 40Hz, and 55Hz, served as input variables. Geobodies were assigned to each cluster formed, and error in facies clustering was quantified by comparing the actual 3D model with the facies grouped by machine learning methods on a voxel-by-voxel basis. This allowed for error quantification and the computation of metrics such as F1 score and accuracy through correlation matrices. Key findings revealed that 1) GTM and SOM exhibited similar performance, with a clustering configuration of 81 for GTM slightly outperforming others. 2)Error rates were approximately 2% for the predominant facies (background shale) but significantly higher for individual channel-related facies, suggesting that channel clusters might represent multiple facies. 3) Resolution and imbalanced data distribution impacted seismic facies predictability, resulting in non-uniqueness in cluster generation. 4) Using synthetic seismic data proved valuable for experimenting with different unsupervised ML, highlighting the need for assessing uncertainty in these methods, given their implications for crucial economic decisions reliant on reservoir interpretation, modeling, and volumetric estimations.

Introduction

Although many machine learning (ML) techniques were created in the 20th Century, only recently have advancements in computer hardware facilitated the widespread adoption of ML among geoscientists. There are broadly two types of ML: supervised and unsupervised. While supervised methods need a label to predict a known output, unsupervised methods have the capacity to find relations between the input data to get an output, usually in the form of clusters or groups.

In seismic interpretation, human interpreters are limited by the number of attributes that can be displayed at once (three or four with co-rendering). Previous studies (e.g., Roy et al., 2013;

Zhao et al., 2015; Roden and Sacrey., 2015) have demonstrated how valuable unsupervised ML methods are for the interpretation of seismic facies, by using multiple seismic attributes as input and finding patterns that were previously overlooked by the interpreter. For example, La Marca (2020) used different seismic attributes as input to Self-Organizing Maps (SOM) and Principal Component Analysis (PCA) to differentiate deepwater seismic facies in the Pipeline 3D seismic dataset in the Taranaki Basin, New Zealand. Infante-Paez and Marfurt (2019) were able to differentiate volcanoes from deepwater deposits in seismic data by using SOMs, and Zhao et al., 2015 present an excellent comparison of unsupervised ML methods which include SOM and GTM to interpret seismic facies in New Zealand channel complexes.

One caveat of unsupervised ML methods is that the interpretation of outputs (clusters) in the absence of well logs relies on the geoscientists' experience and their in-context interpretation. This brings uncertainty when mapping reservoir dimensions, which is critical for the exploration and production of hydrocarbons or geothermal resources and CO_2 storage. Therefore, our motivation is to quantitatively assess unsupervised ML methods, starting with an analysis using synthetic seismic data as a benchmark.

Uncertainty is not knowing or being dubious about a situation, process, or result. In seismic data, from the moment of acquisition to the interpretation of the product of data processing, we have errors, uncertainty, and assumptions being added together (Pampalk, 2001). In this study, we focus on the uncertainty related to the interpretation of clusters derived from two unsupervised ML methods for seismic facies interpretation. Although some authors have addressed uncertainty in supervised ML methods, for example, Ning and Yu (2019) reviewed recent advances in the optimization field to suggest the integration of ML and mathematical programming for decision-making under uncertainty; then, Abdar et al. (2021) updates Ning and Yu's work by providing a

complete and robust review of uncertainty quantification in deep learning, and later Michie et al. (2021) show how performing a sensitivity analysis can aid in assessing uncertainty in fault interpretation using seismic data. As seen, uncertainty in ML methods is a relatively new concept, and based on a search of recent studies, there are none that have focused on assessing uncertainty in the use of unsupervised ML methods applied to seismic interpretation using synthetic data. Skepticism about the reliability of these methods makes this necessary. Therefore, we used synthetic data and a 3D geological model to overcome these concerns.

The geological model and derived synthetic seismic data used in this work were built using outcrop data corresponding to deepwater slope channel systems from the Cretaceous Tres Pasos Formation, Magallanes Basin, Chile, and represent an excellent analog to similar geologic settings (Stright et al, 2014). Five Facies are modeled, which include background shale, channel axis, channel off-axis, channel margin, and Mass Transport Deposits (MTD).

In this manuscript, we begin by presenting the motivation and research objectives. Then, we explain in detail the dataset used and geological setting. Then the selection criteria for the seismic attributes that represent the input data for the methods tested is explained. Subsequently, we show the workflow and details inherent to each ML method. We show how synthetic seismic data allows us to test two unsupervised ML techniques (Self-Organizing Maps and Generative Topographic Maps) with three clustering settings of 36, 81, and 256. We not only qualitatively describe the differences between them, but also quantify the difference between the predicted results and the synthetic data to determine the error in the facies definition and generate confusion matrices with statistical metrics that allow for the understanding of misclassification of facies in each instance. We conclude by providing insights on the use of unsupervised methods and what the interpreter should be aware of when interpreting these results.

Geologic model

The geological model that is the basis of this study is derived from a sandstone-rich deepwater channel system along a progradational slope system (Hubbard et al., 2010). These deepwater slope deposits from the Late Cretaceous (70-80 Mya) Tres Pasos Formation are exposed on approximately 3 km long, 200m thick outcrops near Laguna Figueroa in the Magallanes Basin, (Figure 2.1A) Southern Chile (Macauley and Hubbard, 2013; and Hubbard et al., 2014). The high quality of the stacked channel systems has been used to construct a seismic-scale 3D architectural model of the deepwater channel system (Pemberton et al., 2018; Jackson et al., 2019; Langenkamp et al., 2020; Ruetten, 2021). According to Fildani et al. (2013) the outcrop (Figure 2.1B) is analogous to many slope channel systems globally in stratigraphy and depositional setting, which makes it an excellent benchmark for any study that aims to address problems associated with channel interpretation.

The models are the result of several studies from Macauley and Hubbard (2013), and Southern et al. (2017) combining measured sections, hierarchical stratigraphic interpretations, paleoflow measurements, and thousands of GPS data points that calibrated a drone-derived photomosaic. For these models, the fundamental architectural component are channel elements, defined as distinct, mappable channelized sedimentary bodies (Figure 2.1). Multiple stacked, related channel elements form a channel complex, and two or more complexes form a channel complex set (McHargue et al., 2011; Macauley and Hubbard, 2013, Meirovitz et al., 2020). The outcrops at Laguna Figueroa contain two complex sets, simply referred to as the Upper and Lower Figueroa. The upper complex set consists of eight channel elements and are grouped into four distinct channel complexes. The lower complex set contains twelve channel elements grouped into three channel complexes. Elements are modeled with a standardized width of 400

m and thickness of 25 m. Three additional architectural components are present in the outcrop: mudstone drapes at the base of channel elements, mass transport deposits (MTDs) at the base of channel complexes, and inner-levee thin-bed deposits encasing the channelized elements (Macauley and Hubbard, 2013; Hubbard et al., 2014). The geological models consist of five facies: 1) channel element axis in yellow, 2) channel element off-axis in orange, 3) channel element margin in brown, 4) background shale (inner and outer levee facies) in white, and 5) Mass transport deposits in gray (Figure 2.1C).

Jackson et al. (2019) developed the first fine-scale geocellular model combining channel planforms and vertical stacking for the lower outcrop section (lower channel system) but did not include hierarchical groupings in the architecture. Pemberton et al., (2018) generated forward seismic models using Jackson et al. (2018)'s model and analyzed seismic interpretation of architecture as a function of seismic resolution. Nielson (2018) analyzed the tuning effects of single channel elements. Ruetten (2021) updated Jackson's initial model with new interpretations and added an upper channel system separated from the lower system by a debris flow and studied how stacking patterns impact reservoir connectivity and fluid flow. Finally, Langenkamp (2021) analyzed the influence of stacked channel element architecture on facies classification using Ruetten's model. This work utilizes the geocellular model of Ruetten (2021) and synthetic seismic models from Langenkamp (2021).

The synthetic seismic model used in this study was built using zero-phase Ormsby wavelets of 60 Hz, and 1D convolution (Chile Slope Systems research consortium; Langenkamp et al., 2021) with a reflectivity model. More aspects of each model are found in Langenkamp (2021). Facies-based rock properties, adopted from Stright et al. (2014), show that amplitude

peaks represent an increase in acoustic impedance (Figure 2.1D). In contrast, troughs depict a decrease in acoustic impedance. The synthetic volume has a vertical window of 500 ms. For analysis purposes, we selected the 60Hz volume (Figure 2.1B) as an analog to real datasets (Stright et al., 2014) and cropped the resultant volumes from 120ms to 380ms to avoid dead/blank zones in the reflectivity and focus on the target channel systems.

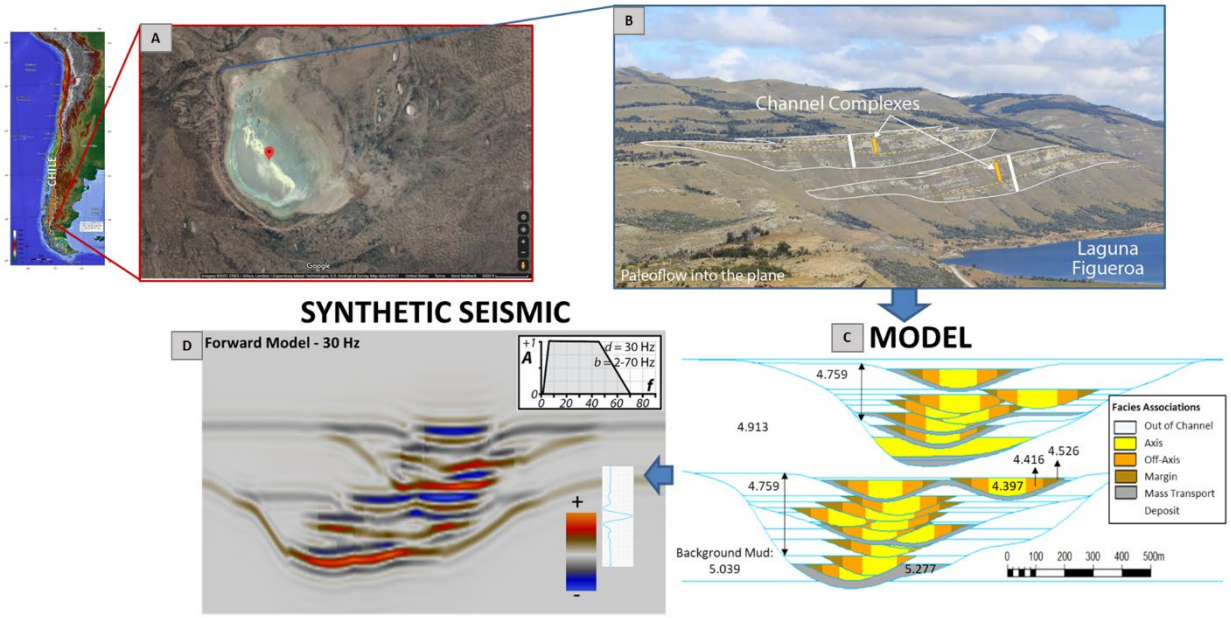


Figure 2.1. Location of outcrop and visual guide showing the steps used to convert the outcrop measurements to a synthetic model. (A) Location of Laguna Figueroa outcrop in the Patagonia-Chile. (B) Upper and Lower Figueroa channel complexes identified in the outcrop, including boundaries and inner channel elements. (C) Creation of a 3D geocellular model using the constraints from (B) augmented by facies associations and corresponding rock properties, including acoustic impedances from GOM (Gulf of Mexico) and Nigerian analogs. (D) The Ormsby wavelet and a representative vertical slice through the 3D synthetic seismic data volume generated from the model are shown in (C). Courtesy of Teresa Langenkamp and Lisa Stright.

Methods

Three major phases were followed to assess the uncertainty in two unsupervised ML methods (Figure 2.2). For Phase 1, we performed an exploratory data analysis to define a vertical window of interest, which would later constrain the results to the objective of the study. We then computed a suite of instantaneous and spectral attributes, which are proven to be effective in defining channel architecture and interpreting deepwater seismic facies (La Marca and Bedle, 2022; Roden and Sacrey., 2015; Zhao et al., 2016). Next, in order to select non-redundant attributes, we calculated the Pearson correlation (Pearson, 1901) to determine the most suitable input seismic attributes.

Phase 2 involved applying the unsupervised ML techniques: Self-Organized Maps (SOM) and Generative Topographic Maps (GTM). We evaluated three clustering cases to perform a sensitivity analysis: 36 clusters, 81 clusters, and a default 256 clusters. Finally, we analyzed the cluster maps and generated geobodies that were representative of each channel facies. Once each geobody was isolated, we resampled it to the model size and quantified the differences between the unsupervised methods and the channel facies (true initial model) on a voxel-by-voxel basis.

Phase 3, the final phase of the study, comprised a quantitative analysis that determined the percentage error per facies and a confusion matrix to understand how accurate the clustering of each facies was. The last phase also involved the calculation of statistical metrics such as precision, recall, F1 score, and accuracy. Each phase's considerations are explained in more detail in the following subsections.

Phase I: Data exploratory analysis, seismic attribute calculation, and input selection

The first step in most seismic interpretation work is the data exploratory analysis, which allows interpreters to identify acquisition footprint, noise, and other features that can influence the interpretation. However, since we are using synthetic models, this step limits to the evaluation of the user-defined windows over which ML algorithm calculation and seismic attribute calculation occur.

Depending on the study focus and data quality it is necessary to constrain or crop the data both horizontally and vertically (below 500 ms for optimal performance) to better focus on the geological objective.

Input seismic attribute selection.

A seismic attribute is a computation made from algorithms applied to seismic data to obtain a more interpretable output. These responses relate to rock physical properties (La Marca, 2020) in rocks and fluids in the subsurface. However, there are tens if not hundreds of seismic attributes (Barnes, 2016), and time constraints do not allow for testing them all.

Although ML algorithms can handle multiple seismic attributes at a time, we should select fewer, more meaningful attributes. According to Chowdhuri and Turin (2020), the principle of parsimony suggests that it is not practical to use a large set of variables in a model, and this would only represent more computation time and complexity. Therefore, as part of our methodology, we reduced the amount of input attributes to use in the ML methods.

There are at least three different strategies that one can use in selecting attributes for use in unsupervised ML methods: 1) Correlating candidate attributes to determine which are statistically independent (Kuhn and Johnson, 2013); 2) using a dimensionality reduction technique such as PCA or Independent Component Analysis (Roden and Sacrey., 2015; Lubo-Robles and Marfurt, 2019); and 3) using the interpreter's geological experience and understanding of the properties measured by different attributes (Infante-Paez and Marfurt, 2019; La Marca and Bedle, 2022).

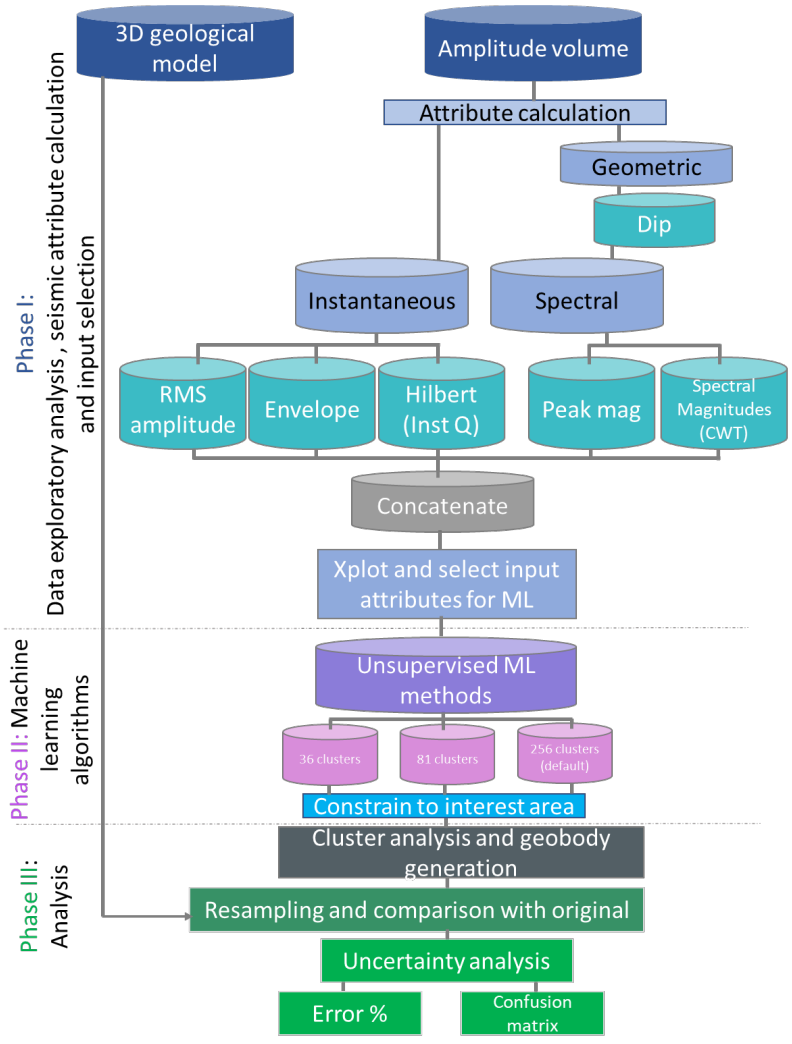


Figure 2.2. Workflow of the study

We used a combination of strategies 1 and 3 in our attribute selection, and selected attributes that are not correlated and that have been proven to be suitable for stratigraphy (Table 2.1), such as instantaneous and dip-corrected spectral attributes. Then, followed the methods in La Marca et al. (2021) and calculated the attributes (Figure 2.3b, 2.3c, 2.3d, and 2.3e) using a 3x3-trace analysis window with a height of 2 ms. After that, we refined our seismic attribute selection by using scatterplots. Careful seismic attribute selection is paramount to generate optimal results when applying ML techniques.

Seismic attribute	Category	Principle	References	Use in channel interpretation
Root Mean Square amplitude	Instantaneous	Is a measure of reflectivity/ Computes the square root of the sum of squared amplitudes divided by the number of samples within the window used	Meek (2015)	To detect sand bodies and mud-filled channels associated with channel belts
Envelope (reflection strenght)	Instantaneous	Represents the envelope (and energy) of the signal	Taner et al (1979)	Highlights changes in lithology, deposition, tuning effect and sequence boundaries
Hilbert (Instantaneous Q)	Instantaneous	Is a linear operator that produces a 90 deg phase shift in a signal, which allows to compute the complex trace attributes	Luo et al (2003)	Detailed visualization of bedding configurations
Peak Frequency	Spectral	Computes the max. value of the absolute value of the amplitudes within a window	Liu (2007)	Thickness variations
Spectral magnitudes (decomposi	Spectral	Applies a suite of constant- bandwidth filters to the seismic data	Partyka et al (1999)	Highlights channels and minor architectural elements (good for thickness changes). Low frequencies often depict thicker architectures and high frequencies thin elements.

Table 2.1. Summary of seismic attributes explored to use as input. Their category, principle, references, and common use in channel architecture interpretation are explained.

Once we followed the geoscientist experience attribute selection criteria, we normalized the attributes using the MinMax scaler (which scales the data between 0 and 1 maintaining skewness) in order to evaluate the correlation between attributes.

Phase II: Unsupervised ML algorithms: Self-Organizing Maps (SOM) and Generative Topographic Maps (GTM)

Machine Learning (ML) is a discipline of artificial intelligence (AI) that aims to find patterns to make predictions with minimal human intervention. ML is composed of broadly two types of methods: supervised (where labels are provided, and the desired output is known) and unsupervised (where labels are not given, and output is unknown). We focused on the latter.

First known as Hebbian Learning, Unsupervised ML models are a category of algorithms that unveil underlying patterns and relationships in the data without any reference, supervision, or labels. Unsupervised methods were first used nearly six decades ago (McCulloch and Pitts, 1943). However, their popularity increased in the geosciences field only in the last decade.

Unsupervised ML can be further divided into two categories: clustering and association. We tested methods that belong to the clustering category. Clustering (Hartigan, 1975) is the grouping of similar objects, where all the objects within the same cluster share a particular property in common. When machine learning uses seismic attributes, clusters can represent geologic information embedded in the data. These clusters can help identify geologic features and geobodies, which sometimes are difficult to interpret by other means (Roden and Sacrey., 2015). In this category, we find methods such as K-means, PCA, and the two we selected for this study: Self Organizing Maps (SOM) and Generative Topographic Maps (GTM).

These two methods were selected over similar techniques because 1) both methods are widely available in many software and have started to be available in code/programming languages, 2) they are relatively easy to understand and define, 3) they are proven to work well

for seismic facies interpretation 4) analyzing data in a particular latent space may reveal data properties that are overlooked in the original/ initial space, and 5) curiosity to contrast methods that are somewhat similar to evaluate their performance against the same input data and dataset evaluated.

There are many case studies that prove that SOMs are robust methods for the technology sector to evaluate competitive technical intelligence (e.g., An and Yu, 2010), in marketing (e.g., Hanafizadeh and Mirzazadeh, 2010), in medicine (e.g., Tuckova et al., 2011) and in environmental applications (e.g., Gibson et al., 2017) to find patterns with similar characteristics within their respective datasets. Roy et al., (2013) describe how SOM has been used since the late 1990s in the oil industry to resolve diverse geoscience interpretation problems, and Chen et al. (2009) present applications to improve signal-to-noise ratio. There are several examples of studies that have used SOMs for facies classification.

For deepwater settings, many studies (e.g., Coleou, 2003; Wallet et al., 2009; Zhao et al., 2015; La Marca and Bedle., 2022) demonstrated that SOMs are useful for defining architectural elements associated with channel complexes. Infante-Paez and Marfurt., (2019) differentiated deepwater deposits from volcanic deposits using SOMs. This method has also been used in unconventional resource plays. Verma et al., (2012), mapped high frackability and high TOC zones in the Barnett Shale, while Sacrey and Roden (2014) located sweet spots in the Eagle Ford shale.

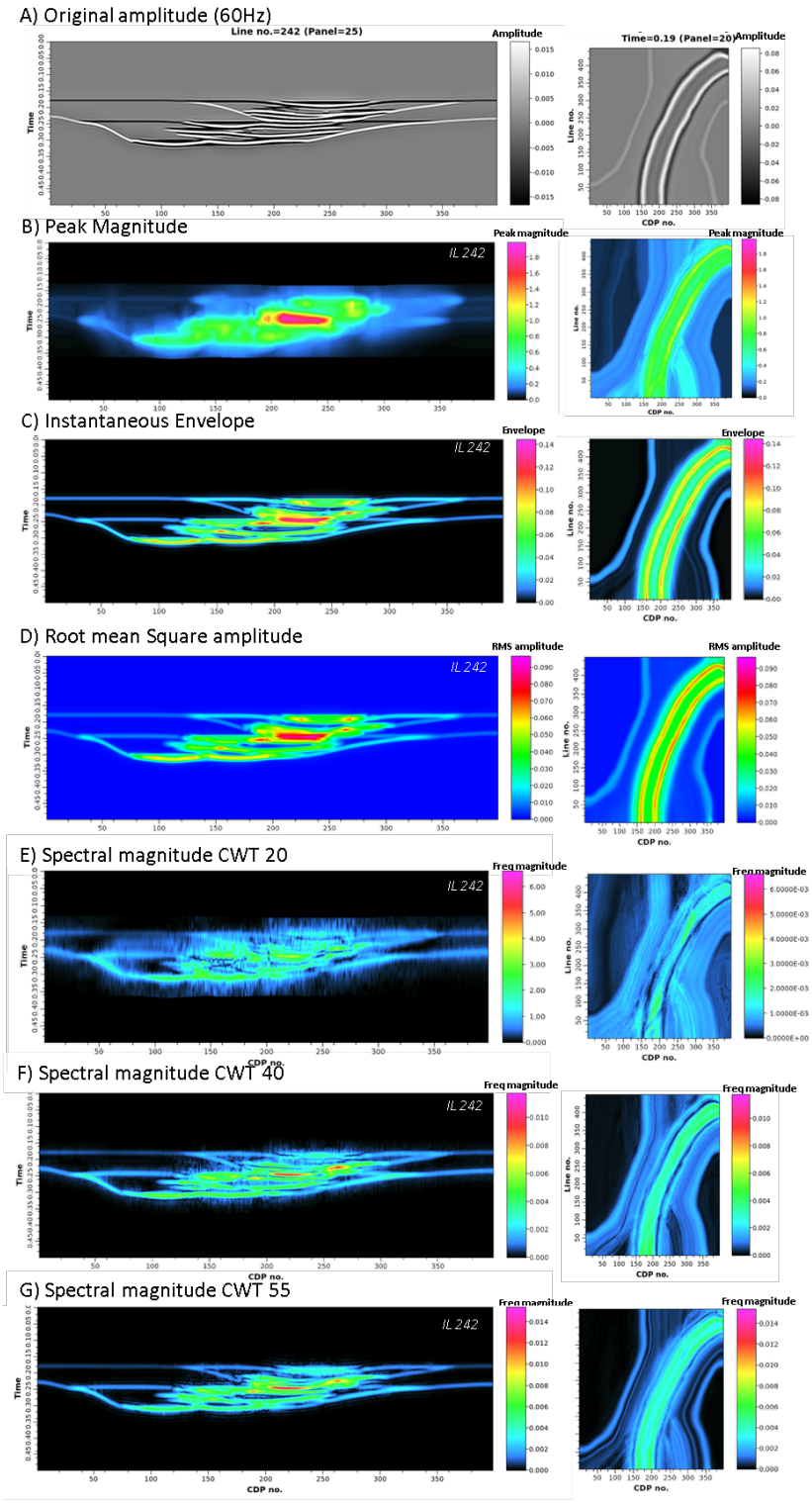


Figure 2.3. (A) A representative vertical slice of the original amplitude volume and seismic attributes used as input: (B) Peak Magnitude, (C) Envelope, (D) RMS, (E) Spectral component 20 Hz, (F) Spectral component 40 Hz, (G) Spectral component 55Hz.

GTM has been used successfully in various fields to model complex high-dimensional data. In the chemical field, Horvath et al. (2019) used it for drug design (chemical space cartography) and Lin (2019) used it for drug design and library visualization. GTM has also been employed in engineering (e.g., aerodynamic configuration presented by Chao et al., 2022), and more commonly in the geosciences field, where we have seen successful studies on channel-related seismic facies (e.g., Wallet et al., 2009, Chopra and Marfurt, 2014) and carbonate seismic facies (Roy et al., 2014).

Assumptions considered for the models selected are: 1) The 3D synthetic seismic volume of 60 Hz dominant frequency was selected to perform the tests due to its similarity to the frequency content of actual seismic data. If we would like to evaluate the response of high-resolution datasets (OBN data or Chirp) we would need to select a synthetic model with higher frequency which would likely offer better resolution. 2) We are using a 60 Hz dominant frequency volume, which is not a high-resolution volume, but it is closer to the data we often use to make seismic interpretation. 3) Our dataset does not include significative noise. We would need to perform a sensitivity analysis to test the ML methods' response in front of different data quality. So, we are presenting one of the best-case scenarios possible. 4) Only five facies were considered and one of them (background shale) represents most of the data in proportion. However, we wanted to test this scenario, since it represents the geological setting that is analogous to other deepwater systems deposited worldwide. 5) Although we acknowledge the use of an elbow plot to determine the optimal number of clusters as shown in La Marca et al. (2021), we decided to test different cluster scenarios to perform a sensitivity analysis, which also falls within uncertainty assessment procedures. 6) Seismic attributes were calculated using an analysis window of 2ms to optimize results as suggested by La Marca et al. (2021)

Before explaining the details for both methods, let us explain the "latent space" concept since we will mention the term in the algorithms used. A latent space is simply a representation of compressed data (in the form of an abstract multi-dimensional space) that contains all the possible outputs and in which similar data points are closer together in space (Roy et al., 2013 and Wallet et al., 2009). Latent space is useful for learning data features and for finding simpler representations of data for analysis. A latent space can be high-dimensional or low-dimensional. A high-dimensional latent space is sensitive to features in the input data and can lead to overfitting if training data is insufficient. On the other hand, a low-dimensional latent space attempts to capture the primary features required to learn and represent the input data.

SOM definition

First introduced by Kohonen, the "Kohonen Self-Organized Map" in 1982 (Kohonen 1982, 1995) is an unsupervised ML technique that transforms a complex high-dimensional input space into a simpler low-dimensional (typically 2D) discrete output space by preserving the topology (relationships) in the data (Bartkowiak, 2004; Asan & Ercan, 2012; Barnes et al., 2016).

SOMs are popular seismic facies clustering technique that extract similar patterns embedded with multiple seismic attribute volumes (Zhao et al, 2016). Since it is an unsupervised technique, the selection of attributes that are thought to contribute the most in the classification is primordial. The neurons involved in the SOM classification seek to identify relationships between the various input datasets, and then grouped similar subsets of the data (Sanger, 1989) into several colors to aid the interpretation. (Figure 2.4A)

SOM process can be summarized in four steps: 1) Input data and initialization: our different seismic attributes are used as input. 2) Weighting and best matching unit (BMU) definition (sometimes referenced as the winning neuron or prototype vector), where a neuron learns by adjusting its position within the attribute space as it is drawn toward nearby data samples in what is known as competitive learning (Asan and Ercan, 2012). Once the learning process has completed, the winning neuron set is used to classify each selected multiattribute sample in the survey (Roden and Sacrey., 2015). 3) Clustering: Given the BMU, each attribute sample in the dataset incrementally approaches towards a similar BMU in each case by the Euclidean distance. A SOM manifold that contains all the possible combinations is then formed, and the algorithm deforms this manifold to better fit the data in each iteration (Zhao et al., 2015). 4) Projection of clusters in a lower 2D dimensional space where colors are assigned. Gao (2007) mentions that if the number of prototype vectors is 256, we will have 256 colors. These are the potential clusters. Our results can form either 256 or considerably a smaller number of clusters (e.g., three or four). Finally, after clusters or classes are obtained, the interpreter uses their knowledge to make geological interpretations of these clusters.

Table 2.2 shows the parameters selected in each case. Notice that the number by the SOM case refers to the number of clusters originally assigned.

Table 2.2. Self-Organizing Maps cases' parameterization specifications

Case	# Attributes used as input	# Clusters	Topology	Std in latent space	Initial neighbourhood radius	Distance	Grid spacing for the cluster projection space	Window size
SOM526	6	256 (36x36)	squared	4	1.2	Euclidean	150	499ms
SOM81	6	81 (9x9)	squared	4	1.2	Euclidean	150	499ms
SOM36	6	36 (6x6)	squared	4	1.2	Euclidean	150	499ms

GTM definition

Generative Topographic Maps (GTM) first introduced by Christopher Bishop in 1997 (Bishop et al., 1998) is a nonlinear dimension reduction, and clustering technique that provides a probabilistic representation of data vectors in latent space (Chopra and Marfurt, 2014) (Figure 2.4B). It is generative because data is assumed to arise by first probabilistically picking a point in a low-dimensional space. Then the mean projection of the posterior probability of the data vector is used in generating the clusters in the high dimensional latent space (Roy et al, 201).

GTM operates following the next steps (Roy et al, 2014): 1) Initial array of grid points (*K*) are on a lower dimensional space. 2) Each of the grid points are nonlinearly mapped onto a similar dimensional non-Euclidean curved surface as a corresponding vector (*mk*). 3) Each data vector (*xk*) mapped into this space is modeled as a suite of Gaussian probability functions centered on these reference vectors (*mk*).4) Components of the Gaussian model are iteratively moved toward data vector that it best represents, hence maximizing likelihood (via the expectation maximization algorithm). 5) Configuration is projected on the latent space.

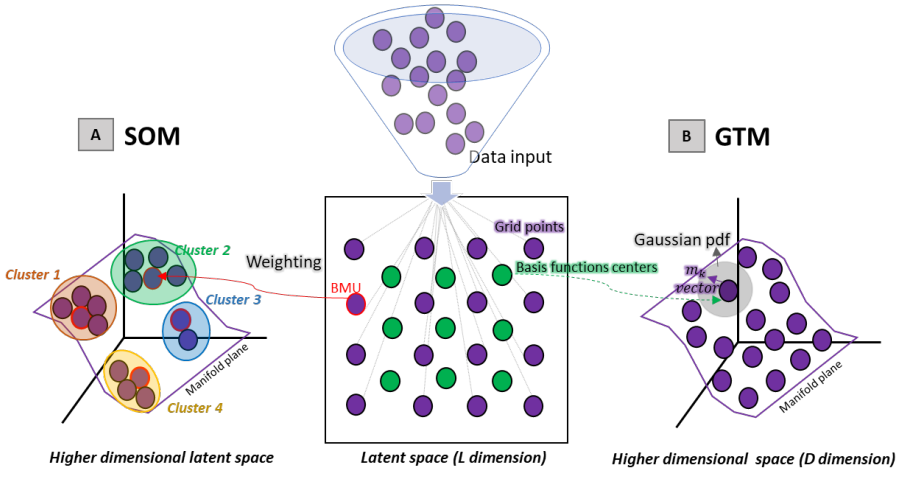


Figure 2.4. Representation of (A) Self-Organizing Maps and (B) Generative Topographic Maps (adapted from Roy et al., 2013).

Given an initial set of datapoints, they will be inputted in a lower dimensional latent space. In both cases the data is later projected in a high dimensional latent space where a manifold space best will fit the arranged data. For (A) the BMU or winning neuron is selected via weighting and each BMU will be representative of the group (cluster) the datapoints that get closer to them will form. For (B) there are a series of basis function centers whose linear combination will allow the projection of the grid points in the non-Euclidean manifold plane as *m_k* vectors. Adaptation to the manifold plane occurs as a product of Gaussian Pdf that occur around each vector. The main difference between the SOM and the GTM regarding their result presentation is that while SOM assigns each data point to exactly one place on the map GTM calculates a distribution for each data point on the map.

Table 2.3 shows the parameters selected in each case. Notice that the number by the GTM case refers to the number of clusters originally assigned.

Table 2.3. Generative Topographic Map's cases' parameterization specifications

Case	# Attributes used as input	# Clusters (samples in latent space)	Topology	# Basis functions (< cluster #)	Weight regularization factor	# Data training iterations	Window size
GTM256	6	256 (36x36)	squared	0.5	0.05	50	499 ms
GTM81	6	81 (9x9)	squared	0.5	0.05	50	499 ms
GTM36	6	36 (6x6)	squared	0.5	0.05	50	499 ms

Advantages and drawbacks of SOM and GTM methods

Although SOM and GTM may result in similar outcomes and both are included within unsupervised ML clustering methods, there are some differences that need to be understood before applying them to our data.

SOM is likely the more popular of the two methods because it is easy to understand, simple to use, produces good intuitive results (Pampalk, 2001), and is available in several computational packages. Moreover, being a neural network that attempts to replicate human brain behavior makes the method attractive to many users. Furthermore, Bishop et al. (1998) explained that the major advantage of the SOM over the GTM algorithm is the processing speed since GTM is a third slower than the SOM algorithm.

However, the SOM method has some drawbacks, including the absence of a quantitative error measure for the convergence acceptance criteria (which could be solved by a cost function) since there is no indication about the vector's probability being well represented by other regions on the manifold (facies) (Kohonen, 1995; Bishop et al., 1998). In other words, the selection of the neighborhood function in each iteration is subjective so that different solutions derive after each iteration. Finally, no probability density is defined that could yield a confidence measure in the final clustering results (Bishop et al., 1998).

GTM comes to solve the SOM drawback of convergence since it allows the selection of a learning rate (Marfurt and Chopra, 2014) and has some other virtues. For example, the method explicitly formulates a density model over the data, uses a cost function that quantifies how well the map is trained, and uses a sound optimization procedure (EM algorithm). In addition to this, GTM presents more information about the mapping than the SOM since it calculates the distribution of single data points, making it easier to visualize. Also, although both methods preserve the topology (relationships) in the data, SOM does not preserve the actual distances (Asan and Ercan, 2012). Therefore, GTM produces a more separable and interpretable map than other methods.

GTM also has some disadvantages to consider, namely: GTM has not been developed in the context of neural networks. Instead, it is embedded in a statistical framework (more insights on this are found in Pampalk, 2001), and the expectation-maximization (EM) method finds only local optimal solutions.

As seen, some of the limitations of the SOM algorithm have been addressed by the GTM algorithm. However, some others cannot be solved since they are inherent to the model of mapping data points from a high-dimensional data-space onto a 2-dimensional map while preserving local distances. Pampalk (2001) recommends using GTM or SOM combined with other data-mining methods, such as pure clustering (K-means) or multi-dimensional scaling techniques.

Phase III: Uncertainty quantification (UQ) and analysis

Geobody definition

A geobody is a 3D object that consists of a series of geoblobs (groups of seeds clustered together) that have been extracted from a seismic volume. The seismic volume can be the original amplitude, a seismic attribute, or as is this case, the result of an ML algorithm applied to seismic data.

Geobody definition depends on interpretation. We isolated each facies by analyzing the clusters map, grouping the cluster numbers that were most representative (Table 2.5), and creating a rule in the calculator of the software assigning a particular number per facies. An example of geobody generation for facies 5 in the GTM case of 36 clusters is presented in equation 1.

We assigned 1 to channel axis facies, 2 to channel off-axis, 4 to channel margin, and 5 to MTDs. Geobodies can also be picked or refined by using opacity thresholds in the interpretation package. If clusters did not fall in the previous facies category, we assigned 0, which means it corresponds to the background shale, which in this geological setting is expected to represent the majority of the facies.

GTM36_Facies5=if((Xplot_GTM36_60HZ = 3) Or (Xplot_GTM36_60HZ = 17) ,5,0) (1)

Once defined our geobodies per facies and each model tested, we performed a resampling of each geobody by using a popular interpretation software. Resampling means sampling a seismic volume (or geobodies interpreted from seismic data) inside the cells of a 3D model. This creates a property that allows to make comparisons and calculations in the same domain as the original 3D model (voxel by voxel). Figure 2.5 shows the original model geobodies that we isolated out of the 3D model. These geobodies are the benchmark for the uncertainty quantification (UQ) analysis.

Uncertainty Quantification (UQ) for unsupervised ML methods

We calculated the confusion matrix to understand which facies were confused (or represented by the same cluster) in each case.

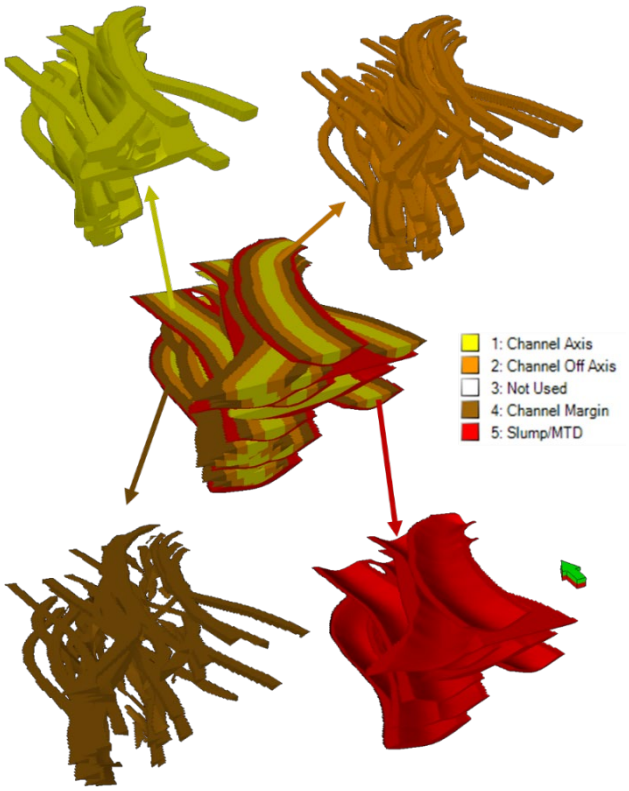


Figure 2.5. Geobodies per facies in the 3D model that serves as basis for the uncertainty quantification analysis.

For this, we used the calculator option in the interpretation software. An example for the case of GTM36 is presented below:

$$DF_GTM36=0$$

$$DF1_GTM36= if\ (GTM36_Facies1=1,1,DF_GTM36)$$

$$DF2_GTM36= if\ (GTM36_Facies2=2,2,DF1_GTM36)$$

$$DF4_GTM36= if\ (GTM36_Facies4=4,4,DF2_GTM36)$$

$$DF5_GTM36= if\ (GTM36_Facies5=5,5,DF4_GTM36) \qquad (2)$$

These operations allowed us to obtain a single file with concatenated facies that was easier to handle for the next step, the uncertainty evaluation via error estimation and confusion matrix generation.

Percent error

Calculation done in a voxel-by-voxel way using equation 3, once we have built the geobodies

*Percent error =((actual facies voxels –ML result voxels)/actual facies voxels) *100* (3)

Confusion matrix and statistical metrics

Since we are testing clustering algorithms (unsupervised ML), and therefore the output is clusters, and knowing the response in the synthetic model, we will use the terms class and clusters interchangeably. Treating these clusters as classes allows us to employ error estimation techniques such as the confusion matrix.

According to Kuhn and Johnson (2013), a common method for describing the performance of a classification model is the confusion matrix. A confusion matrix is a cross tabulation of the observed and predicted classes. It is a NxN matrix employed to measure performance of classification models. Although the methods analyzed in this study are unsupervised and generate clusters, knowing the original response allows for calculation of statistical metrics that help understand misclassification of facies represented by the clusters formed.

Confusion matrices are represented by two axes: the horizontal or X axis represents the predicted facies (clustered), and the vertical Y axis represents the true facies from the original 3D model. The main diagonal in the matrix represents the cases where the classes (In our case facies) were correctly predicted, while the other values off the diagonal represent the number of errors of each possible case.

Confusion matrices allow to determine true positives (TP) when the actual and predicted values are positive, True Negatives (TN) when the actual and predicted values are negative, false positives (FP) when the predicted value is positive while the true value is negative, and false negatives (FN) when the predicted value is negative, but the true value is positive. Although these values are not interpreted individually, they are used to calculate the metrics precision, recall and F-1 score. We can further determine accuracy, macro average, and weighted average to evaluate the performance of the ML methods in our multi-class classification.

Table 2.4 summarizes the different statistical metrics calculated to assess the uncertainty of the ML models here evaluated.

Table 2.4. Metrics to evaluate performance of the ML evaluated. Modified after Kuhn and Johnson (2013).

Metric	Meaning	Equation
Precision	Measures out of all predicted positives how many are actually positive/ Or how good the model is at predicting a category	$Precision = \frac{TP}{TP+FP}$
Recall	Measures how many positive records are predicted appropriately (positive)/ How many times the model was able to detect a specific category	$Recall = \frac{TP}{TP+FN}$
F1-score (or F measure)	This metric allows to evaluate the model performance. It is the harmonic mean of precision and recall	$F1score = 2 * \frac{Precision * Recall}{Precision + Recall}$
Accuracy	Refers to the proportion of correct matches/ How many times the model evaluated was correct overall	$Accuracy = \frac{Correct\ predictions}{Total\ predictions}$
Macro average score	Arithmetic mean of all the pre-class recall or f1 scores. Treats all class equally without considering their support values	$Macro_avg = \frac{\Sigma\ all\ per\ class\ (recall\ or\ F1) scores}{Number\ of\ classes}$
Weighted Average	Takes the mean of per class scores while considering the proportion of each class support	$W_avg = \Sigma\ (perclass\ F1\ score * (support\ of\ class\ /total\ support)$

Results

In the next sub-sections qualitative and quantitative analysis results are presented.

Unsupervised ML results: qualitative analysis

Normally, a seismic interpreter tends to use default parameters when there is a lack of understanding of how to improve the performance of algorithms, whether they lead to seismic attributes or different machine learning techniques. That is the reason why we first show in Figure 2.8 a comparison between the most general (default) case of 256 clusters for both SOM (Figure 2.8A) and GTM (Figure 2.8B). A seismic interpreter would not have any a priori

information that helps them understand the meaning of the clusters formed, and usually, their interpretation relies on their expertise and geological knowledge (La Marca and Bedle, 2022).

Unsupervised results derive four panels as seen in Figure 2.8, which show the cross plotting of main output axes axis 1 and axis 2 and which offers the visual picture of the clusters in our dataset. We also obtain a 2D color legend with as many colors as clusters possible, a 2D histogram that would show the clusters formed and their occurrence or frequency such as the more clusters we have with one color, the more presence this cluster has in the data studied or indicate that more datapoints belong to each cluster formed. Finally, by overlying the histogram and colormap one can correlate which colors in the SOM correspond to each cluster.

There are fewer clusters formed than the numbers we initially set in the input parameters. For example, SOM256 generates 37 clusters, and GTM256 123 clusters. We overall noticed that in all the cases, fewer representative clusters are formed in the SOM as compared to GTM. Also, we observed better details with fewer clusters than with 256, therefore the rest of the analysis are focused on the cases with 81 (high-level clusters) and 36 (low-level clusters).

Because the expected facies and the synthetic model is based on the 3D model that contains them, we can qualitatively evaluate the performance of both methods by contrasting expected (Figure 2.5) vs. obtained facies/ clusters.

It is noticeable that there are more clusters than facies expected (5). This means that either 1) each facies is defined by an aggregate of clusters, 2) some clusters do not have geological meaning or would represent noise, or 3) each cluster can represent more than one facies at a time depending on what features/ patterns were considered in the internal weighting

process that each algorithm used to define each cluster, this may be similar to the overfitting we know in supervised methods.

When visually analyzing Figure 2.6 it is noticed that GTM256 (Figure 2.6B) seems to offer a more robust representation of the expected facies, having Facies 1 (channel axis) shown in color fuchsia, Facies 2 (channel off-axis) by mostly orange clusters, Facies 4 (channel margin) in brownish/ pink, Facies 5 (MTD) in blue/ purple and Facies 0 (Background shale) in green. On the other hand, SOM256 (Figure 2.6A) does not show such discrimination of facies, and clusters' colors are very similar, which can turn the interpretation process into a cumbersome task.

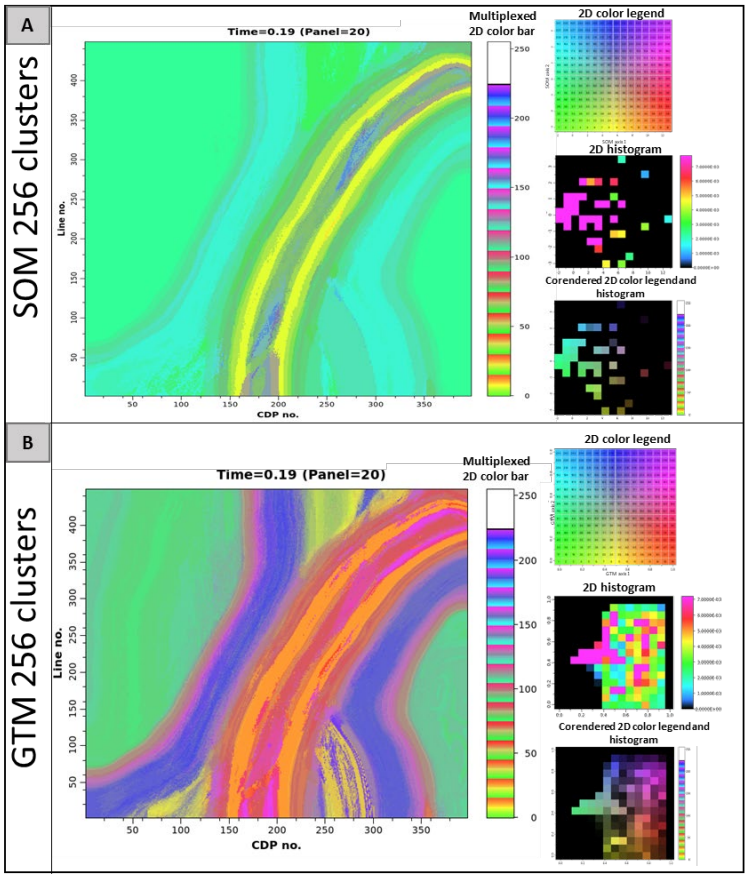


Figure 2.6. SOM and GTM results per cluster arrangement and heat map per case. (A) SOM of 256 clusters, (B) GTM of 256 clusters.

Unsupervised ML results: quantitative analysis

A neuron grid along with the corendered 2D color legend allowed to group clusters interpreted to represent each facies. Table 2.5 presents the clusters we considered per case study.

Table 2.5. Clusters interpreted per case study and facies.

Facies	SOM36	GTM36	SOM81	GTM81
Facies 1	15,20	10,4,3,9	67,78	5,7,14,44,61
Facies 2	22	23,17,28,29	6,13,23	52,53,61,62
Facies 4	4,24	20,16,21,22,27,11,5	30,31,39,40	24,34, 43
Facies 5	33,34	3,17	34, 41,42	33, 41,42,50

Figure 2.7 shows a panel with the SOM and GTM results per case studied. By visual interpretation we observed that GTM (Figures 2.7B and 2.7D) apparently better represents the actual facies than SOM cases (Figures 2.7A and 2.7C) and that facies 0 (background shale) has more data counts and clusters when looking at the histogram. Also, by our interpretation we would slightly prefer the 81 clusters cases since it shows smoother “facies”.

Using the values in Table 2.5 and Equation 2, we generated geobodies per facies and cases. Figure 2.8 shows a graph that allows to compare the original expected geobody vs the obtained in each case. Overall, we notice that the generated geobodies do not mimic exactly the expected ones which may be due to misinterpretation of clusters or errors in the geobody generation method in the software used.

To quantitatively understand and address uncertainty in SOM and GTM methods we estimated a percentage error and built a confusion matrix which we explain in the next subsection.

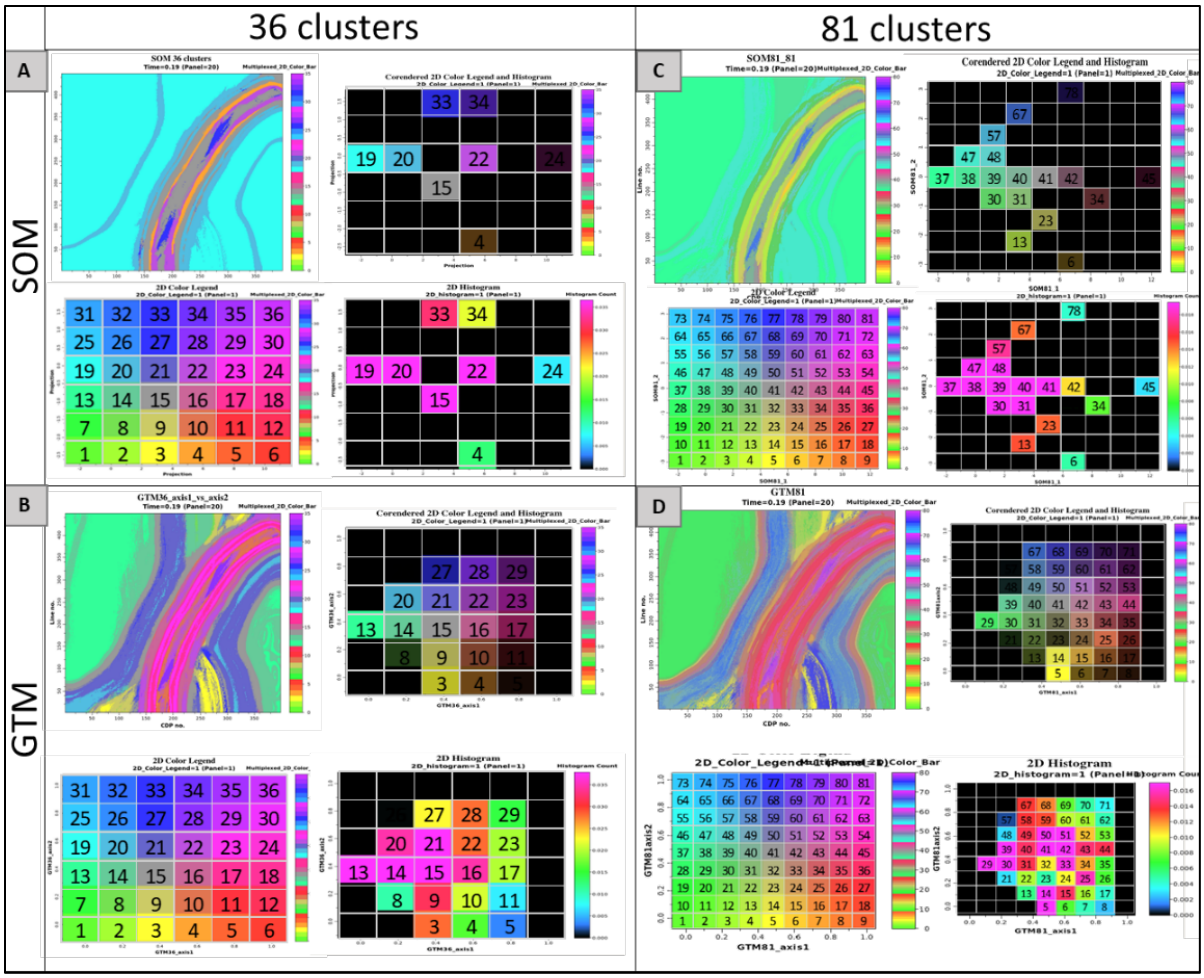


Figure 2.7. SOM and GTM results per cluster arrangement and heat map per case. (A) SOM of 36 clusters, (B) GTM of 36 clusters, (C) SOM of 81clusters, (D) GTM of 81 clusters, (E) SOM of 256 clusters, (F) GTM of 256 clusters

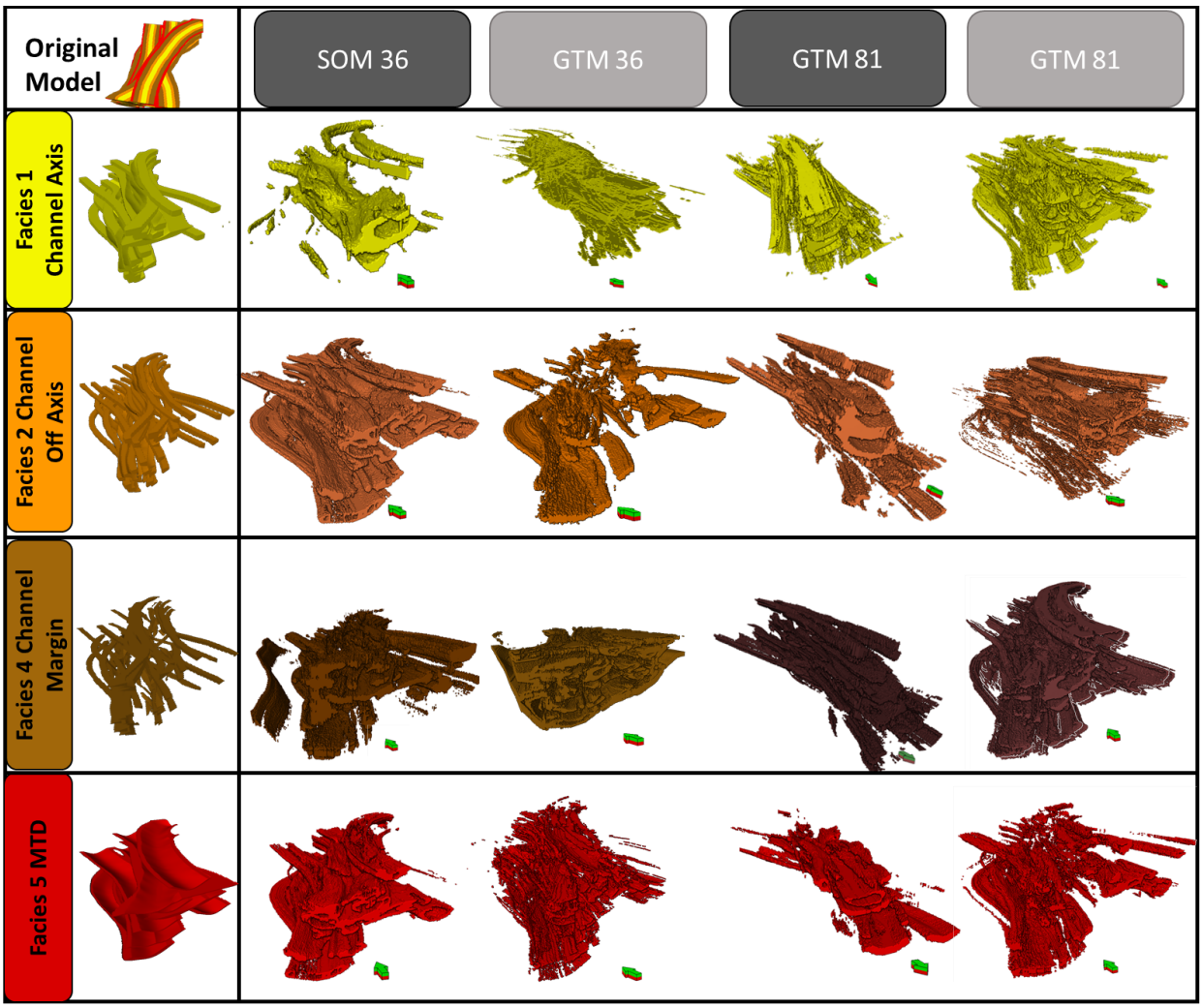


Figure 2.8. Geobody comparison per case and facies compared to the original/ expected geobodies.

Percentage error estimates for UQ

Using equation 4 which considers original voxels vs predicted, error per facies was calculated and Table 2.6 presents these results. Overall, geobodies formed per facies have a high error in predicting individual facies, as opposed to background shale that presented the lowest error in all cases. This implies that we should likely merge two or more facies in a single class (cluster for the unsupervised ML purpose) to correctly predict it.

Table 2.6. Results of percentage error per case and facies

Facies	SOM36 error %	GTM36 error %	SOM81 error %	GTM81 error %
0	1.126	5.658	0.271	2.249
1	91.857	90.108	93.509	82.263
2	99.995	92.514	98.494	95.185
4	86.494	59.589	95.164	82.842
5	58.179	65.907	26.739	44.106

Confusion matrix results for UQ

To address uncertainty in the ML methods studied we developed a confusion matrix per facies and each case tested. Table 2.7 shows the results of the matrices, where the main diagonal represents the correctly predicted data points per facies. We explain one example of how we interpret the results of the matrix. Focusing on the GTM 81, the rightmost column (support) shows the expected number of voxels expected to be predicted per facies.

If we analyze Facies 1, which corresponds to the channel axis, we expected to predict 461,204 voxels. However, when we calculated the confusion matrix (see the main diagonal on the leftmost table) only 81,804 voxels correctly fell in this facies category. Therefore, the remaining 379,400 voxels were misinterpreted by the other facies. If we follow that row where the Facies 2 is (yellow) noticed that 165,356 voxels were misclassified as shale when they belonged to Facies 1. Similarly, 30,309 were misclassified as Facies 2, 92,905 voxels fell in the Facies 4 class, and 90,230 were confused with Facies 5.

We derived some statistical measures that consider the TP, TN, FN, and FP to estimate the error in predicting facies per case, which holistically allows us to address uncertainty in the ML methods used.

Table 2.7. Results of confusion matrix per facies for the SOM36, GTM36, SOM81, GTM81. Precision, recall, F1, and accuracy metrics are shown to the right of each case studied.

A)

							Facies	Precision	Recall	F1-score	support
SOM36	24858571	4462	0	198169	77350		0	0.97	0.99	0.98	25138552
	168730	37554	8	123337	131575		1	0.37	0.08	0.13	461204
	137323	23155	17	81221	99338		2	0.38	0.00	0.00	341054
	167089	8115	18	32813	34912		4	0.07	0.14	0.09	242947
	171985	28175	2	61464	149617		5	0.30	0.36	0.33	411243
							Accuracy			0.94	26595000
							Macro avg	0.42	0.31	0.31	26595000
							Weighted avg	0.94	0.94	0.94	26595000

B)

GTM36	23792424	269643	11340	989036	76109		0	0.98	0.95	0.96	25138552
	126093	45620	32709	186883	69899		1	0.12	0.10	0.11	461204
	104071	20471	25531	146237	44744		2	0.22	0.07	0.11	341054
	104296	14908	4900	98178	20665		4	0.06	0.40	0.11	242947
	140204	23659	39705	191964	15711		5	0.07	0.04	0.05	411243
							Accuracy			0.90	26595000
							Macro avg	0.29	0.31	0.27	26595000
							Weighted avg	0.93	0.90	0.92	26595000

C)

SOM81	25070544	38966	182	17482	11378		0	0.96	1.00	0.98	25138552
	348827	29939	14713	26379	41346		1	0.26	0.06	0.10	461204
	268375	16572	5137	25916	25054		2	0.18	0.02	0.03	341054
	206563	14491	2167	11749	7977		4	0.10	0.05	0.07	242947
	301281	17273	6479	30386	55824		5	0.39	0.14	0.20	411243
							Accuracy			0.95	26595000
							Macro avg	0.38	0.25	0.28	26595000
							Weighted avg	0.92	0.95	0.92	26595000

D)

GTM81	24585623	211703	20889	75880	244457		0	0.97	0.98	0.98	25138552
	165956	81804	30309	92905	90230		1	0.21	0.18	0.19	461204
	133785	50611	16423	63987	76248		2	0.2	0.05	0.08	341054
	137799	24728	5426	41684	33310		4	0.12	0.17	0.14	242947
	229861	27032	8530	66886	78934		5	0.15	0.19	0.17	411243
							Accuracy			0.93	26595000
							Macro avg	0.33	0.31	0.31	26595000
	Facies 0	Facies 1	Facies 2	Facies 4	Facies 5		Weighted avg	0.93	0.93	0.93	26595000

Like the error presented in 5.2.1, when looking at the metrics, we see how just one of the classes (Facies 0 or shale) in each model has the highest precision and recall, and that the models do not show good performance for the other evaluated facies, being Facies 2 and 1 the next best in performance, followed by Facies 5 and 4.

When comparing the tested methods, we see that the models have high accuracy, with SOM36 the one with the highest weighted average (0.94), followed by GTM81 with 0.93.

Overall, all methods present similar performance, but in-depth analysis is provided in the upcoming discussion section.

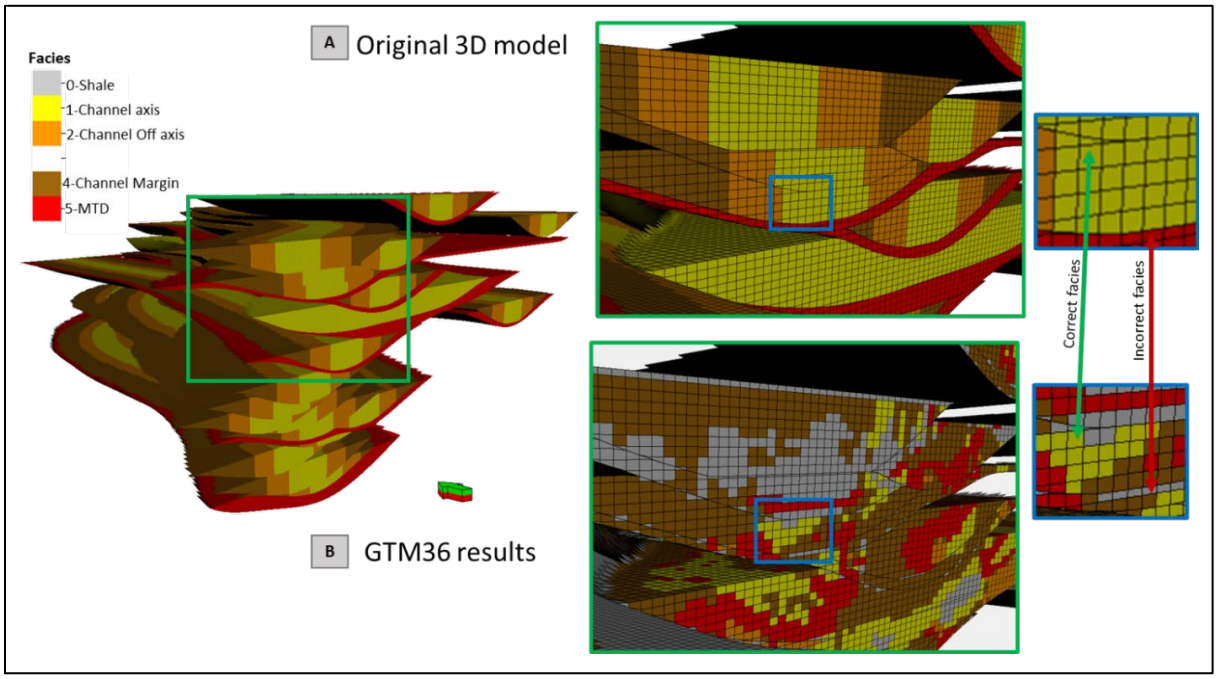


Figure 2.9. Visual representation of uncertainty in unsupervised ML methods. (A) Shows a section of the original 3D model, which has been filtered y the channel facies for visualization purposes. (B)Presents the same section in the results of GTM36 once they have been concatenated and resampled. Notice how some voxels represent the correct facies but others do not. This demonstrates that the use of ML methods will add uncertainty to the reservoir understanding (dimensions, facies, and volumetrics).

Discussion

We used synthetic seismic data and a 3D model originated from outcropping data to address the uncertainty in channel facies and architectures clustering using two popular unsupervised machine learning methods: SOM and GTM. In the following subsections, we discuss the main findings and insights of the study.

Selecting seismic attributes and parameters that are ideal for mapping channel architectures.

In geophysics, and most specifically interpretation of seismic reflection data, the usual input for any ML method are seismic attributes. In this sense, we recommend working only with attributes whose principles the interpreter understands well. In our case, we tested instantaneous and spectral seismic attributes that have been proven to provide good results in the seismic deepwater channel architecture assessment (Zhao et al, 2015; Roden and Sacrey, 2015).

In this study, we are applying attribute selection as if we did not know the expected answer or scenario with the aim of simulating the interpreter's behavior in front of exploration data. Therefore, after applying geological knowledge and statistical relationships we selected six attributes to use as input in the unsupervised machine learning methods: peak magnitude, instantaneous envelope, RMS, and spectral magnitudes of 20 Hz, 40 Hz, and 55 Hz. A good guide that can be followed in the input attribute selection stage is found in Roden and Sacrey. (2015), Infante-Paez and Marfurt (2019), and La Marca and Bedle (2022).

The use of unsupervised machine learning methods for deepwater seismic facies: which method is better?

The choice of an unsupervised ML method over a supervised method depends on the type of data available for the study. In exploration stages, it is often rare to rely on well logs or other data. In this scenario, unsupervised methods provide ways to find relationships among the variables in the data available (Roden and Sacrey, 2015), while reducing the time of interpreting big data associated with multiple input information that comes from the seismic attributes used.

However, unsupervised methods require interpreter evaluation to approve or disapprove the output. In this study, we selected two commonly used unsupervised ML methods (SOM and GTM) to evaluate their performance and better understand the real meaning of each cluster formed as well as compare both somehow similar methods. The advantage of using unsupervised methods relies in that they can identify new relationships between seismic attributes, which are unknown beforehand, assisting the expert in generating a more accurate interpretation.

Although several authors present insights on how GTM is superior to SOM and comes to solve many of its drawbacks (Chopra and Marfurt, 2014), such as the lack of convergence to a local minimum and setting of a learning rate that would optimize the results, we found that in our dataset SOM and GTM performance is very similar. When analyzing visually the results of the different cases: 256,81 and 36 clusters we observed that 1) The reduction of clusters does not necessarily offer an output that is easier to interpret and 2) GTM seemed to offer more distinction between the facies we aimed to interpret. Barkowiak (2004) states that generally, the topology of GTM and SOM looks similar, although GTM yields smaller quantization error than SOM.

When evaluating the metrics comparing each cluster case (e.g., SOM36 against GTM36), we noticed that there is a slight improvement in the overall performance (weighted avg of F1 score) in GTM vs. SOM method. For individual facies prediction, in both cases, Facies 1 and Facies 5 were better predicted than Facies 2 and Facies 4. However, we noticed that the predominance of shale (Facies 0) over other facies, imbalances our dataset, punishing the performance in the prediction of other facies and even causing misclassification of other facies

that fall in the Facies 0 category, which can be observed in the confusion matrix (Table 2.7) and in Figure 2.9.

The main takeaway of these results is that we should determine an optimal number of clusters, likely as suggested in our previous study (La Marca et al, 2021) where better results were achieved with a cluster number of three, which would represent the background shale, MTDs, and Channel facies. The drawback of applying this approach is that we are not letting the algorithm distinguish between economically attractive facies (e.g., Facies 1, or channel axis vs. Facies 4) and results would be too general and non-practical in instances when reservoir modeling and volumetrics are necessary to make important decisions.

Our suggestion would be to make use of data available, for example, if a well exists in a desirable location, we could use it to validate the clustering results and then create labels for training as unsupervised learning. On the whole, we would recommend using supervised methods whenever possible since they have direct performance metrics, and if this is not possible, then use more than one unsupervised ML method (e.g., GTM and Kmeans) with sufficient clusters in the initial setting that would give some freedom to find patterns among the data and consider best practices for each method used (Sun, 2000).

The drawback of both methods tested herein is that they are a reflection of the input and do not have controls for unbalanced data. Both algorithms are projecting the data from the attribute space onto the latent space. They do not have the capability to discern geology, but rather patterns in the data, and that is probably why the shale dominates. The latent space is "mostly shale", just like the attribute space. We prefer the algorithms to understand that shale is one entity, and the channel axis is another, but they only see attributes (datapoints). If the

attributes are similar, they will be close together in the latent space and will fall in the same cluster. This is a limitation of the technique in the method discussed and implemented here.

Uncertainty Quantification (UQ) of unsupervised machine learning methods: working with unbalanced data.

Some deepwater settings as explained in Slatt. 2013, can be mud-dominated, depending on the location and provenance of sediment (Fildani et al, 2013). In our case, slope systems are often comprised of high-quality lobe/ channel systems embedded in a poor reservoir quality shale. Although from the seismic response perspective the acoustic impedance contrast between these different facies (e.g., shale vs channel axis) benefits the imaging, when performing machine learning methods, we would have imbalanced data due to the predominance of shale over other facies and this may impact the accuracy of the methods.

According to Jeni et al. (2013) and Luque et al. (2019), when a dataset is imbalanced, the F1 score (which is a function of precision and recall) and its averages are better metrics to assess model performance as opposed to accuracy. We considered the F1 score and its weighted average in each scenario evaluated. Surprisingly we found that for the 36 clusters arrangement, SOM slightly outperformed the GTM method having a weighted F1 avg of 0.94 as compared to 0.92 in the GTM. On the contrary, when increasing the cluster numbers (81), we noticed how GTM outperforms SOM by 1% although SOM shows higher accuracy (~2% higher).

Another important observation is that weighted values are much higher than the F1 score, which may indicate that the models are good for the dominant class prediction but underperform for the other classes (facies). For example, the highest F1 score of all is always reflected in

Facies 0, and for other facies is always lower than 0.5; the highest was 0.33 shown in Facies 5 for SOM36. This is confirmed when we look at the weighted average which accounts for the proportions of each facies as compared to the total voxels. Therefore, we can say that in the presence of highly imbalanced data, the prediction of other classes would be affected by the predominant one. This is evidenced in the "per class" scores where we see that the models are not good (besides shale), resulting in F1 scores that lay far below 0.5 for any other class. One way to address highly imbalanced data could be testing metrics like the ROC curve or performing a normalization of the values as suggested by Jeni et al, (2013).

When evaluating models, we need to consider all the metrics, since for instance, when looking at the F1 score and macro one may think that SOM36 outperforms its counterparts, however, it presents the worst recall for class 2 from all the cases presented. This means that the model itself is not able to identify all the Facies 2 occurrences. In such case, if we were interested in class 2 and trying to identify potential regions for further investigation, GTM36 would have the best score, although it is still low.

Our interpretation of these results is that there is not a single cluster or set of clusters representing each facies, especially those that are near the resolution of the seismic. A better approach would be to consider background shale vs channel facies to see the actual error in interpreting these facies. In fact, in a previous study (La Marca et al., 2021) where K-means and PCA were evaluated when determining the optimal cluster number for the dataset using the elbow plot (Sum of squares within a method), there were three predominant clusters attributed to shale, channel facies, and MTD. Pampalk (2001) mentions that Since K-means is not concerned

with neighborhoods its clustering results are more accurate compared to the SOM. Therefore, this could be an alternative to improve results.

The advantage of this study is that although previous studies focused on addressing uncertainty in supervised ML methods (e.g., Abdar et al, 2021; and Stracuzzi et al, 2017), and authors like Mathieu, 2017 introduced the concern of addressing uncertainty in unsupervised methods, our study is pioneer in the realm of geophysics and the first that uses synthetic seismic data and outcrop derived models to evaluate the performance of unsupervised ML methods like SOM and GTM. Addressing uncertainty to understand the cluster definition is and how really these ML methods are operating is paramount to be considered prior to making any decisions based on interpretations and gain confidence in their use.

The importance of properly address uncertainty in unsupervised ML methods and implications on the energy sector.

According to Weber (1993), the use of 3D seismic has resulted in some cases, in a remarkable delineation of individual sand bodies and of reservoir architecture. However, even with high-quality 3D seismic data (Figure 2.1), it is still cumbersome to resolve internal reservoir configuration (Coleman et al, 2000) due to poor acoustic impedance contrast or insufficient thickness of the individual beds. As example. Kus et al. (2021) show how sands have changes not only in thickness but also in composition, offering a finger-like planform geometry.

Our analysis attempts to show the error associated with facies interpretation using unsupervised methods like SOM and GTM and allows for interpretations that in cases like the one presented by Kus et al (2021), one should consider selecting more clusters than the optimal

cluster number because some of the isolated clusters may represent those sands that do not necessarily share common features with other sandstones.

Stracuzzi et al (2017) point out that when using seismic data, errors, hence uncertainty, are added in each step. Therefore, uncertainty starts from the acquisition, followed by the effects of data preprocessing, which removes irrelevant signals, but may also alter seismic information. Furthermore, quality of data (resolution, noise, frequency content) may affect the resolvability of different architectures, for example, MTDs (Facies 5 in our study) are often overlooked, and since they represent baffles (Cardona et al, 2020) for fluid flow, they can negatively impact the reservoir assessment if not mapped correctly.

Subsequently, model-form uncertainty arises from the learning process: many plausible model parameterizations exist, and each provides a slightly different output. In addition to this, we noticed that there is uncertainty introduced when selecting attributes that may be not the most suitable/ adequate, to not use the correct parameters or cluster number, followed by the method we use to generate geobodies (how some of the voxels/ geoblobs are not going to be connected if geoblobs are isolated, so we lose important information).And since we assign the clusters a meaning regarding our geological knowledge we may be misinterpreting facies, or interpreting a cluster that actually has no geological meaning (a particular pattern or frequency range). Therefore, it is cumbersome to fully address uncertainty in seismic reservoir characterization.

Conclusion

We introduce a new approach to assess uncertainty in unsupervised ML methods, by using synthetic seismic data and 3D models built upon a real deepwater outcropping. Using synthetic data to quantify the methods' ability to correctly classify facies, provides confidence in

the interpretation of seismic facies using clustering techniques in cases where well data is not available. For our case, and probably commonly evidenced in analog settings, background shale (Facies 0) represented the majority of the data, creating an imbalanced dataset. Although we addressed this issue by focusing on the F1 score instead of accuracy, other metrics outside the scope of this study (e.g., ROC curve), could help to better understand the method's performance. There is no single cluster or set of clusters representing each facies, especially those that are near the resolution of the seismic. A better approach would be to consider background shale vs. channel facies (sum of main channel facies) to see the actual error in interpreting these facies.

Based on the metrics evaluated, overall GTM and SOM have similar performance. When looking at the F1 score and macro one may think that SOM36 outperforms its counterparts, however, it presents the worst recall for class 2 from all. This means that the model itself is not able to identify all the Facies 2 (channel off axis) occurrences. In such case, if we were interested in class 2 and trying to identify potential regions for further investigation, GTM36 would have the best score, although it is still low. As interpreters we need to be aware that there are multiple errors that add uncertainty in the seismic interpretation process, from the acquisition and processing which will condition the quality and resolution of our seismic, to the seismic attributes used as input, number of clusters and parameterization, as well as the human bias in the interpretation process, which if well logs are not available makes it rely on the geoscientist expertise. Due to this, we recommend the use of supervised methods over unsupervised whenever possible, or pure clustering techniques like K-means if unsupervised methods are the only option.

Addressing uncertainty in ML methods, especially unsupervised methods is necessary to properly select algorithms that allow for building more accurate 3D reservoir models, volume estimation, and flow simulation, all of which are critical for the oil and gas industry and in new energy studies like CCUS and geothermal resources development.

Acknowledgments

Authors would like to acknowledge SLB and the AASPI consortium at the University of Oklahoma for the software used for attribute and noise computation and visualization; and the Chile Slope Systems Consortium for providing the synthetic data and 3D model. Special appreciation to Rafel Pires de Lima and Mario Ballinas for their technical support.

CHAPTER 4: UNDERSTANDING UNCERTAINTY IN DEEPWATER CHANNEL SEISMIC FACIES CLASSIFICATION APPLYING RANDOM FOREST ON OUTCROP-CONSTRAINED 3D MODELS AND SYNTHETIC SEISMIC DATA*

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

La Marca, K., H. Bedle., L. Stright., M. Ballinas., K.J. Marfurt,.and D. Devegowda, 2023, Understanding uncertainty in deepwater channel seismic facies classification applying random forest on outcrop-constrained 3D models and synthetic seismic data.

Abstract

In order to understand the uncertainties in seismic interpretation, we apply a random forest classifier, a robust supervised machine learning method, to synthetic seismic data derived from real outcrop data. This novel approach of using outcrop-based synthetic data as ground truth allows us to better understand the influence of classifier hyperparameters (such as number of estimators/ trees, maximum depth of each tree, and others) on prediction accuracy. Based on previous analysis, we choose six seismic attributes that are able to differentiate five deepwater architectural facies: shale (thin-bedded turbidites), channel axis, off-axis, margin, and mass transport deposit (MTD). Our random forest hyperparameter testing indicate that optimization of the random forest classifier is sensitive to the 1) choice of training attributes and their predictive importance, 2) original facies proportions, 3) similarity in the seismic expression of different facies, 4) seismic data resolution, and seismic data signal-to-noise ratio. A simple classifier using common random forest hyperparameters developed for fluid saturation, predicted the facies with only 74% accuracy. Although computationally more expensive, optimizing the random forest hyperparameters provided ~89% accuracy. Unfortunately, when applying the best

model to an unseen portion of the model, although the position of the channel complex set was accurately predicted, our validation accuracy was only 58% , showing the limitations of universal models and persistent uncertainties. Utilizing outcrop-derived ground truth data provides insights into machine learning pitfalls and strengths for subsurface applications, where accurate predictions could impact operational decision-making and safety in both energy production and carbon sequestration.

Introduction

Machine learning (ML) techniques are becoming increasingly popular to solve problems that the naked eye or a human interpreter with limited time would not be able to solve. There are two types of ML: unsupervised and supervised. Supervised methods involve the development of prediction models when a given dataset has the output (labels) available. Supervised methods have many applications in the geophysical field, including the one addressed in this study, seismic facies interpretation. One of the advantages of supervised ML methods is that they can be trained for the labeled multi-class data to perform the task of classification (Ning and Yu, 2019). However, ML models carry uncertainty with them that can be difficult to quantify.

As used herein, uncertainty in machine learning refers to the lack of confidence in the results obtained, or working with imperfect information (Brownlee, 2019). Though no algorithm is perfect, metrics are a good resource to quantify output accuracy, which is critical in fields like medicine and technology where inaccurate predictions have consequences (e.g., self-driving vehicles). In our context, incorrectly predicting facies or fluid presence could lead to drilling wrong targets, causing economic losses and operational risks in the real world where data is often messy. Recent work has assessed uncertainty in ML applications (e.g., Nagl et al, 2022;

Abdar et al, 2021; and Mathieu, 2017). However, little research has examined uncertainty specifically when applying ML models to seismic data for predicting deepwater seismic facies.

Assessing uncertainty is further complicated by the fact that there are two main types of uncertainty to be addressed. Aleatoric and epistemic as described by Hora (1996) and Der Kiureghian and Ditlevsen (2009). Epistemic uncertainty stems from insufficient data representation and is reducible since it can be lowered by providing additional data. Models are expected to exhibit high epistemic uncertainty when the input data is far away from the training data. In practice, nobody can provide an infinite amount of training data, so a model will never be able to reach an epistemic uncertainty of zero. The second type of uncertainty, aleatoric, relates to the data itself. Data are never a perfect representation of reality and are always bandlimited and often contaminated by both random and coherent noise such as multiples and migration artifacts. Aleatory uncertainty, as opposed to epistemic uncertainty, cannot be reduced with additional information since it is not a property of the model but rather an inherent property of the data themselves. Therefore, we need to be aware of the pitfalls that model bias brings to the interpretation process (Posamentier et al, 2007).

According to Phan (2019), uncertainty quantification (UQ) currently underpins many critical decisions, and predictions made without UQ are usually not trustworthy. A common approach to quantify uncertainty is the Bayesian approach (e.g., Langenkamp, 2021), however Bayesian Neural Networks are computationally expensive and "do not scale easily to complex neural network and architectures containing many parameters" as noted by Nagl et al., (2022).

Therefore, we aim to improve the understanding of the uncertainty associated with deepwater seismic facies prediction by comparing a ground truth outcrop-derived 3D model and

derived synthetic seismic data with predictions made using random forest (RF) unsupervised ML algorithm that was developed by Breiman (2001). In this study, RF is applied as a classification technique, since we want to predict facies which are discrete variables and distinguish these classes (facies) from the rest. We then optimize the hyperparameters and apply the best model obtained to unseen data, the answers to which we already know a priori. In reality, we may never have all the observations, in which case we will not need a predictive model. However, in this case, having all the answers allows for a better understanding of the performance, and nuances of the method. As George Box's aphorism states, "all models are wrong, but some are useful", so we aim to understand how useful an optimized RF is to predict deepwater facies.

The RF method takes advantage of the decision trees that ensemble it but alleviates overfitting of training data and is less biased since it integrates individual trees, generally trained via bagging. A key benefit of a bagging algorithm is that it provides a quantitative measure of the importance or redundancy of each attribute in the learning process, resulting in highly accurate prediction (Kim et al., 2018). Another strength of RF is prediction of multiple classes that are available. There are some studies where RF has been applied to seismic data. For example, Kim et al (2018) applied RF algorithm to classify seismic facies and to thoroughly define the limestone and shale facies in a Barnett shale seismic survey. Ballinas et al (2023) discriminated fluid saturation (water, oil and high or low saturations of gas), and Bhattacharya et al (2016) performed lithofacies classification in the Bakken shale. Most of these studies successfully distinguish shale, limestone, or sand prone facies in a general way using seismic facies (e.g., Ao et al, 2019). However, there has been limited work to date that applies RF to understand deepwater seismic facies to the extent of detail in this study.

3D synthetic models have been used to calibrate the sensitivity of new technology to better image geologic features of interest. For instance, Clawson et al. (2003) used 3D synthetics computed from outcrop studies in the Permian Basin to calibrate the mapping Brushy Canyon channels using the relative new coherence and spectral decomposition algorithms. Inspired by the latter, unlike other studies that have a probabilistic approach to quantify uncertainty, this study utilizes ground truth synthetic data to understand the uncertainty associated to the different levels of an unsupervised machine learning method application. This includes aspects such as data understanding and input selection, as well as the analysis of misclassifications, all in the context of comparing with a ground truth model, in what we call "analog-based uncertainty assessment".

We begin our paper with a summary of data preparation, seismic facies selection, and seismic attribute selection. Then we use some of these data to train the random forest algorithm, modifying the RF hyperparameters to improve performance of the model. Then, we use the best model on a portion of the original data that was not used before in a process that we here refer to as validation. Because these are synthetic data generated from a known model, we know the correct answer, allowing us to visualize and numerically assess facies misclassifications. We conclude by summarizing our main findings and how this workflow and the application of random forest classifiers in general can benefit the seismic interpretation community.

Methodology

We employed the random forest algorithm in a supervised manner to classify deepwater seismic facies. To accomplish this, we utilized a 3D model constructed from observations and measurements of outcropping deepwater slope channels (Ruetten, 2021) and a corresponding synthetic model (Langenkamp, 2021) to train and assess the algorithm's performance. Our selection of seismic attributes (features) was guided by previous analysis and correlations (La Marca et al, 2023). We ensured consistency by cropping the input data to the same dimensions and standardized the seismic attributes through Min-Max scaler normalization (Patro and Kumar, 2015) before applying the algorithm.

To achieve optimal results, we fine-tuned the hyperparameters of the random forest algorithm, to find a balance between accurate classification and avoiding overfitting. This optimization process allowed us to harness the potential of RF for classification efficiently and effectively. For a detailed overview of our approach, refer to Figure 3.1, which illustrates the workflow applied in this study. Subsequently, we provide explanations and considerations regarding our methodology.

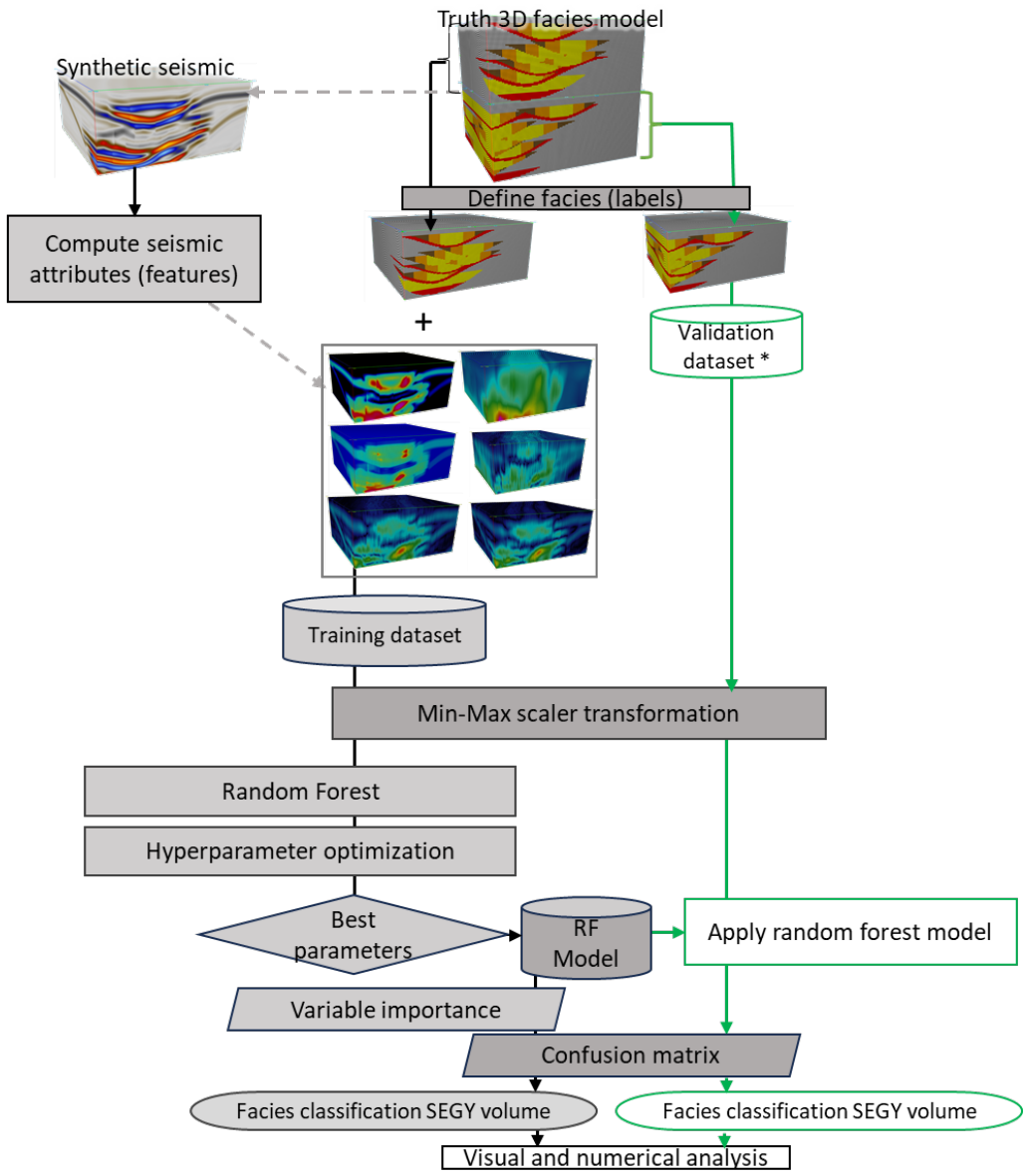


Figure 3.1. Workflow used in the uncertainty assessment of deepwater facies classification using a random forest classification algorithm, applied to an outcrop-derived 3D synthetic seismic model (Langenkamp, 2021). The choice of seismic attributes used in this model is discussed in La Marca et al (2023). * Notice that the original model has been divided into two: upper portion to perform the RF model, and the lower portion (green) to validate or apply in this unseen piece of model. In this study, it's important to note that the terms 'test' and 'validation' are not interchangeable. Validation is employed here to understand uncertainty.

Ground truth model foundation

The Laguna Figueroa outcrop in the Chilean Patagonia, served as the foundation for creating a scaled, GPS-positioned 3D reservoir model that encompasses various channel related facies. These facies serve as our model labels, including 1) channel axis, 2) channel off- axis, 3) channel margin, 4) mass transport deposits (MTD), and 5) thin bedded turbidites (for practical purposes we use the term "shale"). This model, in conjunction with 1D convolution using zero-phase Ormsby wavelets and assigned real acoustic impedances, facilitated the generation of synthetic forward modeled seismic data. For more in-depth information of the model refer to Ruetten (2021) and Langenkamp (2021). From the suite of synthetic data provided, we used the 60Hz dominant frequency volume (Figure 3.2) with random noise added to mimic real data (Stright et al., 2014), and so provide valuable insights into the uncertainty associated with seismic facies classification.

Phase 1. Data preparation and seismic attribute selection

When applying ML techniques, one of the critical initial steps involves data preparation. The choices made regarding data selection, variables (such as seismic attributes), and study window can affect method's robustness. While no single seismic attribute can effectively distinguish all the targeted classes or facies for study, it's important to note that an excessive number of variables does not necessarily provide a solution to this challenge. Therefore, the method for selecting input seismic attributes involves a combination of user expertise, ensuring that they align with common practices for channel interpretation, and assessing their statistical correlation suitability (Barnes, 2007, Kim et al, 2018; La Marca et al., 2023). These attributes were computed across the entire dataset and subsequently cropped within the selected study

window, reducing potential smearing effects or errors associated with trace-based attribute calculations. Figure 3.2 provides a representative 3D visual of each of these attributes.

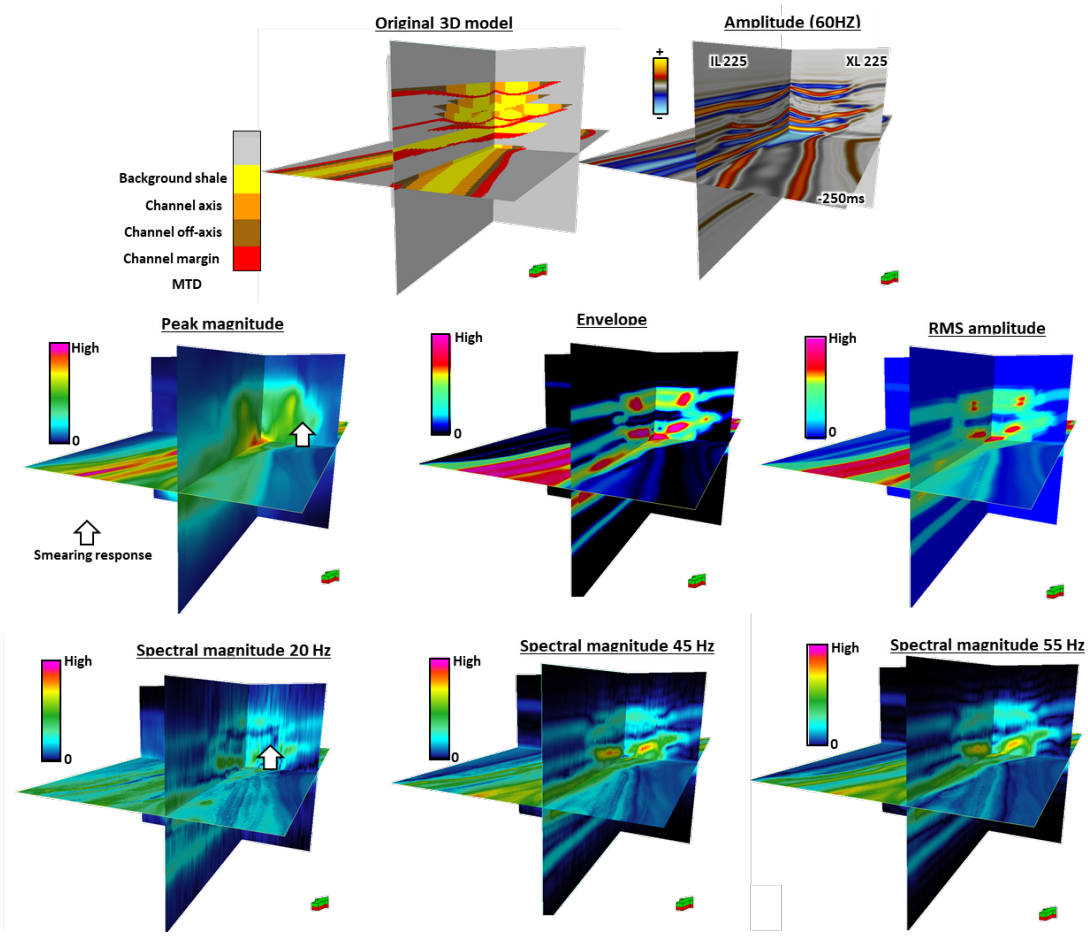


Figure 3.2. First row: 3D Model and synthetic amplitude volume, second and third row: calculated seismic attributes used in the study. Peak magnitude, envelope and Root Mean Square amplitude (RMS) are often used to highlight changes in lithology, whereas spectral magnitudes are used to interpret architectural elements of different thicknesses.

To assess the reproducibility of RF classification, we divided the original volume into two distinct sets as noted in Figure 3.2: one for training/test (upper) and the other for validation (lower). The latter represents an unseen and unused portion reserved for the validation process, to demonstrate how accurately the RF prediction model is performing using "new" data. (note

that test and validation here are not interchangeable and that normally RF does not require validation). Data conditioning was carried out by analyzing the distribution of variables and applying Min-Max scaling, which normalizes the data between 0 and 1 while maintaining skewness (Raju et al, 2020). This scaling process allowed to ensure that both the input data and the validation dataset are in the same analytical space (Ballinas et al, 2023), and scaled, making posterior processing easier.

Phase 2. Random forest method

Numerous supervised ML methods are available for tackling facies classification challenges. Nevertheless, among these options, Random Forest (Ho, 1995; Breiman, 1996, 2001) stands out as one of the most robust supervised classification algorithms as demonstrated by previous studies that compared RF with other methods (Bhattacharya et al, 2016; Kim et al, 2018; Rahimi and Ali, 2022; Nayaran et al, 2023; Ballinas et al, 2023). Therefore, the decision to use RF in our study was driven by its track record of effectiveness in classification tasks, coupled with its robustness in handling various data complexities, including noise, outliers, and data heterogeneity.

Random forest consists of an ensemble of decision trees, with the final classification being determined by the majority within their individual results. In our specific case, the input labels, representing facies, are linked to the input seismic attributes to be used as training dataset. RF is based on bagging (Breiman, 1996), meaning that data in the training dataset is selected with replacement. This approach enhances overall accuracy and reduces overfitting problems (Breiman, 2001).

Feature/ attributes importance

We define the M target facies (labels) one by one, resulting in

$$Ntotal = \sum_{m=1}^{M} Nm$$

*w*here Nm is the number of training voxels for the m^{th} facies. Then we compute a suite of Q seismic attributes (features) that best differentiate the M facies such that the training data consists of N_{total} length-Q vectors. Because of bagging, for each tree constructed within the ensemble, a random subset of data points is selected. This process aims to discern the importance of each feature (seismic attribute) in distinguishing between different classes (facies). The result is a histogram that shows the seismic attribute importance/ contribution in the classification. This unique aspect of random forest greatly contributes to its effectiveness as a classification method.

Hyperparameters of random forest: grid search and cross-validation

A crucial step in ensuring the robustness of RF is the optimization of hyperparameters. These parameters play a vital role in the model's training process and directly influence its accuracy in predicting classes. For this reason, we undertook an exhaustive iteration of grid search to fine-tune the hyperparameters and gain a comprehensive understanding of how each parameter influenced accuracy (score). Grid search creates alternate model configurations by analyzing and

discretizing a target range of values into each hyperparameter of interest and training the model across all hyperparameters for all combinations (Mesafint and Manjaiah, 2021).

During this experiment, we implemented a five-fold cross validation approach during training, maintaining an 80/20 ratio for points used for training compared to those used for the testing stage. Thus, this proportion which is based on Pareto's principle ensures a robust assessment of model performance (Gholamy et al, 2018; and Roshan, 2022).

We performed six tests (Table 3.2) varying one parameter at a time, and the last test (test 6) uses all the optimal parameter values in each of the previous tests. We did this to understand the individual contribution of each parameter to the model performance and to save time, since running a grid search varying all parameters at once, although a common practice, can be very time-consuming. Table 3.2 shows a summary of the parameter evaluated in this study, its meaning and description of the test. We selected the best combination that not only yielded high accuracy but that also was computationally efficient, considering the dataset's scale.

Error estimation and metrics

According to Kuhn and Johnson (2013), a common method for describing the performance of a classification model is the confusion matrix. Therefore, confusion matrices were used to assess the performance of the Random Forest classification model. These matrices compare original (on the Y-axis) versus predicted (on the X-axis) classes in a n x n configuration. The main diagonal of the matrix reveals the correctly predicted classes (facies), while off-diagonal elements represent misclassifications into other classes. Table 3.1 describes the statistical metrics used and how to calculate each one of them.

Although metrics such as precision, recall, accuracy, and weighted average accuracy are used to understand how well classes are classified, in the case of imbalanced data (a greater proportion of one facies over the others) the F1 score becomes our best ally to evaluate the model performance. If you have an imbalanced dataset but want to assign more weight to classes with more samples, consider using weighted average instead of macro average (which treats all classes equally).

Table 3.1. Metrics to evaluate performance of the ML evaluated. Modified after Kuhn and Johnson (2013).

Metric	Meaning	Equation
Precision	Measures out of all predicted positives how many are actually positive/ Or how good the model is at predicting a category	$Precision = \frac{TP}{TP+FP}$
Recall	Measures how many positive records are predicted appropriately (positive)/ How many times the model was able to detect a specific category	$Recall = \frac{TP}{TP+FN}$
F1-score (or F measure)	This metric allows to evaluate the model performance. It is the harmonic mean of precision and recall	$F1score = 2 * \frac{Precision * Recall}{Precision + Recall}$
Accuracy	Refers to the proportion of correct matches/ How many times the model evaluated was correct overall	$Accuracy = \frac{Correct\ predictions}{Total\ predictions}$
Macro average score	Arithmetic mean of all the pre-class recall or f1 scores. Treats all class equally without considering their support values	$Macro_avg = \frac{\Sigma\ all\ per\ class\ (recall\ or\ F1) scores}{Number\ of\ classes}$
Weighted Average	Takes the mean of per class scores while considering the proportion of each class support	$W_avg = \Sigma\ (perclass\ F1\ score * (support\ of\ class\ /total\ support)$

Phase 3. Model validation

After selecting and applying the best RF model, we performed a validation step by applying the algorithm with these chosen settings to a portion of the model that had not been utilized previously. This approach differs from others found in the literature, where wells (that represent a single location in the area) are typically employed for blind testing the model's application. In our case, having access to a complete true model (and consequently, the expected correct facies) provides a unique opportunity to gain a deeper understanding of the uncertainty associated with these facies' predictions using random forest.

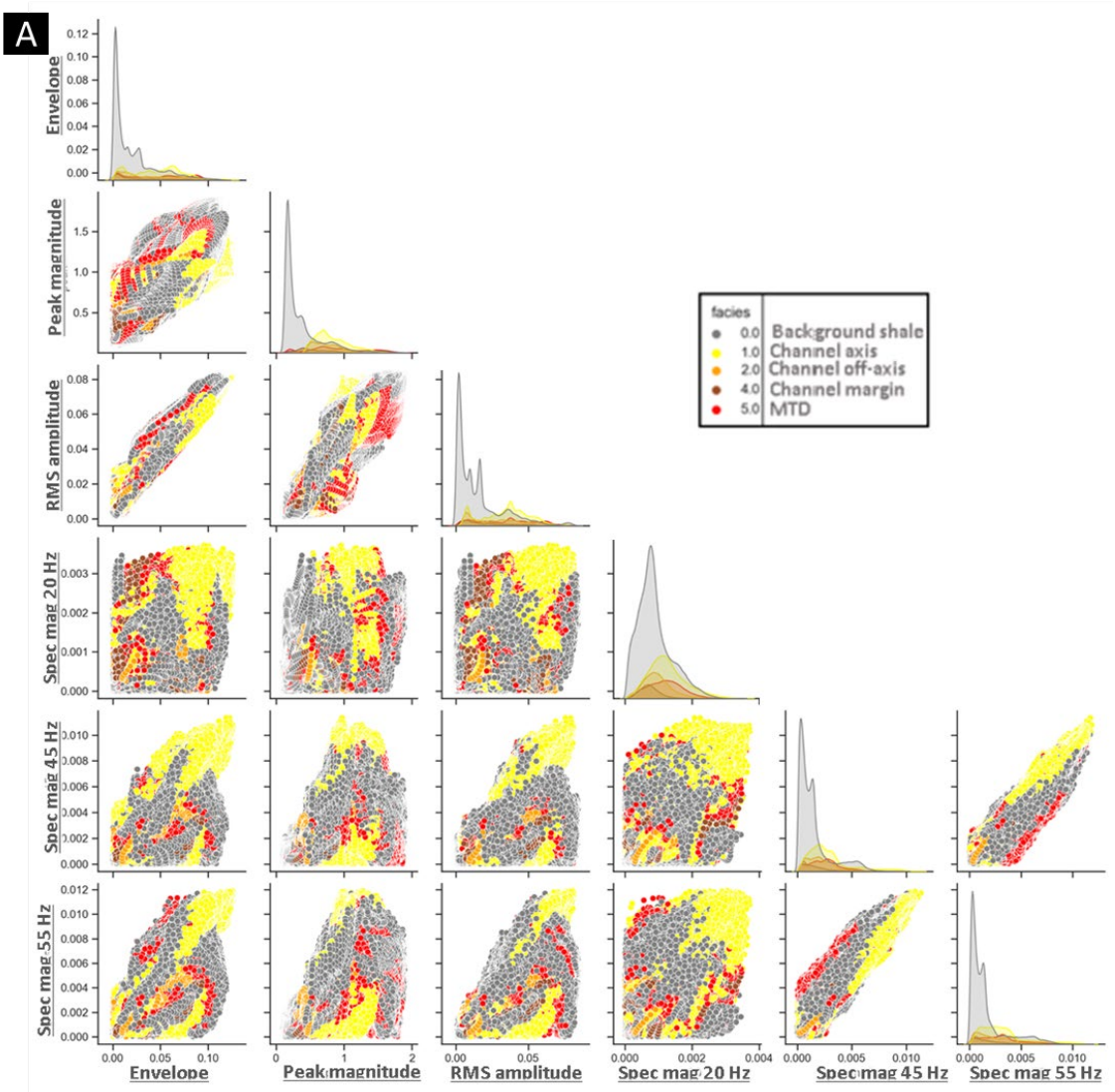
Results and discussion

In order to evaluate the uncertainty associated with deepwater facies classification using random forest supervised ML algorithm, a 3D outcrop-derived model with five facies (labels) was used and six seismic attributes (RMS, instantaneous envelope, peak magnitude and spectral magnitudes of 55 Hz, 45 Hz, and 20 Hz) were employed for training.

The full advantage of utilizing machine learning methods lies in their ability to harness information from multiple seismic attributes. In traditional interpretation practices, interpreters are typically constrained to visualize one or, at most, three attributes at a time. Ironically, this limited approach often introduces more uncertainty into the interpretation process compared to the use of ML methods that can combine and aid in visualizing multi-dimensional data.

Figure 3.3 displays a scatterplot that illustrates the correlation between these seismic attributes and the different classes or facies. Upon analyzing this graph, it becomes evident that uncertainty may arise from the fact that attribute values within each facies exhibit similar ranges,

and each attribute seems to distinguish well between shale facies and channel facies, but differences between individual channel facies may be subtle. Probably, this is partially a result of the choice of features (seismic attributes) within the category of epistemic uncertainty and the resolution and quality of the seismic (aleatoric uncertainty) in which case we have limited to no control.



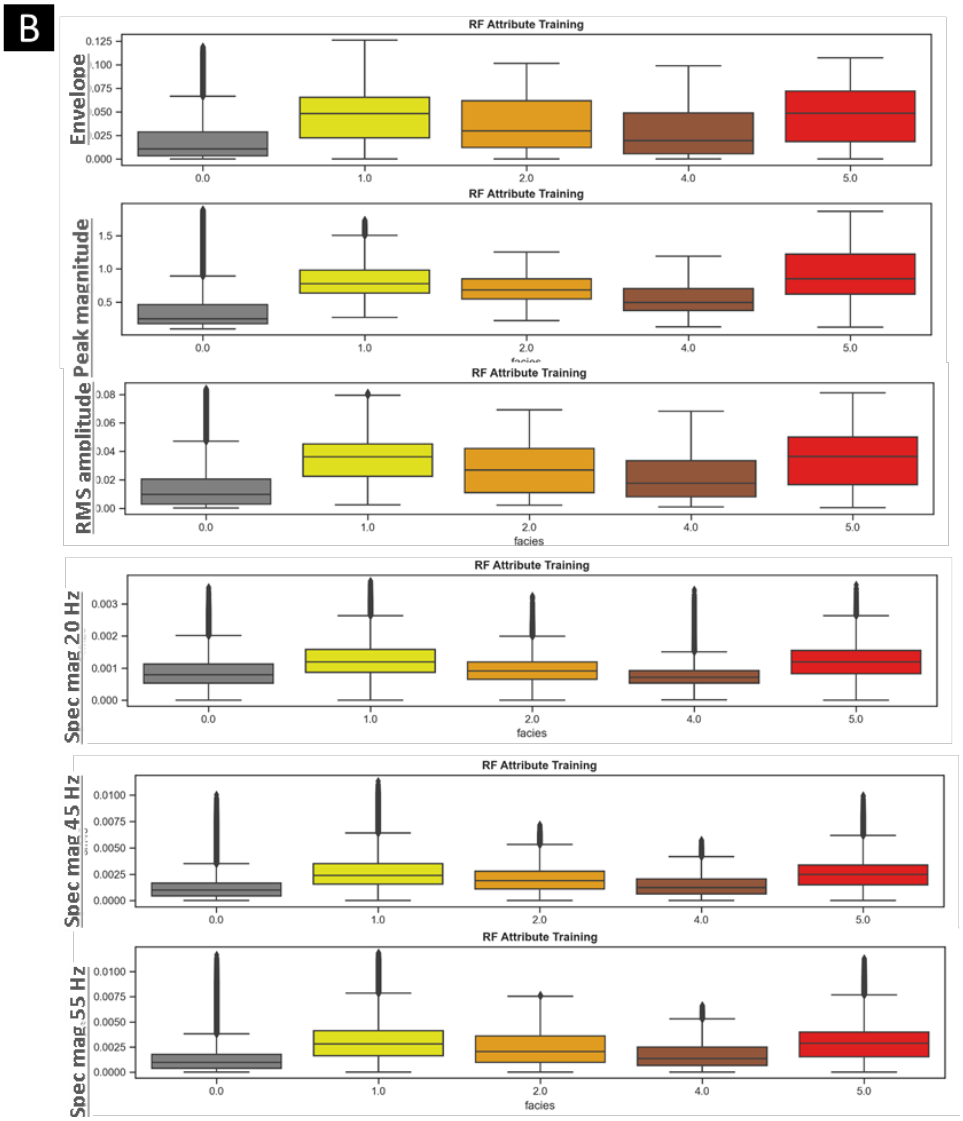


Figure 3.3. (A) Scatter plot displaying the correlation between input seismic attributes used as training data. Data points color code represent the facies defined (Yellow: channel axis, Orange: channel off-axis, Brown: channel margin, Red: Mass Transport deposits (MTD), and Gray (background shale) (B) Box plots showing the uniqueness of each attribute in representing each facies. Notice how shale can always be distinct from the channel facies overall.

Another observation is the prevalence of the gray (shale/non reservoir) facies in comparison to the other facies categories. This aligns to what is depicted in Figure 3.4, which illustrates the dominance of background shale facies over its counterpart, the channel facies. Although this suggests that the used dataset may be imbalanced, it is indeed a representation of

actual deepwater channel facies distribution, and hence, what we will likely find in real seismic data.

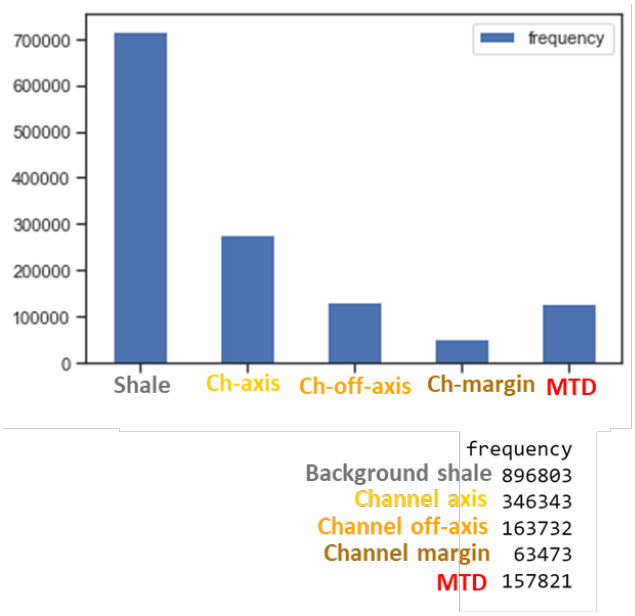


Figure 3.4. Histogram of samples count per facies (label) in the training dataset. Although the shale has greater representation than the other four facies, this imbalance is representative of actual deepwater channel facies distribution.

As stated by Stracuzzi et al. in 2017, even minor alterations in the training or test data can result in substantial and unpredictable variations in performance for a model characterized by uncertainty. Therefore, we conducted a thorough hyperparameter optimization, as outlined in Table 3.2 and Figure 3.1. The outcomes of this optimization process indicated that the algorithm's performance notably improved under certain conditions. Specifically, when the maximum depth (Test #4) was set at a high value, when the maximum number of samples was increased (Test #3), when the minimum number of samples required to split was kept small (Test #2), and when a high number of estimators (Test #5) was utilized. In simple words, the more trees within the model the better. These findings are highlighted in the results from Test #6, as shown in Table 3.2.

Table 3.2. Summary of the best results obtained per hyperparameter optimization test. Each test comprised a grid search with variations of the tested hyperparameter. Test #1 was selected randomly as a start point, and each subsequent test evaluated a different parameter (in bold). Test #6 ultimately shows the best hyperparameter configuration, which is the result of a last grid search using the results/ tendency of Tests 2-6. Test #6 shows the parameters employed in the RF model.

Test #	N estimators	Max depth	Min sample split	Max features	Max samples	Score	Description hyperparameter tested	Result/ tendency that offers best perfomance
1	500	425	10	2	4000	0.790	Original values used as start point	n/a
2	500	425	**5**	2	4000	0.799	Min sample split:Minimum number of data points placed in a node before the node is split	Smaller/ reduce
3	500	425	5	2	**20000**	0.848	Max samples: Determines what fraction of the original dataset is given to any individual tree	Larger/ increase
4	500	**500**	5	2	20000	0.778	Max depth: Maximum number of levels in each decision tree	Reduce or keep
5	**1000**	500	5	2	20000	0.848	Number of estimators: Number of trees	Larger/ increase
6	**1000**	**425**	**3**	**2**	**80000**	**0.882**	**Using the best hyperparameters tendencies**	

Random forest allows to determine variable (seismic attribute) importance

One of the advantages of RF is its ability to determine the importance of attributes in the classification process. The results indicate that peak magnitude significantly outperforms its counterparts (Figure 3.5), accounting for 35% of importance, followed by RMS amplitude at approximately 17%. The remaining variables exhibit a similar level of performance, each contributing around 13%. Notably, the attribute with the least impact on classification was the spectral magnitude of 20 Hz. Spectral seismic attributes generate a smearing response as seen in Figure 3.2 with white arrows, possibly affecting the correct identification of facies, yet are suggested by the authors in the channel architecture interpretation from seismic data. A suggestion would be to try other methods for spectral magnitude calculation different than CWT

(continuous wavelet transform) or combine and test with a suite of attributes less sensitive to smearing.

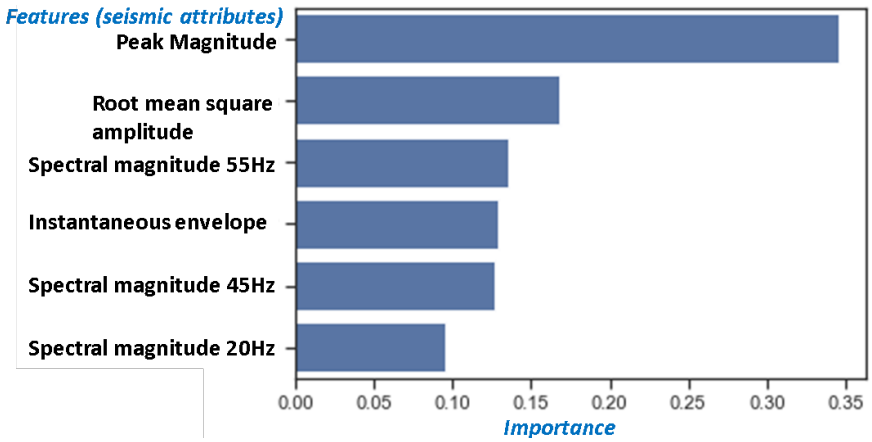


Figure 3.5 A histogram that shows seismic attribute (feature) importance. Notice how peak magnitude considerably outperforms in the analysis. This might be attributed to this attribute sensitivity to changes in tuning which benefits different architectural facies recognition.

After the RF is implemented, we employ two ways to assess its performance: 1) quantitatively by using correlation matrices and performance metrics, and 2) visually, by exporting the 3D prediction generated in SEGY format and comparing it with the original data voxel to voxel. For the first case, Figure 3.6 presents both the confusion matrix (on the left) and the classification report (on the right) for the original (Test 1) RF model versus the optimized one (Test 6).

Precision, expressed as a percentage, measures how accurately particular facies are predicted in comparison to other facies or classes. In the classification report, you can observe that the term "support" refers to the number of samples or data points used in each class. Additionally, the "macro average" considers this support, providing a more reliable training

score, particularly in datasets where there is a significant disparity in the quantity of data points, which is our case with the background shale (as seen in Figure 3.4).

It's worth noting (Figure 3.6) that the macro average value increased from 52% to 81% after applying the optimal hyperparameters and weighted average accuracy went up from 74% to 89%, indicating a substantial improvement in model performance. This emphasizes the importance of hyperparameter optimization in the model accuracy and reduction of uncertainty. To understand how this happens, Figure 3.7 provides a visual representation of the distinctions among the original dataset, the first random forest model, and the optimized RF model. The optimized RF model effectively captures nearly all the facies present in the original dataset. However, it's noteworthy that the margin facies (brown) proved to be the most challenging to predict. As corroborated by the confusion matrix, it was frequently misclassified as shale. Although these facies may be sedimentologically similar, this pattern was observed in relation to other classes as well.

Random forest model performance

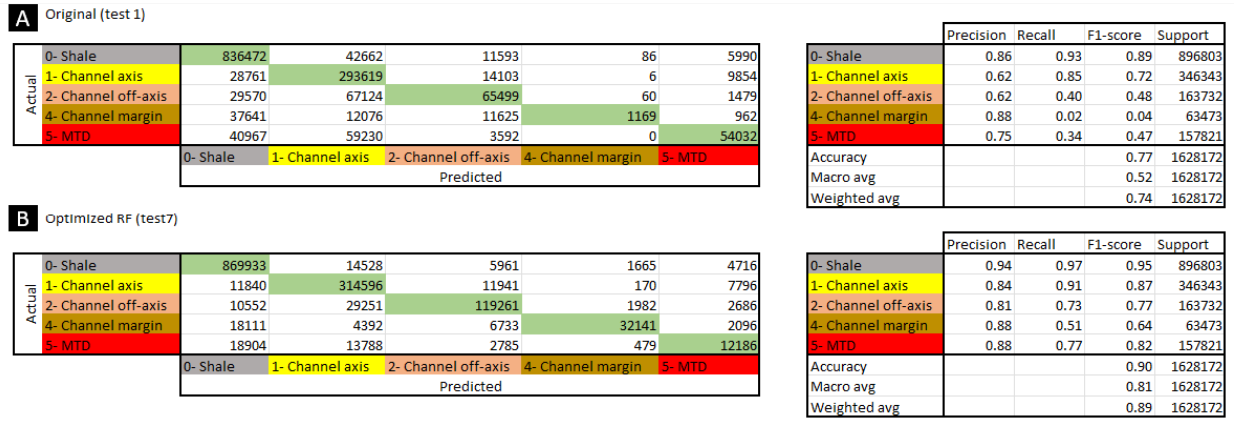


Figure 3.6. Confusion matrices (left) and classification report (right) to evaluate RF classifier performance. Confusion matrices compare original (on the Y-axis) versus predicted (on the X-axis) classes in a n x n configuration. The main diagonal of the matrix reveals the correctly predicted classes (facies), while off-diagonal elements represent misclassifications into other classes. (A)RF results applying test #1 hyperparameters indicated in Table 3.2, (B) Optimized random forest model results, applying test #6 hyperparameter values in Table 3.2 Notice how accuracy increases substantially with the hyperparameter optimization.

Moreover, when delving deeper into the analysis, we find that the highest error rates may be associated in part with the stacking patterns. For example, in Figure 3.7B, note how other channel facies (margin or off-axis) are misclassified as channel axis facies where channels are stacked vertically. Also, high errors are found in channel margin facies. Figure 3.7 shows how the thickness of these elements may suggest that the resolution of the seismic plays an important role in model performance. The fact that margin facies are located in areas where the data is sparser, which is a known downside of the method, results in high errors in classification. These misclassifications have a direct impact on uncertainty in interpretation since geoscientists could be evidencing an increase or decrease in the observed label (facies) which translates directly in over or underestimation of the facies which can impact directly on economics (e.g., overestimate the presence and volumes of prospective facies) and even drilling safety (underestimate the

presence of drilling hazards and or baffles such as MTDs or others in a different geological setting).

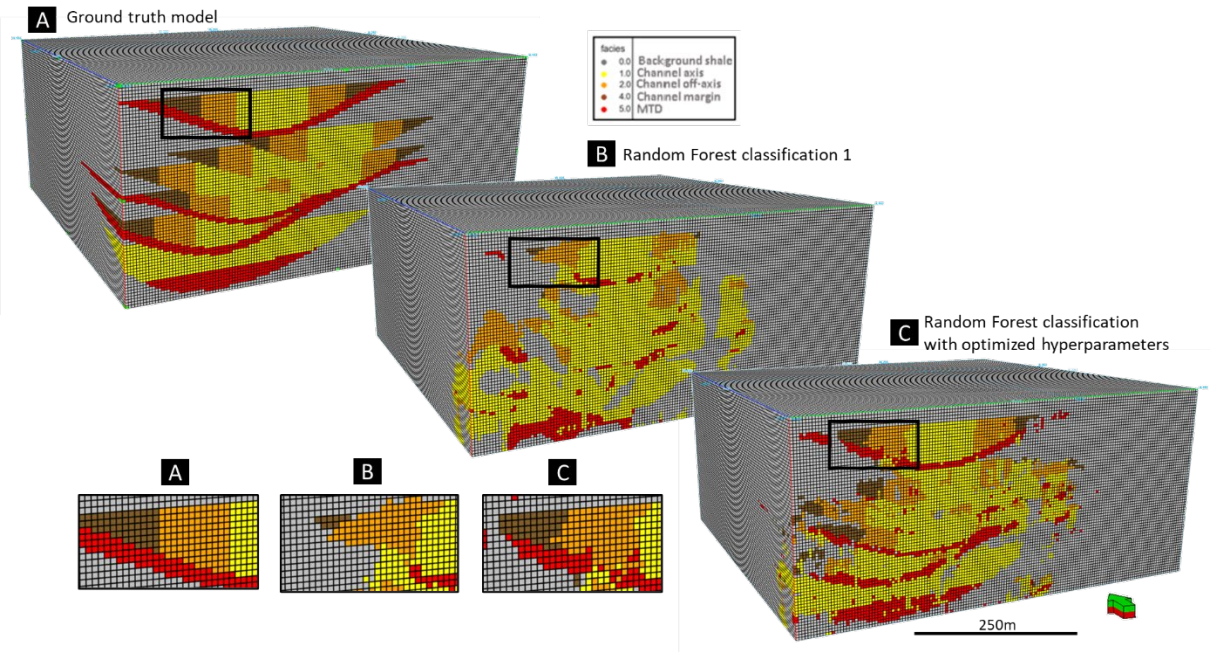


Figure 3.7. Comparison of (A) original model (B) unoptimized RF- test#1 in Table 3.2-, and (C) RF classification with optimized hyperparameters – test #6 in Table 3.2-. Notice how the prediction is highly improved in C, over B as compared to the truth model. If we didn't optimize hyperparameters we would be overestimating the channel axis facies and underestimating thinner facies such as the channel margin and MTD.

Model validation

To evaluate the performance of the RF model, we applied it to a previously untouched portion of the data that had been set aside before the training. This allows us to examine how well the model is predicting the facies. Figure 3.8 shows the confusion matrix and classification report, revealing that the model achieved an accuracy of ~60%. A closer look at the misclassifications occurred, reveals that the label with the highest rate of misclassification was

“channel margin”, which was often mislabeled as “channel axis facies”, followed by “shale”. Notably, this class also had the lowest F1 score.

However, when we examine the entire context visually in Figure 3.9, it is quite impressive to see that the model successfully captures the trends and overall configuration of the channel facies, even though they are located at different depth and orientations from the channels and facies used for training (Figure 7A). The “channel off-axis” and “margin” facies are thin enough to be affected by the resolution of the data. Using seismic attributes, even with optimal parameterization (La Marca, 2023), can lead to some smearing that may impact the labeling process and result in the misclassifications we observe here. Also, having a high accuracy in the training set vs the accuracy obtained in the validation may suggest overfitting, meaning a model learned well the input but that is not applicable to new problems/data.

To have a better understanding of how the channel facies predicted were distributed, a representative cross-section and time slice of the original and predicted validation models are presented in Figure 3.10. Notice how channel position and orientation have been correctly predicted to be able to differentiate between shale (non-reservoir) facies and channel (reservoir) facies. It is observed a predominance or likelihood of prediction of class 1 (channel axis) and class 5 (MTD) over other channel facies. Channel axis facies are more evident where channels stack vertically.

Overall, we notice that the model correctly predicts the location of the channel complex, which is valuable, especially if we extrapolate this capability to applications in exploration and development. The prediction of the correct location of the channel in unseen data reduces uncertainty in the use of RF to understand general position and trends of potential reservoir.

This supports the idea of considering using reliable analogs for developing RF (or other supervised) models that can help to better understand and interpret new frontiers and basins.

Nevertheless, it is evident that some facies within the channel are misplaced. While this might not be a significant issue when considering channel facies as reservoir facies, it could increase the risk if facies such as the "MTD's" (potential baffles) are misclassified. We attribute these misclassifications to the use of a training dataset that has high accuracy but that applied to deeper facies struggles to identify the correct label. This is a real-world problem that warns of an increase in uncertainty when applying a model that learned from training data that suits it but may not be ideal for a different basin or seismic quality.

A

Validation: confusion matrix

Actual	0- Shale	1- Channel	2- Channel	4- Channel	5- MTD
0- Shale	751638	115664	43542	13781	34485
1- Channel axis	55231	180408	30235	1206	43770
2- Channel off-axis	48219	94324	14833	1303	29701
4- Channel margin	16719	27763	4524	520	4899
5- MTD	30370	59150	10799	273	14815

Predicted

B

Classification report

	Precision	Recall	F1-score	Support
0- Shale	0.83	0.78	0.81	959110
1- Channel axis	0.38	0.58	0.46	310850
2- Channel off-axis	0.14	0.08	0.10	188380
4- Channel margin	0.03	0.01	0.01	54425
5- MTD	0.12	0.13	0.12	115407
Accuracy			0.59	1628172
Macro avg			0.32	1628172
Weighted avg			0.58	1628172

Figure 3.8. (A) confusion matrix and (B) classification report for the validation model. Here we see a decrease of ~ 20% in accuracy when applying the optimized RF model to unseen data. See Figure 3.6 for reference. Support refers to the number of voxels per facies.

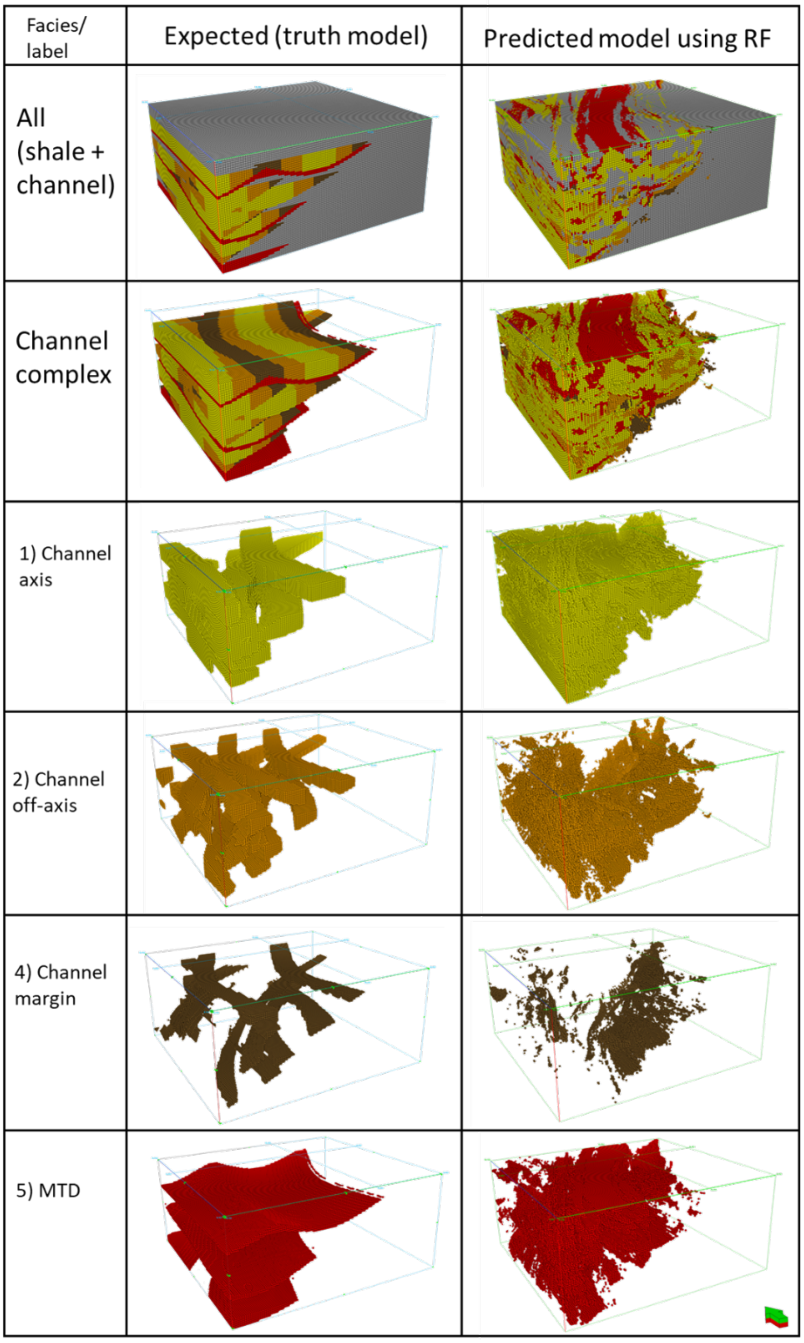


Figure 3.9. Comparison of (A) original vs (B) predicted RF validation. In this chart, facies have been isolated via opacity to compare areal distribution and proportions visually. Channel complexes' location and orientation are predicted well. Chanel axis, shale. and MTD facies show predominance over the other channel facies, which has been misclassified into either of these predominant channel facies or shale.

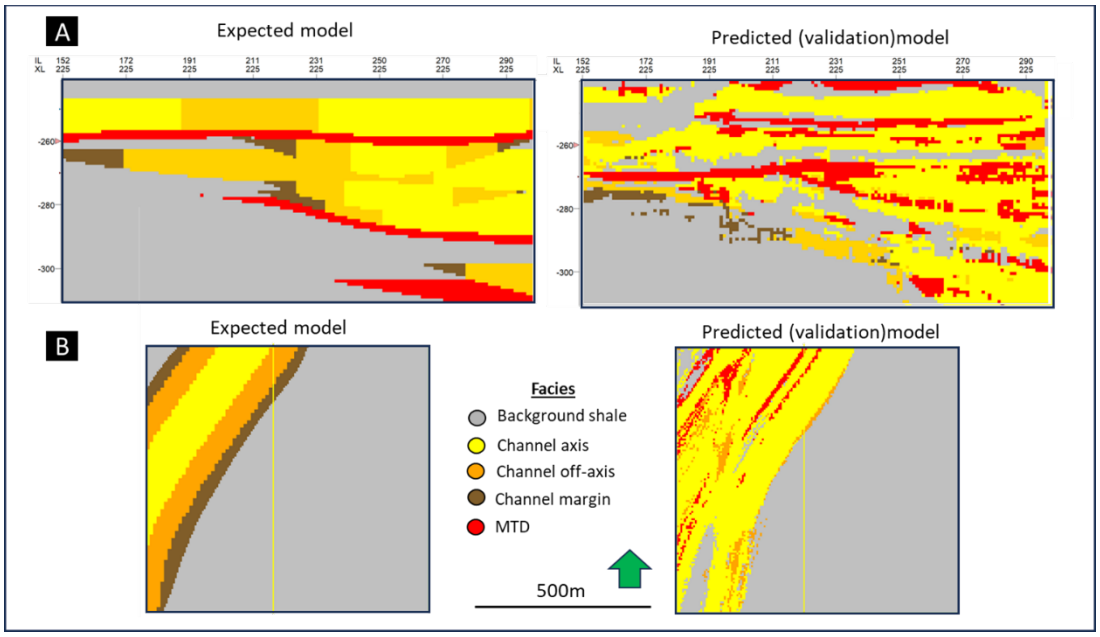


Figure 3.10. Comparison of expected (truth) vs predicted in the validation model in (A) Representative crossline, (B) representative time slice. Notice how channel position and orientation has been correctly predicted to be able to differentiate between shale (non- reservoir) facies and channel (reservoir) facies. It is observed a predominance or likelihood of prediction of class 1 (channel axis) and class 5 (MTD) over other channel facies. Channel axis facies are more evident were channels stack vertically.

Changes in lithology, stacking pattern, channel distribution, level of compaction and presence of other elements like potential salt are expected to affect the results, which is suggested also by Langenkamp (2021). There is not a one- size that fits all algorithm or method, or parameters. However, we want to underscore the importance of fine-tuning hyperparameter to boost the accuracy of model predictions and address epistemic uncertainty. On the other hand, we realized that aleatory uncertainty will always be a factor, as it is intrinsic to variables like data quality, facies proportions and the presence of outliers which we can be aware of but have limited control on.

This study reveals several key takeaways: 1) there is no perfect model that will reproduce with fidelity all facies in an uncalibrated basin. 2) Model optimization requires a) understanding

original data limitations like resolution, tuning, quality, noise (aleatoric uncertainty); b) using appropriate seismic attributes; c) exhaustive hyperparameter tuning, albeit time-consuming; and d) balanced, representative training data via resampling. 3) Despite the robustness of supervised methods, expert geoscience input is essential for contextualizing and incorporating prior knowledge into interpretations. Ultimately, machine learning assists geoscientists' work, but cannot eliminate uncertainty - only address it.

Conclusions

RF is a robust, although computationally expensive method that offers high accuracy, but this method is sensitive to input variables, labels, parametrization, and proportions of data. We used synthetic seismic data generated from an outcrop-derived 3D model (ground truth data) to eliminate interpreter errors in defining seismic facies (labels) used in training a random forest facies classifier. Eliminating such errors provided insight into misclassification due to the selection of hyperparameters and the choice of seismic attributes used in the classification. The use of hyperparameter optimizers such as the grid search allows for an increase in the classification accuracy which solves for the epistemic uncertainty, however, variables such as noise and data that may be imbalanced and aleatoric uncertainty are intrinsic to the data and may recover reprocessing. Nonetheless, the availability of the complete answer/ground truth model in this study offers a unique opportunity to understand the level of misclassification between facies. RF offers a way to understand seismic attribute importance. In this aspect, an evaluation of input data that better fits the initial facies interpretation is paramount to optimize predictions using RF. The RF algorithm was effective since tested accuracy values were high and channel facies were visibly distinguished. For the validation process, visually the expected locations of facies within the channels were predicted, although the overall performance score decreased. Overall

shale/non-reservoir facies were distinguished from the channel/reservoir facies. Misclassifications occurred mostly from channel facies being classified as shale, which is attributed to the imbalance in data proportion. We attribute misclassifications to possibly: 1) overfitting (training accuracy is high and validation error is low), 2) limited resolution of the seismic data that results in stratigraphic mixing where thinner facies are overpowered by nearby thicker facies, 3) imbalanced data (we have more shale than any other facies) so the model will prefer to pick shale over other label/ facies. This imbalance can be solved by resampling the data for training, 4) RF does not work well with sparse data, so in edges where channel margin or off axis facies are present these were most likely to be misclassified, 5) Need of more hyperparameter tuning or the use of alternate features (attributes).

When the model was applied to a validation dataset (an unseen piece of seismic data), the accuracy decreased. This speaks about the uncertainty we will have in finding the right facies if we apply a RF model that learned from the training dataset but that fails in accommodating properly to a new set of facies that are subject to different configuration and properties (epistemic uncertainty), and the different quality and resolution of seismic (aleatory uncertainty). Therefore, the level of certainty in the outcome of an ML model will be directly proportional to the quality of the training and input datasets.

A significant finding of the study is that although there are misclassifications, RF showed to be a powerful tool to interpret correct position and overall deepwater channel facies configuration, which reduces uncertainty, especially in uncalibrated basins. We suggest that the robustness of predictions are a function of the resolution and the signal-to-noise ratio of the seismic data, type of seismic attributes used, and the similarity in the seismic expression of

different geologic facies. Two things are paramount: an exhaustive hyperparameter tuning, and the geoscientist analysis of the predictions. We highlight the importance of the use of analogs to understand subsurface seismic data, bearing in mind that there is not a single model that fits all. Uncertainty has a tremendous impact on decision making which can represent not only economical but human losses if decisions are not taken appropriately and accurately, so creating awareness on the potential misinterpretations associated with these uncertainties remains necessary.

Acknowledgments

The authors thank the AASPI research group for their software and support, and SLB for the Petrel license donated to the University of Oklahoma. Similarly, we thank the Chile Slope Systems Consortium for providing the data used. Thanks to Dr. Devegowda for the foundation to perform the coding and to Marcus Maas for the insights provided.

Data and materials availability

Synthetic data, model, and python code developed for this study are available and can be shared via request to the corresponding author and subsequent approval.

SPECIAL SECTION: WORKING WITH IMBALANCED DATA IN MACHINE LEARNING ALGORITHMS: BRIDGING THE GAP BETWEEN REAL AND OPTIMAL IN SEISMIC FACIES INTERPRETATION *

*This section will be sent to AAPG explorer-non peer reviewed, informative digital journal- at the end of 2023 (an initial draft was already sent for evaluation, and it does not intend to follow a science chapter format, but rather a quick review of the unbalanced data problem):

La Marca, K., and H. Bedle, 2023. Working with imbalanced data in machine learning algorithms: bridging the gap between real and optimal in seismic facies interpretation

While scientists and workforces are becoming increasingly comfortable with the term 'machine learning', and the reluctance to employ these methods is gradually fading, there are still many uncertainties regarding their correct application and understanding of the output. Therefore, there is a need to explain what might be occurring within the so-called 'black boxes' and what could be causing high errors and misclassifications. If not well understood, these issues can lead to incorrect interpretations and economic losses.

We would dare to estimate that 80% of any machine learning method, especially unsupervised ones, relies on proper input data preparation. There is a concept that 'garbage-in'-results in 'garbage-ou'. Thus, we not only need to understand the kind of data we have but also how we can optimize it. The amount of data, their relationships, and quality are just a few aspects we need to consider in this regard. We will focus on the first aspect here, as disparities in data (imbalance) can lead to significant errors.

In reality, most geological settings we study tend to be imbalanced. Consider a deepwater setting, where channel complexes are often surrounded by shales in a major proportion, or where salt tectonics dominate, and salt bodies represent the negative class. Since a perfect dataset exists only in a utopian world, we need to understand our data and how to optimize results without biasing or overfitting our models.

The term 'imbalance' in data refers to the differences in proportions between classes. In this context, the class with the majority of records or instances will be called the 'negative class,' while the underrepresented or minority class will be named the 'positive class.'

In reality, most geological settings we study tend to be imbalanced. Consider a deepwater setting, where channel complexes are often surrounded by shales in a major proportion, or where salt tectonics dominate, and salt bodies represent the negative class. Since a perfect dataset exists only in a utopian world, we need to understand our data and how to optimize results without biasing or overfitting our models.

To understand the impact of class imbalance, we need to first comprehend: 1) the degree of imbalance between classes and 2) the overlap between classes. The degree of imbalance (also known as Imblance Ratio or IR) is calculated by relating the total number of negative class examples to the number of positive class examples. Figure 4.1 provides an example of how shale facies make up the majority of a training dataset used to predict deepwater channel facies. In this scenario, five facies are used as labels. Relating background 'shale' to any of the other facies results in doubling or tripling them, creating an imbalanced dataset.

Another consideration is the type of machine learning technique we intend to use. Supervised methods rely on error metrics because both the input and output are known. In the case of unsupervised ML techniques, only the input is known, and the selection of the correct number of clusters is still under debate. Therefore, we will provide general guidance based on our experience and common practices regarding what to do with each type of method for seismic facies interpretation.

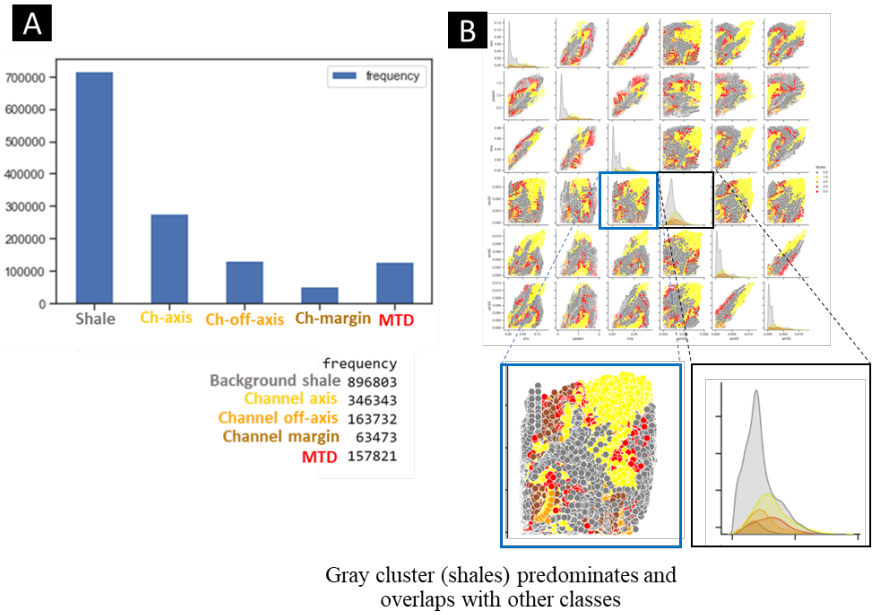


Figure 4.1. (A) samples per facies histogram in a training dataset showing an imbalance between predominant shale facies and minoratary channel facies. (B) Feature relation vis scatterplot showing that gray (shales) predominates and in some cases its range is wide so it overlaps with other classes.

However, when we sum up all the channel facies and MTD, they will be almost proportional to the background shale, suggesting that we should consider simplifying our labels.

Unsupervised methods and imbalanced data

When using clustering algorithms, such as K-means, SOM, and GTM , it is suggested to use seismic attributes that are suitable for the geological target. Analyze their statistical relationships, and refine them if necessary using a dimensionality reduction technique (e.g., PCA, ICA, Shap values). Additionally, consider using a method such as an elbow plot to determine the optimal number of clusters. An example is shown in Figure 4.2, which depicts results and error using a ML technique (such as GTM) with 3 clusters where 5 clusters were expected. Notice how error is reduced by using a optimal cluster number determined by an elbow plot. This suggests that altough we intend to depict 5 facies, the data patterns end up forming 3 major clusters.

Interpreting these visually, we can see that one corresponds to the majoritary shale facies, another represents the MTDs, and the third cluster represents the channel facies (axis, off-axis and margin) combined. It is important to remember that while we aim to identify all discrete facies, the properties of the seismic data (frequency, noise, resolution etc.) and the type of seismic attributes used, as well as their parameterizations, play a fundamental role in the detailed or non- definition of different seismic facies.

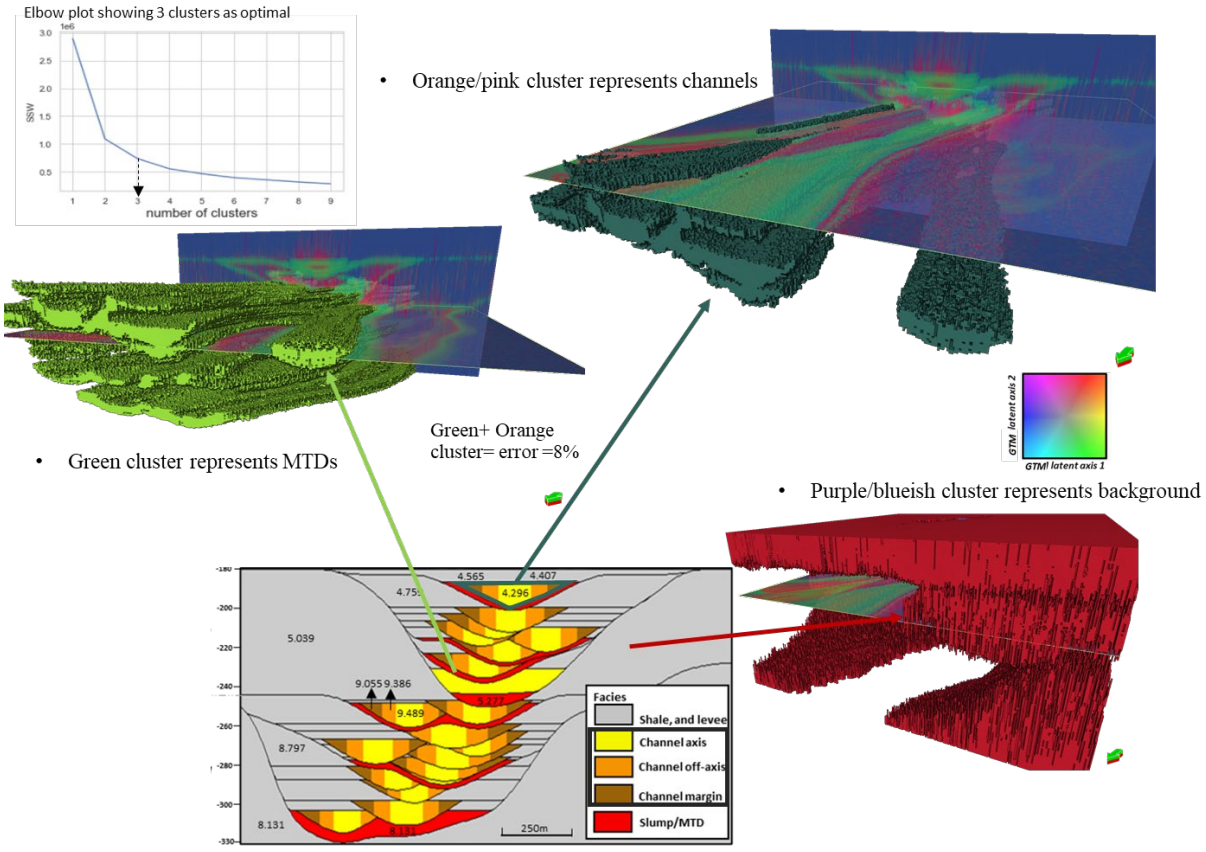


Figure 4.2. Elbow plot indicates that the optimal number of clusters is 3, and GTM results reveal MTDs in a light green geobody (GTM green), Shale in red geobody (Purple/blue in GTM), and channel facies combined in dark green geobody (orange in GTM). This demonstrates that not every cluster has to represent a single facies. The elbow plot is a valuable tool for estimating the optimal number of clusters. The amount of clusters may be related to the quality and resolution of the data and distintive patterns found.

Supervised methods and imbalanced datasets

Supervised ML needs an input (labels+features) to train the model, allowing us to know a priori the expected facies and enabling numerical error estimation or performance evaluation.

Figure 4.2 displays a distribution of a training data. By contrasting seismic attributes used as features, it becomes apparent that there is an overlap between shale and most of the channel facies. This overlap will result in missclassifications of these facies with shale. One potential solution is to explore other features that create greater distinction between classes or, to perform a resampling of the data.

Some of the methods we can employ to address imbalanced datasets include:

a) The use of **simple algorithms** such as DBscan (e.g. Piegari et al, 2008) or K-means with realistic labels or minimum classes. Another option is the use of **boosting algorithms** (such as random forest) which assign different weights to the training distribution in each iteration. After each iteration, boosting increases the weights associated with the incorrectly classified examples and decreases the weights associated with the correctly classified examples.

b) **Oversampling or undersampling**: oversampling refers to appending to the original data set, while undersampling involves the removal of data from the original dataset, typically from the majority or negative set, to achieve the same proportion or balance. However, this method could introduce its own set of problematic consequences, which can potentially hinder learning, as mentioned by He and Garcia (2009).

c) The use of **correct statistical metrics**: A confusion matrix is a popular tool for understanding and evaluating classification problems. It compares actual or original vs

predicted values. The main diagonal shows the samples that were classified correctly, while the other fields help us understand where and to what extent missclassifications occur. While accuracy is normally the metric evaluated in confusion matrices, it places more weight on common classes than on rare classes. This can make it challenging for a classifier to perform well on rare classes when evaluating imbalanced datasets, as indicated by Chawla (2009). In such cases, it is recommended to use the F score, which is a weighted harmonic mean between precision and recall. There is also a G score that, instead of a harmonic mean, uses a geometric mean.

d) ROC (Receiver Operating Characteristic) curve and AUC: ROC is used to analyze classifier performance by comparing the False Positive Rate (FPR) on the x-axis vs. the True Positive Rate on the y-axis. The closer the curve is to the upper-left corner, the better the classifier is. By calculating AUC (Area Under the Curve), we can obtain a score for the classifier. A higher AUC score, closer to 1, indicates a good classifier with a top-left ROC curve. AUC lower than 0.5 can be considered as indicating a poor classifier. It's important to note that these methods are designed for binary problems and not multi-class scenarios, which are common in seismic facies. When dealing with multi-class scenarios, we may need to evaluate ROC or AUC per class, and it becomes sensitive to class skew, as the negative class would be a combination of N-1 classes (He and Garcia, 009)."

To summarize the recommendations here provided, we have created an easy-to-follow workflow (Figure 4.3) that could be helpful for interpreters who are beginners in dealing with imbalanced data.

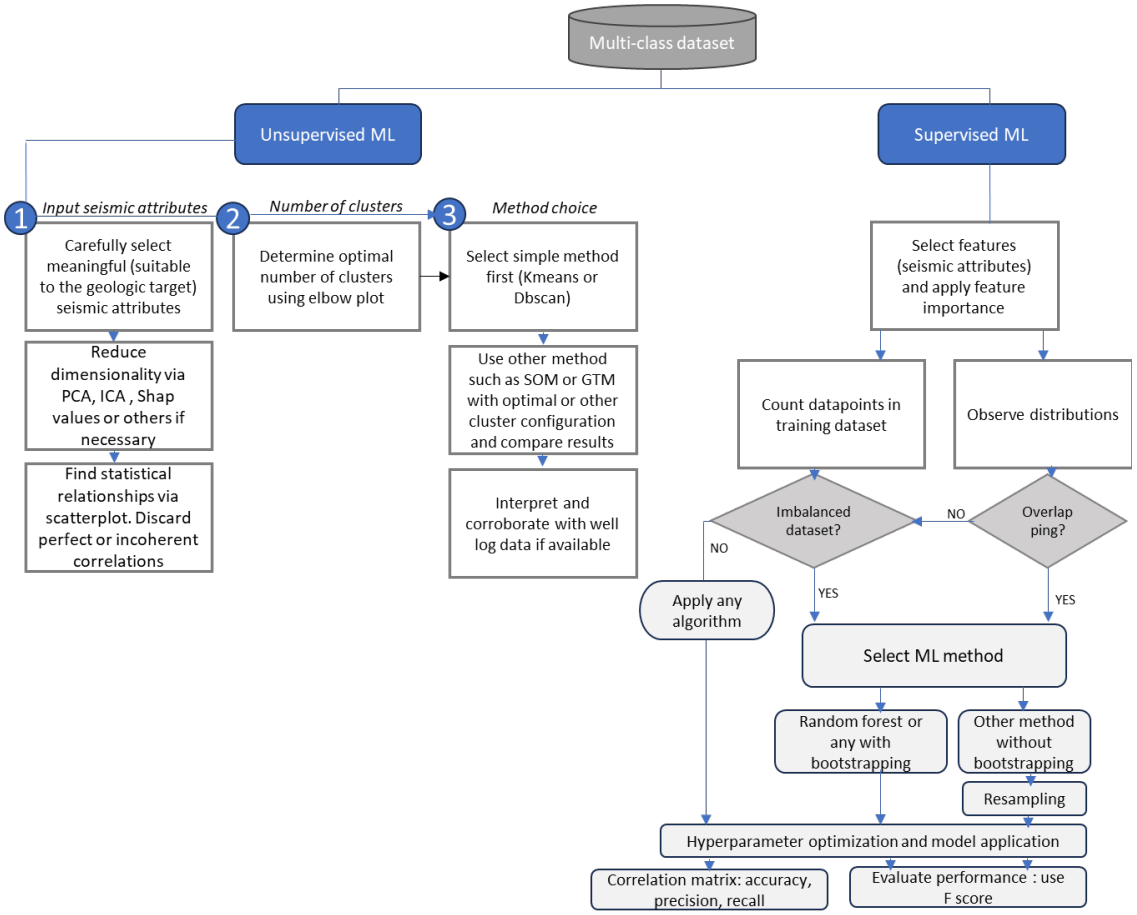


Figure 4.3. Workflow to optimize machine learning results for facies interpretation in the face of inbalanced datasets

No dataset will be ever perfect and you should be suspicious of very accurate models that will probably reflect bias or overfitting problems. Also, not every imbalanced dataset will necessarily result in poor training data. As geoscientists, we need to use the tools such as seismic attributes and ML techniques intelligently, be aware of our data's limitations, apply best practices, including parameter optimization, and, most importantly, recognize that in geology, nothing is perfect. Our understanding of the geological context and subsequent inferences still need to be carried out. We believe that machines can't fully replace us, at least not yet.

CHAPTER 5: CONCLUSIONS

In this dissertation the overarching theme is the application of synthetic seismic data and machine learning techniques to understand and address uncertainty in the interpretation of deepwater geological features, particularly channel facies. Each chapter addressed provides valuable insights and highlights the challenges and opportunities in this field.

Chapter 2 focused on the sensitivity analysis of 3D synthetic seismic volumes, emphasizing the importance of various parameters such as frequency content, attribute selection, analysis window size, and noise content. The study introduced the concept of "stratigraphic mixing" and demonstrates its impact on geological feature interpretation, with implications for economic decisions in exploration and reserves estimation. The text also discussed the sensitivity of different architectural elements to seismic attributes, providing a guide for best practices in seismic interpretation.

Chapter 3 introduced a new approach to assess uncertainty in unsupervised machine learning methods (SOM and GTM) using synthetic seismic data. It addressed the challenge of classifying facies in imbalanced datasets and the importance of using metrics like F1 score. I presented the limitations and uncertainties associated with unsupervised methods and recommended supervised methods preferable, or clustering techniques if well data is not available. In the latter, the use of meaningful and optimized attributes as input is paramount. The study highlighted the need to address uncertainty in machine learning methods for building more accurate 3D reservoir models, which are critical for the energy industry.

Chapter 4 explored the use of synthetic seismic data to evaluate uncertainty in supervised machine learning methods, specifically random forest (RF). It discussed the challenges of data

imbalance, noise, and the effect of seismic resolution on classification accuracy. The text underscores the importance of optimizing input data and the potential for misclassification when applying models to new datasets (proved by validation). It acknowledged the inherent uncertainties associated with data quality and seismic resolution (known as aleatoric uncertainty) and stresses the role of geoscientist analysis in reducing uncertainty and giving sense to the results regardless of the method employed.

To better explain the implications of imbalance datasets, which will likely be the case in a deepwater geological setting where usually shale predominates over other facies, I addressed this topic briefly in the section presented after Chapter 4, explaining what it is, the implications and a guide to handle it based on literature review.

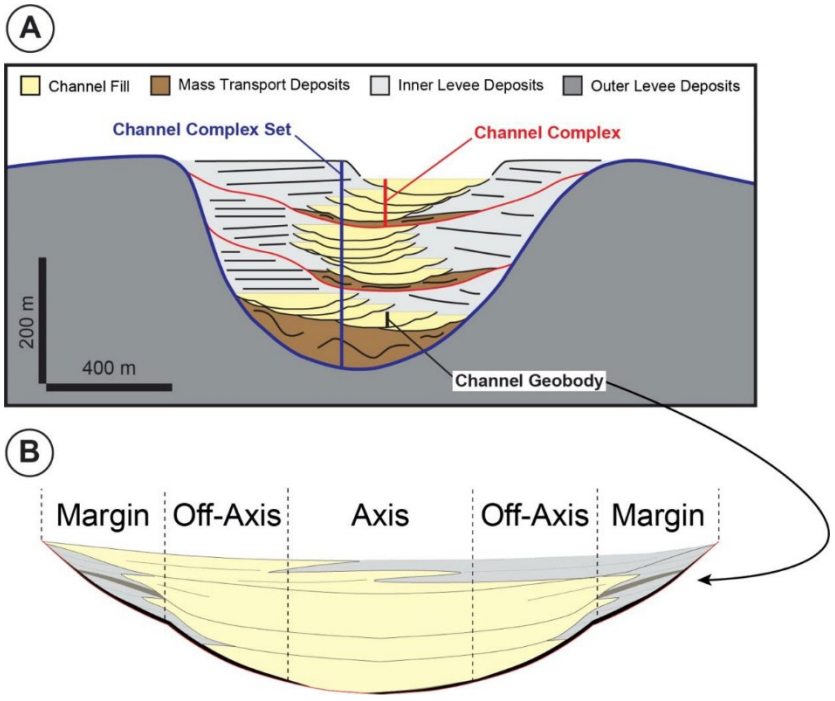
This dissertation has the significance of using synthetic data and machine learning methods to address uncertainty in deepwater geological interpretation. Therefore, a guide to optimize seismic attributes and ML methods to aim this is provided throughout the text. It is also underscored the economic and decision-making implications of uncertainty in the energy industry and advocate for best practices, hyperparameter tuning, and geoscientist expertise to reduce uncertainty and improve the accuracy of geological interpretations. Overall, these studies provide valuable insights for geoscientists and researchers working in the field of seismic interpretation, deepwater geological analysis and reservoir characterization.

APPENDIX

APPENDIX A: specs for each synthetic seismic volume

Dominant Frequency (Hz)	Length (ms)	Sample rate (ms)	Low-cut Frequency (Hz)	Low-Pass Frequency (Hz)	High-Pass Frequency (Hz)	High-cut Frequency (Hz)
15	200	1	1	3	23	35
30	200	1	2	6	45	70
60	100	1	4	12	90	140
90	100	1	6	18	135	210
180	26	1	10	30	225	350

APPENDIX B: detailed explanation of channel architectural facies (source: Jackson et al., 2019)



REFERENCES

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, l., Ghavamzadeh, M., Fieguth, P., Cao, X.,Khosravi, A., Acharya, U., Makarenkov, V., and Nahavandi, S, 2021, A review of uncertainty quantification in deep learning: Techniques, applications and challenges: Information Fusion, **76**, 243-297.

Alderton, D., and S. Elias, 2021. Encyclopedia of Geology. Second edition. Elsevier ISBN 978-0-08-102909-1

An, L., and C. Yu, 2010, Self-organizing maps for competitive technical intelligence analysis: International journal of computer information systems and industrial management applications. **4**. 83-91.

Ao, Y., L. Hongqi ., L. Zhu., S. Ali., and Z. Yang, 2019, Identifying channel sand-body from multiple seismic attributes with an improved random forest algorithm: Journal of Petroleum Science and Engineering, **173**, 781-792.

Asan, U., and S. Ercan, 2012, An Introduction to Self-Organizing Maps, *in* Kahraman, C. (eds) Computational Intelligence Systems in Industrial Engineering. Atlantis Computational Intelligence Systems, **6**. Atlantis Press, Paris.

Ballinas, M R., H. Bedle., and Devegowda, D, 2023, Supervised machine learning for discriminating fluid saturation and presence in subsurface reservoirs: Journal of Applied Geophysics, **217**, Elsevier.

Barnes, A., 2016, Handbook of Poststack Seismic Attributes: Geophysical References Series, **21**: The Society of Exploration Geophysicists, 254 p

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

Bartkowiak, A, 2004, Visualizing large data by the SOM and GTM methods - what are we obtaining? : Intelligent Information Processing and Web Mining, Proceedings of the International IIS: IIPWM'04 Conference .

Bhattacharya, S., T.R. Carr., and M. Pal., 2016, Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA: Journal of Natural Gas Science and Engineering, **33**, Elsevier.

Bishop, M., M. Svensen., and C. Williams, 1996, GTM: A principled alternative to the Self-Organizing Map: Advances in Neural information processing systems **9**, NIPS. Conference paper

Bishop, M., M. Svensen., and C. Williams, 1998, GTM: The generative Topographic mapping. Neural Computation. **10** (1): 215-234.

Brasileiro Troccoli, E., A. Guerra., J. Brian., and M. Holz, 2022, K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis: Journal of applied geophysics, **198**. Elsevier.

Breiman, L., 2001b, Random forests: Machine Learning, **45**, 5–32.

Breiman, L, 1996. Bagging Predictors: Mach. Learn. **24** (2), 123–140. https://doi.org/ 10.1007/BF00058655. August.

Cardona. S., L. Wood., D. Brandon., Z. J., and Strachan. L, 2020. Characterization of the Rapanui mass-transport deposit and the basal shear zone: Mount Messenger Formation, Taranaki Basin, New Zealand: Sedimentology, **67**, 2111-2148.

Chao. S., L. Hongyang., Z. Zhou., L. Xiao., and L. Weibin, 2022, Inverse design of aerodynamic configuration using generative topographic mapping: Journal of Northwestern Polytechnical University, **40** (4): 837-844.

Chawla, N, 2009, Data mining for imbalanced datasets: An overview: Data mining and knowledge discovery handbook, pp. 875-886, 2009

Chen, Y., M. Zhang., M. Bai., and W. Chen, 2019. Improving the signal-to-noise ratio of seismological datasets by unsupervised machine learning: Seismological research letters. **90**, 4. Doi: 10.1785/0220190028.

Chopra, S., and K.J. Marfurt, 2020. Adopting multispectral dip component for coherence and curvature computations: The leading edge. **39** (8). 593-596, doi: 10.1190/tle39080593.1.

Chopra S., and K.J. Marfurt, 2007. Seismic attributes for prospect identification and reservoir characterization. Society of Exploration Geophysicists.

Chowdhury. MZI., and T. Turin, 2020, Variable selection strategies and its importance in clinical prediction modelling: Fam Med Com Health, **8**.

Clawson, S., H.Meng., M.Sonnenfeld.,M. Uland., S.Atan., M.Batzle, M.Gardner., and M. Uman, 2003, The value of 3D seismic attributes for illuminating deep water deposits by seismic forward modeling of the Brushy Canyon Formation: SEG technical program expanded abstracts. 2431-2434.

Coleman Jr., F. C. Sheppard., and T. K. Jones, 2000. Seismic resolution of submarine channel architecture as indicated by outcrop analogs, *in* A. H. Bouma and C. G. Stone., Fine-grained turbidite systems, AAPG Memoir 72/SEPM Special Publication **68**, p. 119–126.

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

Der Kiureghian A., and O. Ditlevsen, 2009, Aleatory or epistemic? does it matter?: Structural Safety 31:105–112. Elsevier.

Enwenode, O., 2014. Understanding Noise in Seismic Record: Seismic Data Analysis Techniques in Hydrocarbon Exploration. 63-68. Doi: 10.1016/B978-0-12-420023-4.00004-6.

Fashagba, I., P. Enikanselu., L. Ademola., and M. Olabode, 2020. Seismic reflection pattern and attribute analysis as a tool for defining reservoir architecture in 'Sabalo' field, deepwater Niger Delta: Journal of Petroleum Exploration and Production Technology. Vol **10**. Issue 3. Springer. https://doi.org/10.1007/s13202-019-00807-1

Fedorova, E.V., N.V. Bukhanov., and V.E. Baranov, 2016. Synthetic Seismic Models Construction for Detailed Geological Outcrop. 7th EAGE Saint Petersburg International Conference and Exhibition. DOI: https://doi.org/10.3997/2214-4609.201600270

Ferreira, S.L.C., R.E. Bruns., H.S. Ferreira., G.D. Matos., J.M. David., G.C. Brandao., E.G.P. da Silva., L.A. Portugal., P.S. dos Reis., A.S. Souza., and W.N.L. dos Santos, 2007. Box-Behnken design: an alternative for the optimization of analytical methods: Analytica chimica Acta 597, 179-186.

Fildani, A., S. M. Hubbard., J.A. Covault., K.L. Maier., B.W. Romans., M. Traer., and J.C. Rowland, 2013. Erosion at inception of deep-sea channels: Marine and Petroleum Geology, **41**, pp.48-61.

Fletcher, S., 2013. Stratigraphic characterization of a Cretaceous slope channel complex in the Tres Pasos Formation, Arroyo Picana-Laguna Figueroa outcrop belt, Chilean Patagonia: MS Thesis, University of Calgary.

Gholamy, A., V. Kreinovich., and O. Kosheleva, 2018, Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation: Departmential technical reports. Computer science. UTEP.

Gibson, P.B., S.E.Perkins-Kirkpatrick., P. Uotila., A.S.Pepler., and L.V.Alexander, 2017, On the use of self-organizing maps for studying climate extremes: Geophys. Res. Atomos., **122**, 3891-3903.

Hanafizadeh, P., and M. Mirzazadeh, 2010, Visualizing market segmentation using self-organizing maps and fuzzy delphi method- ADSL market of a telecommunication company. Expert systems with applications, **38**, Elsevier. 198-205

Hartigan, H, 1975, Clustering Algorithms. John Wiley & Sons.

He, H., and E.A. Garcia, 2009, Learning from imbalanced data: IEEE Transactions on Knowledge & Data Engineering, vol. **9**, pp.1263-1284.

Herron, D, 2011. First steps in seismic interpretation: Geophysical monograph series, **16**. SEG.

Ho, T. K, 1995, Random decision forests: Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, Canada, (IEEE), 278–282.

Hora S, 1996, Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management: Reliability Engineering and System Safety, **54**(2–3):217–223. Elsevier.

Horvath, D., G. Marcou., and A. Varnek., 2019, Generative topographic mapping in drug design: Drug Discovery Today: Technologies, **32–33**, pp 99-107.

Hossain, S, 2020. Application of Seismic Attribute Analysis in Fluvial Seismic Geomorphology: Journal of Petroleum Exploration and Production Technology **10**, no. 3: 1009–19. https://doi.org/10.1007/s13202-019-00809-z.

Hubbard, S. M., Fildani, A., Romans, B. W., Covault, J. A., & McHargue, T. R, 2010, High-relief slope clinoform development: Insights from outcrop, magallanes Basin, Chile. Journal of Sedimentary Research, **80**(5–6), 357–375.

Hubbard, S. M., Covault, J. A., Fildani, A., & Romans, B. W, 2014, Sediment transfer and deposition in slope channels: Deciphering the record of enigmatic deep-sea processes from outcrop. Bulletin of the Geological Society of America, **126**(5–6), 857–871.

Hüllermeier, E.,and W. Waegeman, 2019, Aleatoric and epistemic uncertainty in machine learning: A tutorial introduction, arXiv preprint arXiv:1910.09457.

Infante-Paez, L., and K.J. Marfurt, 2019, Using machine learning as an aid to seismic geomorphology, which attributes are the best input? Interpretation, **7** (3), SE1-SE18.

Jackson, A., L. Stright., S.M. Hubbard., and B.W. Romans, 2019. Static connectivity of stacked deepwater channel elements constrained by high-resolution digital outcrop models: AAPG Bulletin, **103**(12), 2943–2973, doi: 10.1306/03061917346

Jeni, L., C. Jeffrey., and F de La Torre, 2013, Facing Imbalanced Data- Recommendations for the use of performance metrics: International conference on affective computing and intelligent interaction.

Kim, Y., R. Hardisty., E. Torres., and K. Marfurt, 2018, Seismic facies classification using random forest algorithm: SEG 88th annual meeting expanded abstracts, 2161-2165.

Kohonen, T, 1982, Self-organized formation of topologically correct feature maps. Biological Cybernetics ,**43**, 59-69.

Kohonen, T, 1995, Self-Organizing Maps. Berlin: Springer-Verlag.

Kuhn, M., and K. Johnson, 2013, Applied predictive modeling: Springer

Kumar D., and I. Ahmed, 2021. Seismic Noise. *in*: Gupta H.K. Encyclopedia of Solid Earth Geophysics. Encyclopedia of Earth Sciences Series. Springer, Cham. https://doi.org/10.1007/978-3-030-58631-7_146

Kus, M., and Z. Jobe., F. Laugier., W. Walker., and M. Sullivan, 2021, Quantifying lateral heterogeneity of distant submarine lobe deposits, Point Loma formation, California: Implications for subsurface lateral facies prediction: The depositional record. **8**, Issue 2.

La Marca, K., H. Bedle., L. Stright., and K.J. Marfurt, 2023, Sensitivity analysis of seismic attributes parametrization to reduce misinterpretations: Applications to deepwater channel complexes: Marine and Petroleum Geology, **153**, 106309, Elsevier.

La Marca, K. and H. Bedle, 2022. Deepwater seismic facies and architectural element interpretation aided with unsupervised machine learning techniques: Taranaki basin, New Zealand. *Marine and Petroleum Geology*, **136**, p.105427.

La Marca, K., H. Bedle., L. Stright., R. Pires de Lima., and K. J. Marfurt, 2022, Quantifying uncertainty in unsupervised machine learning methods for seismic facies using outcrop-derived 3D models and synthetic seismic data. In SEG International Exposition and Annual Meeting . SEG. Expanded abstracts, 1312-1316.

La Marca, K., K.J. Marfurt, H. Bedle, L. Stright, and T. Langenkamp, 2021, Sensitivity analysis of seismic attributes parametrization for interpretation of a multi-story deepwater channel system: Tres Pasos Formation, Magallanes Basin Chile: First International Meeting for Applied Geoscience &Energy, Expanded Abstracts, 1191–1195

La Marca, K., 2020. Seismic attribute optimization with unsupervised machine learning techniques for deepwater seismic facies interpretation: Users vs machines: M.S. thesis, The University of Oklahoma

Langenkamp, T., L. Stright., H. Bedle., M. Hubbard., and B. Romans, 2020. Quantifying heterogeneity preserved in inverted seismic data: implications for reservoir prediction and geostatistical modeling: presented at the ACE 2020 annual meeting, AAPG.

Langenkamp, T, 2021. Evaluating the impact of deepwater channel architecture on the probability of correct facies classification using 3D synthetic seismic data. MS thesis. Colorado State University.

Lin, T., J. Guo., B. Zhang., K.J. Marfurt., and D. Chang, 2014. Seismic Attributes Estimation Using a Self-Adaptive Window: SEG Technical Program Expanded Abstracts 2014, 1654–58. Denver, Colorado: Society of Exploration Geophysicists.

Liu, J., and K.J. Marfurt, 2007, Instantaneous spectral attributes to detect channels: GEOPHYSICS. **72**, 2.

Lubo-Robles, D., K. Marfurt, 2019, Unsupervised seismic facies classification using ICA: Interpretation, **7**, 3, Society of Exploration Geophysicists.

Luo Y, Al-Dossary S, Marhoon M, Alfaraj M, 2003, Generalized Hilbert transform and its applications in geophysics: The leading Edge.

Luo, Y., W. G. Higgs, and W. S. Kowalik, 1996, Edge detection and stratigraphic analysis using 3D seismic data: 66th Annual International Meeting, SEG, Expanded Abstracts, 324–327.

Luque, A., A. Carrasco., A. Martin and A. De las Heras, 2019, The impact of class imbalance in classification performance metrics based on the binary confusion matrix: Pattern recognition. Elsevier.

Lyu, B., J. Qi, F. Li., Y. Hu, T. Zhao, S. Verma, and K.J. Marfurt, 2020, Multispectral coherence: which decomposition should we use?: Interpretation. **8**(1), T115- T129.

Macauley, R. V., and S.M. Hubbard, 2013. Slope channel sedimentary processes and stratigraphic stacking, Cretaceous Tres Pasos Formation slope system, Chilean Patagonia: *Marine and Petroleum Geology***, 41**(1), 146–162, doi: 10.1016/j.marpetgeo.2012.02.004

Macqueen, J., 1967, Some methods for classification and analysis of multivariate observations, in proceedings of the 5th Berkeley Sympossium on mathematical statistics and probability, 1: statistics, 281-297.

Manning, C.D., 2009, An introduction to information retrieval: Cambridge university press

Marfurt, K., and T. Alves, 2015, Pitfalls and limitations in seismic attribute interpretation of tectonic features: Interpretation, **3**, SEG

Marfurt, K. J., R. L. Kirlin, S. H. Farmer, and M. S. Bahorich, 1998. 3D seismic attributes using a running window semblance-based algorithm: *Geophysics*, **63**, 1150–1165, doi: 10.1190/1.1444415

Mathieu, M, 2017, Unsupervised learning under uncertainty: PhD. Dissertation, New York University.

McCulloch. W; and W. Pitts, 1943, A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics. **5** (4): 115–133.

McGregor. L., L. Scott., R. Cooper., and J. Nichols, 2022. Ultra-High resolution seismic: applications of P-Cable in the energy transition: First Break, **40**, 67-70.

McHargue, T., Pyrcz, M. J., Sullivan, M., Clark, J. D., Fildani, A., Levy, M., … Covault, J. A, 2011. Event-based modeling of turbidite channel fill, channel stacking pattern, and net sand volume. *Outcrops Revitalized: Tools, techniques and applications*, **10**, 163-173

Meek. T, 2015. Applications of 3D seismic attribute analysis workflows: a study from Ness County, Kansas, USA. Ms. C. thesis, Kansas State University.

Meirovitz, C. D., L. Stright., S.M. Hubbard., and B.W. Romans, 2020. The Influence of Inter- and Intra-channel Architecture on Deep-water Turbidite Reservoir Performance: *Petroleum Geoscience*, **27**, doi: 10.1144/petgeo2020-005.

Mesafint D., and H. Manjaiah D, 2021, Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results: International Journal of Computers and Applications.

Michie, E. A. H., Mulrooney, M. J., and Braathen, A, 2021, Fault interpretation uncertainties using seismic data, and the effects on fault seal analysis: a case study from the Horda Platform, with implications for CO2 storage: Solid Earth, **12**, 1259–1286.

Nagl, M., M. Nagl, and D. Risch, 2005, Quantifying uncertainty of machine learning methods for loss given default: Frontiers in applied mathematics and statistics, Sec. mathematical finance, **8**-2022.

Nayaran, S., S. Konka., A. Chandra., K. Abdelrahman., P. Andras., and A.M. Eldosouky, 2023, Accuracy assessment of various supervised machine learning algorithms in litho-facies classification from seismic data in the Penobscot field, Scotian basin: Frontiers in earth science. Frontiers.

Nielson, A., 2018. Using RMS amplitudes from forward seismic-reflectivity modeling of channelized deepwater slope deposits to inform stratigraphic interpretation and sub-

seismic scale architecture, Tres Pasos Formation, Magallanes Basin, Patagonia, Chile: MS Thesis, Colorado State University.

Ning, C., and You. F, 2019, Optimization under uncertainty in the era of big data and deep learning: when machine learning meets mathematical programming: Computers and chemical engineering, **125**. Elsevier. 434-448

Pampalk, E, 2001, Limitations of the SOM and the GTM: project work for the course "Connectionism Laboratory" Department of medical cybernetics and Artificial Intelligence, University of Viena.

Partyka G., J. Gridley., and M. Lopez, 1999, Interpretational applications of spectral

Patro, S., and K. Kumar Sahu, 2015, Normalization: A preprocessing stage: arXiv preprint arXiv:1503.06462

decomposition in reservoir characterization: The Leading Edge,**18**, 3. SEG.

Pearson, K, 1901, On lines and planes of closest fit to systems of points in space: The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **2**:11, 559-572.

Pemberton, E., L. Stright., S. Fletcher., and S. Hubbard, 2018. The influence of stratigraphic architecture on seismic response: reflectivity modeling of outcropping deepwater channel units: Interpretation, **6** (3), T783–T808. doi: 10.1190/INT-2017-0170.1

Phan, Buu Truong, 2019, Bayesian deep learning and uncertainty in computer vision. Ms. thesis. University of Waterloo.

Pickering, K.T., Corregidor, J., Clark, J.D., 2015. Architecture and stacking patterns of lower-slope and proximal basin-floor channelized submarine fans, Middle Eocene Ainsa System, Spanish Pyrenees: an integrated outcrop-subsurface study. Earth-sci. Rev. 144, 4-81.

Piegari, E., M. Hermann., and W. Marzocchi, 2022, 3-D spatial cluster analysis of seismic sequences through density-based algorithms, Geophysical journal international, **230**, 2073-2088

Posamentier, H.W., and V. Kolla, 2003. Seismic geomorphology and stratigraphy of depositional elements in deepwater settings: Journal of Sedimentary Research **7**3, 367-388.

Posamentier, H.W., Davies, R.J., Cartwright, J.A., Wood, L., 2007, Seismic geomorphology-an overview. Geolo. Soc. London, Special public. **277** (1), 1-14.

Prather, B.E., 2003. Controls on reservoir distribution, architecture, and stratigraphic trapping in slope settings. Marine and Petroleum Geology ,**20**, 529-545.

Rahimi, M., M. Ali Riahi, 2022, Reservoir facies classification based on random forest and geostatistic methods in an offshore oilfield: Journal of applied geophysics, **201**, Elsevier.

Raju, V.N.G., K. P. Lakshmi., V. M. Jain., A. Kalidindi., and V. Padma, 2020, Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification, in: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, pp. 729– 735.

Ray, T., and S. Verma, 2020, Seismic facies classification using generative topographic mapping- a case study from offshore nova scotia: SPE conference.

Roden, R., T. Smith., and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing map: Interpretation, 3, SAE59–SAE83.

Roshan, V, 2022, Optimal ratio for data splitting: Statistical analysis and data mining, **15**, Wiley.

Roy, A., 2013, Latent space classification of seismic facies: Ph.D. thesis, University of Oklahoma.

Roy, A., A. Romero., T. Kwiatkowski., and K, Marfurt, 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico

Ruetten, A., 2021. Evaluating the impact of hierarchical deepwater slope channel architecture on fluid flow behavior, Cretaceous Tres Pasos Formation, Chile MS Thesis, Colorado State University.

Sacrey, D., and R. Roden., 2014, Understanding attributes and their use in the application of neural analysis- case histories both conventional and unconventional: Search and discovery article #41473.

Sanger, T., 1989, Optimal Unsupervised learning in a single-layer linear feedforward Neural Network: Neural Networks, **2**, pp 459-473. Pergamon press.

Slatt, R., 2013, Stratigraphic reservoir characterization for petroleum geologists, geophysicists, and engineers. Origin, recognition, initiation, and reservoir quality: Developments in petroleum science, **61**. Second edition. Elsevier.

Slatt, R., E. Eslinger., V. Dyke., and K. Staffan, 2009. Acoustic and petrophysical properties of a clastic deepwater depositional system from lithofacies to architectural elements' scales: GEOPHYSICS, **74**. 2. Doi: 10.1190/1.3073760.

Southern, S.J., Stright, L., Jobe, Z.R., Romans, B., and Hubbard, S., 2017. The stratigraphic expression of slope channel evolution: insights from qualitative and quantitative assessment of channel fills from the Cretaceous Tres Pasos Formation, southern Chile: AAPG Annual Convention, Houston, TX, April 3-5, 2017.

Stracuzzi, D., G. Maximillian., M. Chen., M. Peterson and C. Vollmer, 2017, Uncertainty quantification for Machine Learning. Sandia report SND2017-6776. Sandia National Laboratories.

Stright, L., Stewart, J., Campion, K. and Graham, S., 2014. Geologic and seismic modeling of a coarse-grained deepwater channel reservoir analog (Black's Beach, La Jolla, California) Seismic Modeling of a Deep-Water Reservoir Outcrop Analog, California: AAPG bulletin, **98**(4), pp.695-728.

Subrahmanyam, P., and P.H. Rao, 2008. Seismic attributes- a review: 7th. International conference and exposition on petroleum geophysics.

Sun, Y., 2000, On quantization error of self-organizing map Network: Neurocomputing

Schwab, A. M., B.T. Cronin., and H. Ferreira, 2007. Seismic expression of channel outcrops: Offset stacked versus amalgamated channel systems: Marine and Petroleum Geology. **24**, 6-9, 504-514. Doi: 10.1016/j.marpetgeo.2006.10.009.

Taner, M.T., F. Koehler, and E.E. Sheriff, 1979, Complex seismic trace analysis: Geophysics, **44**, 1041-1063.

Tanimoto, T., K. Heki., and J. Artru-Lambin, 2015. Interaction of Solid Earth, Atmosphere, and Ionosphere. Treatise on Geophysics: Second Edition. **4**. 421-443. Doi: 10.1016/B978-0-444-53802-4.00083-X

Tuckova J., M. Bartu., P. Zetocha., and P. Grill, 2011, Self-organizing maps in medical applications. In proceedings of the international conference on neural computation theory and application.**1**. Special session of challenges in Neuroengineering, 422-429.

Verma S., A. Roy, R. Perez, and K.J. Marfurt, 2012, Mapping high frackability and high TOC zones in the Barnett Shale: Supervised Probabilistic Neural Networks vs. unsupervised multi-attribute Kohonen SOM: 82nd SEG Annual Meeting, 1-5.

Wallet, B. C., M.C.de Matos., J.T. Kwiatkowski., and Y. Suarez, 2009, Latent space modeling of seismic data: An overview: The Leading Edge, **28**(12), 1454-1459.

Weber, K, 1993, The use of 3-D seismic in reservoir geological modeling: in The Geological Modelling of Hydrocarbon reservoirs and outcrop analogues. Special publication International Association of sedimentologists, **15**, 181-188. In

Zhao, T., V. Jayaram., A. Roy., and K.J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation. **3**, no. 4, SAE29-SAE58.

Zhao, T., J. Zhang., F. Li., and K.J. Marfurt, 2016, Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps: Interpretation, **4**, no.1, SB79–SB89.