University of Oklahoma

Graduate College

UNSUPERVISED SEISMIC FACIES CLASSIFICATION IN A DEEPWATER CHANNEL SYSTEM: DIMENSIONALITY REDUCTION, CLUSTER VALIDATION, AND STRATIGRAPHIC FRAMEWORK DEVELOPMENT FOR MACHINE LEARNING INTERPRETABILITY.

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

April DeAnn Moreno-Ward

Norman, Oklahoma

2026

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/342417

UNSUPERVISED SEISMIC FACIES CLASSIFICATION IN A DEEPWATER CHANNEL SYSTEM: DIMENSIONALITY REDUCTION, CLUSTER VALIDATION, AND STRATIGRAPHIC FRAMEWORK DEVELOPMENT FOR MACHINE LEARNING INTERPRETABILITY.

A DISSERTATION APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Deepak Devegowda

Dr. Matthew J. Pranter

Dr. Hao Hu

© Copyright by APRIL DEANN MORENO-WARD 2026

All Rights Reserved.

The path through a PhD looks different from the inside than it does from the outside. From the inside, it can feel like a tangled mess; lost, blocked, and wondering if everyone else has an understanding you will never reach. Most of us feel exactly that way, even if nobody says so. The obstacles are real. The confusion is real. But so is the solution you haven't found yet, the creative path around the block, the moment something finally clicks. Finishing this dissertation does not mark the end of my learning or my self-discovery; it simply showed me that I am capable of both. If you are just beginning or lost somewhere in the middle, know that the tangled mess you are living in right now is not a sign that you cannot do it. This is what doing it looks like.

Acknowledgements

This work would not have been possible without the support of many people who have helped shape me, challenged me, and accompanied me through this journey.

To my husband and son, Dustin and Grey Ward, you have been on this journey with me every step of the way. I know my attention was often divided over these last three years, and I am profoundly grateful for your patience, understanding, and your willingness to share in this pursuit. Your support has meant the world to me.

To my parents, Adan and Cipriana Moreno, the resilience you instilled in me gave me the courage to never give up on my goal of pursuing and earning a PhD. I am deeply grateful for the person you have helped me become.

My growth as a geoscientist owes much to mentors who invested in me at critical moments. Dr. John Holbrook took me under his wing from the beginning of my geologic discovery and mentored me through my master's degree. Gary King, Bill Manthey, and Gene Behrens, my first industry mentors, helped me to think like an applied scientist and pushed me to develop my skills in visual communication. Their honest, constructive feedback was exactly what I needed, and it was largely through these friendships that I began to see myself as a geoscientist.

To my dissertation committee, Drs. Matt Pranter, Deepak Devegowda, and Hao Hu, I am grateful for your time and energy helping me to understand the scope of this project and giving me the critical feedback I needed and the support necessary to make it through this process. Dr. Kurt Marfurt, whom I had the good fortune of knowing for nearly a decade before completing this PhD, has been a consistently encouraging and intellectually curious presence; it has been a true pleasure to work with him. Dr. David Lubo-Robles has served as an exceptional model of what a research career can look like, from PhD student to postdoc to research scientist within this group; he brings a depth of knowledge in combining machine learning and geophysics that has been invaluable.

And to my advisor, **Dr. Heather Bedle, the right mentor at the right time**. She understood how I think and process information and gave me the space to do so while also holding me to a standard that pushed me well beyond my comfort zone. I truly believe that her mentorship was pivotal in my ability to successfully complete this program.

Finally, to my extended family and friends, your love, support, and belief in me have been a blessing. And the broader geoscience community I have had the privilege of knowing over the past many years; your warmth, encouragement, and sense of community have played a larger role in shaping me, both professionally and personally, than you may realize. There are far too many of you to name, but please know that each of you has played a part in getting me here. Thank you.

Table of Contents

List of Figures xiii
Chapter 2: xiii
APPENDIX 2.A: GEOLOGIC, ATTRIBUTES, AND SEISMIC DATA xvi
APPENDIX 2.B: SCATTERPLOTS AND CLUSTER MODELS xvi
Chapter 3: xvii
APPENDIX 3.A: Data Tables xx
APPENDIX 3.B: Histograms xxi
APPENDIX 3.C: Clustered Models xxii
Chapter 4: xxiii
Chapter 5: xxvi
Chapter 6 xxix
List of Tables xxxiii
Abstract xxxiv
Chapter 1: Introduction 1
Motivation 1
Research Question 2
The algorithm answers what you ask, not what you mean 2
Dissertation Structure 3
References 5
Chapter 2: Improving Interpretability: The Role of Dimension Reduction in Multiattribute Seismic Facies Identification* 7
ABSTRACT 7
INTRODUCTION 8
Geologic background 11
Data 14
Attributes for channels, an overview 15
Dimension Reduction 16
Dimension Reduction techniques 17
Research Objective 19
DATA PREPARATION METHOD 20
Step 1a: Data preprocessing 21
Step 1b: Well tie for seismic correlation 22
Step 1c: Lithology and facies identification 22
Step 1d: Seismic attribute selection 24
Step 2: Data Preparation for Machine Learning 26
RESULTS & DISCUSSION 28
Attributes for architectural elements 28
Attribute selection: 28

Seismic Amplitude and Lithologic Indicators 32
Differential Compaction as an Indicator 33
Channel Complex Morphological Evolution 33
DR for interpretability of geologic features 36
DR for architectural elements 37
CONCLUSION 41
ACKNOWLEDGEMENTS 42
REFERENCES 43
Chapter 3: enhancing multiattribute seismic facies analysis: the effects of dimension reduction on unsupervised clustering* 51
ABSTRACT 51
INTRODUCTION 52
Geologic Background 52
Data 53
Machine learning in seismic workflows 55
The algorithms, a quick review 55
Attributes for channels: an overview 57
ML Considerations in Geoscience 59
METHODOLOGY 60
Rationale for k-means 61
Silhouette Score and Davies-Bouldin Index (DBI) 63
Overview 63
RESULTS AND DISCUSSION 65
Quantitative Evaluation (Statistical) 65
DR Algorithm measures explained (Statistical) 68
Quantitative Higher Performers 69
Qualitative Evaluation (Visual) 72
Qualitative Dimensional Perspective 72
Quantitative vs. Qualitative Summary 74
Clustering without DR 78
Clustering with DR 78
Post-DR Cluster Results Summary 80
Rationale for post-DR performance overview 82
CONCLUSION 85
ACKNOWLEDGEMENTS 86
REFERENCES 87
APPENDIX A: Data Tables 93
APPENDIX B: Histograms 100
APPENDIX C: Clustered Models 107

Chapter 4: Beyond Amplitude Alone: A Corendering Technique for Enhanced Deepwater Channel Visualization – a Tutorial Part 1: Systematic Optimization Using Cosine Instantaneous Phase and Relative Acoustic Impedance* 114

ABSTRACT 114
INTRODUCTION 114

GEOLOGIC SETTING 116
METHODS: HOW WE OPTIMIZED THE DISPLAY 119
STEP ONE: ATTRIBUTE SELECTION 121
PHASE 123
STEP 2: CORENDERING THE COSINE OF INSTANTANEOUS PHASE AND RELATIVE ACOUSTIC IMPEDANCE 127
STEP 3: THE CHOICE OF COLOR BARS 127
STEP 4: INTERPOLATION TO A FINER SAMPLE INCREMENT 128
STEP 5: SPECTRAL BALANCING 130

RESULTS: WHAT THE OPTIMIZED DISPLAY HELPS REVEAL 132
CONCLUSIONS 137
ACKNOWLEDGEMENTS 139
REFERENCES 140

Chapter 5: Beyond Visualization Alone: A Systematic Interpretation Tutorial for Deepwater Channel Stratigraphy Identification Part 2: Framework application and Romney 3D Case Study* 146

ABSTRACT 146
INTRODUCTION 147
METHODS: HOW TO APPROACH INTERPRETATION OF CHANNEL STRATIGRAPHY 150
Foundation: Understanding Key Seismic Attributes 150
The Interpretation Workflow: Integrating Geological and Geophysical Perspectives 151

Initial Target Definition and Attribute Application 151
Well Calibration and Data Integration 152
Integration Phase: Where Geological Thinking Takes Center Stage 153
Critical Iteration and Validation 158

The Geological Mindset: Beyond Pattern Recognition 159

IMPROVED FEATURE VISIBILITY AND IDENTIFYING AES 163
MEASURED GEOMETRIES AND RELATIONSHIPS 164

RESULTS: WHAT IS THE GEOLOGIC MEANING? 165

IDENTIFICATION OF DEEPWATER CHANNEL ARCHITECTURAL ELEMENTS 165
MIGRATION AND FLOW OF DEEPWATER CHANNEL SYSTEMS 174

CONCLUSIONS 175
ACKNOWLEDGEMENTS 176
REFERENCES 177
Appendix 4A: Channel Architectural elements and controls 179

Chapter 6: FROM CLUSTERS TO FACIES: INTEGRATING DOMAIN KNOWLEDGE INTO ML-BASED SEISMIC INTERPRETATION IN DEEPWATER CHANNELS* 186
ABSTRACT 186
1. INTRODUCTION 186
2. Consideration for incorporating ML 188
2.1 How to approach including ML 188
2.2 Attributes: single attribute, multiattribute, ML? 191
2.3 Facies Calibration Strategy 194
2.4 GEOLOGICAL AND DATA CONTEXT 197
2.4.1 Study Area and Depositional Setting 197
2.4.2 Available Data 198
3. METHODOLOGY 200
3.1 Seismic Attribute Analysis and Dimension Reduction 200
3.2 Unsupervised Clustering and Model Generation 200
3.3 Integrated Model Evaluation Framework 201
3.4 Cluster Normalization and 3D Visualization 204
4. RESULTS: SEISMIC FACIES ARCHITECTURE 204
4.1 Response-Type Facies: Seismic Facies Characteristics 204
4.2 Stratigraphic Interval Facies 206
4.3 Architectural Element Mapping 210
4.4 Direct Model Comparison 213
5 Discussion: Seismic Facies as a foundation for understanding the geologic setting 215
5.1 The Value of Integrated Model Evaluation 215
5.2 The irreplaceable role of domain knowledge and geologic reasoning: 215
Managing interpreter bias 216
Practical Recommendations 216
5.3 Uncertainties and Path Forward 217
5.4 Workflow Transferability 218
6.CONCLUSIONS 220
ACKNOWLEDGEMENTS 221
REFERENCES 222
Chapter 7: Conclusion 225
Key findings: 226
Dimensionality Reduction Improves Interpretability 226
Dimensionality Reduction Improves Clustering performance: 226
Geologic Foundation Must precede ML evaluation: 227
Answering the Research Question: 228
That framework consists of several critical components: 228

Implications for Practice 229
Limitations and future work 229
Final reflections 230

Appendix 1: Enhancing seismic delineation of obscured geologic architectural elements in a deepwater channel complex through multi-attribute analysis: A study from the Taranaki Basin* 232

Summary: 232
Geologic Location: 233
Motivation: 233
Methods: 234

Model Selection and Visualization: 236

Conclusion: 239

Appendix 2: Finding the balance: Human insight vs. machine learning techniques to optimize deepwater channel facies identification* 243

Summary 243
Rationale for Machine Learning: 243
Geologic Location: 244
Workflow: 244

Where Do We Start: 245
Importance of Attribute Selection: 245
Understanding our Algorithms 246
Qualitative Evaluation (QLE): 246
Quantitative Evaluation (QE): 247
Clustering– pre and post-DR 247

Conclusion 247
References: 249

Appendix 3: The Algorithm Doesn't Know Geology — A Practitioner's Guide for ML-Assisted Seismic Facies Interpretation* 253

Summary: 253
Introduction: 254
Input/Output Evaluation for Geologic Implementation: 254
Geologic Location: 257
Workflow: 258
Defining Facies: 258
Conclusion 261
References: 263

Appendix 4: Fences and Fenceposts - Choosing the right attributes for the application of machine learning in seismic interpretation* 266

Introduction: 266
On the Usefulness of CSP Here 267

CSP versus Variance 269
Algorithm Selection..... 271
Application: 272
Acknowledgements..... 274

List of Figures

Chapter 2:

Figure 2.1: a) Showing the attribute sweetness flattened above channel of interest. This provides additional resolution to map out channel migration and timing. b) Inline 1900, X-line 4850, and Z-slice — 2800 ms. With amplitude. C) Inline view at 1840, X-line 5330 – 5210, with the Romney-1 lithology log on flattened sweetness. D) Inline 1900, from X-line 4970 – 3890, zoomed to the cross-sectional view of the channel would be in part b.

Figure 2. 2: A) regional study location showing major faults, tectonic plates, and topographic features. B) Romney 3D seismic survey location in white box; overlaid to show incised valleys off platform.

Figure 2. 3: Portion of the Mississippi River: yellow indicating scroll bars, blue channel and channel extent outlines, pink channel migration, and green arrow indicating general direction of channel/bar migration. Topographic maps from the National Map OnDemand Topo site; Wickliffe SW, Arlington, Wolf Island, and Oakton 7.5 minute quadrangle maps west of Clinton, KY

Figure 2. 4: A) mapped surface incised by study channel at approximately — 3.008 s. B) overlain with seismic showing channel complex features.

Figure 2. 5: Project workflow. 1) Initial step includes pre-processing prior to advanced analysis. Followed by 2) statistical analysis, 3) data visualization, and lastly, 4) analysis and interpretation.

Figure 2. 6: a) Well data with 5 main units identified and labeled as well facies. Seismic data at the well evaluated to identify seismic response corresponding to identified well facies; labeled as

seismic facies. b) Well and seismic facies extrapolated into the channel of interest. Of the 5 identified facies, only 4 are found in the channel.

Figure 2. 7: Unsupervised clustering of cosine instantaneous phase, envelope, GLCM dissimilarity, k1 and k2 curvature, and sweetness before DR, 2 clusters. A) Scatterplot with cosine phase (column indicated by red arrow) shows a distinct separation in data; this trend is shown clearly in the histogram. Other attribute scatterplots show little to no cluster separation; the corresponding histograms show definitive overlap of data. B) Elbow method indicating 13 clusters and C) silhouette score indicating 4 clusters an optimal number.

Figure 2. 8: Using inline 1842, x-line 4200, and z-slice — 3204, image with A) uninterpreted and B) interpreted selected attributes indicating interpretability of larger scale channel features. Yellow arrows indicate areas with coarser grained material and blue arrows indicate areas with finer grained material. Pink arrows indicating well defined channel edges. Co-rendering in B) viii and ix further enhance channel edges.

Figure 2. 9: Inline 1840. a) amplitude b) sweetness c) envelope d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) k1 overlain on cosine instantaneous phase i) k2 overlain on cosine instantaneous phase. Same image on right, yellow arrows indicating younger cutoff channel with a straighter morphology. Blue arrows indicate older channel edge.

Figure 2. 10: a) Extrapolating well and seismic facies into the channel for lithology indicators. b) Using a combined x-line, inline, and z slice view, lithology indicators can be applied. Channel migration indicator outlines can begin as a starting stage to overall geologic interpretation. Comparison with inline view will help determine which channels are younger. Light pink arrow indicating differential compaction. Purple arrow: as the top edge of the cutoff loop is interrupted, we can infer that the cutoff loop is older than the portion of the channel outlined in green.

Figure 2. 11: Images show z-slice at — 3260 ms. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. While each attribute can highlight geologic features, individually, they will not be able to highlight all detail. i) Channel flow direction indicated with pink arrow; channel migration and vertical progradational history show channel straightening at yellow arrow.

Figure 2. 12: Images show z-slice at — 3236 ms. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. While each attribute can highlight geologic features, individually, they will not be able to highlight all detail. i) Channel flow direction indicated with pink arrow; channel migration and vertical progradational history can be estimated once channel morphology is mapped.

Figure 2. 13: Z-slice — 3236, Inline 1902, X-line 4280. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. Individually each attribute may highlight apparent geologic features, however, in conjunction they tend to draw out more detail. i) shows the progression of the channel and migration direction. Internal channel complex architecture shows upward changes in channel morphology.

Figure 2. 14: Yellow arrow indicating younger cutoff channel and blue arrow indicating channel edge migration. Orange arrow highlights deviation from other models in highlighting channel edge morphology; this deviation could affect channel architecture interpretation and is worthy of note. While all models delineate main channel scours, K-PCA and UMAP highlight smaller-scale features better. a) ICA, 2 components b) ICA, 3 components c) PCA, 2 components d) PCA, 3 components e) K-PCA, 2 components f) K-PCA, 3 components g) UMAP, 2 components h) UMAP, 3 components

Figure 2. 15: A) Time 3.2s, blue arrow indicating splay location, pink arrow showing variation in delineating smaller scale channel migration, and green arrow indicating scroll bar. While all models delineate main channel features, K-PCA and UMAP highlight smaller-scale features better. i) ICA, 2 components ii) ICA, 3 components iii) PCA, 2 components iv) PCA, 3 components v) K-PCA, 2 components vi) K-PCA, 3 components vii) UMAP, 2 components viii) UMAP, 3 components. 1 indicates black lines outline main channel boundaries, 2 indicates the cutoff channel, and 3 indicates the scroll bar.

B) Time 2.8s, blue arrows indicating internal scroll bar and pink arrows indicating splay. While all models delineate main channel features, K-PCA and UMAP highlight smaller-scale features better. i) ICA, 2 components ii) ICA, 3 components iii) PCA, 2 components iv) PCA, 3 components v) K-PCA, 2 components vi) K-PCA, 3 components vii) UMAP, 2 components viii) UMAP, 3 components

APPENDIX 2.A: GEOLOGIC, ATTRIBUTES, AND SEISMIC DATA

Table 2.A: Brief overview of attribute taxonomy. For a more in-depth discussion of attribute calculations and taxonomy see: Subrahmanyam and Rao 2008; Dewett et al. 2021; Chopra and Marfurt 2005; Taner et al. 2001 and 1994; Liner 2016; Marfurt 2018, Brown 1996, Barnes 1997; and Carter 1995

Figure 2.A: Summary of logs available for the Romney-1

APPENDIX 2.B: SCATTERPLOTS AND CLUSTER MODELS

Figure 2.A: Initial scatterplot for data distribution analysis. Yellow circles indicate possible outliers; however, these data points could represent a unique geologic feature. Caution should be used when handling seemingly outlying data.

Figure 2.B: ranking if sorted on silhouette score

Figure 2.C: ranking if sorted on mean distance to cluster centers

Chapter 3:

Figure 3.1: Romney 3D seismic survey location. A) Area of study located in deepwater off the northwest shelf of New Zealand's north island. With main faults for reference. Light pink rectangle, zoom to: B) black rectangle indicating Romney 3D seismic volume extents. C) A white circle indicates the location of the Romney-1 well and a dark pink polygon indicates cropped volume extents. D) Using Romney-1 well log, well facies are identified, E) and are used to identify seismic facies. F) Next, the seismic facies are interpolated into the channel of interest.

Figure 3.2: A) Attributes selected for ML are displayed in a cross-sectional view to visualize their ability for channel feature interpretation. a) amplitude b) sweetness c) envelope d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity. h) cosine instantaneous phase with k1 curvature and i) cosine instantaneous phase with k2 curvature demonstrate co-rendering attributes to improve interpretation. B) Table for a summary of attribute use and reference for more in-depth review.

Figure 3.3: Cluster analysis and interpretation workflow.

Figure 3.4: Unsupervised clustering (2 clusters) of sweetness, cosine instantaneous phase, envelope, GLCM dissimilarity, and k1 and k2 curvature before DR. Scatterplots with cosine phase (column indicated by red arrow) show distinct separation in data; this trend is shown clearly in the histogram. Other attribute plots show little to no cluster separation. Histograms

indicate a similar distribution of cluster data. The plot for sweetness and envelope may indicate related data; however, they are not correlative, rather, they plot with similar trends. Sweetness is defined as the envelope (or instantaneous amplitude) ÷ √instantaneous frequency, while envelope is a measure of the magnitude of the amplitude. Using data trends alone could result in excluding data that may appear to be redundant.

Figure 3.5: Workflow for quantitative evaluation. A) Evaluation for pre-DR clustering and B) Evaluation for post-DR clustering.

Figure 3.6: DR algorithms explained. Each algorithm uses the same set of input seismic attributes for reducing dimensionality within the dataset; however, the evaluation of attributes is not the same across all models. Evaluators should understand how each algorithm weights the inputs for the proper determination of attribute importance.

Figure 3.7: Using the silhouette and DBI score for all cluster models and both 2-and-3-components, we evaluate the top 10 scorers. #1 indicates the highest score overall. A pink star indicates models identified as silhouette high performers, and a green star indicates DBI high performers within the post-DR group. Red arrows indicate data overlap. 1) K-PCA 2 components, 8 clusters. 2) K-PCA optimized 2 components, 8 clusters. 3) K-PCA 2 components, 5 clusters. 4) K-PCA optimized 2 components, 5 clusters. 5) K-PCA 2 components, 4 clusters. 6) K-PCA optimized 2 components, 4 clusters. 7) K-PCA 2 components, 3 clusters. 8) K-PCA optimized 2 components, 3 clusters. 9) K-PCA 2 components 2 clusters. 10) UMAP optimized 3 components, 8 clusters.

Figure 3.8: Using visual inspection, the top 10 performers across all models are displayed here. A blue star indicates models previously identified as visually having unique clusters, a pink star indicates models identified as silhouette high performers, and a green star indicates DBI high performers within the post-DR group. High performers were chosen from within the DR group.

Red arrows indicate data overlap. 1) UMAP optimized 2 components, 5 clusters. 2) UMAP 2 components, 8 clusters 3) K-PCA optimized 2 components, 5 clusters. 4) UMAP 2 components, 4 clusters. 5) K-PCA 2 components, 8 clusters. 6) UMAP optimized 3 components, 8 clusters. 7) K-PCA 2 components, 8 clusters. 8) UMAP 2 components, 5 clusters. 9) K-PCA 2 components 5 clusters. 10) K-PCA optimized 3 components, 5 clusters.

Figure 3.9: UMAP 3 components, A) without optimized hyperparameters and B) with optimized hyperparameters. In A), if we visualize the clusters in 2D only, there appears to be an overlap between the green and orange clusters (red arrow), however, if we render this model in 3D we see there is in fact no overlap in the data. B) With the optimized hyperparameters, there is no indication of cluster overlap in either the 2D or 3D view and the clusters have more separation between them.

Figure 3.10: K-PCA 3 components, a) without optimized hyperparameters and b) with optimized hyperparameters. In both examples, when viewed as a 2D graph, we note areas where clusters appear to be overlapping (red arrows), however, when viewed as a 3D graph we begin to see that clusters may not in fact be overlapping. With this in mind, the 3D view in a) still has data closely spaced (pink arrows), and determining actual overlapping would take significantly longer for the human eye than the statistical evaluation from ML. UMAP 3 components c) without optimized hyperparameters and d) with optimized hyperparameters. c) When viewed as a 2D graph, we note areas where clusters appear to be overlapping (red arrows), however, when viewed as a 3D graph we note good cluster separation (black arrows). d) With optimized hyperparameters, the clusters appear to have both good cohesion and separation with minimal appearance of perceived overlap indicated by red arrows. Areas of good separation are noted with black arrows and well-separated clusters with a black dashed polygon.

Figure 3.11: Summary of all post-DR 2 components. A) 4 clusters B) 5 clusters. Areas with better cluster separation are indicated by black arrows and areas with more overlap are indicated

by a red arrow. The dashed-black polygon shows well-separated and easy-to-identify clusters and open brackets around the cluster with wide-spread.

Figure 3.12: Summary of all post-DR 3 components. A) 4 clusters B) 5 clusters. Areas with better cluster separation are indicated by black arrows and areas with more overlap are indicated by red arrows. Those clusters that are unique and show good separation from surrounding clusters are indicated by a black dashed polygon. Dashed-black polygon showing well separated and easy to identify cluster and open bracket around cluster with wide-spread.

APPENDIX 3.A: Data Tables

Appendix 3.A Table A: Romney 3D seismic survey data and Romney-1 well data

Appendix 3.A Figure A: DR algorithms with 2- and 3-component data input weights for evaluation. PCA evaluates input data with feature importance and explained variance %. The first 3 principal components capture 92% of the variance. K-PCA evaluates percent contribution and component magnitude. The component(s) magnitude is relative to each other with higher values indicating more importance. UMAP evaluates the feature importance (FI), and the authors have included the permutation feature importance (PI). While the FI may indicate one attribute has high importance to the projection magnitude, the PI indicates which data have a significant impact on the structure of the projection.

Appendix 3.A Figure B: Data distribution comparison for seismic attributes used for ML workflow. Columns are as follows: 1. original attribute data, 2. data after min-max scaling, 3. data after min-max scaling and outlier removal from GLCM, 4. data after robust scaling, and 5. data after robust scaling and outlier removal from GLM. Data distribution remains unchanged across columns, except for a minimal change for GLCM with outlier removal.

Appendix 3.A Table B: All models ranked (grouped by clustering method), highest to lowest, by silhouette score (left) and DBI (right). A pink cell indicates the highest performer per clustering method, where silhouette and DBI agree. A light orange cell indicates a high silhouette performer, and a green cell indicates a high DBI performer, per each clustering method.

Appendix 3.A Table C: All models ranked, highest to lowest, by silhouette score (left) and DBI (right). A pink cell indicates the highest performer per clustering method, where silhouette and DBI agree. A light orange cell indicates a high silhouette performer, and a green cell indicates a high DBI performer, per each clustering method. Color coding is held from the previous figure.

Appendix 3.A Table D: Statistics for chosen attributes before scaling, after scaling, after scaling, and outlier removal. Scaling conducting using min-max.

Appendix 3.A Table E: Statistics for chosen attributes before scaling, after scaling, after scaling, and outlier removal. Scaling was conducted using both min-max and robust scaling for comparison.

APPENDIX 3.B: Histograms

Appendix 3.B Figure A: Data distribution prior to scaling

Appendix 3.B Figure B: Data distribution after min-max scaling

Appendix 3.B Figure C: Data distribution after min-max scaling and outlier removal, from GLCM.

Appendix 3.B Figure D: Data distribution after robust scaling

Appendix 3.B Figure E: Data distribution after robust scaling and outlier removal, from GLCM

Appendix 3.B Figure F: Top- shows all models and the number of clusters graphed using the silhouette score to provide a visual for which DR technique performs better for cluster performance. Bottom- shows all models and the number of clusters graphed using the DBI to provide a visual for which DR technique performs better for cluster performance.

APPENDIX 3.C: Clustered Models

Appendix 3.C Figure A: 3D perspective for K-PCA with 3 components. Areas with easy-to-recognize cluster separation are indicated with black arrows. Areas where cluster spacing is still difficult to determine are indicated with pink arrows. Well-defined clusters are indicated with black dashed polygons.

Appendix 3.C Figure B: 3D perspective for K-PCA with 3 components and optimized hyperparameters. Areas with easy-to-recognize cluster separation are indicated with black arrows. Clusters with widespread indicated with black open brackets.

Appendix 3.C Figure C: 3D perspective for UMAP with 3 components. Areas with easy-to-recognize cluster separation are indicated with black arrows. Clusters with widespread indicated with black open brackets

Appendix 3.C Figure D: 3D perspective for UMAP with 3 components and optimized hyperparameters. Areas with easy-to-recognize cluster separation are indicated with black

arrows. Clusters with widespread indicated with black open brackets. Well-defined clusters are indicated with black dashed polygons.

Appendix 3.C Figure E: All post-DR models, 2 components, and 8 clusters are shown here. a) PCA b) K-PCA c) UMAP. The model(s) with the more unique clusters are marked with a blue star, Silhouette high performer(s) are noted with a pink star, and DBI high performer(s) are noted with a green star (as ranked within the post-DR group). Areas with more data overlap are indicated by a red arrow. Clusters with good separation are indicated by a black arrow and well-separated clusters are noted with black-dashed polygon. Well-defined clusters are indicated with black dashed polygons.

Appendix 3.C Figure F: Cluster distribution for post-DR K-PCA and UMAP 3 components with and without optimization. Shown are 8 clusters only. A pink star indicates a high performer per silhouette score, and a green star indicates a high performer based on DBI score (as ranked within the post-DR group). Areas with more data overlap are indicated by a red arrow. Clusters with good separation are indicated by a black arrow and well-separated clusters are noted with black-dashed polygon.

Chapter 4:

Figure 4.1. (a) Area of study is off the northwest side of the north island of New Zealand, courtesy Google Earth. (b) Small white polygon indicates the location of the Romney 3D seismic survey, courtesy of The GEBCO grid. (c) Romney timeslice at -3.24 and (d) inline at 1950.

Figure 4.2. Cartoon illustrating some of the key architecture elements of the turbidite channel addressed in this paper

Figure 4.3: Workflow for improved interpretability, which consists of five major steps through an iterative process.

Figure 4.4: (a) Time slice at t=3.21 s through a coherence volume showing the edges of individual channels that comprise a turbidite system. Line AA' showing a complex channel through the (b) coherence, and the original seismic amplitude volumes in (c) plan view and (d) cross-sectional view. Although coherence does a good job of delineating channel edges when seen on the time slice in (a), it does a poor job on the vertical slices, providing less detail than the original seismic amplitude data in (d). For this example, the seismic data have a 4 ms sample increment and a 12.5 m by 25 m bin size. The coherence was computed using an analysis window containing 9 traces and 3 samples along the structural dip. Seismic data courtesy of NZP&M.

Figure 4.5: Line AA' through (a)) corendered amplitude and the cosine of instantaneous phase, cosφ (b) instantaneous phase, φ, (c) the cosine of instantaneous phase, cosφ,, and (d) relative acoustic impedance corendered with the cosine of instantaneous phase, cosφ. Corendering the cosine of instantaneous phase with the original seismic amplitude failed to better delineate stratigraphic boundaries. The color bar used for the cosφ overlay in (c) displays only the strongest values with white for cosφ > +0.8 and black for cosφ < -0.8. (d) In contrast, corendering cosφ with RAI resampled to 1 ms after spectral balancing, better delineates stratigraphic boundaries.

Figure 4.6. (a) The major components of the color bar initially considered. Clearly, the white value of cosφ=1 cannot be seen on white, but it also shows up poorly on yellow. In contrast, the black value of cosφ=-1 shows up well on white and yellow but poorly on blue. (b) A modified color bar, which we will use in the next figures to display relative acoustic impedance that allows corendering of cosφ. The triplets indicate the RGB values of each color.

Figure 4.7. (a) Line AA' through the relative acoustic impedance (RAI) volume, which rotates the seismic amplitude data by -90° and increases the low-frequency component of the data. (b) RAI corendered with cosφ. This image shows that the white cosφ≈+1 anomalies show up well against the dirty yellow high positive RAI response, and the black cosφ≈-1 anomalies show up well against the light blue moderately negative RAI anomalies. These two attributes were computed from the original 4 ms sample increment data. (c) Line AA' but now through RAI and cosφ volumes computed from data that have been interpolated to a sample increment of 1 ms. For a period of 50 Hz, there are only six samples per cycle, such that cosφ is under-sampled at the 4 ms sample increment in (b). Zoomed images corresponding to the green rectangles of (d) the 4 ms volume shown in (b), and (e) the 1 ms volume shown in (c).

Figure 4.8. Time-variant spectra of the seismic amplitude averaged over the entire survey (a) before and (b) after spectral balancing. The spectrum in (b) was subjected to a 0-4-80-100 Hz Ormsby filter to suppress high and low frequency artifacts. Notice the spectra in (b) are nearly constant (appearing as the same color) for a given time level. A single time-variant operator is applied to the entire survey, thereby preserving relative amplitude.

Figure 4.9. Line AA' through the amplitude volume (a) before and (b) after spectral balancing, where the colored arrows indicate improved vertical resolution within the turbidite system. The spectra of the two images are shown in the previous figure. Line AA' through volumes of (a) amplitude, (b) h cosφ, (c) relative acoustic impedance, and (d) relative acoustic impedance corendered with cosφ. Stratigraphic boundaries mapped guided by (e) cosφ alone, and (f) relative acoustic impedance corendered with cosφ. Using our understanding of geologic processes and the environment of deposition, we can push our interpretation further and hypothesize areas of finer- (blue arrows) and coarser-grained material (red arrows).

Figure 4.10. Line AA' through the (a) cosφ volume and the (b) corendered RAI and cosφ after spectral balancing of the 1 ms sample increment data. Arrows indicate areas with the channel that exhibit higher resolution than the comparable images in Figure 9a and 9c.

Figure 4.11: Time slice at t=3.21 s through the corendered RAI and cosφ volumes (left) sampled at a 1 ms increment after spectral balancing and through the corendered RAI and coherence volumes (right). Because the reflectors seen in Figure 9 are relatively flat, the time slice is parallel to the extrema of cosφ in many locations, resulting in the broad white and black smears between the positive and negative RAI anomalies compared to the sharp edges in coherence shown in Figure 4a.

Figure 4.12: Vertical slice through Line AA' n (a) and (d) with areas of interest marked with arrows. The pink arrow indicates an area of lower RAI, whereas the yellow arrow indicates an area of higher RAI. Because the stratigraphy is relatively flat, timeslices (b) and (c) through these same features exhibit broad cosφ anomalies. Displaying these images a chair diagrams in (c) and (f) allow us to better link the stratigraphic edges seen in the vertical section to those seen on the time slices.

Chapter 5:

Figure 5.1: Considerations of expected morphology based on more modern analogues. a) is a portion of the Mississippi River (near Ten Mile Pond Conservation Area), zoomed in to focus (b) on abandoned meander loops and subsequent architectural elements. c) and d) are samples taken from the Missouri River near Yankton, SD. These samples were taken using hand augers across the river valley to note changes in sediment deposition both vertically and laterally in response to river migration. The dashed black arrows indicate the equivalent location these sediment samples would have been found in the abandoned channel features. e) Shows a cross-section of the Romney channel, with f) showing a quick sketch of the expected grain size

Figure 5.2: Interpretation workflow for identifying stratigraphic responses using both cosφ and relative acoustic impedance. The 14-step framework separates geometry and facies identification

and then helps us combine these as we integrate through steps 8-12. If the overall interpretation is inconsistent or does not represent geologic meaning, reiteration to step 9 should be considered until the interpretation is consistent, considering geometric, lithologic, and depositional constraints

Figure 5. 3: Inline 1980 (a) amplitude, (b) instantaneous phase, and (c) cosine of instantaneous phase. Comparison of how each attribute helps delineate stratigraphic features.

Figure 5.4: Using the same 14-step workflow, interpreters should assess whether observations are consistent and geologically reasonable before proceeding. There is the risk of moving too quickly without considering alternative reasons for what is being observed.

Figure 5.5: Line AA' through (a)) corendered amplitude and the cosine of instantaneous phase, cosφ (b) instantaneous phase, φ, (c) the cosine of instantaneous phase, cosφ,, and (d) relative acoustic impedance corendered with the cosine of instantaneous phase, cosφ. Corendering the cosine of instantaneous phase with the original seismic amplitude failed to better delineate stratigraphic boundaries. The color bar used for the cosφ overlay in (c) displays only the strongest values with white for cosφ > +0.8 and black for cosφ < -0.8. (d) In contrast, corendering cosφ with RAI resampled to 1 ms after spectral balancing, better delineates stratigraphic boundaries.

Figure 5.6. (a) The major components of the color bar initially considered. Clearly, the white value of cosφ=1 cannot be seen on white, but it also shows up poorly on yellow. In contrast, the black value of cosφ=-1 shows up well on white and yellow but poorly on blue. (b) A modified color bar, which we will use in the next figures to display relative acoustic impedance that allows corendering of cosφ. The triplets indicate the RGB values of each color.

Figure 5.7. (a) Line AA' through the relative acoustic impedance (RAI) volume, which rotates the seismic amplitude data by -90° and increases the low-frequency component of the data. (b) RAI corendered with cosφ. This image shows that the white cosφ≈+1 anomalies show up well against the dirty yellow high positive RAI response, and the black cosφ≈-1 anomalies show up well against the light blue moderately negative RAI anomalies. These two attributes were computed from the original 4 ms sample increment data. (c) Line AA' but now through RAI and cosφ volumes computed from data that have been interpolated to a sample increment of 1 ms. For a period of 50 Hz, there are only six samples per cycle, such that cosφ is under-sampled at the 4 ms sample increment in (b). Zoomed images corresponding to the green rectangles of (d) the 4 ms volume shown in (b), and (e) the 1 ms volume shown in (c).

Figure 5.8. Time-variant spectra of the seismic amplitude averaged over the entire survey (a) before and (b) after spectral balancing. The spectrum in (b) was subjected to a 0-4-80-100 Hz Ormsby filter to suppress high and low frequency artifacts. Notice the spectra in (b) are nearly constant (appearing as the same color) for a given time level. A single time-variant operator is applied to the entire survey, thereby preserving relative amplitude.

Figure 5.9. Line AA' through the amplitude volume (a) before and (b) after spectral balancing, where the colored arrows indicate improved vertical resolution within the turbidite system. The spectra of the two images are shown in the previous figure. Line AA' through volumes of (a) amplitude, (b) h cosφ, (c) relative acoustic impedance, and (d) relative acoustic impedance corendered with cosφ. Stratigraphic boundaries mapped guided by (e) cosφ alone, and (f) relative acoustic impedance corendered with cosφ. Using our understanding of geologic processes and the environment of deposition, we can push our interpretation further and hypothesize areas of finer- (blue arrows) and coarser-grained material (red arrows).

Figure 5.10. Line AA' through the (a) cosφ volume and the (b) corendered RAI and cosφ after spectral balancing of the 1 ms sample increment data. Arrows indicate areas with the channel that exhibit higher resolution than the comparable images in Figure 9a and 9c.

Figure 5.11: Time slice at t=3.21 s through the corendered RAI and cosφ volumes (left) sampled at a 1 ms increment after spectral balancing and through the corendered RAI and coherence volumes (right). Because the reflectors seen in Figure 9 are relatively flat, the time slice is parallel to the extrema of cosφ in many locations, resulting in the broad white and black smears between the positive and negative RAI anomalies compared to the sharp edges in coherence shown in Figure 5.4a.

Figure 5.12: Vertical slice through Line AA' n (a) and (d) with areas of interest marked with arrows. The pink arrow indicates an area of lower RAI, whereas the yellow arrow indicates an area of higher RAI. Because the stratigraphy is relatively flat, timeslices (b) and (c) through these same features exhibit broad cosφ anomalies. Displaying these images with chair diagrams in (c) and (f) allow us to better link the stratigraphic edges seen in the vertical section to those seen on the time slices.

Chapter 6

Figure 6.1: Simplified workflow illustrating the two potential pathways for selecting how to implement ML into the seismic interpretation workflow. Path 1 utilizes all six seismic attributes as direct inputs for unsupervised ML via k-means++ clustering: A) GLCM dissimilarity, B) envelope, C) cosine instantaneous phase, D) sweetness, E) k1 curvature, and F) k2 curvature. Path 2 first employs these same six attributes as inputs for DR, where the resulting output is evaluated using qualitative methods to assess improvements in interpretability before clustering. The DR output then serves as the input for unsupervised k-means++ clustering. The statistical performance of each k-means++ output can be directly compared; however, this comparison alone does not constitute a final determination of geologic interpretability.

Figure 6.2 The six input seismic attributes, derived from amplitude, are displayed in both cross-sectional(inline 2100) and planview (timeslice 3.2s). Of the three attributes that show sensitivity to internal architectural elements within the channel complex, cosφ, sweetness, and envelope, cosφ is the only one that consistently resolves those features with sufficient clarity to support the identification and interpretation of these features. The remaining attributes capture laterally variations across the system, but may demonstrate limited diagnostic utility when evaluated independently for a full geologic history.

Figure 6.3: When evaluating this deepwater channel system, we have to acknowledge that while we do not have lithologic truth via a well, we can use other indicators to understand the channel fill. A) The major surface that the channel incised was mapped, and mappable horizon above the channel complex was also mapped. B) Show the same surfaces, but with a slightly different vantage. What we see is a confirmed “structural” high. C) The same two mapped surfaces again, but this time with an inline using amplitude shown. The interpretation for this feature is differential compaction. There is recorded claystone outside of the channel. As this channel is tied to transgressive-regressive cycles we can also infer that the material filling the channel are more coarse material coming from the shelf.

Figure 6.4: A) Romney dataset location offshore NW of New Zealand, approximately B) 60 km from the continental shelf edge and 160 km NW of the Taranaki Peninsula. C) Inline 1866 with pink polygon noted channel system A, our primary channel of interest, and green polygons indicating channel system B.

Figure 6.5: When evaluating this deepwater channel system, we have to acknowledge that while we do not have lithologic truth via a well, we can use other indicators to understand the channel fill. A) The major surface that the channel incised was mapped, and mappable horizon above the channel complex was also mapped. B) Show the same surfaces, but with a slightly different vantage. What we see is a confirmed “structural” high. C) The same two mapped surfaces again, but this time with an inline using amplitude shown. The interpretation for this feature is

differential compaction. There is recorded claystone outside of the channel. As this channel is tied to transgressive-regressive cycles we can also infer that the material filling the channel are more coarse material coming from the shelf.

Figure 6.6: Cross-section A-A' showing seismic facies evaluation. Uninterpreted and interpreted panels are shown for: amplitude(A/B); Post-K-PCA clustering model (8 clusters, 2 components) (C/D); K-PCA 2-component clustering corendered with cosφ (E/F); Post-K-PCA clustering model (8 clusters, 3 components) (G/H); K-PCA 3-component clustering corendered with cosφ (I/J). Six facies packages are identified from base to top (oldest to youngest). These packages are distinct from the originally picked seismic facies (A-C) and are here interpreted in the context of the broader depositional environment. Facies (1*) marks a basal calcareous claystone that begins an upward fining sequence into Facies (1), characterized by increased siltstone content with sandstone stringers. Facies (2) represents a major limestone marker (transgressive interval), Facies (3) calcareous claystone-dominated with minor sandstone and trace limestone, and Facies (4) silty to calcareous claystone transition. Below the mud log start depth of 2369 m (~ -2887 ms), Facies (5) is interpreted as potential gravity flow deposits, and Facies (6) as a mass transport deposit (MTD). Facies succession reflects transgressive-regressive cyclicity with episodic coarse clastic delivery to the basin floor.

Figure 6.7: Seismic facies characterization of channel architecture shown in timeslice (1, A-H) and chair view (2, A'-H'). Uninterpreted attributes include amplitude (A/A'), amplitude with bump map (B/B'), K-PCA 2 components with 8 clusters (C/C'), and K-PCA 3 components with 8 clusters (D/D'). Interpreted equivalents are shown in (E/E'), (F/F'), (G/G'), and (H/H'), respectively. The consistency of geometric features across both the attribute space and the model space demonstrates that K-PCA clustering preserves geologic structure rather than introducing artifacts. Black arrows mark the oldest channel phase, characterized by high amplitude and straight geometry, interpreted as high-energy flow with coarse-grained material along channel edges. Orange arrows indicate an older meandering channel with lower amplitude, interpreted as mud-filled with lower-energy flow. Pink arrows highlight a zone of higher-energy flow with

preserved sediment distribution shape. Yellow arrows identify preserved sediment distribution within the interchannel area.

Figure 6.8: Comparative evaluation of post-K-PCA clustering models with 8 clusters: (1) 2-component model and (2) 3-component model. Panels (A/B) show cross-sectional views of models (1) and (2), respectively, with (C/D) showing zoomed insets of the annotated features in (A/B). Panels (E/F) show timeslice views and (G/H) show chair views of models (1) and (2). Solid black circles highlight scroll bar features associated with response-type B facies; dashed polygons highlight response-type A facies characterized by parallel reflectors. Red circles identify an area above the channel of interest showing response-type B grading to type C at its base. Red dashed polygons highlight the limestone marker at Romney-1, associated with response type A. In Panels (E/F/G/H), black arrows identify an older, low-amplitude, mud-fill channel phase's clustered response; bold red arrows highlight sharp lateral discontinuities that allow for better channel boundary identification; and short red arrows in the cross-section portion of the chair view highlight discontinuous reflectors that define the channel boundary. Overall, model (1) provides more consistent lateral continuity, crisper feature delineation, and better separation of geologically meaningful cluster distributions than model (2).

List of Tables

Table 2.1: Seismic survey and well information.

Table 2.2: Overview of attributes chosen for this analysis including category and general use. For more information on each attribute, see recommended text.

Table 3.1: Summary of algorithms discussed in this manuscript.

Table 3.2: Summary of statistical scoring methods for clustering results. Common measures include the silhouette score and the Davies-Bouldin Index to validate or invalidate the silhouette score.

Table 3.3: Summary of DR algorithms discussed in this manuscript.

Table 3.4: Model label reference table

Abstract

Machine learning (ML) techniques are becoming increasingly embedded in geoscience workflows, from seismic attribute analysis to facies classification and stratigraphic interpretation. As a result, geoscientists are expected to rapidly adopt and apply these methods in their daily work while simultaneously defending their methodology and delivering geologically robust answers. The pressure to implement ML quickly has created a dangerous temptation: to assume the algorithm understands geology, and that statistical outputs are inherently defensible as geologic interpretations. They are not.

As with any transformative technology, the adoption of ML in geoscience requires more than technical proficiency. It requires a disciplined framework that preserves critical thinking and honors the fundamentals of geologic interpretation. That framework must be rooted in geologic knowledge: understanding the study area, the processes that shaped the depositional environment, and the stratigraphic and structural context within which the data exists. It must also include a working understanding of how ML algorithms function, what questions they can and cannot answer, and how inputs shape outputs. Without this foundation, practitioners risk conflating statistical similarity with geologic meaning, producing models that are mathematically sound but geologically indefensible.

This dissertation addresses that gap by developing and demonstrating a practical framework for integrating unsupervised ML into seismic facies interpretation within a geologically informed context. Using a deepwater channel system in the Taranaki Basin, New Zealand, as a case study, I evaluate how dimensional reduction (DR) and k-means ++ clustering can be strategically incorporated into a traditional seismic interpretation workflow without sacrificing geologic rigor. The research examines critical decision points in the workflow: attribute selection, choice of DR technique and clustering algorithm, statistical model evaluation, facies definition and identification, and then demonstrates how quantitative ML outputs can be balanced with qualitative geologic knowledge to produce interpretations that are both computationally efficient and geologically meaningful.

Central to this approach is recognizing that the interpreter should not assume that the best statistical model necessarily produces the best geologic model. Making that determination

requires a robust evaluation framework. This dissertation establishes that framework by treating ML as a tool rather than as a geoscientist, and by developing systematic methods for evaluating outputs against known seismic attribute characteristics, stratigraphic position, and depositional context.

The first two chapters establish the ML workflow itself, examining how DR improves interpretability in multiattribute seismic facies identification, and developing both quantitative and qualitative metrics to evaluate DR performance and clustering outcomes. These chapters demonstrate that post -DR clustering consistently outperforms pre-DR approaches, but they also reveal that statistical performance alone does not guarantee geologic validity.

Before finalizing the interpretation framework, it is necessary to step back and strengthen the geologic foundation upon which the ML models will be evaluated. I work through a systematic tutorial for enhancing deepwater channel visualization through the corendering of cosine instantaneous phase and relative acoustic impedance, moving beyond amplitude-based interpretation alone. I then extend this approach into a comprehensive interpretation framework for identifying deepwater channel stratigraphy, applied to the Romney 3D seismic volume. Together, these chapters provide critical insight into what channel architecture and stratigraphy actually mean in terms of seismic expression, ensuring that ML-derived facies models can be evaluated against a robust, defensible understanding of the depositional system. While the complete facies interpretation will follow this dissertation work, the methodology presented here establishes the interpretive rigor necessary for that final step.

Chapter 1: Introduction

Machine learning (ML) techniques are being rapidly adopted in seismic interpretation workflows (Bhattacharrya, 2021; Dramsch, 2020), driven by the promise of computational efficiency, pattern recognition at scale, and the ability to handle high-dimensional datasets that challenge traditional interpretation methods (Wang et al., 2018; Zhao et al., 2017; Marfurt, 2018; Chopra and Marfurt, 2007;). However, this rapid adoption has created a critical problem: ML outputs are often treated as geologic answers rather than a step in the interpretation workflow. The result is a growing gap between what the algorithm produces and what the geology actually represents, and what the interpreter can defensibly explain. What is missing is a logical framework for understanding how to evaluate whether ML outputs are geologically meaningful, and more fundamentally, a disciplined approach to determining when and how ML should be incorporated into interpretation workflows in the first place.

Motivation

Deepwater channel systems are characterized by complex stratigraphic architectures that present significant challenges for seismic interpretation (Posamentier and Kolla, 2003; Kolla et al., 2007). The primary geophysical approach to resolving this complexity is through seismic facies identification and classification, yet doing so effectively requires a disciplined integration of quantitative statistical methods and qualitative geologic reasoning. The tools available to the modern interpreter have grown significantly more powerful, yet the frameworks for evaluating whether their outputs are geologically sound have not kept pace. The question is not whether ML can be applied to seismic interpretation; it demonstrably can, but whether it should be applied in a given context. And if it is determined that it should be, how to implement it in a way that enhances rather than mistakenly tries to replace geologic reasoning. Just because an algorithm can partition data does not mean the partitions are geologically meaningful. The challenge, then, is twofold: to develop a framework that preserves interpretive rigor while leveraging the computational power of modern ML techniques, and to make that framework practical for

practitioners working in an industry where time is rarely a luxury and the pressure to deliver results quickly can lead to an overreliance on computational tools.

Research Question

The central research question guiding this work is: How can we better incorporate ML/AI techniques into seismic interpretation workflows to improve interpretation resolution and increase confidence in geologic interpretations and models?

This overarching question breaks down into several critical sub-questions that must be addressed systematically:

- Why implement ML? What specific interpretive challenge does it solve that conventional methods cannot address as effectively?
- What are the inputs? How do we select seismic attributes that are geologically meaningful, not just mathematically convenient for the algorithm?
- How do we evaluate the outputs? What criteria determine whether a model output is geologically important versus statistically interesting but geologically irrelevant?
- How do we define facies? The algorithm identifies patterns in attribute space, but it is the interpreter who must define what those patterns mean in terms of depositional processes, stratigraphy, and geologic history. Further complicating this process, there could be multiple ways to define and identify facies within the same area of evaluation.

The algorithm answers what you ask, not what you mean

A fundamental challenge in applying ML to geoscience is that algorithms respond to the question posed, not the question inferred. If we ask a clustering algorithm to find similarity in seismic attribute space, it will do exactly that. Identify samples with similar attribute responses and group them accordingly. However, similarity in attribute responses does not automatically imply similarity in depositional processes, lithology, or stratigraphic significance. This

distinction is critical and cannot be overlooked for an effective application of ML to interpretation.

The interpreter must bridge that gap by defining facies based on geologic processes and stratigraphic context, not simply on cluster membership. We should not forget that the algorithm is a tool to aid in delineating patterns; geologic interpretation is the responsibility of the practitioner (Bond et al., 2007; Polson and Curtis, 2010).

Dissertation Structure

This dissertation is organized into four chapters and four appendices, each addressing a specific aspect of the research question.

Chapter 2 examines how dimension reduction (DR) can improve the interpretability of multiattribute seismic data for facies identification. Working with a deepwater channel system in the Taranaki Basin, New Zealand, this chapter establishes the foundation for understanding how DR techniques can reveal geologic patterns that can be obscured when working with high-dimensional attribute space. Several DR techniques are evaluated, and the results are compared to identify which technique best improves visual results, and whether DR genuinely improves our ability to see geologically meaningful features, not just whether it reduces computational complexity.

Chapter 3 evaluates the effects of DR on unsupervised clustering performance by comparing 55 different model configurations using both quantitative (statistical) and qualitative (visual) evaluation criteria. This chapter demonstrates that post-DR clustering consistently outperforms pre-DR approaches across multiple statistical metrics. However, it also serves to remind us that statistical performance alone should not be the final evaluation metric. The best-performing model statistically may not be the best-performing model for a geologically informed model. Making that determination requires more than metrics, it will require geologic reasoning. The algorithm does not have that.

Chapters 4 and 5 take a step back from the ML workflow to strengthen the geologic

foundation upon which the ML-derived models will ultimately be evaluated. **Chapter 4** presents a systematic tutorial for enhancing deepwater channel visualization through the corendering of cosine instantaneous phase and relative acoustic impedance, but first with careful consideration of spectral balancing, phase considerations, and resampling of the data. **Chapter 5** extends this visualization approach into a comprehensive interpretation framework for identifying deepwater channel stratigraphy. Together, these chapters establish what channel architecture, stratigraphy, and facies variations actually mean in terms of seismic expression. This work is essential because it provides the interpretive baseline against which ML outputs can be evaluated. Without a robust understanding of the geologic system and its seismic expression, there is no way to assess whether a ML model is capturing geology or meaningless patterns – noise.

Chapter 6 does the hard work of carrying these statistical models through to the final evaluation, which is the geologic interpretation. While the statistical metrics can suggest which model has the best performance based on the algorithm's ability to separate data in patterns, the ultimate metric should be based on the preservation of geologic features AND improved interpretability of the model outputs.

The appendices present related work that supports and extends the core dissertation findings. **Appendix 1** demonstrates how multiattribute analysis can enhance the delineation of obscured architectural elements (AEs) within the Taranaki Basin study area. **Appendix 2** addresses the balance between human insights and ML techniques in optimizing deepwater channel facies identification. **Appendix 3** presents the initial results for the culmination of this larger body of dissertation work. It presents a practitioner's guide for ML-assisted seismic facies interpretation, emphasizing the central caveat that the algorithm does not in fact know geology and should not be treated as though it does. **Appendix 4** discusses strategies for attribute selection in ML applications, using the analogy of "fences and fenceposts" to consider which attributes are telling us something about what is happening within the feature (lithology or reservoir characterizing attributes) versus which attribute is telling us that we have arrived at a boundary.

Together, these chapters and appendices provide a comprehensive framework for thoughtfully integrating ML into seismic interpretation workflows. The goal is not to replace geologic knowledge with algorithms, but to demonstrate how ML, when properly applied within

a geologically informed framework, can enhance interpretive capabilities, improve resolution in facies identification, and increase confidence in geologic interpretations. The value lies not in the algorithm itself, but in how critically and thoughtfully it is implemented.

References

Bhattacharya, S. (2021). A brief review of popular machine learning algorithms in geosciences. *A Primer on Machine Learning in Subsurface Geosciences*, 81-121.

Bond, C. E., Gibbs, A. D., Shipton, Z. K., & Jones, S. (2007). What do you think this is?``Conceptual uncertainty"in geoscience interpretation. *GSA today*, *17*(11), 4.

Chopra, S., & Marfurt, K. J. (2007). *Seismic attributes for prospect identification and reservoir characterization*. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in geophysics*, *61*, 1-55.

Kolla, V., Posamentier, H. W., & Wood, L. J. (2007). Deep-water and fluvial sinuous channels—Characteristics, similarities and dissimilarities, and modes of formation. Marine and Petroleum Geology, 24(6-9), 388-405

Marfurt, K. J. (2018). *Seismic attributes as the framework for data integration throughout the oilfield life cycle*. Society of Exploration Geophysicists.

Polson, D., & Curtis, A. (2010). Dynamics of uncertainty in geological interpretation. *Journal of the Geological Society*, *167*(1), 5-10.

Posamentier, H. W., & Kolla, V. (2003). Seismic geomorphology and stratigraphy of depositional elements in deep-water settings. Journal of sedimentary research, 73(3), 367-388

Wang, Z., Di, H., Shafiq, M. A., Alaudah, Y., & AlRegib, G. (2018). Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. *The Leading Edge*, *37*(6), 451-461.

Zhao, T., Li, F., & Marfurt, K. J. (2017). Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. *Interpretation*, *5*(2), T163-T171.

Chapter 2: Improving Interpretability: The Role of Dimension Reduction in Multiattribute Seismic Facies Identification*

*This paper is published in SEG Interpretation as

Moreno-Ward, A., Vera-Arroyo, A., & Bedle, H. (2025). Improving interpretability: The role of dimension reduction in multiattribute seismic facies identification. *Interpretation*, *13*(3), T475-T496. https://doi.org/10.1190/INT-2024-0109.1

ABSTRACT

Machine learning (ML) is increasingly recognized for its potential to enhance seismic interpretation by improving the analysis of geologically complex and obscured subsurface features. Integrating ML into seismic processing and interpretation enhances multi-attribute analysis, increases resolution, and leads to more precise data interpretations. We emphasize best practices in ML workflows by evaluating these techniques in the Taranaki Basin, located offshore northwest New Zealand. Building on foundational research in deepwater channel systems that have used either/or both synthetic and real data, our project advances these methodologies to assess newer dimension reduction (DR) techniques. We aim to refine our workflow for the identification of seismic channel facies and their associated depositional architectures. This effort involves integrating advanced algorithms in conjunction with K-means such as independent component analysis (ICA), principal component analysis (PCA), kernel principal component analysis (K-PCA), and uniform manifold approximation and projection (UMAP)—to improve classification methods and address the limitations encountered in previous studies. We detail an advanced methodological approach that combines DR techniques with seismic attribute selection to achieve clearer visualizations of channel signatures and more accurate delineation of architectural elements (AE). This enhanced clarity is crucial for deepening understanding of channel complexes and depositional environments, thereby improving mapping of reservoir extent and connectivity. Our research highlights ML utility in defining geologic features and its potential to improve seismic interpretation. The project is structured around two main objectives, the first discussed here and the second, a comprehensive statistical analysis, in a subsequent paper. Our primary objective is to verify that DR techniques improve the interpretability of geologic features within this dataset.

Project findings show an improvement in geologic interpretability; in terms of internal structure, K-PCA with 3 components outperforms other assessed DR techniques for this dataset. When evaluating time-slice geologic interpretability, UMAP delineates more features.

INTRODUCTION

The implementation of new dimension reduction (DR) techniques for improved interpretability in 3D seismic data is necessary to reduce the impact of opportunity cost. While many machine learning (ML) techniques have become commonplace in seismic analysis, several have yet to be tested. Within the area of study, we will focus on two DR techniques that are used frequently and two that are novel for seismic facies interpretation in deepwater channel systems which can be difficult to interpret due to their dynamic morphology and exacerbated by limited well control. The task becomes more challenging as we seek to investigate and identify more detailed channel architecture and lithology indicators, which are vital for assessing the heterogeneity of porosity, and permeability variations within the reservoir. In response to these challenges, our project leverages a blend of seismic attributes and advanced statistical and algorithmic methods to refine the process of geologic interpretation. Our approach centers on the strategic use of DR algorithms to improve the resolution and interpretability of seismic features while preserving geologic accuracy. Conducting a systematic assessment of the effectiveness of these algorithms, especially in terms of preserving geologic significance and augmenting feature discernibility, stands as a fundamental initial phase. This approach helps us to understand the limitations of these methods, therefore improving confidence in the interpretation stage, and bolstering our capacity to delineate geologic characteristics.

The rationale for employing multi-attribute analysis in complex channel systems is straightforward; co-rendering multiple attributes, for example, with each showing improved visibility of geologic features, can help interpreters identify otherwise hard-to-see features. Additionally, it can address difficult complex interpretations, i.e. migration direction, material types, and compaction effects, that may not be readily understandable without the evaluation of multi-attribute combinations. Despite the possible limitations of seismic resolution, using ML serves as a critical initial step in evaluating general lateral migrations and clear reworking of sediments. Considerations include whether other channel migrations intersect the bar or if

deposited splays or stacked bars can be delineated. Minor features like mud draping, though small-scale, are vital for understanding fluid movement. Seismic amplitude, dependent on data quality, offers a foundational method for identifying geologic features. Enhancements like adjusting color scale bars, modifying lighting, or adding bump mapping are strategies to expose more subtle features. A more detailed analysis would include calculating targeted attributes, mapping horizons, and implementing horizon flattening techniques. A practical example of this approach in our project is the application of the sweetness attribute in conjunction with horizon flattening above the channel, as demonstrated in Figure 1. This method effectively emphasizes the main flow and migration directions of the channel, along with precise delineation of scroll bars. Using techniques such as these, an interpreter could begin mapping relative channel ages and with an attribute such as sweetness start to identify areas in the reservoir that may be good targets.

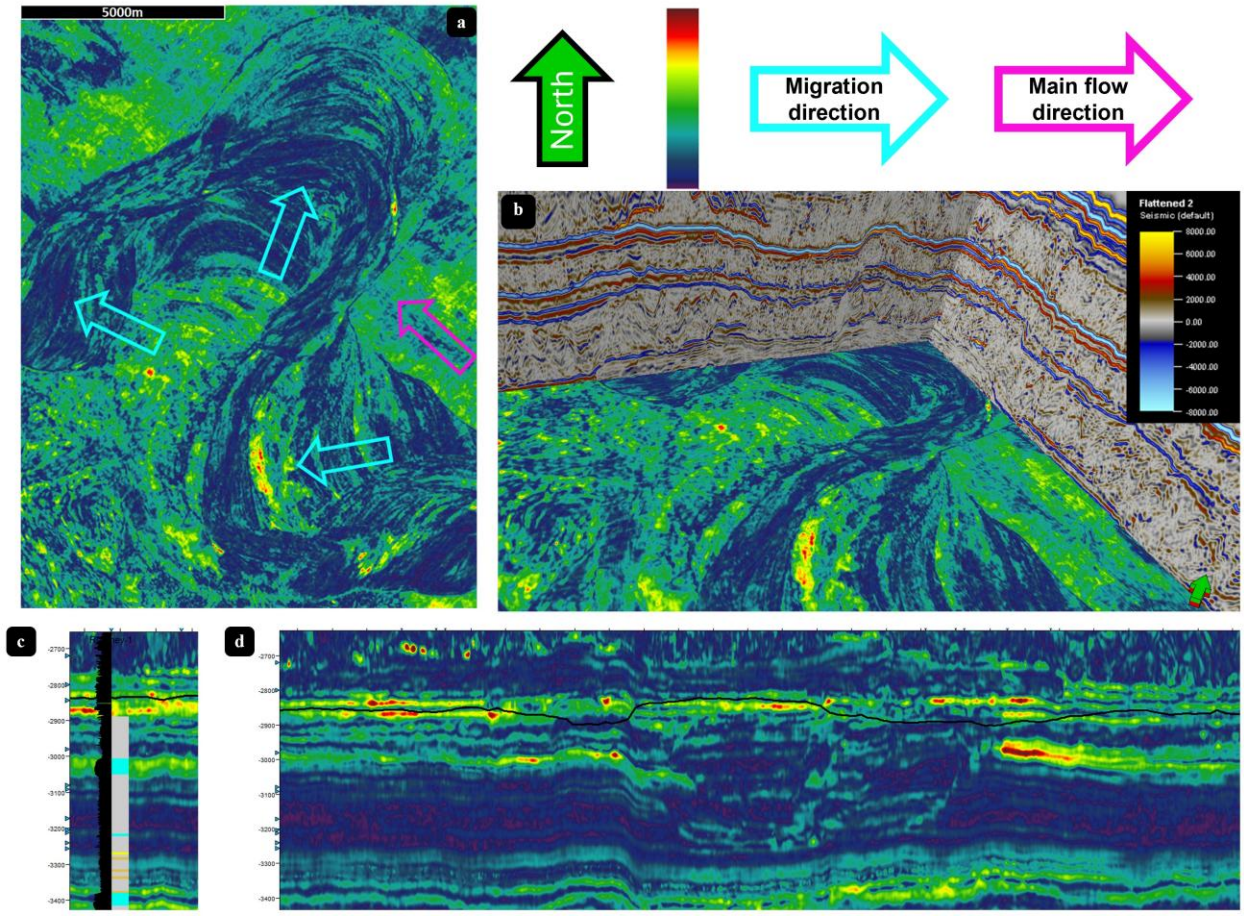


Figure 1: a) Showing the attribute sweetness flatten above channel of interest. This provides additional resolution to map out channel migration and timing. b) Inline 1900, X-line 4850, and Z-slice — 2800 ms. With amplitude. C) Inline view at 1840, X-line 5330 – 5210, with the Romney-1 lithology log on flattened sweetness. D) Inline 1900, from X-line 4970 – 3890, zoomed to the cross-sectional view of the channel would be in part b.

Evaluating and interpreting channel morphology requires an interpreter to include principles of stratigraphy. We will build on insights from Posamentier et al.'s (2022) work which demonstrates various techniques, including optimized lighting, attribute calculations (e.g., curvature, similarity), and strategic co-rendering, which demonstrated the enhancement of channel features for improved interpretation. Keeping this framework in mind, we will further refine our approach to seismic interpretation, emphasizing the integration of seismic stratigraphy, geomorphology, and reflector characteristics for better channel pattern recognition and interpretation. For the Taranaki Basin application, attention will be paid to the interpretability of architectural elements (AE), and marked improvements provided by multi-attribute analysis will be recorded. Additional recent work in the AOI from Chopra et al. (2022), illustrates effective results from the use of strategic attributes, specifically coherence, spectral decomposition, sweetness, relative acoustic impedance, total energy, and reflector convergence in conjunction with self-organizing maps (SOM) and generative topographic mapping (GTM) for the elucidation of channel levees, channel fill, and over-bank deposits in the Taranaki Basin. Using single seismic attributes for interpretation can limit the interpreter to what the attribute can delineate. While co-rendering multiple attributes for analysis can improve visualization, using too many attributes can increase the dimensionality of datasets, leading to decreased effectiveness in visualization and human analysis, even when leveraging ML techniques. To address this challenge, we use DR techniques in conjunction with ML algorithms to refine and enhance the geologic interpretation process.

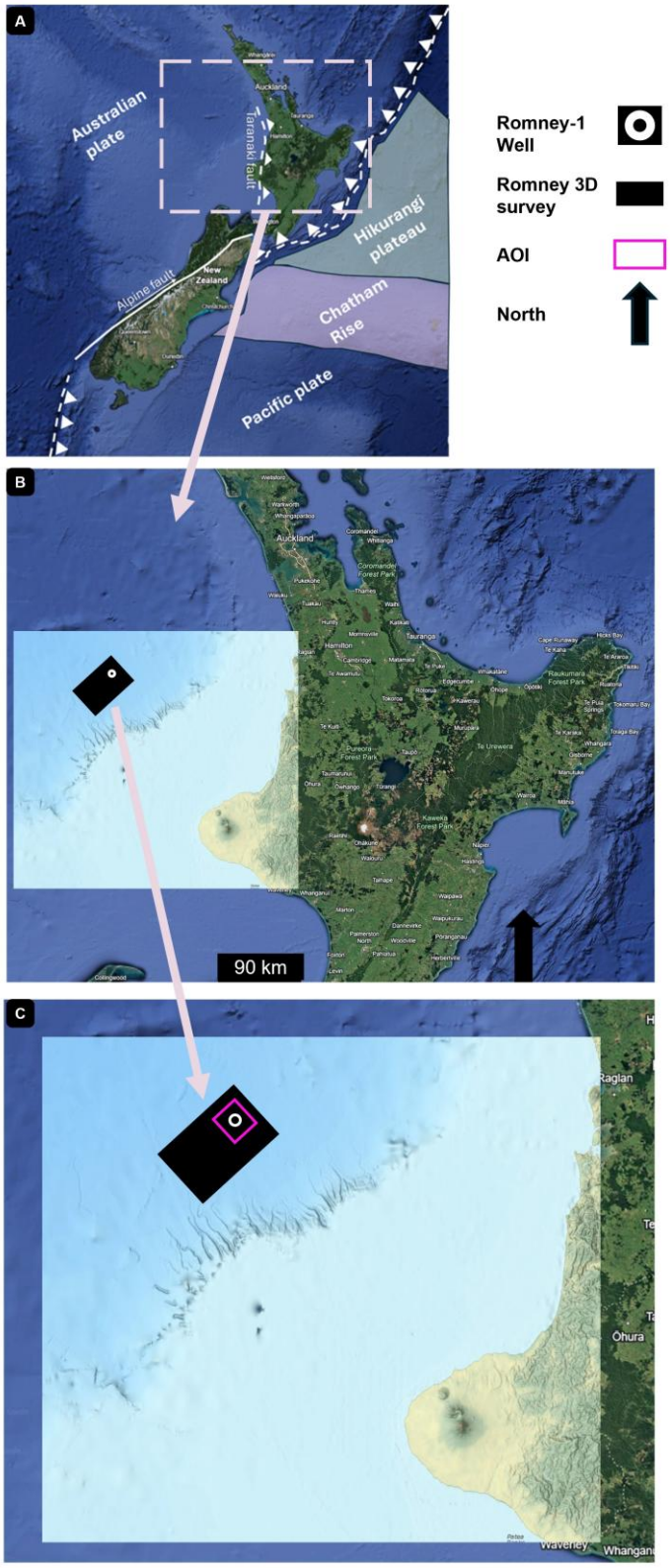


Figure 2: A) regional study location showing major faults, tectonic plates, and topographic features. B) Romney 3D seismic survey location in white box; overlaid to show incised valleys off platform.

Geologic background

The Taranaki Basin, a Mesozoic rift basin located off the northwest coast of New Zealand's North Island (Figure 2) has been shaped by complex processes of extension, subsidence, and compression from the Late Cretaceous to the Neogene period (Leila et al. 2022; Kamaruzaman et al. 2023; Grahame 2015). Spanning approximately 100,000 km² with sediment layers up to 10 km thick (Grahame, 2015), it presents unique opportunities and challenges for geologic interpretation. The basin's extensive geologic research history makes it an ideal location for our investigation into the use of ML and DR techniques. Key studies include those by King and Robinson (1988), Thrasher (1992), Kamp et al. (2004), Nicol et al. (2004), Cathal et al. (2015), Grahame (2015), Reilly et al. (2015), Li et al. (2017), Quan et al. (2017), Strogen et al. (2017), Lubo-Robles and Marfurt (2019), Strogen et al. (2019), Di et al. (2020), Naewboonnien (2021), Chopra et al. (2022), and La Marca and Bedle (2022), along with Kamaruzaman et al. (2023).

The deepwater Miocene channel imaged in the Romney 3D seismic volume within our area of interest (AOI) can be described as a nested channel complex with well-preserved channel migration architecture and accumulation of hydrocarbons (Stroud et al. 2022). There are several areas where channel movement timing can be inferred, however, due to the active progression of the channel during deposition and the AE identification is complex. By using modern examples of channel migration for ground-truthing, we can compare seismic responses of known features such as channel loops, scroll bars, and cutoff loops with topographic maps and corresponding geologic studies. Figure 3 shows an interpreted section of the Mississippi River west of Clinton, KY. High-resolution 7.5-minute quadrangle maps reveal traces of paleo channel loops, scroll bars, and a modern cutoff channel (see Holbrook et al. 2006) for a full study on the portion of the Mississippi River as seen in Figure 3). While features on the topographic maps may appear faint without additional filtering, the geometry of the paleochannel loops and the concentric semi-circular patterns of scroll bars are clearly discernible. In the map's northeast section, the channel migration geometry, highlighted in pink, is also easily identifiable. We can use these distinctive patterns and geometries for the recognition of similar features in the seismic response of our analogous feature of the Romney channel.

Ensuring accurate interpretation of fluvial features is crucial to avoid misidentifying AE and misinterpreting geological scales. For example, although Figure 4 initially appears to show a single channel at the base of the channel complex, enhanced seismic overlays reveal a more intricate internal structure. While map views provide some insight into channel scours, fully understanding migration patterns and sediment deposition trends remains challenging. Comprehensive interpretation demands extensive use of seismic attributes, coupled with deep knowledge in stratigraphy and fluvial geomorphology. Additionally, seismic interpretation can also be validated by comparison to modern analogues for enhancing confidence in these models.

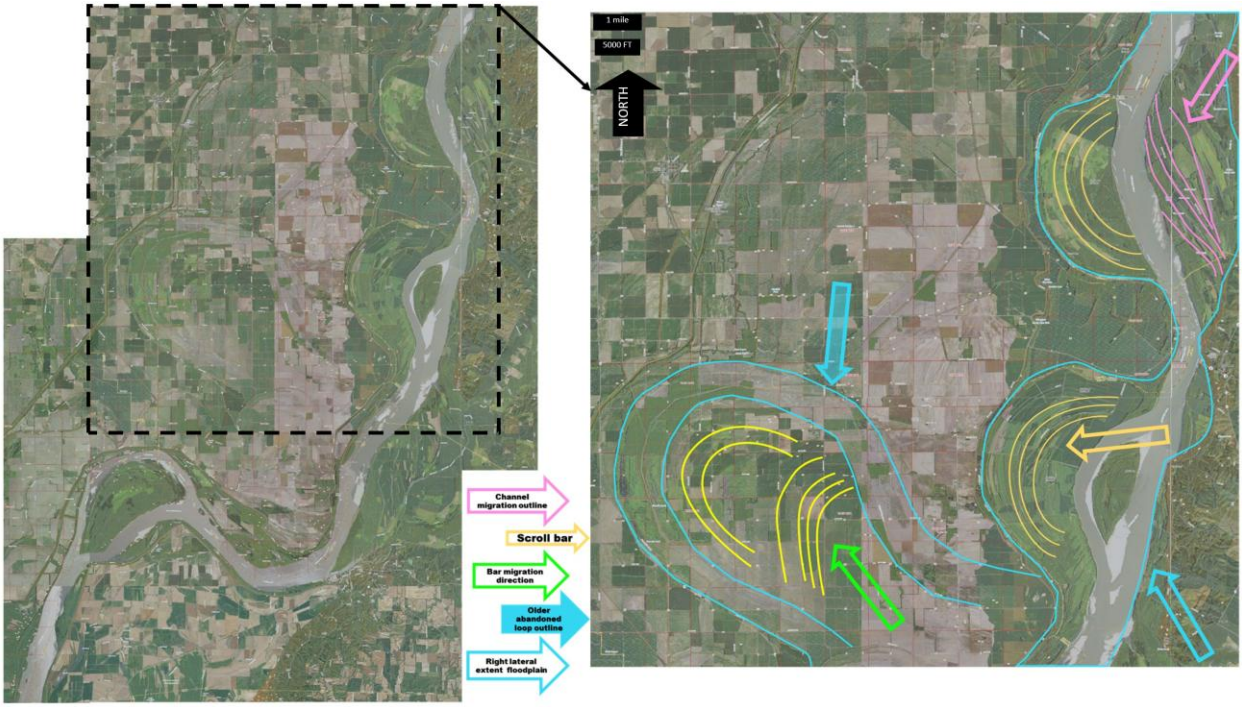


Figure 3: Portion of the Mississippi River: yellow indicating scroll bars, blue channel and channel extent outlines, pink channel migration, and green arrow indicating general direction of channel/bar migration. Topographic maps from the National Map OnDemand Topo site; Wickliffe SW, Arlington, Wolf Island, and Oakton 7.5 minute quadrangle maps west of Clinton, KY

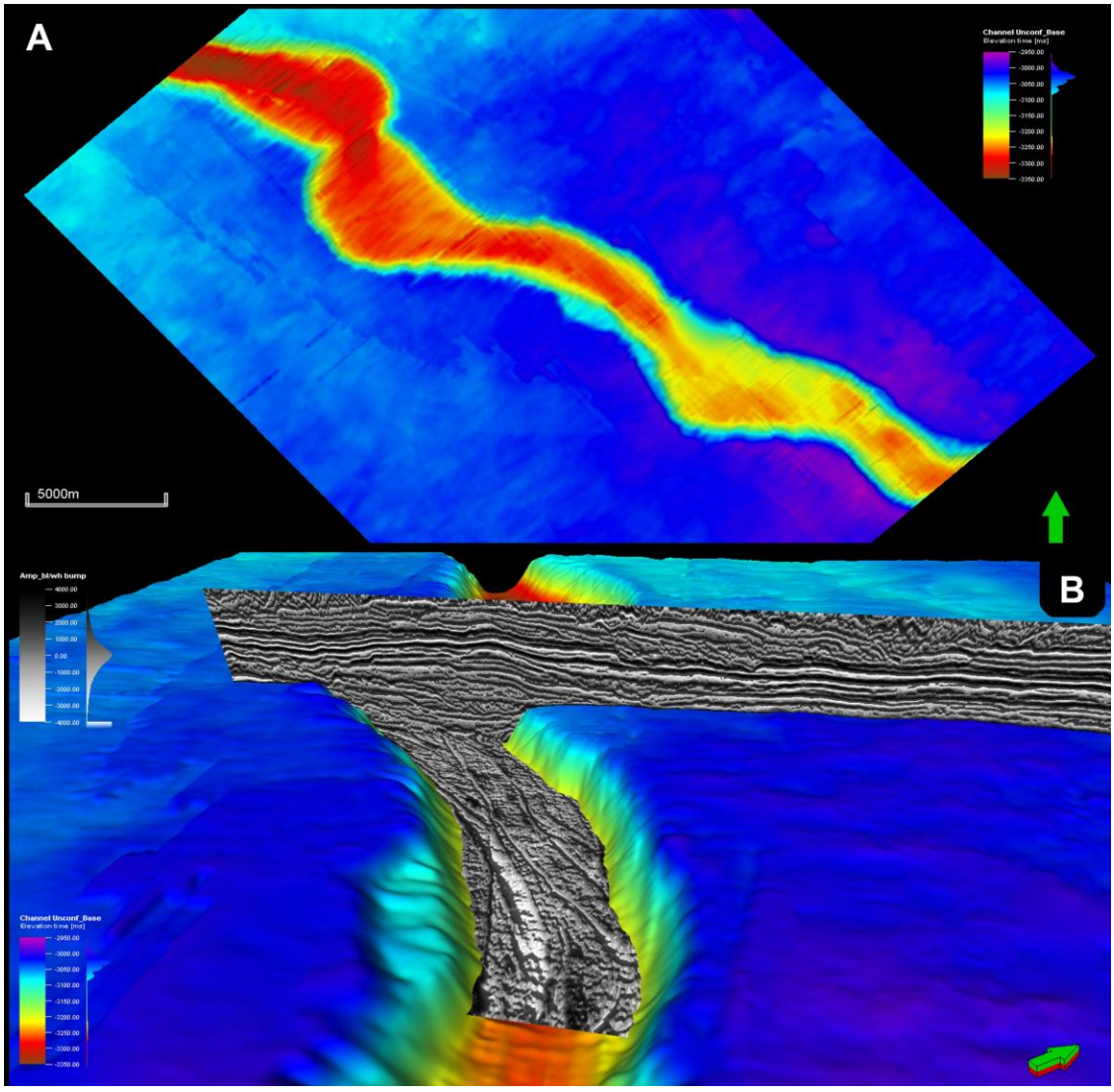


Figure 4: A) mapped surface incised by study channel at approximately — 3.008 s. B) overlain with seismic showing channel complex features.

Data

The Romney 3D data set was acquired in 2011, with an area covering 35 km by 55km, a bin size of 12.5 m by 25 m, and a sample rate of 4ms. There is one well, the Romney-1, located in the volume near, but outside, the channel of study. The recorded water depth in this area at drilling is approximately 1546 m at the lowest astronomical tide, as per the drilling report. The Romney-1 resulted in a dry hole. For a summary, see Table 1.

Seismic Survey		Well	
Name	Romney 3D	**Name**	Romney-1
Survey type	3D	**Well Type**	Vertical wildcat exploration
Environment	Marine	**Spud date**	26-Nov-13
Acquisition	2011	**Completion date**	20-Jan-14
Area	35x55 km	**Rig type**	Drillship
Bin size	12.5x25 m	**Latitude**	37° 53' 39.337" S
Sample rate	4 ms	**Longitude**	172° 43' 52.719" E
Record length	7.48 s	**Drilling floor**	25 m above sea level
Dominant frequency	38 Hz	**Water depth**	1546.6 m lowest astronomical tide
Dominant Wavelength	76 m	**Total depth**	4619 m measured depth
Vertical Resolution (l/4)	19 m	**Result**	Dry hole
Lateral resolution	24 m	**Status**	Plugged and abandoned

Table 1: Seismic survey and well information.

Attributes for channels, an overview

Multi-attribute analysis has been extensively used for the enhanced delineation of channel features in seismic data analysis, with numerous studies exploring its efficacy. In the Taranaki Basin, Chopra et al. (2022) employ multi-attribute analysis for improved channel characterization using coherence, spectral decomposition, sweetness, relative acoustic impedance, total energy, and reflector convergence. Chenin et al. (2021) utilize variance, GLCM average energy, curvature, and spectral decomposition to better image potential basal channel scours. Li et al. (2017) co-rendered sweetness and semblance to improve visualization of channel edges and channel

migration features. Numerous other studies beyond our AOI have investigated attribute combinations for effective multi-attribute analysis in visualizing and interpreting channel systems: Ortiz-Sanguino et al. (2022), spectral decomposition and multispectral coherence; Ray et al. (2022), coherence, coherent energy, structural and amplitude curvature, and spectral decomposition; and La Marca et al. (2021), dip magnitude and azimuth, coherence, RMS amplitude, broadband Laplacian filtering, and coherent gradient, among others. While each example used different attributes, generally these are from the following classifications: wavelet, multi-trace, geometrical, and instantaneous. With the abundance of potentially useful attributes for the identification and interpretation of channel-related features, there is a need to limit the project list of attributes to those that are most useful; the initial selection begins with appropriate attribute categories. Utilization of DR means we are not limited to three attributes; 1) reviewing the seismic volume and 2) the nature of the channel, we select six attributes from categories that are most relevant to our study.

Dimension Reduction

Multi-attribute analysis, while widely practiced and effective, faces constraints regarding the number of attributes that can be simultaneously compared, co-rendered, and clustered. This limitation underscores the significance of employing DR techniques, as they can enhance the resolution and interpretability of seismic features by reducing the dimensionality of the attribute space while preserving essential geologic information. Although principal component analysis (PCA) and independent component analysis (ICA) are commonly utilized DR techniques in seismic analysis, the pursuit of newer, more advanced techniques is an active area of research within geophysics

Similar to the process of multi-attribute selection, DR techniques play a critical role in revealing which combination of attributes helps define more subtle geologic features in seismic data. We want to determine which DR techniques effectively reduce the complexity of seismic data, simplifying it into more manageable and interpretable forms while maintaining a high-quality representation of the original dataset. This simplification allows geoscientists to focus on the most

informative aspects of the data, enhancing the visibility of subtle geologic features that might otherwise remain obscured within the noise and complexity of the original data.
While numerous DR algorithms exist, their application in seismic interpretation varies based on their effectiveness in managing the unique characteristics of seismic data, such as high dimensionality, and the nature of the original dataset. If that data is linear, then either ICA or PCA would be a potential good option; if the data is non-linear, we may consider kernel principal component analysis (K-PCA), or uniform manifold approximation and projection (UMAP). PCA and ICA are examples of DR techniques previously applied to seismic data (Lubo-Robles and Marfurt 2019; Chopra and Marfurt 2014; Honorio et al. 2014; Hyvarinen and Oja 2000; Roden et al. 2015; Zhao et al. 2015). These algorithms help in the identification of patterns, trends, and relationships in the data that are not readily apparent facilitating a deeper understanding of subsurface geology; these algorithms are discussed next.

Dimension Reduction techniques

DR techniques play a crucial role in enhancing the interpretability of complex seismic data by reducing the number of variables while preserving the most important information. These techniques help to identify and highlight hidden patterns, trends, and relationships within the data, enabling geoscientists to gain clearer insights into the underlying geologic structures. In this section, we will discuss four commonly used dimension reduction techniques in seismic interpretation: PCA, ICA, K-PCA, and UMAP.

PCA is used to emphasize variation and capture strong patterns through identification of principal components (orthogonal) with maximal variation, thereby highlighting the most significant patterns and trends. It simultaneously simplifies the data and retains those characteristics that contribute most to its variance, contributing to a clearer understanding of the underlying geologic structures.

Lubo-Robles and Marfurt (2019) utilized ICA and PCA to evaluate which technique would improve visualization of channel features near our AOI in the Taranaki Basin. PCA breaks down the data into principal components while reducing the dimensionality, however, in doing this it may inadvertently remove some of the meaningful geologic data.

In their subsequent work, Lubo-Robles et al. (2023) used PCA with self-organizing maps to aid in identifying bottom-simulating reflectors (BSRs) in their study area. PCA was applied initially to identify the features or the attributes that represent the most variability followed by the application of SOM. The project results found that PCA improved the SOM results for the identification of BSRs.

Chopra et al. (2022) utilized PCA techniques to reduce the dimensionality of the data set before employing SOM and generative topographic mapping (GTM) for unsupervised classification for the identification of seismic characterization. The results indicate that pre-processing the data with DR techniques improved the results in both classification methods. La Marca and Bedle (2022) utilized k-means, PCA, and GTM for seismic facies.

ICA is a computational method for separating a multivariate signal into additive, independent non-Gaussian signals, emphasizing statistical independence. This method is used to identify and separate superimposed geologic features and is particularly effective in distinguishing overlapping seismic signals, enabling us to isolate distinct geologic phenomena that might be less apparent in the raw data. Given the variability inherent in channel systems, this technique is important. It excels in transforming complex seismic data into a new feature space, providing a more interpretable representation that enhances the accuracy of subsequent processes, such as clustering.

Lubo-Robles et al. (2019) utilized ICA to evaluate if this technique would improve the visualization of channel features near our AOI in the Taranaki Basin. Chopra et al. (2022) utilized ICA techniques to reduce the dimensionality of the data set before employing SOM and generative topographic mapping for unsupervised classification for the identification of seismic characterization. The results indicate that pre-processing the data with DR techniques improved the results in both classification methods.

K-PCA: While PCA is adept at uncovering linear relationships, K-PCA extends this capability to nonlinear relationships in the data. It uses kernel functions to project the data into a higher-dimension space where nonlinear relationships can be linearized (Kaib et al. 2023 and Mika et al. 1998). This is particularly valuable in seismic analysis, as geologic formations often exhibit complex, nonlinear relationships.

Mika et al. (1998) used K-PCA to enhance clarity in the data set by removing noise and provided a higher dimension approximation of the initial data set through pre-imaging. Their findings show the effectiveness of K-PCA when used to de-noise the dataset, accurately assign data points, and reconstruct approximate pre-image vectors or data points. Using similar processes, Wang (2014) utilizes K-PCA to help show complex data structures and improve classification accuracy compared directly to PCA performance. They concluded that K-PCA is effective in reconstructing pre-images, outperforming PCA.

UMAP is used for improved visualization and excels in preserving both the local and global structure of high-dimension data. It effectively displays a broad overview of the data's structure while maintaining detailed relationships between closer data points. In seismic analysis, UMAP provided a comprehensive view of the data's structure, enhancing the ability to interpret complex geologic features.

Allaoui et al. (2020) utilized UMAP for preprocessing non-linear data before clustering, demonstrating enhanced cluster performance and reduced computational time compared to k-means. This study evaluated five commonly used datasets to validate performance improvements. Similarly, McInnes et al. (2018) employed UMAP for preprocessing non-linear data in clustering, achieving improved scalability and computational speed across nine publicly available datasets.

Designed to handle more complex non-linear data, similar to K-PCA (Wang 2012; McInnes et al. 2018), UMAP is particularly important given the complexities of channel systems. These systems are characterized by dynamic processes such as channel avulsion and migration, and features including point bars, mid-channel bars, splays, levees, and variable channel fill. By enhancing data quality and maintaining data structure, UMAP results in more interpretable and meaningful geologic data.

Research Objective

ICA and PCA are traditional DR techniques used for preprocessing data before clustering, while K-PCA and UMAP show strong potential for handling non-linear datasets with complex

dimensionality. However, K-PCA and UMAP are not yet well-tested in seismic analysis. Direct comparisons among these methods are necessary to fully understand their capabilities, limitations, and effects on the geologic interpretability of channel architecture.

Our approach focuses on strategically testing and assessing the selected DR algorithms to analyze how each method improves feature discernibility for geologic interpretability. By implementing DR techniques, we can simultaneously analyze high-dimensional data (attributes) and reduce dimensionality for improved interpretability. A systematic assessment of these algorithms' effectiveness in preserving geologic significance and augmenting geologic feature separability is a step in assessing DR techniques. This approach helps us understand each method's limitations, increasing confidence in the interpretation stage and enhancing our ability to accurately delineate reservoir characteristics.

Overall, our objective herein is to assess the improvement in interpretability of these four DR models: ICA, PCA, K-PCA, and UMAP. Next, we detail our methodology to further explain our approach.

DATA PREPARATION METHOD

To properly evaluate multiple DR techniques, we have designed a replicable workflow consisting of two main steps (Figure 5). In **Step 1**, we preprocess and optimize the data for advanced analysis by defining a smaller area of interest (AOI), cropping the seismic volume, calculating selected attributes, creating a well-tie, and incorporating facies. In **Step 2**, we concatenate the attributes into a single project data frame, apply min-max scaling, and machine learning algorithms. We then visualize the DR results to analyze the effectiveness of each technique for both internal and map view channel AE. We will discuss these steps in more detail next.

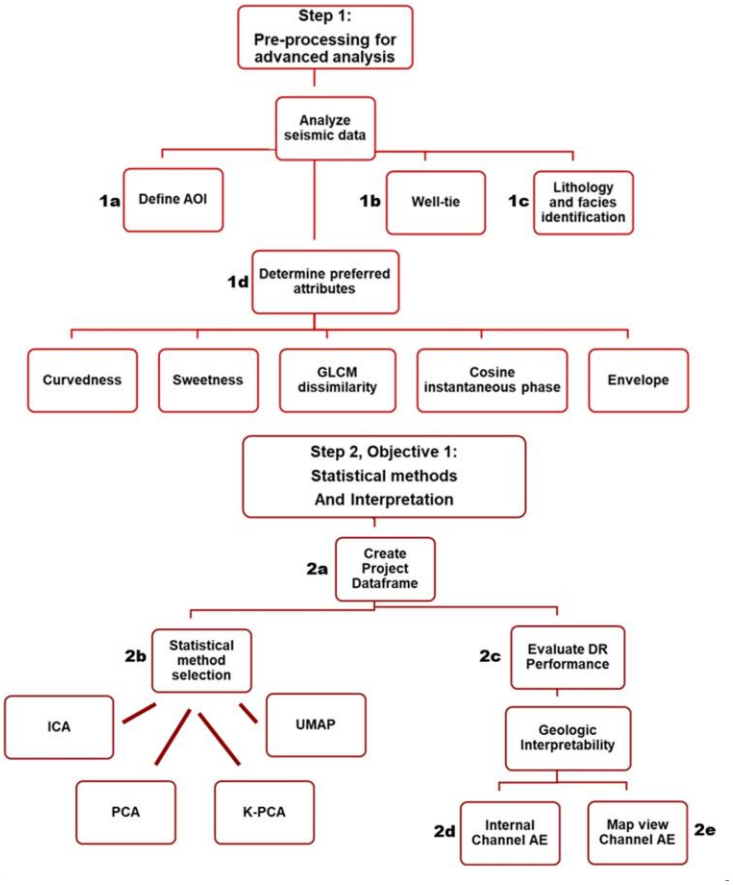


Figure 5: Project workflow. 1) Initial step includes pre-processing prior to advanced analysis. Followed by 2) statistical analysis, 3) data visualization, and lastly, 4) analysis and interpretation.

Step 1a: Data preprocessing

To prepare the data for advanced statistical analysis, we evaluate the seismic volume for quality and useability, specifically assessing interpretability in the original data. This ensures that the AOI does contain geologic detail relevant for ML analysis. The seismic amplitude displayed the channel of interest with striking clarity. The seismic volume is cropped to the determined AOI, which includes a portion of the channel and the Romney-1. Selected attributes (curvedness (k1 and k2), sweetness, GLCM dissimilarity, cosine instantaneous phase, and envelope) are calculated on the cropped volume, and a well-tie is conducted. Facies identification for well and seismic facies is performed using the well tie for validation.

Step 1b: Well tie for seismic correlation

Subsurface measurements are non-unique, i.e. an observed measurement can be the result of mineralogic composition, fluid, and/or depth which means lithology cannot be determined without a direct measure. Wireline logs provide an opportunity to truth seismic observations and are evaluated against well cuttings prior to integration into seismic amplitude; the interpolation of cuttings is conducted at wireline log sampling rate (5 m to .15 m) and we smooth the sonic transit time (DT) and bulk density (RHOB) by subsampling curves to 1 data point every 10 meter to estimate the acoustic impedance (AI) (1/DTC*RHOB). The subsampled AI is used to well-tie the Romney-1 borehole using the Roy-white methodology, extracting the wavelet at the borehole location to minimize the misfit between borehole seismic synthetic and observed seismic. The correct vertical location of the synthetic seismic included stretch and squeeze around a 20 ms bulk shift, which yielded a correlation (synthetic seismic * observed seismic) of 0.79.

Once the borehole and smoothed AI are in the correct vertical location we correlate the sampled lithology log to the seismic amplitude. There are three main anchored points with seismic and wireline logs: 1) Ocean floor, wireline acquisition report indicates GR started recording 1600 m below sea level. 2) The positive high amplitude low-frequency parallel reflector at 3400 ms and 3) borehole total depth. These three reference points are used for correlating the seismic and the wireline log data, allowing for confidence in the lithologic interpretation of the seismic.

Step 1c: Lithology and facies identification

The lack of abundant borehole data presents challenges to the evaluation and identification of deepwater channel seismic facies. To provide a baseline that can be verified, we focus on the Romney-1 borehole, drilled in 2010. The available wireline and borehole information allows us to calibrate gamma ray, p-velocity, and bulk density to rock types and facies, and then extend this interpretation to the 3D seismic response. Utilizing available well-logs (Appendix A, Figure A), and seismic data, the interpreter defines facies. Figure 6 illustrates the selection of both well facies

and corresponding seismic facies. Well facies are determined by identifying unique GR response of sedimentary units and cross-checked with the defined rock types. We identified 5 unique lithology units; unit A corresponds to a distinct drop in gamma ray (GR), and an increase in both resistivity and the compressive transit time (DTC) log reading. Evaluation of the mud log verifies that this marks a transition from shale to a clearly defined lime unit. Through evaluation of the correlated seismic response, we see high amplitude, continuous parallel reflectors. Unit B in the well log corresponds to an increase in GR and a decrease in both RES and DTC; this section is recorded in the mud log as a shale unit. At the base of the shale unit, we note an additional change in both the GR increase and a smoothing of the RES response and mark this as unit C. Moving to the seismic response; unit B can be described as a low-amplitude, discontinuous, and sub-chaotic response, while unit C can be describes as having a low amplitude, discontinuous, and chaotic response. Unit D in the well log has a well-defined marked decrease in GR while the RES does increase at the top of the unit with a bench towards the top and then slowly decreases towards the bottom. The seismic response for D can be described as high-amplitude, continuous, parallel reflectors. Compared with the mud log, this is a well-defined limestone capped above and below by claystone. And lastly, unit E is marked by an increase in GR, though the GR values are varied in this section. The seismic response shows low amplitude, continuous to sub-continuous to sub-parallel reflectors. This unit is reported in the mud log as a mixed claystone/siltstone interlayered.

Once these units are identified in both well and seismic data at the borehole, they are interpolated into the channel as seen in Figure 6 b; when cross-checked with amplitude, higher amplitudes within the channel are assigned as coarser grain material (sands) and the lower amplitudes to finer grain material (silts and muds). This interpolation of material within the channel is supported in the lithology log, however, the lack of well log data directly in the channel necessitates assigning deepwater channel seismic facies rather than lithology. Before fully interpreting these facies, we will visually evaluate the DR model for geologic interpretability.

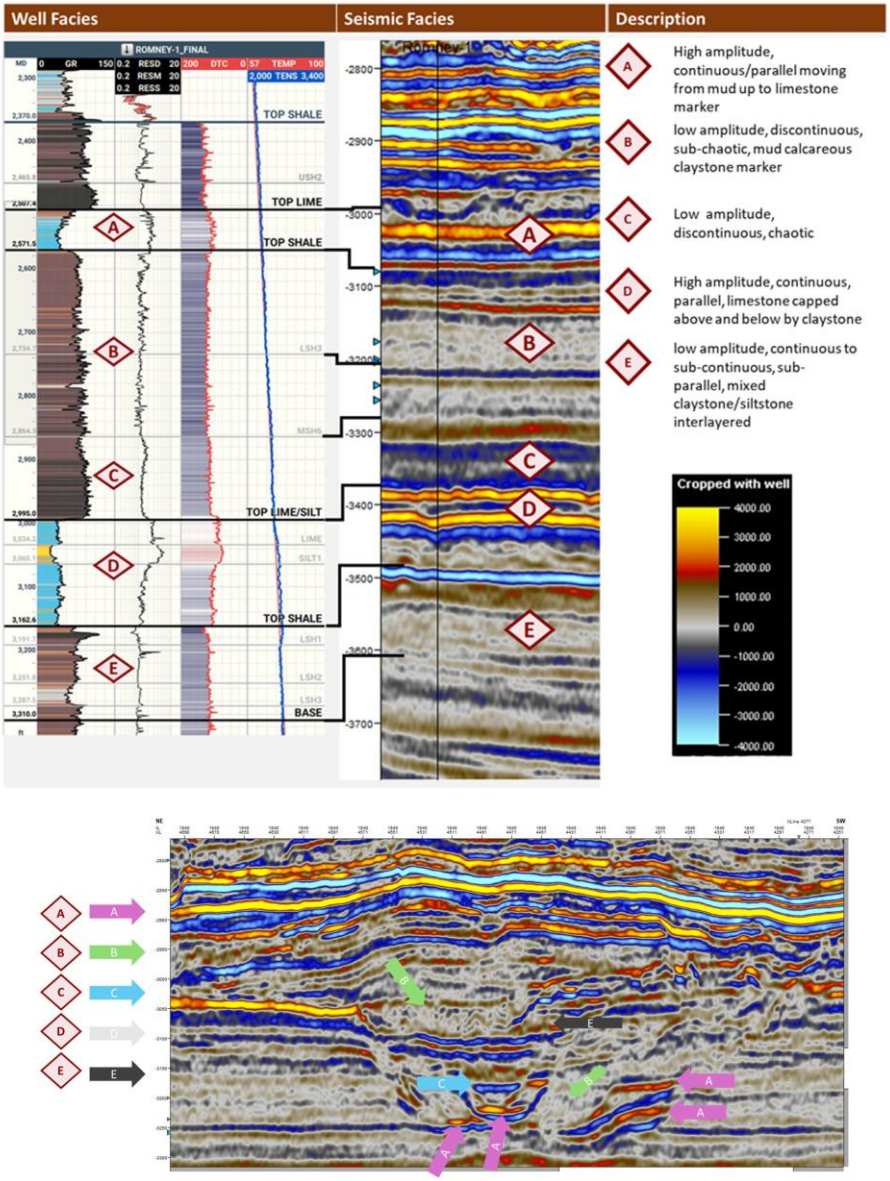


Figure 6: a) Well data with 5 main units identified and labeled as well facies. Seismic data at the well evaluated to identify seismic response corresponding to identified well facies; labeled as seismic facies. b) Well and seismic facies extrapolated into the channel of interest. Of the 5 identified facies, only 4 are found in the channel.

Step 1d: Seismic attribute selection

Seismic attributes have been classified according to the attribute taxonomy as discussed in (Subrahmanyam 2008; Deweet 2021; Chopra and Marfurt 2005; Taner 2001 and 1994; Liner 2016;

Marfurt 2018, Brown 1996, Barnes 1997; and Carter 1995), and a summary Table can be found in Appendix A, Table A. As this project seeks to classify seismic channel facies, we need to select attributes from categories that are useful for the identification of differences within the seismic response in the channel and adjacent to the channel and those that can delineate channel boundaries. For this study, we selected attributes from the following categories, instantaneous, multi-trace/textural, and geometrical: these are briefly summarized in Table 2.

Attribute	Attribute Category	Use	Additional Information
Cosine instantaneous phase	Instantaneous	illustrates continuity of the seismic traces	Barnes 2007
Envelope	Instantaneous	measures the strength or intensity of the seismic amplitude	Subrahmanyam 2008; Chopra 2007
GLCM dissimilarity	Multi-Trace/ Textural	measures changes in seismic amplitude often associated with structural changes and/or lithologic boundaries	Yenugu et al. 2010; Qi et al. 2016
K-Curvedness	Geometrical	quantifies seismic refection curvature, with respect to 0	Chopra 2007
K1-Curvedness	Geometrical	Most positive. Identification of positive geologic relief	Roberts 2001, Verma et.al. 2022
K2-Curvedness	Geometrical	Most negative. Identification of negative geologic relief	Roberts 2001, Verma et.al. 2022
Sweetness	Instantaneous	Instantaneous attribute to enhance bright spots	Hart 2008, Li 2017

Table 2: Overview of attributes chosen for this analysis, including category and general use. For more information on each attribute, see the recommended text.

Step 2: Data Preparation for Machine Learning

The data are compiled into a project data frame (**Step 2a**), with each attribute representing a dimension in the feature space. The data are scaled to a minimum of 0 and a maximum of 1 to ensure the analysis is not skewed. Attributes are cross-plotted (Figure 7) to visually inspect relationships and identify outliers for possible removal or awareness of unique features. An initial unsupervised clustering algorithm (k-means) is conducted and shows non-unique clustering, indicating the need for DR techniques before k-means.

With many DR methods available we chose four to test on this database; ICA, PCA, K-PCA, and UMAP (**Step 2b**)**.** Post-DR k-means clustering shows an overall improvement in performance; prior to conducting a full review of cluster results, we will review the DR performance for geologic interpretability (**Step 2c**)**.** This review is conducted on both internal channel AE (cross-sectional view, **Step 2d**) and in map view (**Step 2e**) for laterally observable channel AE and will be discussed in more detail next.

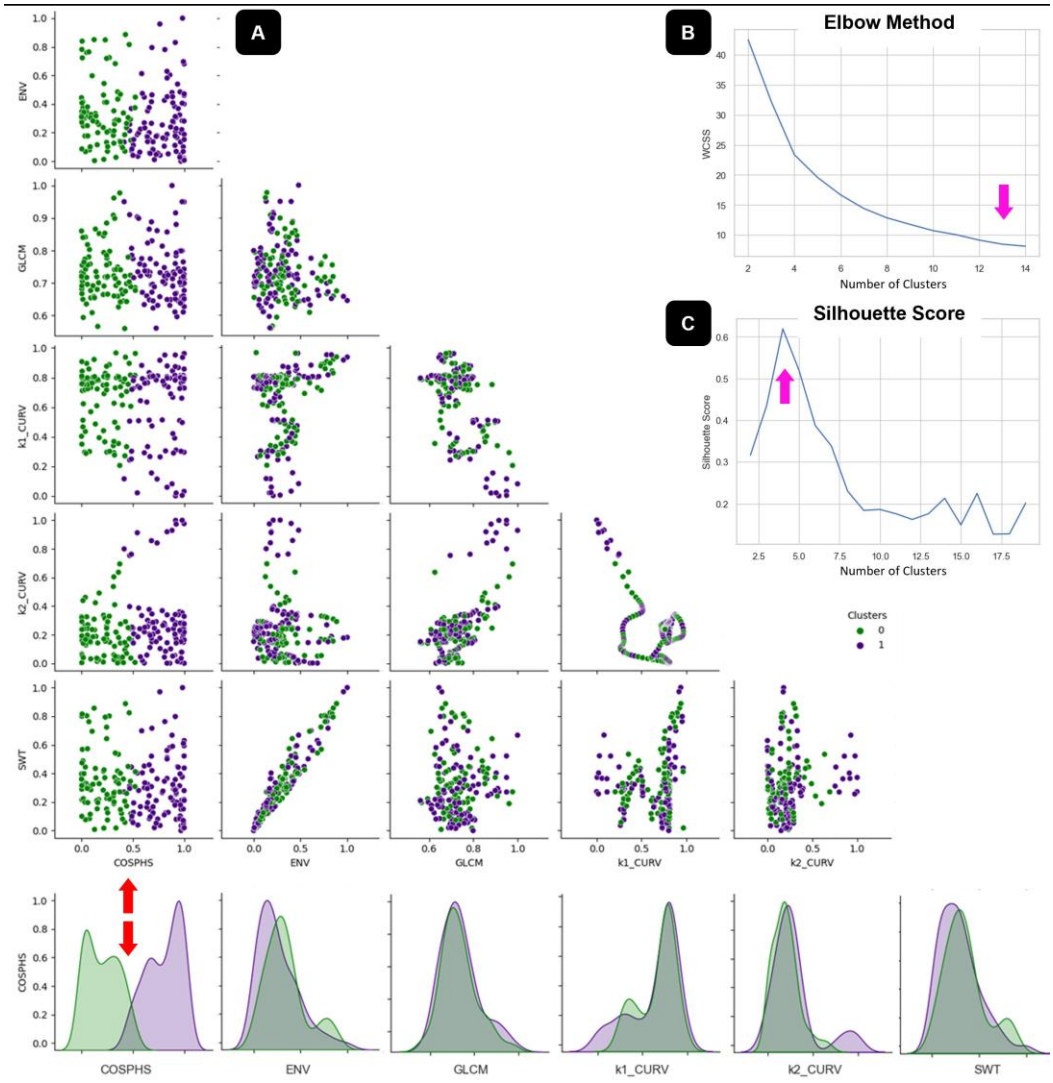


Figure 7: Unsupervised clustering of cosine instantaneous phase, envelope, GLCM dissimilarity, k1 and k2 curvature, and sweetness before DR, 2 clusters. A) Scatterplot with cosine phase (column indicated by red arrow) show a distinct separation in data; this trend is shown clearly in the histogram. Other attribute scatterplots show little to no cluster separation; the corresponding histograms show definitive overlap of data. B) Elbow method indicating 13 clusters and C) silhouette score indicating 4 clusters an optimal number.

RESULTS & DISCUSSION

Attributes for architectural elements

Seismic resolution, both vertical and lateral, impacts an interpreter's ability to identify and delineate features of interest. While resolution is tied directly to the frequency, sampling interval, and process of data acquisition, the interpretability of the data can be improved by post-acquisition filtering, attribute calculations, and ML processes. Though the post-stack 3D Romney effectively highlights large-scale channel features, continued workflow improvements are important to further enhance geologic interpretation and identify areas of interest, i.e. hydrocarbon traps. Let us briefly review the attribute selection process for this project.

Attribute selection:

In **Step 1** of the workflow, we carefully selected seismic attributes that can help define essential geologic features including lithologic change indicators, channel edges, and internal structures. However, utilizing each attribute individually for mapping channel morphology to provide unique interpretation precision enhancement may not be as effective as multi-attribute analysis. Amplitude effectively demarcates the main channel edges and scroll bars in some instances, providing a foundational visual guide, as seen in Figures 6-8. Conversely, the cosine instantaneous phase brings to light significant lateral changes across reflectors and emphasizes the delineation of cross-cutting channels, Figures 6, 7, and 9-11. While attributes such as sweetness and envelope highlight lithological transitions, they alone do not refine the resolution of AE to the desired degree. Likewise, K1 and K2 curvature attributes signal reflector deviation from 0 but do not capture the detailed nuances essential for identifying smaller-scale channel AE.

Figure 8 illustrates the challenge of relying solely on single attributes to map geological features at a detailed resolution. Outline overlays, compiled by tracing features from each attribute, facilitate pattern recognition, however, without a general channel outline, individual attributes alone struggle to convey a coherent interpretation. Figure 8B b and c illustrate sweetness and envelope with channel boundaries mapped; yellow arrows indicate areas with

coarser-grained material while blue arrows indicate areas with finer-grained material. While each of these two attributes highlights some apparent difference within as opposed to adjacent to the channel, it becomes very difficult to map more complex AE inside the channel boundary. Figure 8B a and d illustrates amplitude and cosine instantaneous phase; using these two attributes, channel boundaries and internal structure become clearer. Yellow arrows mark coarser material, while blue arrows mark finer-grained material; the pink arrows in Figure 8B d highlight clearly defined channel edges. A closer look at mapping channel features using one attribute at a time helps to underscore the need for a multi-attribute integration approach to comprehensively map the intricate morphology of channel systems.

Figure 9 shows a cross-sectional view of each attribute, blue arrows indicate older channel edges while yellow arrows indicate younger cutoff morphology; channel edges are clearly defined in a) amplitude and d) cosine instantaneous phase. Channel edges and some internal structure can be inferred from both b) sweetness and c) envelope, however, using these attributes in conjunction with amplitude and cosine instantaneous phase would improve confidence in the determination of bounding edges. When evaluating e) k1, f) k2, and g) GLCM dissimilarity individually, it becomes much more difficult to define channel features; g) has better definition in demarcating internal channel features, however, the resolution is not high enough to clearly map out full internal channel structure. Now that we have assessed what channel features can be identified with each attribute independently, let's focus attention on seismic amplitude and lithologic indictors.

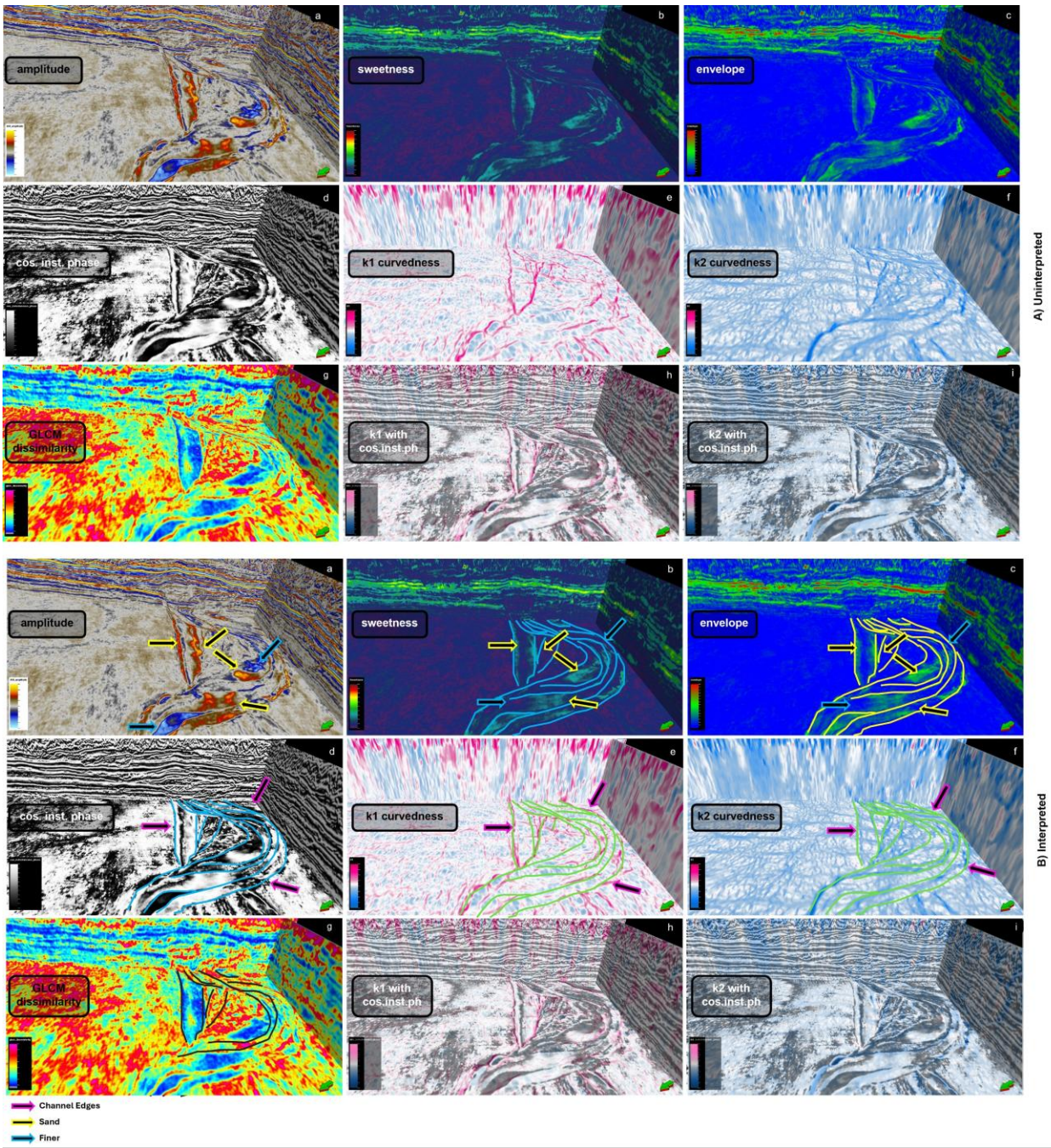


Figure 8: Using inline 1842, x-line 4200, and z-slice — 3204, image with A) uninterpreted and B) interpreted selected attributes indicating interpretability of larger scale channel features. Yellow arrows indicate areas with coarser grained material and blue arrows indicate areas with finer grained material. Pink arrows indicating well defined channel edges. Co-rendering in B) viii and ix further enhance channel edges.

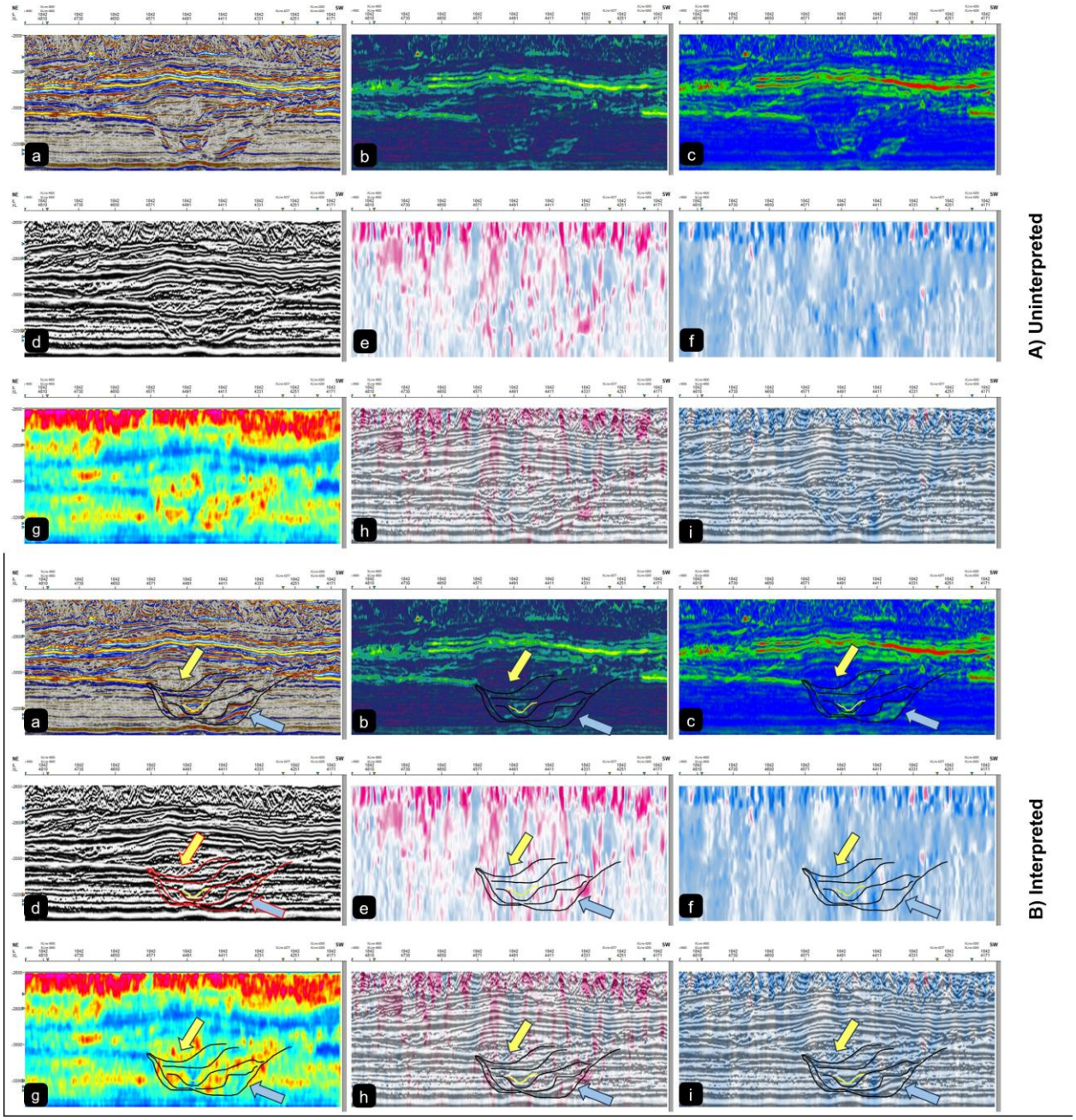


Figure 9: inline 1840. a) amplitude b) sweetness c) envelope d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) k1 overlain on cosine instantaneous phase i) k2 overlain on cosine instantaneous phase. Same image on right, yellow arrows indicating younger cutoff channel with a straighter morphology. Blue arrows indicate older channel edge.

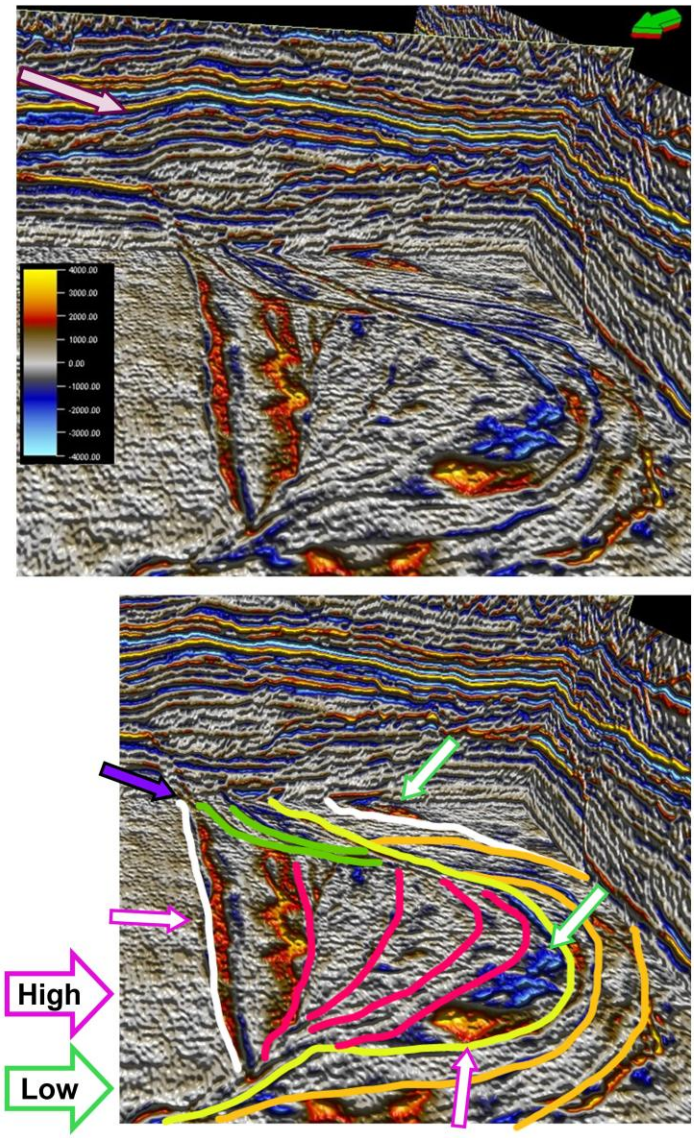


Figure 10: a) Extrapolating well and seismic facies into the channel for lithology indicators. b) Using a combined x-line, inline, and z slice view, lithology indicators can be applied. Channel migration indicator outlines can begin as a starting stage to overall geologic interpretation. Comparison with inline view will help determine which channels are younger. Light pink arrow indicating differential compaction. Purple arrow: as the top edge of the cutoff loop is interrupted, we can infer that the cutoff loop is older than the portion of the channel outlined in green.

Seismic Amplitude and Lithologic Indicators

Detailed seismic amplitude evaluation informs interpreters about depositional environments using lithologic indicators (Figure 10). By analyzing areas of high and low amplitude and integrating inline and x-line slices with time slices, initial channel architecture can be delineated. Understanding channel morphology and sedimentary processes is crucial for identifying predictable depositional patterns, such as coarser-grained sediments near channel margins and finer-grained sediments toward the center and adjacent floodplains. Channel migration leads to sediment reworking, as illustrated by the pronounced cutoff channel in Figure 10 (purple arrow), indicative of accelerated flow velocities and increased sediment deposition. The age relationship between the cutoff channel and the large-scale meander loop is inferred through cross-cutting relationships, suggesting a shift from higher to lower flow velocities and changes in sediment composition. Higher resolution

interpretation can be achieved by considering additional information beyond amplitude even in the absence of well logs, such as lithological indicators within the channel fill.

Differential Compaction as an Indicator

Figure 10 also reveals evidence of differential compaction. A structurally higher elevation, marked by a light pink arrow, suggests compaction variation within the subsurface strata. Such thickness variations typically result from ongoing depositional processes that compact the underlying layers differently. The finer materials surrounding the highlighted channel are more prone to compaction, contrasting with the channel's varied sediment sizes, ranging from mud through silt to sand. This grain size variability within the channel complex contributes to its resistance against compaction compared to the adjacent finer sediments. Beyond the identification of lithology indicators, the interpretation of channel systems is further complicated by their dynamic evolutionary processes.

Channel Complex Morphological Evolution

By examining time-slices and inline/crossline views, we can delineate the main shape of the channel complex and map its progression over time as seen in Figures 9-11.
At –3260 ms (Figure 11) a channel with low sinuosity is highlighted by a light pink arrow in d) and variable migration in the channel loop marked by a light orange arrow. These observations are clearly visible using cosine instantaneous phase alone, while with amplitude (a) these features are less distinctive; with sweetness (b) and envelope (c) we do see evidence to support this interpretation. Moving to –3236 ms (Figure 12) the channel becomes more complex in the section marked by the light pink arrow as does the scroll bar and cutoff channel marked by a light orange arrow. These changes are likely due to increased energy in the system. Additionally, it becomes more difficult to identify these features using a single attribute. The diagram sketched in part i) was created by overlaying multiple attributes. Figure 13 enhances the understanding of channel morphology changes over time by incorporating both inline and x-line views.

Mapping a channel's morphologic changes through time offers valuable insights into its depositional history. By delineating the channel sequence into a rough chronological order, we can better analyze changes in channel shape, flow rates, and sedimentation patterns. Studying these changes can provide crucial information about the reservoir's characteristics, such as porosity and permeability, which are essential factors in assessing potential.

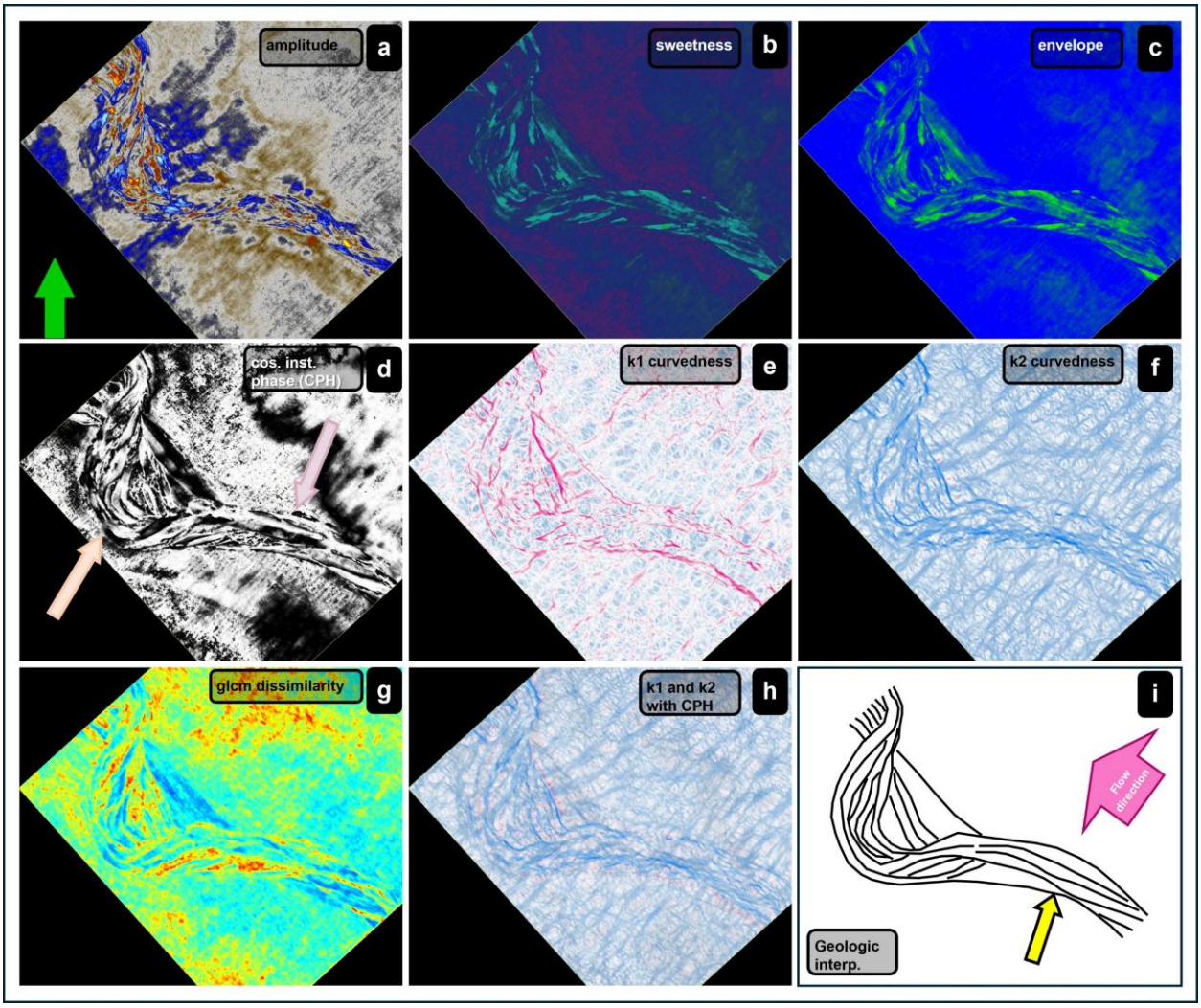


Figure 11: Images show z-slice at — 3260 ms. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. While each attribute can highlight geologic features, individually, they will not be able to highlight all detail. i) Channel flow direction indicated with pink arrow; channel migration and vertical progradational history show channel straightening at yellow arrow.

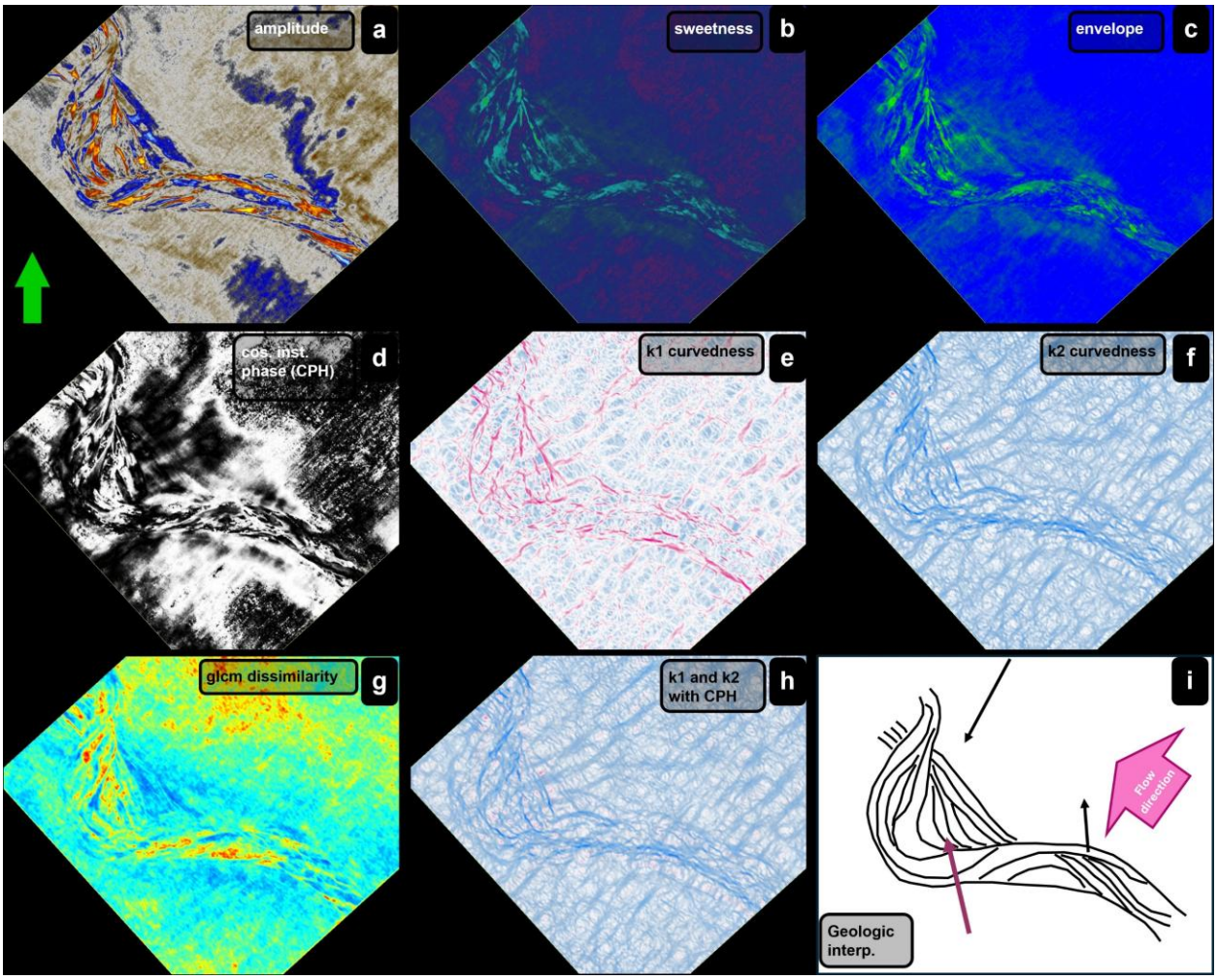


Figure 12: Images show z-slice at — 3236 ms. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. While each attribute can highlight geologic features, individually, they will not be able to highlight all detail. i) Channel flow direction indicated with pink arrow; channel migration and vertical progradational history can be estimated once channel morphology is mapped.

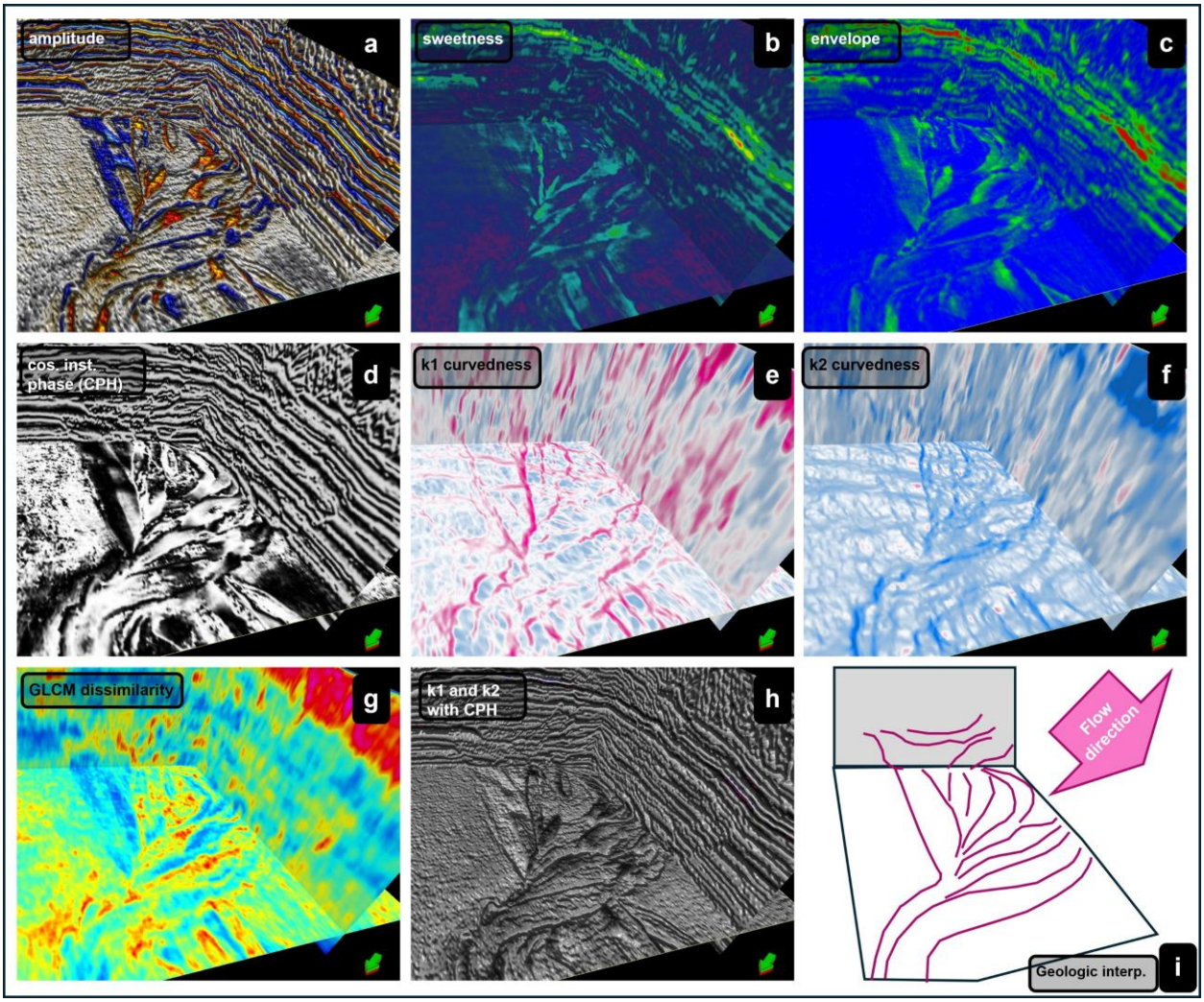


Figure 13: Z-slice — 3236, Inline 1902, X-line 4280. a) amplitude b) sweetness c) envelop d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity h) amplitude with black/white color and bump map i) channel architecture. Individually each attribute may highlight apparent geologic features, however, in conjunction they tend to draw out more detail. i) shows the progression of the channel and migration direction. Internal channel complex architecture shows upward changes in channel morphology.

DR for interpretability of geologic features

Now that we have established a need for DR, pre-processed the data for statistical analysis (**Step 2a**) and selected the statistical methods (**Step 2b**), each output must be evaluated to determine the performance in preserving geologic importance with visual inspection, **Step 2c**. While co-rendering of attributes is used effectively for feature delineation, the number of attributes that can be co-rendered is a limitation. Co-rendering too many attributes can obscure data from

each layer, causing confusion. With this project's focus on identifying deepwater channel facies, it is paramount to ensure the selected attributes will highlight channel features, i.e. channel boundaries and the distinction between channel fill and the surrounding facies. With the application of DR, we can simultaneously use more appropriate attributes and decrease the complexity of the data while preserving geologic importance for improved interpretability. This should hold for the smaller-scale architectural features associated with channel systems. These DR models are ready for visual analysis and will discuss this next.

DR for architectural elements

After initially evaluating each DR for statistical performance and rendered for geologic interpretation, the models are visually assessed to evaluate the preservation and interpretability of the channel elements, **Steps 2d and 2e**. Each DR model is calculated to reduce the dimensions from six (each attribute in the project data frame represents 1 dimension) to 2 and 3 components. ICA, PCA, K-PCA, and UMAP models are compared for 2 and 3 components and 2 vs. 3 components.

Figure 14 shows inline 1840 for all models, 2 and 3 components, with internal channel features outlined. The yellow arrow indicates the youngest portion of this channel complex; the seismic response in this feature is chaotic, corresponding to seismic facies B, Figure 6 b. This response is evident in ICA and PCA with 3 components; however, in terms of preserving the AE associated with the morphology of this channel, K-PCA does a better job. Again, looking at Figure 14, the blue arrow indicates one of the oldest bounding geometries for the channel complex. We see well-defined sub-parallel reflectors with what appears to be a possible levee. This area corresponds to seismic facies A seen in Figure 6 b. This seismic response is easily recognizable in all models, except as indicated by the orange arrow for K-PCA with 2 components. While we do see the sub-parallel reflectors, the bench is not visible leading to a misinterpretation excluding the levee and tracing that channel edge to the top of the complex.

Figure 15 shows the DR models for time 3.2s and 2.8s and illustrates the changing channel morphologies over time. Figure 15 A showing a time slice at 3.2s, illustrates an older portion of the channel complex. We observe a sinuous to meandering deepwater channel; each

DR model highlights the main outline of the channel (1), a cutoff channel (2), and a scroll bar (3). While each model effectively highlights the main channel boundaries, we do see that c and d have a loss of resolution in the area indicated by the pink arrow. In the scroll bar, we see variability in the resolution, as indicated by the green arrow. PCA, with 2 components, does not preserve high variability; while PCA with 3 components does show more heterogeneity, there are no real geologic elements to interpret. Both K-PCA and UMAP preserve more of the heterogeneity within the scroll bar and the geologic interpretability to determine how that scroll bar migrated over time.

Figure 15 B shows a time slice at 2.8s, a younger and visually complex section. We have focused on a large channel loop with an internal scroll bar. The pink arrow indicates a feature interpreted as a large splay. K-PCA and UMAP models delineate the lateral extent of the splay and highlight internal complexity within the scroll bar as well.

For a visual comparison of channel elements discussed for Figures 13 and 14 in the original seismic attribute display see Figures 9 and 10. For a visual comparison of channel elements discussed for Figures 13 and 14 in the original seismic attribute display see Figures 9 and 10.

Reviewing the DR model performance reveals that K-PCA performs best in preserving geologic features in a cross-sectional view as we are better able to delineate internal nested channel geometries and channel levees. UMAP is the next best model for this dataset better delineating geologic features when in map view.

Through visual comparisons with original seismic attribute displays (Figures 9 and 10), it is evident that K-PCA excels in this case with preserving geologic features in cross-sectional views, facilitating the delineation of internal nested channel geometries and channel levees. Meanwhile, UMAP emerged as the next best model for this dataset, providing better delineation of geologic features in map views.

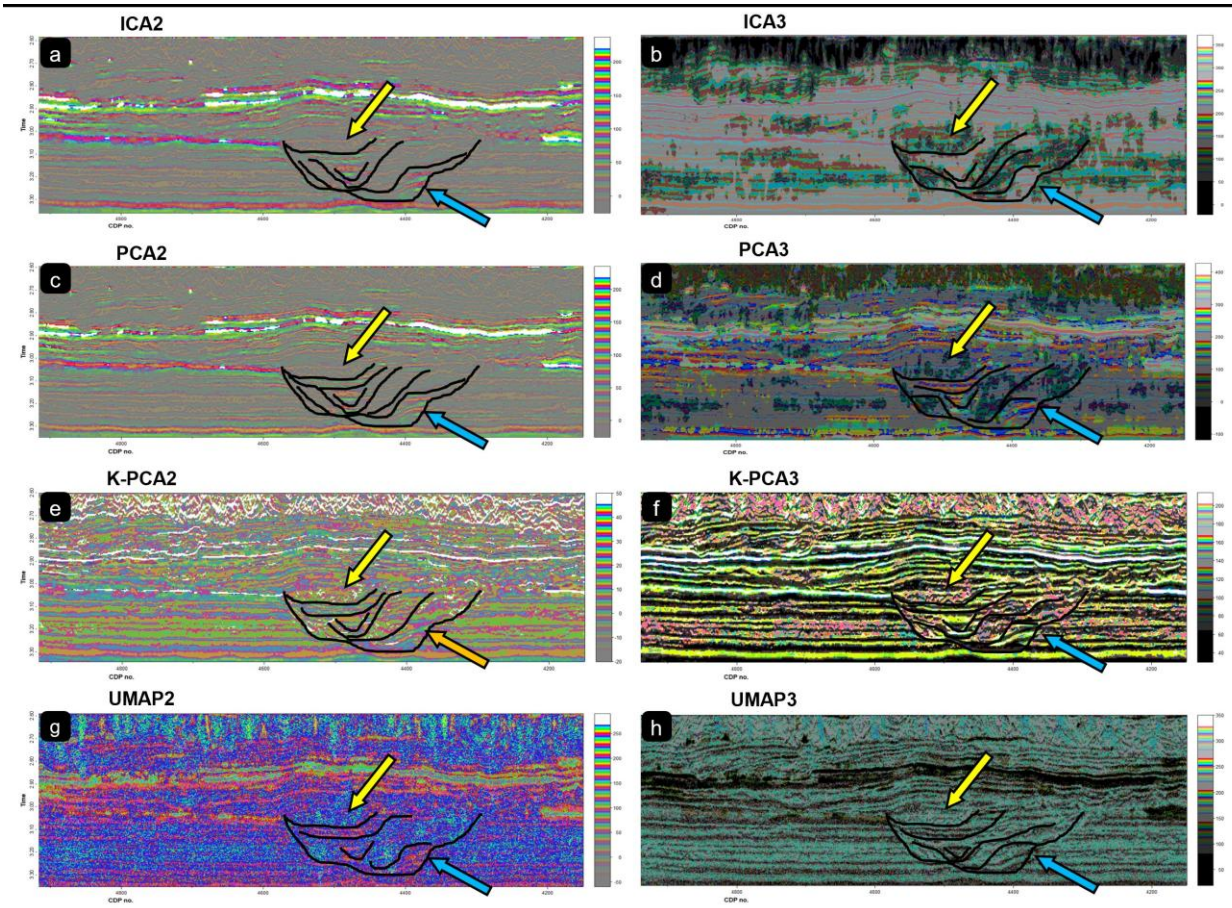


Figure 14: Yellow arrow indicating younger cutoff channel and blue arrow indicating channel edge migration. Orange arrow highlights deviation from other models in highlighting channel edge morphology; this deviation could affect channel architecture interpretation and is worthy of note. While all models delineate main channel scours, K-PCA and UMAP highlight smaller-scale features better. a) ICA, 2 components b) ICA, 3 components c) PCA, 2 components d) PCA, 3 components e) K-PCA, 2 components f) K-PCA, 3 components g) UMAP, 2 components h) UMAP, 3 components

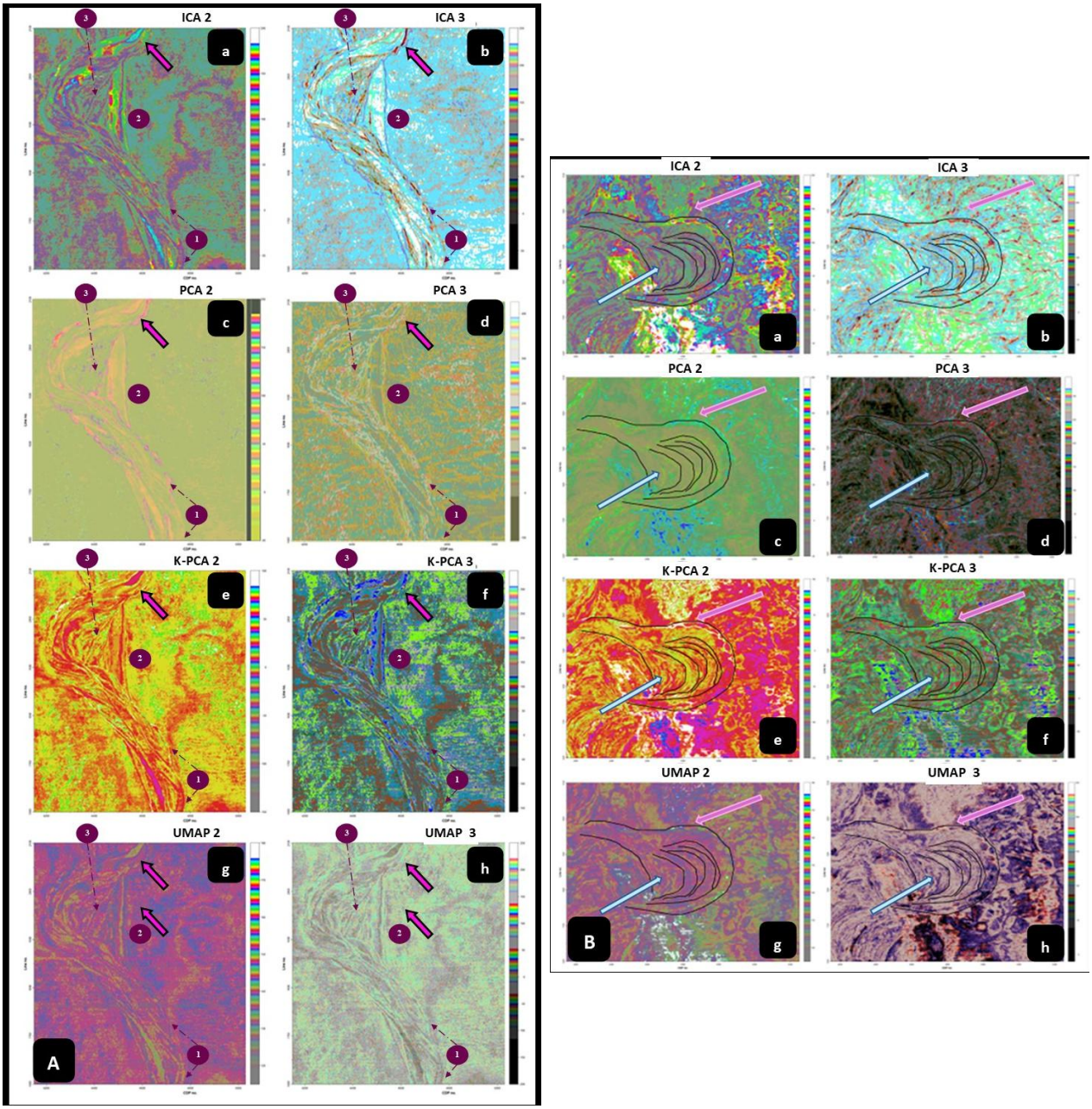


Figure 15: A) Time 3.2s, blue arrow indicating splay location, pink arrow showing variation in delineating smaller scale channel migration, and green arrow indicating scroll bar. While all models delineate main channel features, K-PCA and UMAP highlight smaller-scale features better. a) ICA, 2 components b) ICA, 3 components c) PCA, 2 components d) PCA, 3 components e) K-PCA, 2 components f) K-PCA, 3 components g) UMAP, 2 components h) UMAP, 3 components. 1 indicates black lines outline main channel boundaries, 2 indicates the cutoff channel, and 3 indicates the scroll bar.

B) Time 2.8s, blue arrows indicating internal scroll bar and pink arrows indicating splay. While all models delineate main channel features, K-PCA and UMAP highlight smaller-scale features better. a) ICA, 2 components b) ICA, 3 components c) PCA, 2 components d) PCA, 3 components e) K-PCA, 2 components f) K-PCA, 3 components g) UMAP, 2 components h) UMAP, 3 components

CONCLUSION

We have demonstrated several key principles in the application of DR techniques to seismic interpretation: 1) the effectiveness of each DR technique can vary based on viewing perspective. For improved visibility of internal channel structure: K-PCA is identified as having greater performance for the selected attributes in both preserving geologic features and delineating internal channel structure. For map view, UMAP performs best for identifying AE, such as splays and scroll bar migration features. 2) Our results indicate that a combination of statistical evaluation and geologic interpretation provides the most effective overall model selection, balancing quantitative performance with geologic meaningfulness and underscoring the importance of integrating computation methods with an expertise in the geologic problem. 3) We found that careful attribute selection and data preparation are critical first steps in the ML workflow, directly impacting both the quality and interpretability of the results highlighting the importance of pre-processing for ML application and seismic data analysis.

These findings provide interpreters with specific tools to better visualize and understand complex channel morphologies in the Taranaki Basin, which can be adapted to similar deepwater channel systems elsewhere. This approach emphasizes the value of combining statistical evaluation with visual geological interpretation, as the most effective models balance quantitative performance with geologic meaningfulness.

It is important to note some limitations in our approach. No single model proved fully effective across all viewing perspectives, as mentioned, highlighting the need for a multi-model approach for a comprehensive interpretation. Additionally, a subject matter expert is still necessary for the selection of appropriate attributes (which can be time-consuming) and to evaluate channel architecture. The author's background in fluvial geomorphology provided valuable context for the interpretation of this deepwater channel setting.

Interpreters conducting future studies in similar or diverse geological settings should consider implementing K-PCA for cross-sectional analysis and UMAP for map view interpretations. This approach can be adapted to various structural and stratigraphic scenarios, helping to reveal the most effective methods for each unique geologic context. While the findings and workflows discussed in this paper serve as a valuable starting point, effective implementation depends on carefully evaluating and selecting attributes, preparing datasets for

multi-attribute analysis, creating multiple models for comparison, and assessing these models both statistically and visually. By balancing quantitative performance and geological meaningfulness, interpreters can significantly enhance their ability to identify and interpret complex geologic features in seismic data. This approach improves the immediate interpretation of channel systems and contributes to advancing seismic geomorphology through refined multi-attribute analysis techniques.

ACKNOWLEDGEMENTS

The authors would like to thank the AASPI consortium for providing the software used for the calculation of attributes and for their financial support of this research.
We also thank the New Zealand government for making the Romney data available for research.

REFERENCES

Allaoui, M., M. L. Kherfi, and A. Cheriet, 2020, Considerably improving clustering algorithms using UMAP dimensionality reduction technique: A comparative study: International Conference on Image and Signal Processing, Springer International Publishing, 317-325.

Barnes, A. E., 1997, Genetic classification of complex seismic trace attributes: SEG Technical Program Expanded Abstracts, 1151-1154.

Barnes, A. E., 2007, A tutorial on complex seismic trace analysis: GEOPHYSICS, **72**, no. 6, W33-W43.

Brown, A., 1996, Seismic attributes and their classification: The Leading Edge, **15**, 1090-1090.

Carter, D. C., 1995, Use of Windowed Seismic Attributes in 3D Seismic Facies Analysis and Pattern Recognition: International Symposium on Sequence Stratigraphy, 73-87.

Chenin, J., C. Silver, and H. Bedle, 2021, Seismic geomorphology anomalies within a Pliocene deepwater channel complex in the Taranaki Basin, offshore New Zealand: Interpretation, **9**, no. 1, C1-C10.

Chopra, S., R. K. Sharma, K. J. Marfurt, H. Bedle, and S. Verma, 2022, Attempts at seismic characterization of a deepwater turbidite channel in Taranaki Basin, New Zealand: First Break, **40**, no. 10, 27-39.

Chopra, S., and K. J. Marfurt, 2014, Churning seismic attributes with principal component analysis: SEG Expanded Abstracts, 2672-2676.

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

Chopra, S., and K. J. Marfurt, 2005, Seismic attributes—A historical perspective: GEOPHYSICS, **70**, no. 5, 3SO-28SO.

Dewett, D. T., J. D. Pigott, and K. J. Marfurt, 2021, A review of seismic attribute taxonomies, discussion of their historical use, and presentation of a seismic attribute com

munication framework using data analysis concepts: Interpretation, **9**, no. 3, B39-B64.

Di, H., L. Truelove, C. Li, and A. Abubakar, 2020, Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand: The Leading Edge, **39**, no. 10, 727-733.

Grahame, J. L., 2015, Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity.

Hart, B. S., 2008, Channel detection in 3-D seismic data using sweetness: AAPG Bulletin, **92**, no. 6, 733-742.

Holbrook, J., W. J. Autin, T. M. Rittenour, S. Marshak, and R. J. Goble, 2006, Stratigraphic evidence for millennial-scale temporal clustering of earthquakes on a continental-interior fault: Holocene Mississippi River floodplain deposits, New Madrid seismic zone, USA: Tectonophysics, **420**, no. 3-4, 431-454.

Honorio, B. C. Z., A. C. Sanchetta, E. P. Leite, and A. C. Vidal, 2014, Independent component spectral analysis: Interpretation, **2**, SA21-SA29.

Hyvärinen, A., and E. Oja, 2000, Independent component analysis: Algorithms and applications: Neural Networks, **13**, 411-430.

Kaib, M. T. H., A. Kouadri, M. F. Harkat, A. Bensmail, and M. Mansouri, 2023, Improving kernel PCA-based algorithm for fault detection in nonlinear industrial process through fractal dimension: Process Safety and Environmental Protection, **179**, 525-536.

Kamp, P. J., A. J. Vonk, K. J. Bland, R. J. Hansen, A. J. Hendy, A. P. McIntyre, ..., and C. S. Nelson, 2004, Neogene stratigraphic architecture and tectonic evolution of Wanganui, King Country, and eastern Taranaki Basins, New Zealand: New Zealand Journal of Geology and Geophysics, **47**, no. 4, 625-644.

Kamaruzaman, E. H., A. D. La Croix, and P. J. Kamp, 2023, Quantitative seismic geomorphology of sediment conduits on an evolving Miocene slope in Taranaki Basin (New Zealand): The influence of increasing slope gradient through time: Marine and Petroleum Geology, **152**, 106233.

King, P. R., and P. H. Robinson, 1988, An Overview of Taranaki Region Geology, New Zealand: Energy Exploration & Exploitation, **6**, no. 3, 213-232.

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**, 105427.

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 Geoscience & Energy, SEG, Expanded Abstracts, 1191-1195.

Leila, M., I. El-Sheikh, A. Abdelmaksoud, and A. A. Radwan, 2022, Seismic sequence stratigraphy and depositional evolution of the Cretaceous-Paleogene sedimentary successions in the offshore Taranaki Basin, New Zealand: implications for hydrocarbon exploration: Marine Geophysical Research, **43**, no. 2, 23.

Li, Q., S. Yu, W. Wu, L. Tong, and H. Kang, 2017, Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: A case study from the Taranaki Basin, New Zealand: New Zealand Journal of Geology and Geophysics, **60**, no. 3, 199-208.

Liner, C., 2016, Elements of 3D Seismology: Society of Exploration Geophysicists.

Lubo-Robles, D., and K. J. Marfurt, 2019, Independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: Interpretation, **7**, no. 3, SE19-SE42.

Lubo-Robles, D., H. Bedle, K. J. Marfurt, and M. J. Pranter, 2023, Evaluation of principal component analysis for seismic attribute selection and self-organizing maps for seismic facies discrimination in the presence of gas hydrates: Marine and Petroleum Geology, **150**, 106097.

Marfurt, K. J., 2018, Seismic Attributes as the Framework for Data Integration Throughout the Oilfield Life Cycle: Society of Exploration Geophysicists.

McInnes, L., J. Healy, and J. Melville, 2018, Umap: Uniform manifold approximation and projection for dimension reduction: arXiv preprint arXiv:1802.03426.

Mika, S., B. Schölkopf, A. Smola, K. R. Müller, M. Scholz, and G. Rätsch, 1998, Kernel PCA and de-noising in feature spaces: Advances in Neural Information Processing Systems, **11**.

Naewboonnien, K., 2021, The Use of Seismic Attributes for Submarine Channel Interpretation in Deepwater Taranaki Basin, New Zealand: Bulletin of Earth Sciences of Thailand, **11**, no. 2, 134-143.

Nicol, A., V. Stagpoole, and G. Maslen, 2004, Structure and petroleum potential of the Taranaki fault play: New Zealand Petroleum Conference Proceedings, 7-10.

Ortiz-Sanguino, L., H. Bedle, M. J. Pranter, and S. Verma, 2022, Lithology variability and channel geomorphology of the mixed carbonate-siliciclastic San Andres and Grayburg interval using frequency-based seismic attributes, southwestern Midland Basin: SEG International Exposition and Annual Meeting, SEG, Expanded Abstracts, D011S023R001.

Posamentier, H. W., V. Paumard, and S. C. Lang, 2022, Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data: Earth-Science Reviews, **228**, 103963.

Qi, J., T. Lin, T. Zhao, F. Li, and K. Marfurt, 2016, Semisupervised multiattribute seismic facies analysis: Interpretation, **4**, no. 1, SB91-SB106.

Ray, A. K., R. Khoudaiberdiev, C. Bennett, P. Bhatnagar, A. Boruah, R. Dandapani, ..., and S. Verma, 2022, Attribute-assisted interpretation of deltaic channel system using enhanced 3D seismic data, offshore Nova Scotia: Journal of Natural Gas Science and Engineering, **99**, 104428.

Reilly, C., A. Nicol, J. J. Walsh, and H. Seebeck, 2015, Evolution of faulting and plate boundary deformation in the Southern Taranaki Basin, New Zealand: Tectonophysics, **651**, 1-18.

Roberts, A., 2001, Curvature attributes and their application to 3 D interpreted horizons: First Break, **19**, no. 2, 85-100.

Roden, R., T. Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **3**, no. 4, SAE59-SAE83.

Strogen, D. P., H. Seebeck, A. Nicol, and P. R. King, 2017, Two-phase Cretaceous–Paleocene rifting in the Taranaki Basin region, New Zealand; implications for Gondwana break-up: Journal of the Geological Society, **174**, no. 5, 929-946.

Strogen, D. P., K. E. Higgs, A. G. Griffin, and H. E. Morgans, 2019, Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression: Geological Magazine, **156**, no. 10, 1751-1770.

Stroud, B., Barber, G., & Bedle, H., 2022. Multiattribute analysis of channel architectural elements within the deepwater Taranaki Basin, 3D Romney data. Interpretation, **10**, no. 3, C13-C17.

Subrahmanyam, D., and P. H. Rao, 2008, Seismic attributes-a review: 7th International Conference and Exposition on Petroleum Geophysics, Expanded Abstracts.

Taner, M. T., J. S. Schuelke, R. O'Doherty, and E. Baysal, 1994, Seismic attributes revisited: SEG Technical Program Expanded Abstracts, **36**, 1104-1106.

Taner, M. T., J. Walls, and G. Taylor, 2001, Seismic Attributes, Their Use in Petrophysical Classification: 63rd EAGE Conference & Exhibition, Extended Abstracts.

Thrasher, G. P., 1992, Late Cretaceous Geology of Taranaki Basin, New Zealand: Ph.D. thesis, Victoria University of Wellington.

Verma, S., and S. Bhattacharya, 2021, Unsupervised Machine Learning for Interpreting Shelf-to-Basin Seismic Geomorphology and Paleoclimate: 82nd EAGE Annual Conference & Exhibition, EAGE, Extended Abstracts, 1-5.

Wang, Q., 2012, Kernel principal component analysis and its applications in face recognition and active shape models: arXiv preprint arXiv:1207.3538.

Yenugu, M., K. J. Marfurt, and S. Matson, 2010, Seismic texture analysis for reservoir prediction and characterization: The Leading Edge, **29**, no. 9, 1116-1121.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29-SAE58.

Chapter 3: enhancing multiattribute seismic facies analysis: the effects of dimension reduction on unsupervised clustering*

*This study is in review in SEG Interpretation

ABSTRACT

The integration of machine learning (ML) into 3D seismic workflows has shown improved geologic interpretability of deepwater channels in the Taranaki Basin, New Zealand. The integration of two novel dimension reduction (DR) techniques, kernel principal component analysis (K-PCA) and uniform manifold approximation and projection (UMAP), has shown improved interpretability for these channel systems. K-PCA yielded higher-resolution seismic images for identifying internal channel architectural elements (AE), whereas UMAP demonstrated improved resolution for identifying lateral AE when compared to an established DR technique, principal component analysis (PCA). Building on these findings, we shift our focus to determine how the implementation of dimension reduction (DR) on seismic attributes before conducting unsupervised clustering affects cluster performance. Our results demonstrate that pre-DR unsupervised k-means clustering produces non-unique clusters that exhibit low statistical performance, while post-DR clustering with K-PCA and UMAP shows noticeable improvement in both statistical metrics and visual separation of clusters. The interpretation of cluster performance requires careful consideration of dimensional perspective; data that appears highly overlapped in 2D visualizations often reveals distinct cluster separation and better cohesion when examined in 3D space. We propose a balanced evaluation framework that integrates both quantitative metrics (silhouette score and Davies-Bouldin Index) with a qualitative geologic assessment to ensure meaningful interpretation. We aim to present a workflow for adopting and utilizing advanced statistical methods from a geoscience perspective. As ML adoption grows within the field, workflows and interpretational considerations presented by geoscientists for geoscientists become increasingly valuable.

INTRODUCTION

Implementing novel dimension reduction (DR) techniques to 3D seismic data shows marked improvement in the interpretability of complex deepwater channel systems (Moreno-Ward et al. 2025). We build upon these findings through a detailed comparison of an established DR technique (PCA) against two novel methods (K-PCA and UMAP), for seismic facies interpretation (see Carter 1995 for a detailed description of seismic facies), including their effect on unsupervised clustering performance.

While seismic amplitude serves as a foundational method for the initial identification of geologic features, and readily available enhancements such as modification of color scales, lighting, and bump mapping can accentuate subtle features, more advanced techniques are essential to continue improving accurate and detailed interpretations. Co-rendering attributes is a natural next step; however, the co-rendering of multiple seismic attributes is constrained by practical limits, including the number of attributes overlain and the potential compromised clarity of visualization that can impede an individual's analysis.

These limitations further indicate the need to incorporate DR into the seismic interpretation workflow. Having demonstrated the rationale and results of improved interpretability in Moreno-Ward et al. 2025, we will delve into a more comprehensive discussion of the clustering results for three DR techniques in this paper.

Geologic Background

The area of study, in the Taranaki Basin located off the northwest coast of New Zealand (Figure 1, A-C), has undergone extensive geoscience study. More detailed information on this Mesozoic age rift basin can be found by reviewing the geologic research (Bull et al. 2020; Di et al. 2020; Grahame 2015; Kamp et al. 2004; King and Thrasher 1992, King 1988; Kroeger et al. 2019; Naewboonnien 2021; Quan et al. 2017; Reilly et al. 2015; Strogen et al. 2017; Strogen et al. 2019) or machine learning research (Chopra et al. 2022; Lubo-Robles 2019; Li. et al. 2023; La Marca 2022; La Marca 2020; Chopra et al. 2019; Verma 2021). Using a well-researched geologic

location as a test site for unsupervised machine learning minimizes uncertainties since the area's geologic history, processes, and features are well characterized.

Data

The Romney 3D survey, acquired in 2011 and covering an area of 35 km by 55km, has a bin size of 12.5 m by 25 m with a sample rate of 4ms. There is one well located within the seismic volume, but outside the channel of interest: the Romney-1.

The Romney-1 well was evaluated, and 5 distinct lithologic packages were identified, A-E. These packages were identified in the seismic data, post well-tie, and the seismic response was described. Using these well- and seismic-facies, we interpolate into the channel of interest (Figure 1, D-F). With the interpolation and our understanding of channel morphology, we can interpret channel architecture and use this information to evaluate the clustered models in subsequent steps. Next, we will review the attributes chosen.

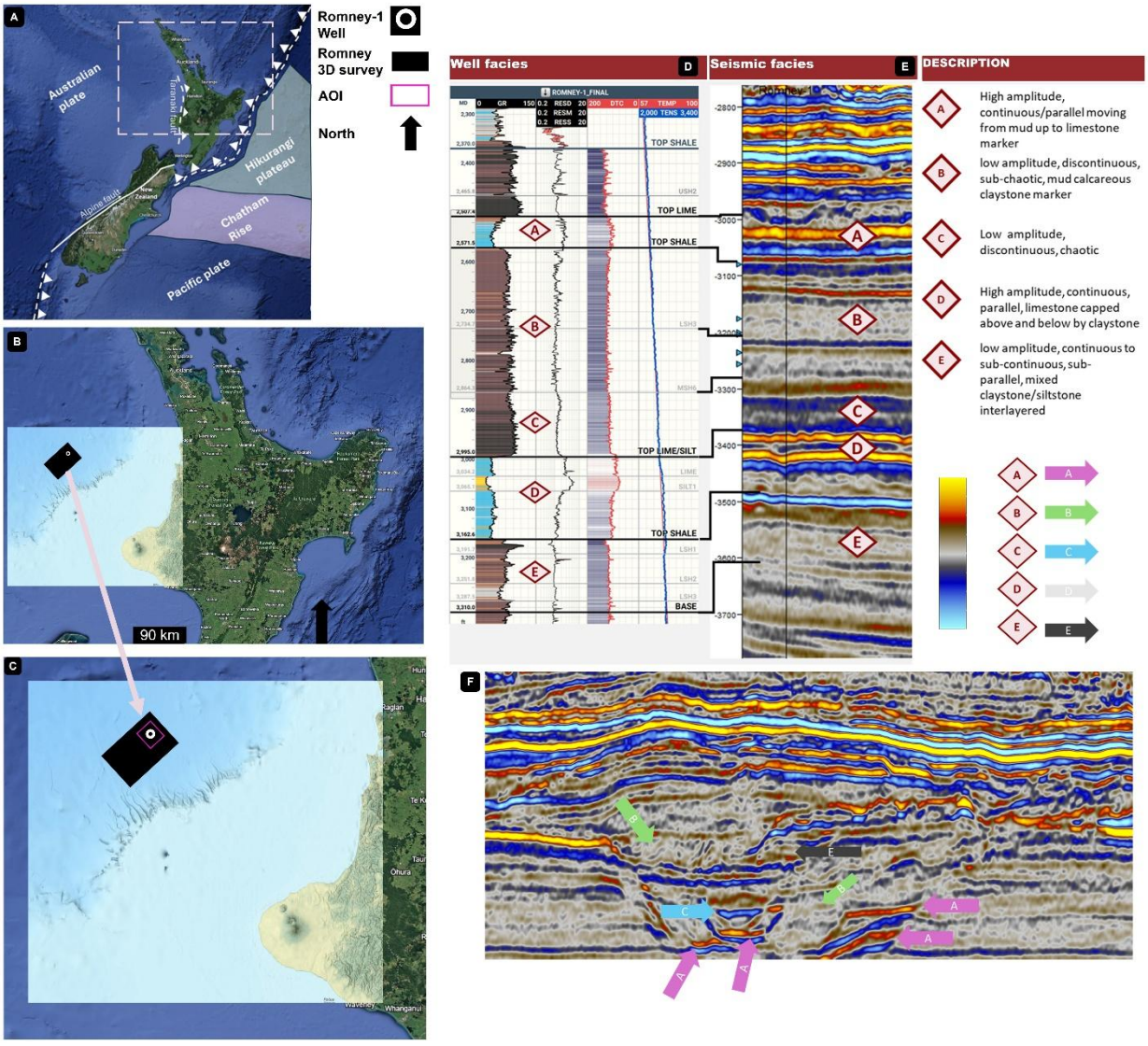


Figure 1: Romney 3D seismic survey location. A) Area of study located in deepwater off the northwest shelf of New Zealand's north island. With main faults for reference. Light pink rectangle, zoom to: B) black rectangle indicating Romney 3D seismic volume extents. C) A white circle indicates the location of the Romney-1 well and a dark pink polygon indicates cropped volume extents. D) Using Romney-1 well log, well facies are identified, E) and are used to identify seismic facies. F) Next, the seismic facies are interpolated into the channel of interest.

Machine learning in seismic workflows

As multi-attribute analysis is limited by the number of attributes that can be simultaneously compared, co-rendered, and clustered, DR techniques play a vital role. They help to uncover attribute combinations that define subtle geologic features in seismic data and help to identify patterns, trends, and relationships that may not be readily apparent, facilitating a deeper understanding of subsurface geology. This reduction of dimensional space and enhancement of pattern identification is apparent when evaluating cluster results. We will briefly recap the DR techniques that will be discussed.

The algorithms, a quick review

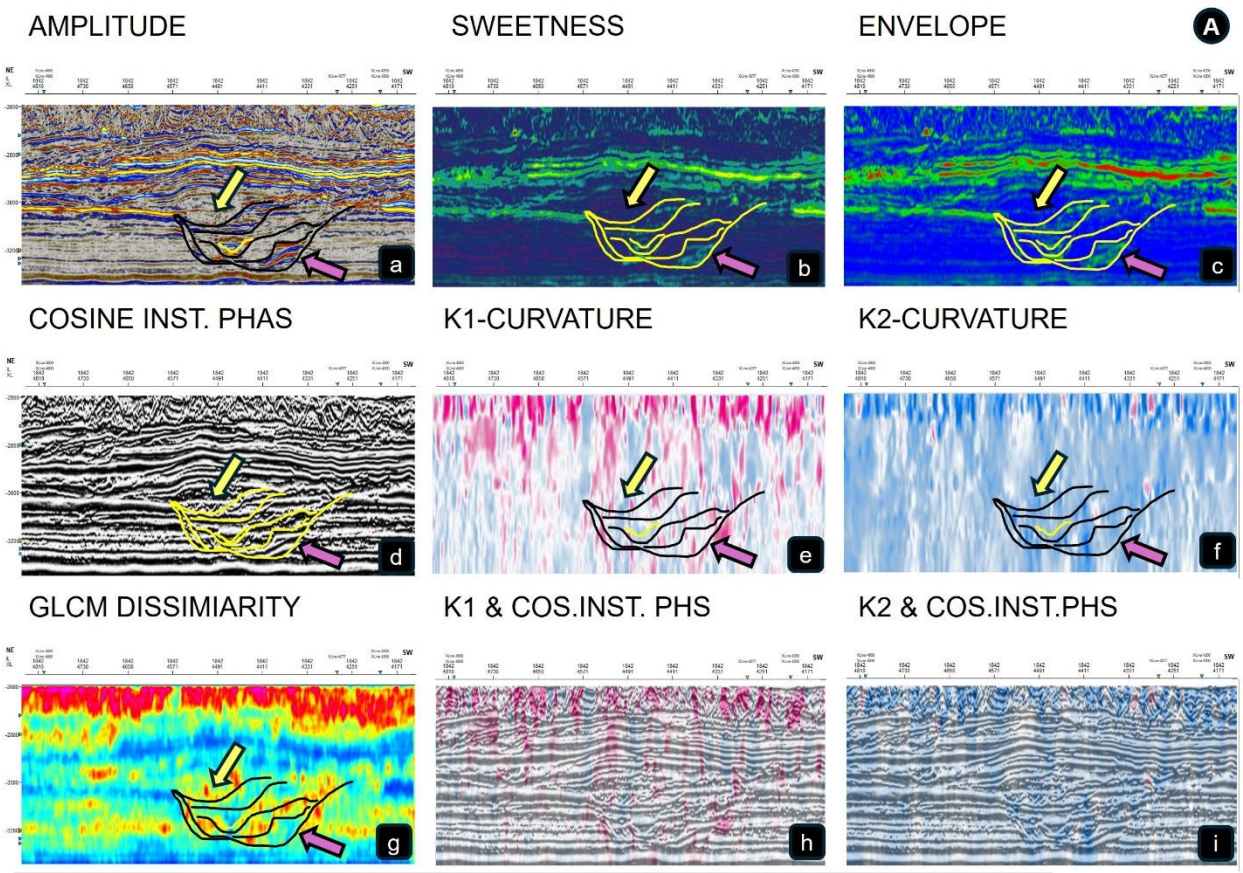
PCA emphasizes variation and captures strong patterns in data by identifying orthogonal principal components. At the same time, **K-PCA and UMAP** excel in transforming complex seismic data into a more interpretable representation, thereby enhancing the accuracy of subsequent processes, such as clustering, and resulting in more meaningful geologic data (Table 1). We previously evaluated the effectiveness of each technique for improving and/or preserving geologic interpretability; we now focus on how these techniques improve cluster performance. We will use this information to address our second project goal: to evaluate the effectiveness of pre-clustering DR in enhancing cluster performance. Let's briefly review the chosen seismic attributes next.

	PCA	KPCA	UMAP
Full Name	**Principal Component Analysis**	**Kernel Principal Component Analysis**	**Uniform Manifold Approximation and Projection**
Type	Dimensionality Reduction		
Function	Reducing dimensionality while preserving variance	Reducing dimensionality using kernel trick	Reducing dimensionality while preserving local structure
Output	Principal components (linear combinations of features)	Principal components in a transformed space	Lower-dimensional embedding of data
Assumptions	Linear relationships in data	Data can be mapped to higher dimension	Data lies on a low-dimensional manifold
Scalability	Good with large datasets	Moderate	Good with large datasets
Interpretability	High	Moderate	Moderate to low
Non-linear handling	Poor	Good	Excellent
Parameters	# of components	# of components, kernel type, kernel parameters	# of neighbors, Min distance, # of components
Strengths	Easy to interpret, reduces overfitting	Captures complex patterns, non-linear relationships	Preserves local data structure, effective on complex data
Weaknesses	Assumes linear relationships, can miss non-linear structures	Computationally intensive, sensitive to kernel choice	Sensitive to hyperparameters, can be computationally intensive

Table 1: Summary of algorithms discussed in this manuscript.

Attributes for channels: an overview

Each attribute was chosen for its ability to help delineate channel features, from edges (cosine instantaneous phase, k1 and k2 curvature) to within/outside channel features (sweetness, envelope, and GLCM dissimilarity). Figure 2A illustrates each attribute's ability to guide the identification of geologic features in a cross-sectional view of our channel of interest and the application of simple co-rendering. Figure 2B summarizes the use and additional reference material for each attribute.



B

Attribute	Use	Additional Information
Cosine instantaneous phase	illustrates continuity of the seismic traces	Barnes 2007
Envelope	measures the strength or intensity of the seismic amplitude	Subrahmanyam 2008; Chopra 2007
GLCM dissimilarity	measures changes in seismic amplitude often associated with structural changes and/or lithologic boundaries	Yenugu et al. 2010; Qi et al. 2016
Curvature: Most positive and negative	Measure the lateral angle change in the reflector; can indicate presence of channels	Chopra and Marfurt 2007
Sweetness	Instantaneous attribute to enhance bright spots	Hart 2008, Li 2017

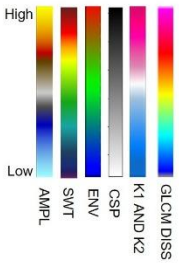


Figure 2: A) Attributes selected for ML are displayed in a cross-sectional view to visualize their ability for channel feature interpretation. a) amplitude b) sweetness c) envelope d) cosine instantaneous phase e) k1 curvature f) k2 curvature g) GLCM dissimilarity. h) cosine instantaneous phase with k1 curvature and i) cosine instantaneous phase with k2 curvature demonstrate co-rendering attributes to improve interpretation. B) Table for a summary of attribute use and reference for more in-depth review.

ML Considerations in Geoscience

As geoscientists become more versed in ML applications, further studies are needed to determine how these advanced techniques can improve interpretation workflows. In our quest to adopt these into our workflows, can we maintain the critical balance between statistical performance and geologic interpretability? While ML may appear as a quick solution to analyzing seismic data, interpreters must first have background knowledge of the geologic system and a fundamental understanding of the application and output of the methods they choose to use.
Preliminary questions and considerations before the implementation of ML should be, 1) what the feature of study is, 2) what the quality of the input data is, 3) have the appropriate seismic attributes been selected, 4) are the input data linear or non-linear, and 5) how the output data can be effectively evaluated.

The continued improvement of geoscience workflows is a critical part of ensuring understanding and interpretation. As ML becomes more widely available and mainstream for geoscientists, we would like to offer a perspective by geoscientists for geoscientists. We aim to develop an accessible workflow that provides a framework to integrate both quantitative and qualitative evaluations, making advanced statistical algorithms more approachable for geoscientists. While the use of algorithms such as PCA is quite common in current seismic workflows, considering the complexity of the data, are these the best algorithms for the application? The goal for this deep dive into K-PCA and UMAP applications in seismic data is to advance our understanding of ML's role in future seismic workflows and make the use of these advanced techniques more accessible for geoscientists.

Ensuring proper utilization of these ML techniques in our workflows requires an understanding of how they work, their limitations, and proper application to develop a more informed group of users within our field. Additionally, there must be a balance between utilizing ML outputs and incorporating the in-context interpretation from the interpreter's knowledge and training. Next, we will discuss the methodology used to evaluate the performance of clusters pre- and post-DR from both the quantitative and qualitative perspectives.

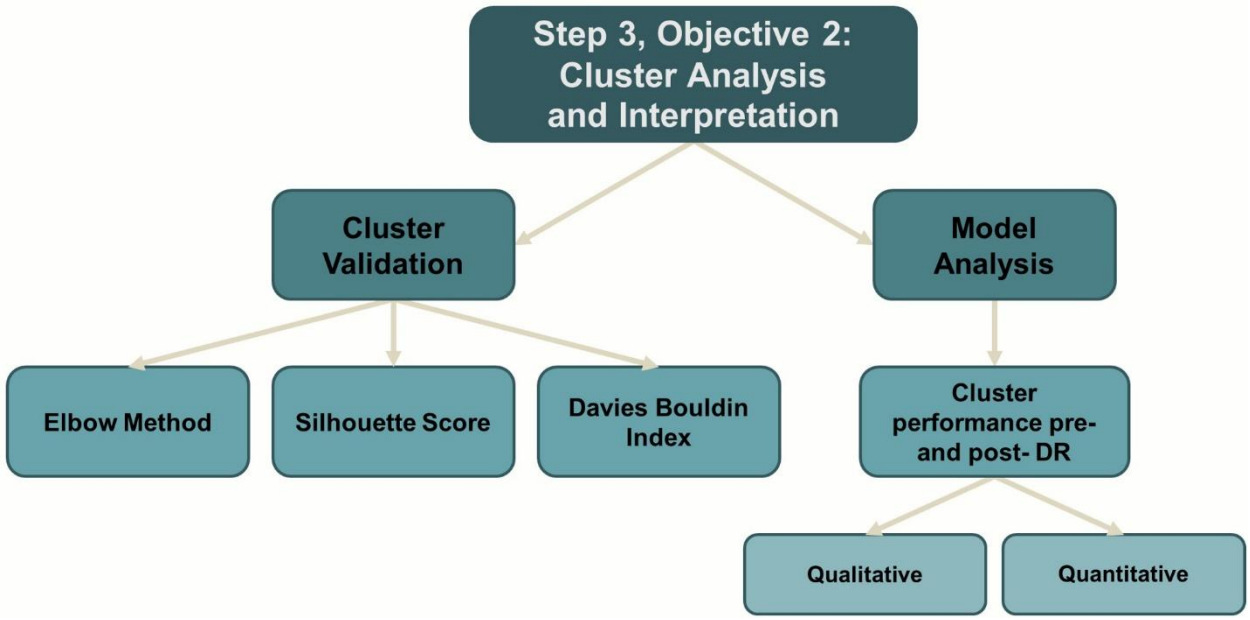


Figure 3: Cluster analysis and interpretation workflow.

METHODOLOGY

Let us review some of the steps necessary prior to beginning unsupervised clustering. Pre-processing for advanced analysis includes the following: define the area of interest (AOI), well-tie(s), lithology and facies identification (where data allow), and proper attribute selection. Next, we determine those ML algorithms best suited to the project objective and conduct an initial evaluation of cluster results using the original selection of attributes only. Observing less optimal k-means clustering, DR techniques selected for this AOI were applied to the original seismic attributes, and those results were evaluated for geologic interpretability in cross-sectional and map view (Moreno-Ward et al. 2025).

Building on these initial steps, we proceed with the next stage of our workflow with a thorough evaluation of unsupervised k-means clustering results (Figure 3). This evaluation includes the rationale for selecting k-means for clustering, the suboptimal performance pre-DR necessitating the implementation of DR methods, and the post-DR effect on clustering results. A comprehensive analysis of cluster results will follow; for now, let's discuss each step of the machine learning workflow.

Rationale for k-means

We selected unsupervised k-means clustering, a popular method for partitioning datasets into k-distinct, non-overlapping clusters (for additional information on the use of k-means, see Haibin 2018; Talebi 2020; Zhao 2015; and Troccoli 2022) where each data point is assigned to the cluster with the closest centroid.

Scaling the Data: Before dimension reduction and clustering, the attribute data is plotted to understand the data distribution. With the large variance in scales across the selected attributes, the data is scaled to avoid over- or under-weighting each data point. The data distributions were reviewed under five conditions: (1) pre-scaling; (2) post min-max scaling; (3) post min-max scaling with an outlier removed from GLCM dissimilarity; (4) post robust scaling; (5) post robust scaling with an outlier removed from GLCM dissimilarity. The data distributions for each do not change from method to method, apart from a slight shift from removing the outlier. For the statistical analysis being conducted, a min-max scaler was chosen. Let's continue with the clustering algorithm next.

For the implementation of k-means, we utilize the k-means++ algorithm, as it improves upon the standard k-means by strategically choosing initial centroids (Arthur and Vassilvitskii 2007). This approach significantly reduces the chances of suboptimal cluster assignments by lowering the within-cluster variance from the outset, as it enhances the efficiency and efficacy of the clustering process, decreasing the likelihood of poor cluster assignments. This technique is particularly effective for the unsupervised exploration of facies and their boundaries in seismic data, which may not be initially clear. Next, the clustering technique outcomes are evaluated to help determine the optimal number of clusters.

Scatterplots: We utilize scatterplots to understand cluster distribution patterns and outliers before clustering. We noted a single data point in the GLCM dissimilarity attribute that plotted away from the rest of the spread and removed it as an outlier. Post initial k-means clustering, pre-dimension reduction, the scatterplot cluster analysis revealed suboptimal cluster performance, prompting the implementation of DR techniques into our workflow (Figure 4). Applying k-means clustering post-DR shows a marked improvement in clustering efficacy; those results will be evaluated and discussed later. For each clustering approach, both pre- and post-DR, statistical

evaluation using the silhouette score and Davies-Bouldin Index (DBI) is performed and discussed next.

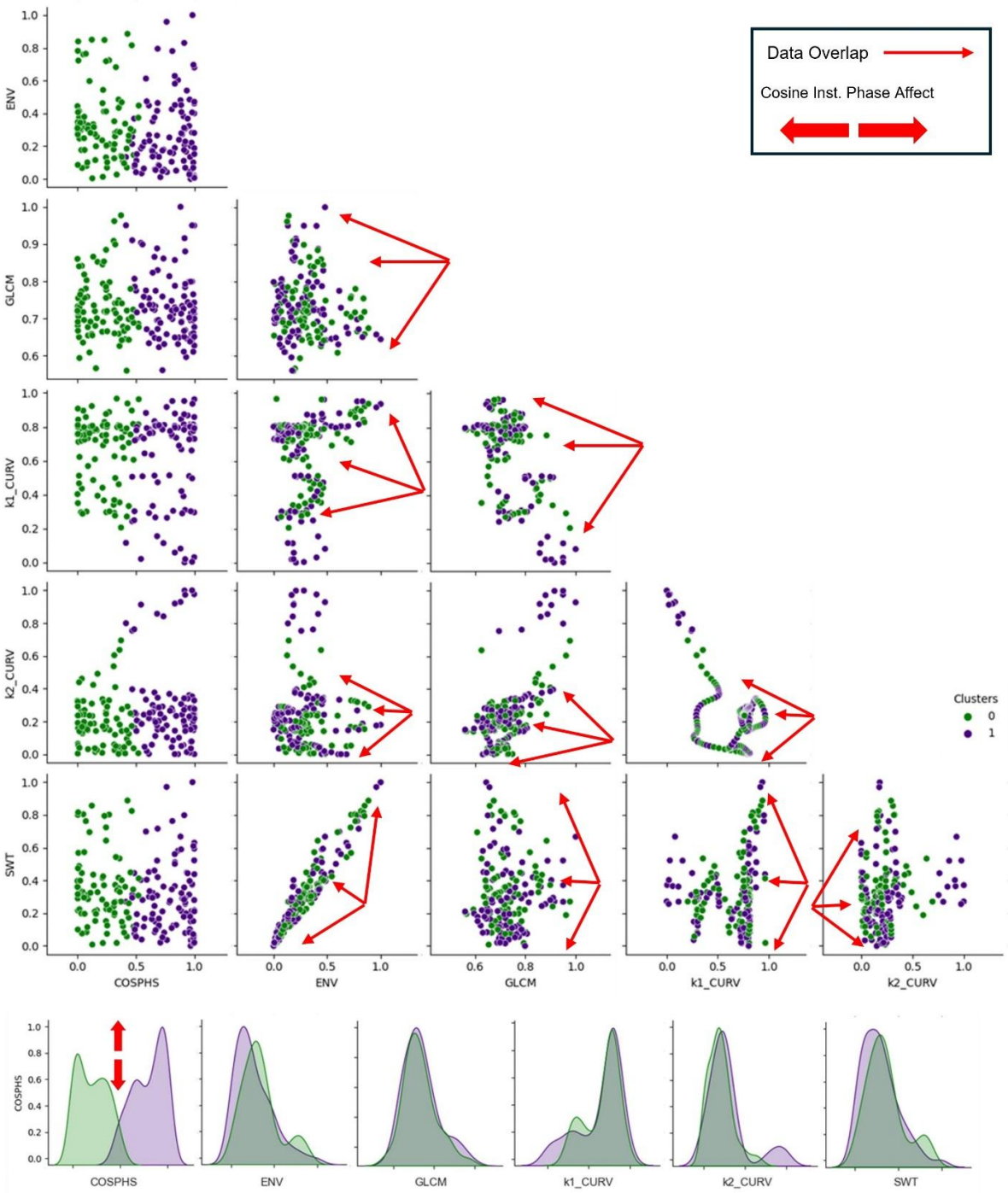


Figure 4: Unsupervised clustering (2 clusters) of sweetness, cosine instantaneous phase, envelope, GLCM dissimilarity, and k1 and k2 curvature before DR. Scatterplots with cosine phase (column indicated by red arrow) show distinct separation in data; this trend is shown clearly in the histogram. Other attribute plots show little to no cluster separation. Histograms indicate a similar distribution of cluster data. The plot for sweetness and envelope may indicate related data; however, they are not correlative, rather, they plot with similar trends. Sweetness is defined as the envelope (or instantaneous amplitude) ÷ √instantaneous frequency, while envelope is a measure of the magnitude of the amplitude. Using data trends alone could result in excluding data that may appear to be redundant.

Silhouette Score and Davies-Bouldin Index (DBI)

To optimize the number of k-means clusters, we applied two principal methods: the silhouette score and DBI. These approaches have a critical role in identifying the optimal number of clusters, although the ultimate decision lies with the interpreter. While both methods provide a robust statistical framework for cluster selection, the final choice should prioritize the clarity and geologic relevance of the results.

Interpreters must carefully evaluate statistical characteristics derived from dimension reduction in conjunction with the geologic features of seismic data, given their inherent complexity and nuance. This ensures that the selected clusters are not only statistically justified but also geologically meaningful, accurately reflecting the subsurface heterogeneity and aiding in effective exploration and development activities.

Overview

The silhouette score evaluates how close an object is to its cluster (cohesion) relative to other clusters (separation) (Rousseeuw 1987), while the DBI evaluates the compactness of each cluster and intra-cluster similarity scoring with positive values (Davies and Bouldin 1979; Mughnyanti et al. 2020). The silhouette score can suggest a smaller number of clusters if it finds that a few broad groups have high internal similarity and clear distinction from each other. We chose DBI as a validation metric to help ensure the clusters selected as high performers are statistically robust. See Table 2 for a summary of these two methods.

Metric	Silhouette Score	Davies-Bouldin Index (DBI)
Purpose	Measures how well a point fits in its cluster vs. other clusters	Measures cluster compactness and inter-cluster similarity
Range	**-1 to 1**	**Postive values**
	-1 = wrong cluster 0= edge of cluster 1 = fits well within the cluster	Lower Values = Better Clusters
Takeaway	Evaluates cluster cohesion and separation	Evaluates cluster compactness and separation
Key Focus	Quatlity of cluster	Quality of cluster
Difference	Evaluates each data point and how it fits within its cluster	Evaluates the overall structure by averaging across all clusters rather than focusing on each datapoint.

Table 2: Summary of statistical scoring methods for clustering results. Common measure is the silhouette score; the authors have included the Davies-Bouldin Index to validate or invalidate the silhouette score.

Following the statistical evaluation, scatterplots and boxplots were generated to enhance the visualization of the models' clustering efficiency. These visual aids, coupled with the Silhouette Score and DBI, serve as essential tools for confirming the models' statistical precision before advancing in the workflow.

We will now move into a discussion of results, beginning with reviewing results for the statistical evaluation of clustered models, followed by an explanation of the DR input measures, and ending with a review of top performers. Next, we will review the results from a visual analysis and discuss the role of dimensional perspective before moving to the observations of clustering pre- and post-DR. We will end our results and discussion section by reviewing the performance of K-PCA and UMAP and providing the rationale for including DR before unsupervised clustering. Let us move to the results and discussion now.

RESULTS AND DISCUSSION

Quantitative Evaluation (Statistical)

Quantitative evaluation (QE) is a multi-step process. First, the proper selection of algorithms for the task is key, as is understanding what each measure is and how to interpret the data (Table 3). Next, we need to establish a need for each method; Figure 5 (A and B) illustrates our process for pre- and post-DR. A) The QE for initial clustering resulted in non-unique clusters, which is likely due to the complexity of the seismic data, indicating a need for DR to improve cluster performance. B) Establishing a need for DR, we determine how each algorithm uses the input data: PCA uses loadings and variance, K-PCA looks at percent contribution and component magnitude, and UMAP uses feature importance. By applying K-PCA and UMAP to resolve our seismic data (reducing dimensionality and enhancing subtle geologic features) before clustering, we expect improved k-means clustering performance, which we evaluate using silhouette scores

Quantitative Evaluation			
Algorithm	**Measure**	**What it means**	**Key Point**
PCA	PCA Loadings	How much each original feature contributes to a principal component	Shows which features are most important.
	PCA Variance	How much of the data's spread is captured by each principal component.	Higher variance = more information captured.
KPCA	Percent Contribution	The % each input contributes to that component	The higher the %, the more that feature contributes
	Component Magnitude	Represents the average absolute magnitude of the transformed data on each component	The values are relative to each other. Higher value indicates more importance.
UMAP	Feature Importance	Which input has the most influence on data projection	Features with **higher scores** → shape UMAP structure the most.
	Permutation Importance	How does the systematic shuffling of data affect the	Higher value may indicate importance for

Table 3: Summary of DR algorithms discussed in this manuscript.

and DBI for validation. The results indicate improved performance. We will look at the QE in more depth now, see Table 4 to reference model labels.

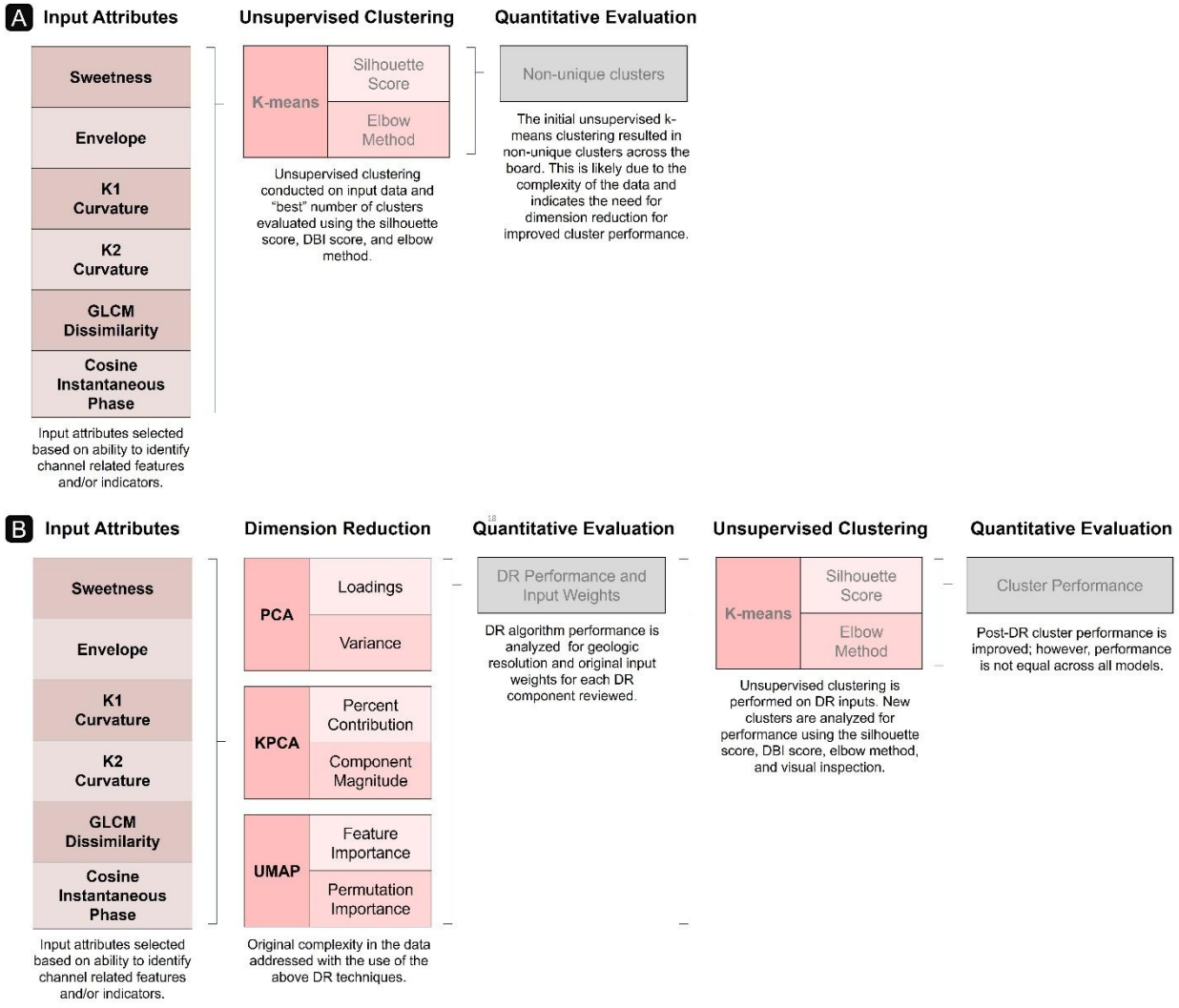


Figure 5: Workflow for quantitative evaluation. A) Evaluation for pre-DR clustering and B) Evaluation for post-DR clustering.

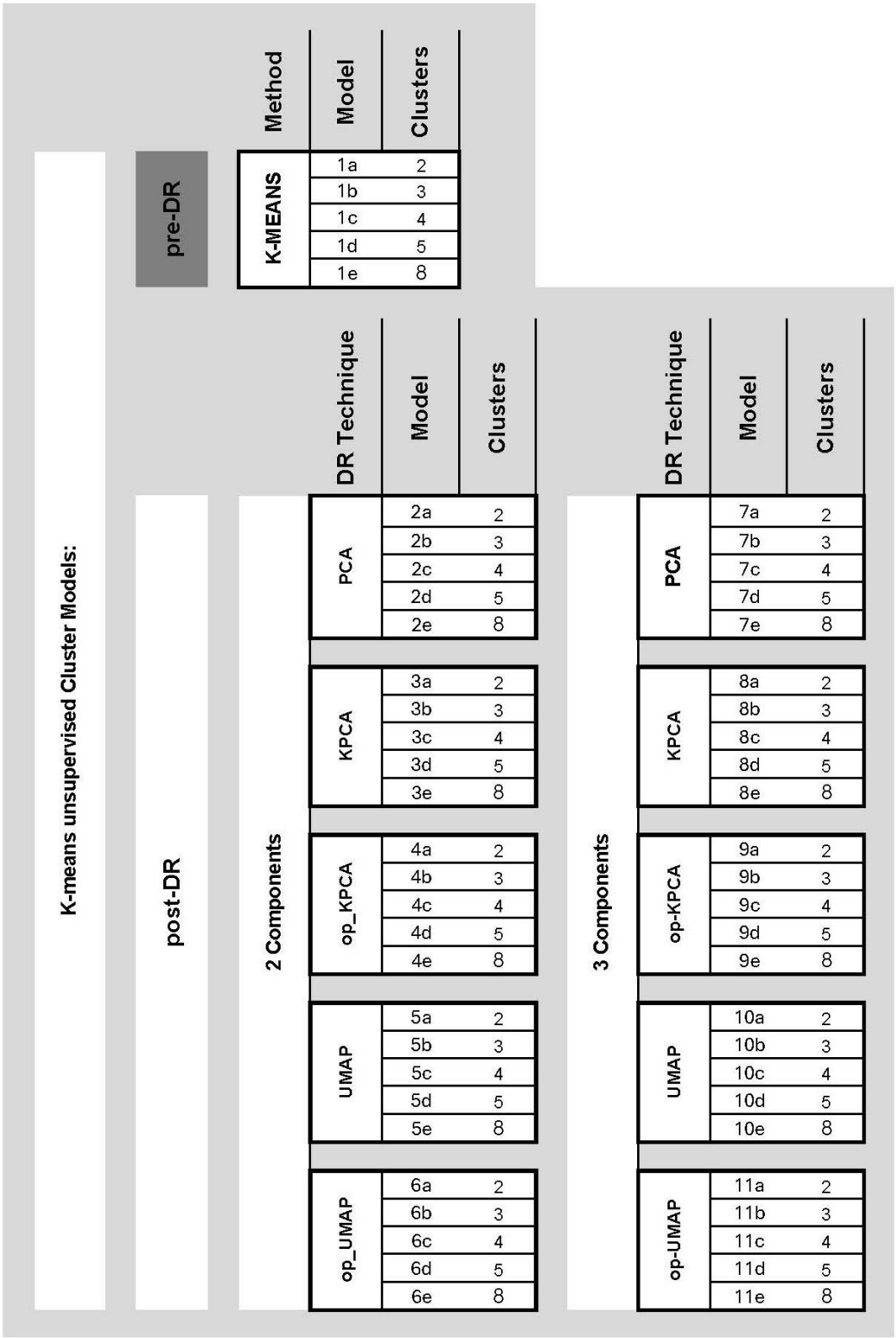


Table 4: Model label reference table.

DR Algorithm measures explained (Statistical)

UMAP

Measure	Feature Indication	Feature Interpretation
Small FI Value	low direct contribution to the **UMAP** projection magnitude.	does not strongly influence the position of data points in the projection.
Large PI Value	crucial for maintaining the overall structure of the projection.	Shuffling this feature disrupts the relationships between data points, high importance for preserving structure.
Overall Meaning	may have a subtle but important role in maintaining the data structure, despite not having a large direct influence.	likely essential for **preserving relationships** in the data (local or global), even if it doesn't appear to contribute large values in the projection.
Possible Cause	might be involved in complex or non-linear relationships between data points.	could maintain **local/global distances** even if its absolute values are small in the projection.

K-PCA

Measure	What It Means	Key Point
Percent Contribution	What % the attribute contributes to each component.	Higher % value indicates a higher importance
Component Magnitude	Indicates the overall strength of a attribute to the component.	Values are relative to each other. Larger magnitudes indicates more importance for preserving patterns and/or variability in non-linear space.

PCA

	Measure	What It Means	Key Point
VARIANCE	***PCA Variance***	How much of the data's spread is captured by each principal component.	Higher variance = more information captured.
	Explained Variance Ratio	The percentage of total variance each principal component captures.	Tells you how much of the data each component explains.
LOADINGS	***PCA Loadings***	How much each original feature contributes to a principal component.	Shows which features are most important.
	Positive Loadings	The feature increases as the principal component increases.	Feature and component move together.
	Negative Loadings	The feature decreases as the principal component increases.	Feature and component move in opposite directions.
	Magnitude of Loadings	The size of the loading (positive or negative) tells how strong the feature's influence is.	Larger magnitude = more influence.

Figure 6: DR algorithms explained. Each algorithm uses the same set of input seismic attributes for reducing dimensionality within the dataset; however, the evaluation of attributes is not the same across all models. Evaluators should understand how each algorithm weights the inputs for the proper determination of attribute importance.

Figure 6 provides a summary of each DR measure.

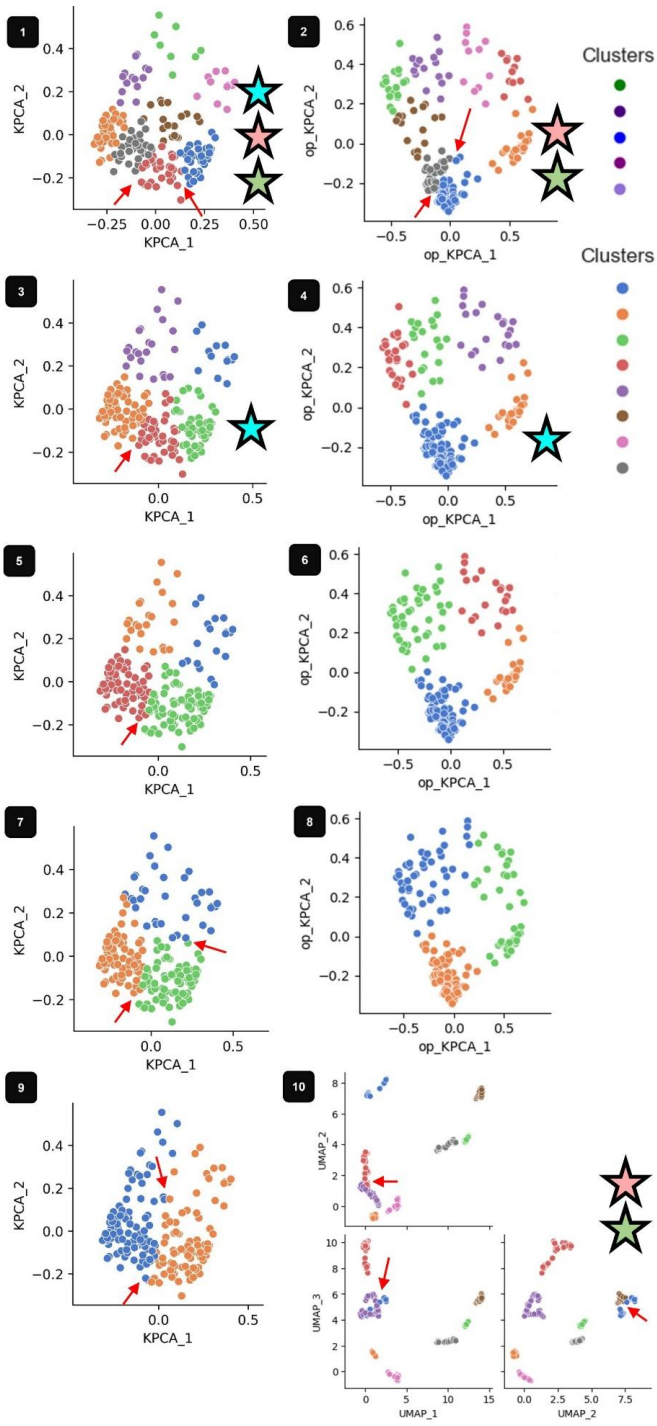
UMAP uses a feature importance (FI) to weight the input attribute data; low FI values indicate that the input has a low overall contribution to the projection magnitude and does not have a strong influence on where the data is projected onto the new dimension. We have cross-validated the FI with permutation feature importance (PI), which measures how the small shuffling of input data can impact projection. Large PI values indicate that the input data is crucial for the structure of the projection. Using these two measures, we can identify inputs that are important for preserving both local and global relationships despite not scoring high in terms of direct influence. K-PCA evaluates data based on the percent contribution and component magnitude. The percentage contribution looks at each attribute's overall contribution to a given component, while the component magnitude indicates more importance for preserving patterns and/or variability in the non-linear space.

PCA evaluates variance and loadings. The variance indicates the amount of data captured by each principal component, with higher variance indicating more information captured. The loadings, which represent the eigenvector of the covariance/correlation matrix, indicate how much each original attribute contributes to a given principal component. Large positive or negative loadings indicate a strong influence. Positive loadings indicate that the attribute and principal component are positively correlated; an increase in the original variable corresponds to an increase in the principal component. Negative loadings indicate an inverse correlation; an increase in the original variable corresponds to a decrease in the principal component. We will review the statistical higher performers next.

Quantitative Higher Performers

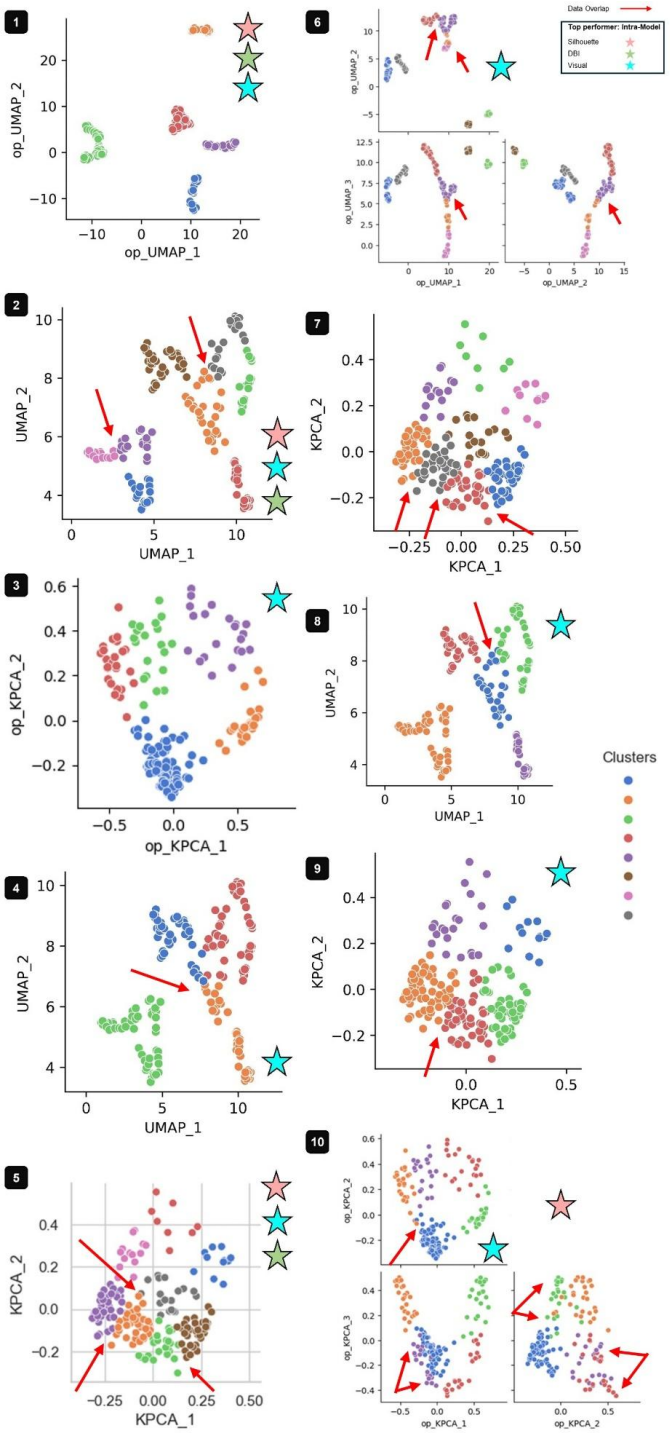
Using both the silhouette and DBI scoring, we constructed 2 sets of ranking tables to evaluate statistical performance. First, each clustering model is grouped per the DR used, and the number of clusters is ranked per the silhouette and the DBI, noting where the two metrics showed agreement and deviated. Overall, the number of clusters with the highest score was 8, followed closely by 4 and 5. Second, all cluster models are combined and ranked highest to lowest, once

using the silhouette score and then with DBI; score rankings with agreement and deviations are noted. Overall, post-DR clustering outperforms pre-DR performance; K-PCA and UMAP outscored both PCA and k-means pre-DR. The highest overall score, using silhouette, was 0.904 for model 3e, while the first highest scoring PCA had a value of 0.451 (model 2b), and the k-means' highest score was 0.352 (model 1c). Using DBI as the metric, the highest performer had a value of 0.145 (also model 3e), with the highest performing PCA model at 0.728 (model 2c) and k-means with a score of 0.906 (model 1d) (DBI higher performance shown by lowest score).



With these metrics in mind, we select the top 10 overall performers, noting where optimization shows distinct improvements, and will discuss these now (Figure 7). The highest-performing DR and clustering combination was observed as model 3e. This model showed high statistical metrics in both silhouette and DBI scoring (denoted by pink and green stars, respectively) and displayed distinct cluster separation without visual overlapping. The second highest performing model, model 8e, which included optimized

Figure 7: Using the silhouette and DBI score for all cluster models and both 2-and-3- components, we evaluate the top 10 scorers. #1 indicates the highest score overall. A pink star indicates models identified as silhouette high performers, and a green star indicates DBI high performers within the post-DR group. Red arrows indicate data overlap. 1) K-PCA 2 components, 8 clusters. 2) K-PCA optimized 2 components, 8 clusters. 3) K-PCA 2 components, 5 clusters. 4) K-PCA optimized 2 components, 5 clusters. 5) K-PCA 2 components, 4 clusters. 6) K-PCA optimized 2 components, 4 clusters. 7) K-PCA 2 components, 3 clusters. 8) K-PCA optimized 2 components, 3 clusters. 9) K-PCA 2 components 2 clusters. 10) UMAP optimized 3 components, 8 clusters.



hyperparameters (OP), indicates that default parameters yielded better statistical performance for K-PCA models according to scoring metrics. Within the top 10 models, UMAP appeared once at rank 10 (model 11e), demonstrating effective cluster separation despite minor data overlapping.

For UMAP models, optimization mostly showed marked improvement. While the highest-performing UMAP model (ranked #10) utilized OP model 11e (silhouette: 0.739, DBI: 0.33), the second-best UMAP performer used default settings, model 5e (ranked 11; silhouette: 0.736, DBI: 0.370). After rank 11, optimization consistently improved performance, as indicated by the OP model 11d (ranked #13; silhouette: 0.695), and then beginning at rank 15, the subsequent five highest ranked models, all consisting of UMAP with OP.

After evaluating performance using statistical metrics, it is important to conduct a visual inspection, ensuring that we are not

Figure 8: Using visual inspection, the top 10 performers across all models are displayed here. A blue star indicates models previously identified as visually having unique clusters, a pink star indicates models identified as silhouette high performers, and a green star indicates DBI high performers within the post-DR group. High performers were chosen from within the DR group. Red arrows indicate data overlap. 1) UMAP optimized 2 components, 5 clusters. 2) UMAP 2 components, 8 clusters 3) K-PCA optimized 2 components, 5 clusters. 4) UMAP 2 components, 4 clusters. 5) K-PCA 2 components, 8 clusters. 6) UMAP optimized 3 components, 8 clusters. 7) K-PCA 2 components, 8 clusters. 8) UMAP 2 components, 5 clusters. 9) K-PCA 2 components 5 clusters. 10) K-PCA optimized 3 components, 5 clusters.

using statistical performance as our only indicator. We will discuss the visual high performers next.

Qualitative Evaluation (Visual)

Qualitative Evaluation (QLE) of cluster uniqueness should focus on two key criteria: distinct separation between clusters (inter-cluster) and compact spacing within clusters (intra-cluster). The top 10 models were selected based on their effectiveness in meeting both requirements (Figure 8).

UMAP, with OP, model 6d is selected as the top performer: additionally reporting high silhouette and DBI scores (denoted by pink and green stars, respectively) within model 6. The second-highest QLE performer, UMAP model 5e (default settings), also scored well within its model group 5 for both silhouette and DBI scores, with minor data overlap noted (red arrows). Models ranked 4, 5, and 7-10 also show well-defined inter-cluster separation with minor data overlap.

Qualitative Dimensional Perspective

In this discussion regarding QLE, it is important to note the concept of dimensional perspective when visually evaluating cluster performance. While seasoned evaluators instinctively consider the dimensionality of data, newer users may default to basic 2D plots, missing the true relationships of data that higher-dimensional visualizations can reveal. Figure 9 illustrates the difference in perceived cluster performance based on visual perspective: A) shows the post-DR for UMAP model, while B) shows UMAP with OP model. With a quick visual perspective of the 2D plot for A), we see what appears to be a significant overlap indicated by a red arrow. Looking at the 3D plot, we see that there is, in fact, no data overlap, as indicated by the black arrows. B) While there was no indicated data overlap in the 2D visualization, the 3D perspective should still be evaluated.

Figure 9 illustrates an easy-to-see difference between 2D and 3D perspectives, but what if the model is more complicated? Figure 10 evaluates K-PCA and UMAP, 3 components and 8 clusters. In parts a and b, K-PCA, we initially see clusters that have distinct overlaps (red arrows),

leading to the conclusion that both the intra- and inter-cluster performance is low. When we re-evaluate in 3D, we are better able to delineate a truer sense of the separation in clusters; however, again, depending on the angle we view the data, we may get a perception of data overlap. In parts c and d, we observe perceived data overlap (red arrow) for our non-optimized model; however, again, when we re-evaluate in 3D, we see that there is, in fact, no data overlap. Using OP (part d), we see clusters that are fairly compact and separate, but we still want to view them in 3D to verify. We will summarize our findings of QE vs. QLE next.

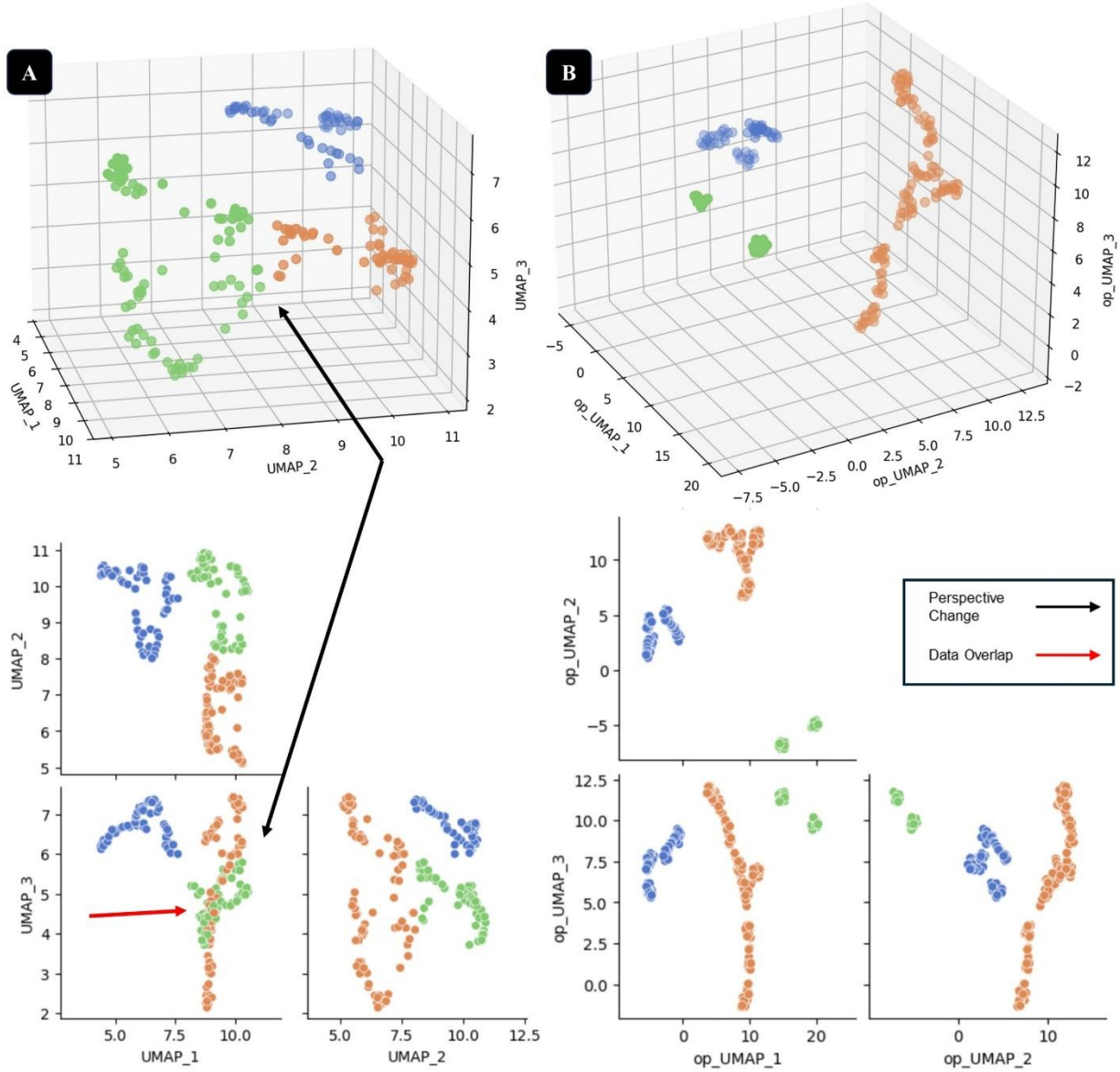


Figure 9: UMAP 3 components, A) without optimized hyperparameters and B) with optimized hyperparameters. In A), if we visualize the clusters in 2D only, there appears to be an overlap between the green and orange clusters (red arrow), however, if we render this model in 3D we see there is in fact no overlap in the data. B) With the optimized hyperparameters, there is no indication of cluster overlap in either the 2D or 3D view and the clusters have more separation between them.

Quantitative vs. Qualitative Summary

When comparing QE vs. QLE high performers, we do see a significant difference in the DR models chosen. While K-PCA dominated the QE rankings (occupying 9 of the top 10 spots),

UMAP models show stronger performance in visual assessment. UMAP model 5d with OP ranked first in visual quality and thirteenth in statistical metrics, achieving high silhouette and DBI scores within its model group. Similarly, UMAP model 5e (default hyperparameters) ranked second visually and eleventh in statistical metrics, also showing strong statistical performance within its model group.

The contrast between quantitative and qualitative rankings highlights some additional complexity and considerations required for comprehensive cluster evaluation. Full cluster performance evaluation from a human interpreter using QLE methods alone would require additional time. Each data point would need to be viewed in multiple dimensions, and the spacing between data points should be measured, as would the spacing between clusters, to get a true sense of inter-, intra-, and dimensional perspective spacing of data points and clusters. The ability to leverage ML in this workflow decreases the time required to make those calculations; however, we must still consider the human interpreter's contextual knowledge when selecting the models that are deemed as the highest performers. Are these models in fact improving our ability to interpret geologic features, and do they help validate our interpretations?

We will summarize the pre- and post-cluster results for each set of models next.

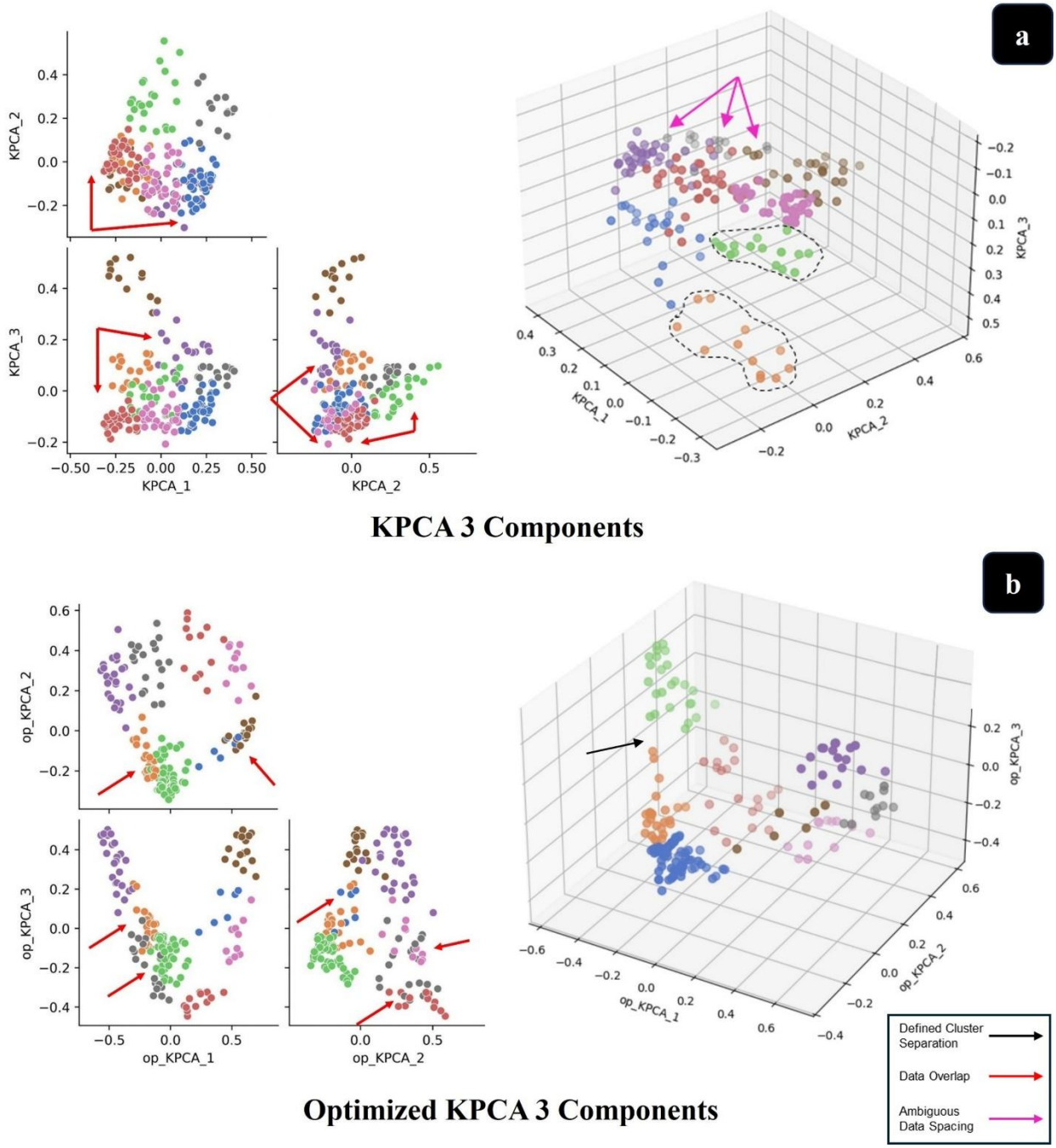


Figure 10: K-PCA 3 components, a) without optimized hyperparameters and b) with optimized hyperparameters. In both examples, when viewed as a 2D graph, we note areas where clusters appear to be overlapping (red arrows), however, when viewed as a 3D graph we begin to see that clusters may not in fact be overlapping. With this in mind, the 3D view in a) still has data closely spaced (pink arrows), and determining actual overlapping would take significantly longer for the human eye than the statistical evaluation from ML.

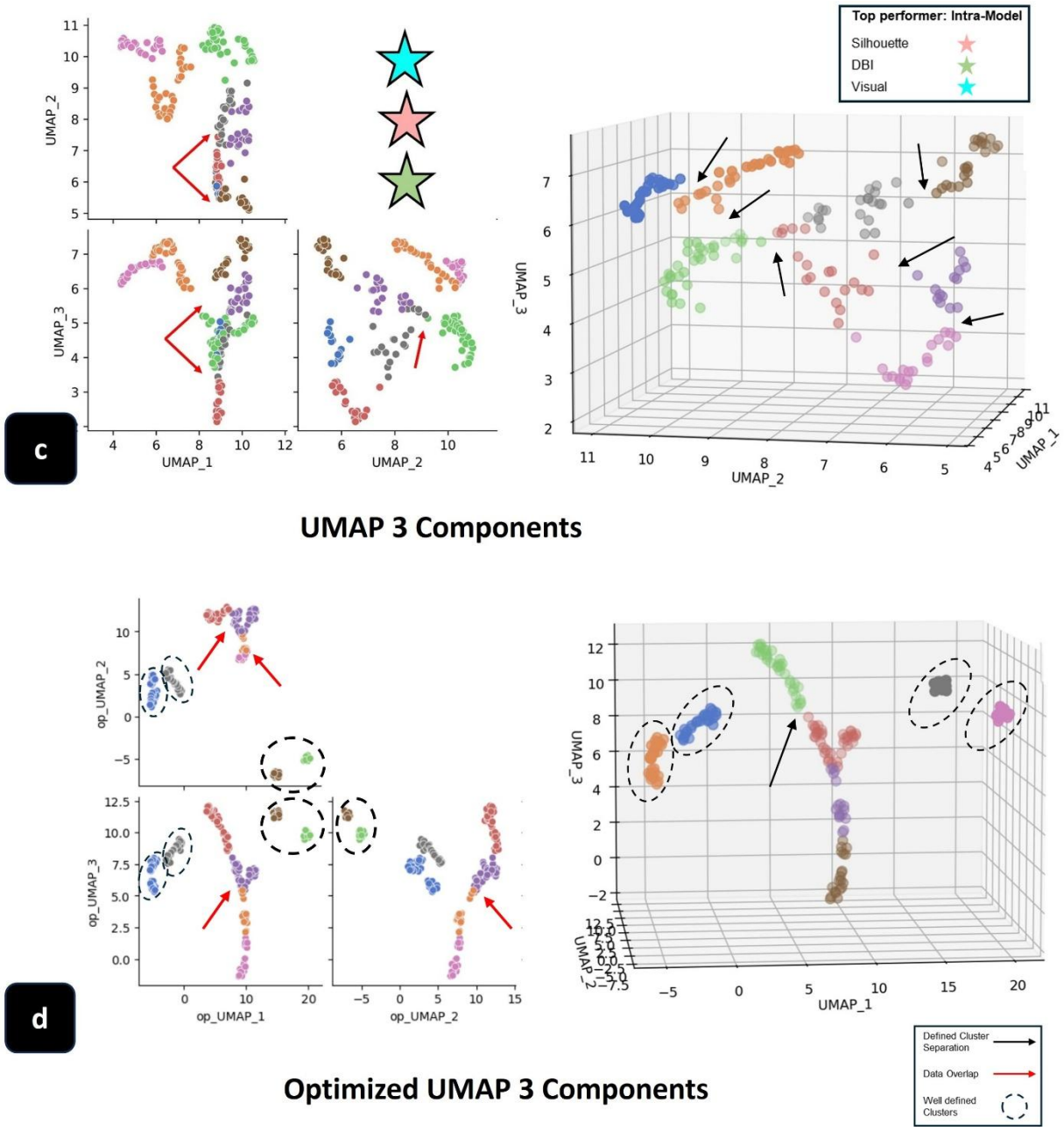


Figure 10 cont.: UMAP 3 components c) without optimized hyperparameters and d) with optimized hyperparameters. c) When viewed as a 2D graph, we note areas where clusters appear to be overlapping (red arrows), however, when viewed as a 3D graph we note good cluster separation (black arrows). d) With optimized hyperparameters, the clusters appear to have both good cohesion and separation with minimal appearance of perceived overlap indicated by red arrows. Areas of good separation are noted with black arrows and well-separated clusters with a black dashed polygon.

Clustering without DR

Figure 4 shows an initial k-means cluster model before the implementation of DR, model 1a. Visual inspection shows significant overlapping of data for all attributes except where plotted against cosine instantaneous phase, which shows overlap just along the center line. When evaluating the pre-DR scatterplots for 3-5 and 8 clusters, there is a continued lack of unique clustering results. Using quantitative evaluation (statistical metrics), the highest-performing pre-DR k-means model ranks as number 49 out of 55 models discussed in this paper, with a silhouette score of 0.352 and a DBI score of 0.965; for perspective, our top-performing model has a silhouette score of 0.904 and a DBI score of 0.145. This significant overlapping in the clusters from a qualitative evaluation (visual) and low performance from a quantitative evaluation (statistical) reinforces that the data's dimensionality is too high for k-means to handle directly; the curse of dimensionality (Allaoui 2020; Bengio 2005; Bellman 1956, 1964, 1969). When the data becomes too complex or has too many dimensions, the distance between the data points can become less meaningful, resulting in non-unique clusters. With this poor performance for pre-DR, we include DR into the workflow and evaluate post-DR cluster results to determine QE and QLE improvements.

Clustering with DR

Cluster performance after DR involves implementing each selected technique (PCA, K-PCA, and UMAP) on the dataset, followed by conducting k-means clustering. Evaluating the silhouette score and visual analysis, we focus on 2-5 and 8 cluster results. We summarize our findings for all models next, beginning with models reduced to 2 components (n=2), followed by 3 components (n=3).

For models with n=2:

- **PCA, model group 2**: All model clusters outperform k-means before DR. Improved cluster uniqueness is observed with higher cluster numbers; the

silhouette score suggests model 2b for optimal performance, while DBI suggests model 2c.

- **K-PCA, model group 3**: All model clusters outperform k-means before DR and indicate improvement as compared to PCA; the silhouette score suggests model 3e for optimal performance with DBI in agreement.
- **UMAP, model group 5**: All model clusters outperform k-means before DR. Visual inspection indicates better inter-cluster separation as compared to PCA and K-PCA. Silhouette score using UMAP's default setting indicated 18 clusters for optimal performance, with DBI in agreement.

For models with n=3:

- **PCA, model group 7**: All models displayed either intra-cluster wide spreads and/or non-unique inter-clusters. Within this set of models, group 7, model 7d is indicated as the top-performer using silhouette and DBI scores. Qualitative analysis also shows poor cluster performance overall.
- **K-PCA, model group 8**: Visual inspection of the clusters indicates improvement when compared directly to k-means pre-DR. It is also noted that K-PCA with 3 components' metrics, are sub-optimal as compared to K-PCA with 2 components. Similar to PCA, 5 clusters are indicated as the top performer using silhouette and DBI scores within this set of models; however, visual inspection again shows both wide intra-cluster spreads and/or non-unique inter-clusters.
- **UMAP, model group 10**: When reduced to 3 components, shows improvement as compared directly to k-means pre-DR. Hyperparameterization adjustments for UMAP models do have a noted effect on model performance.

Reviewing the performance of K-PCA and UMAP after optimization, we find the following:

- **K-PCA:**
  - **With 2 components:** Optimizing hyperparameters does not improve the QE scoring. Silhouette score and DBI both indicate model 3e as optimal.
  - **With 3 components:** More wide data-spread and overlapping clusters are noted. Silhouette score and DBI indicate models 8d and 8c clusters as optimal.
- **UMAP:**

- **With 2 components:** There is both good inter- and intra-cluster representation, with only limited indications of data overlap. In both 2 and 3 clusters, we note some widespread data plotting. Silhouette score and DBI agree with model 5d as the optimal.
- **With 3 components,** we still note some widespread data plotting in all models. Silhouette score and DBI agree with model 10e as the optimal.

Post-DR Cluster Results Summary

To summarize cluster results for post-DR models, we will discuss Figures 11 and 12, which focus on post-DR clustering for 4 and 5 clusters, across all models. In Figure 11 (2 components), we see an improvement in intra-cluster spread as we move from 4 to 5 clusters. This improvement is expected as the distance to the centroid is reduced each time an additional cluster is added; however, using the distance to the centroid as a metric of performance will quickly lead to over-fitting the data.

Intra-model QE comparisons show a difference in scores for 2 components, ranging from 0.001 to 0.084, and a difference in DBI scores ranging from 0.015 to 0.219. Only 1 model indicates 4 clusters to be QE better performers using the silhouette score: PCA 2 components. Only 1 model indicates 4 clusters to be QE better performers using DBI, also, PCA 2 components.

In Figure 12 (3 components), we see an improvement in intra-cluster spread as we move from 4 to 5 clusters. This improvement is expected as the distance to the centroid is reduced each time an additional cluster is added; however, using the distance to the centroid as a metric of performance will quickly lead to over-fitting the data.

Intra-model QE comparisons show a difference in scores for 3 components, ranging from 0.001 to 0.065, and a difference in DBI scores ranging from 0.011 to 0.087. Only 1 model indicates 4 clusters to be QE better performers using the silhouette score: UMAP 3 components. Only 1 model indicates 4 clusters to be QE better performers using DBI: PCA 3 components, OP. K-means pre-DR indicates 4 clusters are optimal to 5 clusters based on the silhouette score, with

a difference of 0.003, and indicates 5 clusters are optimal to 4 clusters based on the DBI with a difference of 0.059.

Using the silhouette score, higher score values indicate better performance, and DBI scores, lower values indicate better performance. We can quickly quantify and compare statistical measures of good inter- and intra-cluster results. When evaluating the performance based on mean distance, smaller values indicate more compact clusters, i.e., each data point in the cluster is closer to the cluster center; the higher the number of clusters, the more compact the clusters become. However, this could quickly lead to choosing a high number of clusters as the highest performer and essentially overfitting the data.

Using silhouette scores to assess performance and DBI to validate, we find that K-PCA 2 components consistently outperform other DR techniques, followed closely by UMAP 2 and 3

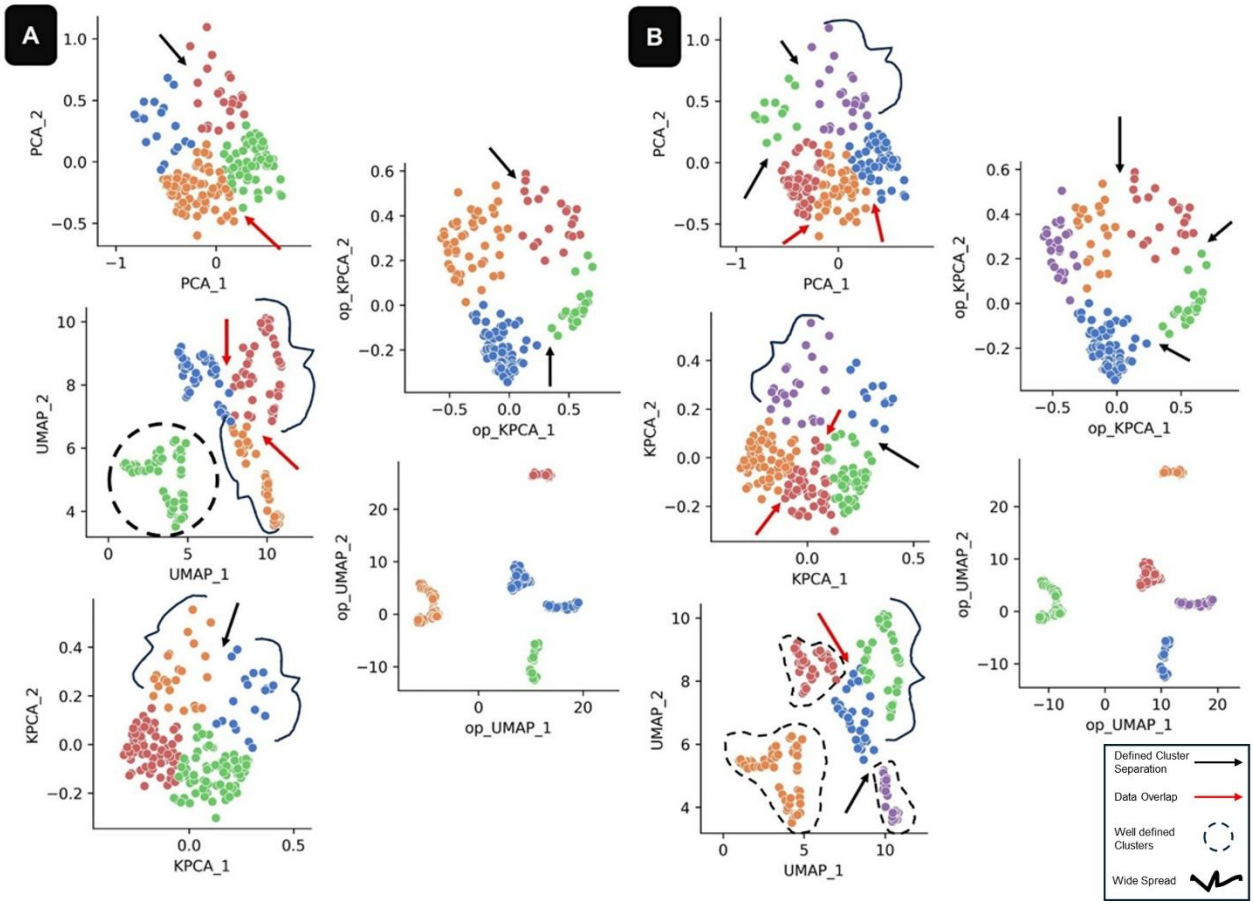


Figure 11: Summary of all post-DR 2 components. A) 4 clusters B) 5 clusters. Areas with better cluster separation are indicated by black arrows and areas with more overlap are indicated by a red arrow. The dashed-black polygon shows well-separated and easy-to-identify clusters and open brackets around the cluster with wide-spread.

components with OP. With these results in mind, let's discuss the factors that may contribute to improved performance for post-DR.

Rationale for post-DR performance overview

Several factors influence the performance of clustering when applying DR techniques to complex seismic data:

- Lower post-DR cluster performance due to the use of techniques that focus on the identification of maximized variance, such as PCA. This may result in clusters that are disproportionately influenced by strong patterns in the data that may not be representative of the geologic environment.
- If the DR technique discards data that is critical for distinguishing between potential clusters, performance may be diminished.
- The application of linear DR methods on non-linear data, such as seismic information, may result in suboptimal outcomes.

To help address some of these challenges, we explored two novel DR techniques for seismic workflows, K-PCA and UMAP. Additionally, K-PCA and UMAP have key parameters that can be optimized, based on the dataset, to potentially improve the efficacy of the DR applied. For K-PCA, the kernel parameter is critical and can be linear, polynomial, radial basis function (rbf), or sigmoid. RBF is identified as the best option for handling non-linear data, and after comparing its function against the polynomial kernel, we proceeded with rbf. The gamma parameter determines the radius of influence for each data point, with larger values creating more detailed but compact mapping and smaller values focusing on more general features. We tested our data set using both K-PCA default settings (kernel rbf) and optimized hyperparameters: rbf kernel, a gamma value of 10, and reduced to 2 and 3 components for later comparison.

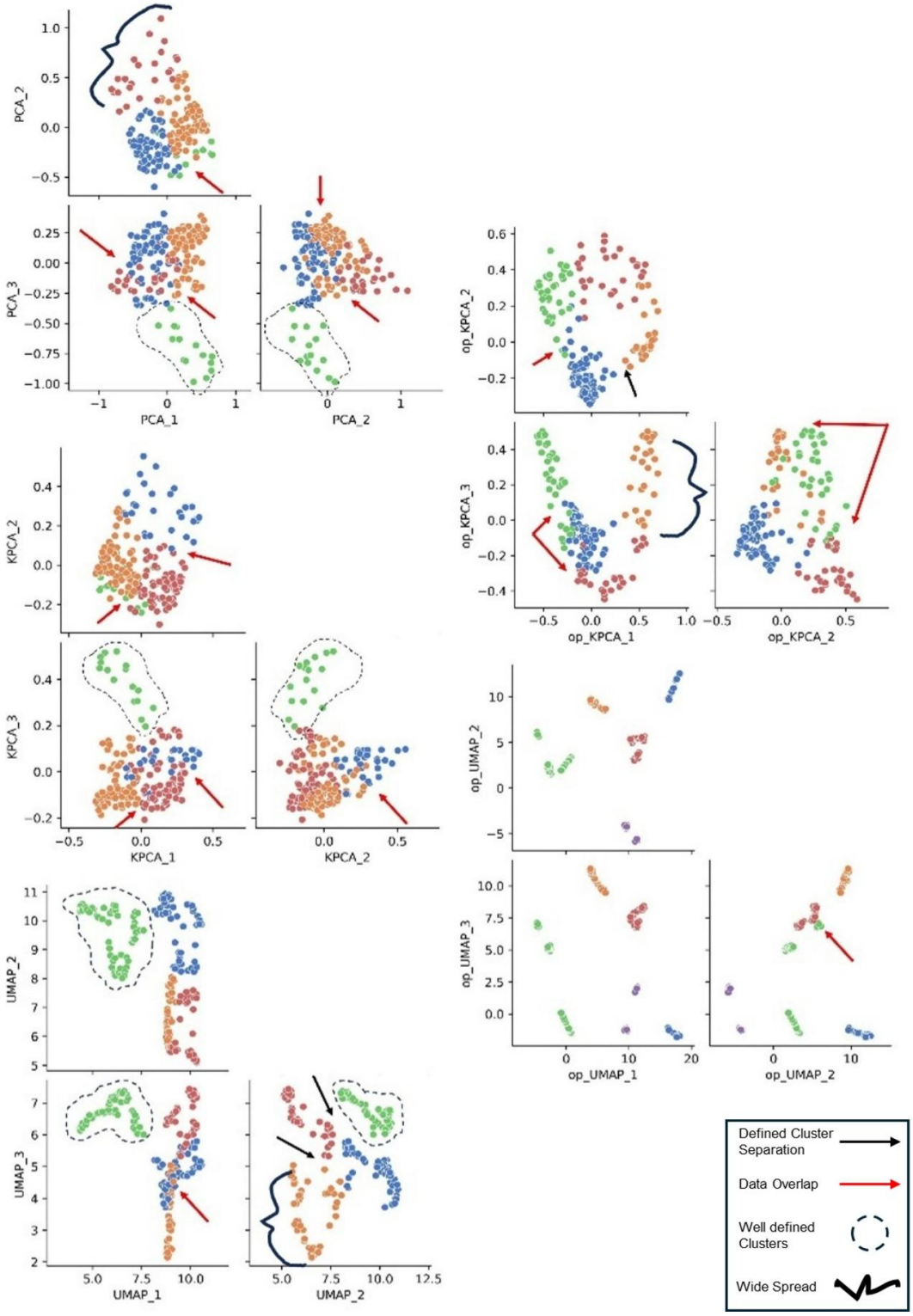


Figure 12: Summary of all post-DR 3 components, 5 clusters. Areas with better cluster separation are indicated by black arrows and areas with more overlap are indicated by red arrows. Those clusters that are unique and show good separation from surrounding clusters are indicated by a black dashed polygon. Dashed-black polygon showing well separated and easy to identify cluster and open bracket around cluster with wide-spread.

For UMAP, one of the key parameters is n_neighbors, which ranges from 5-50 and has a default of 15. The min-dist parameter controls whether the clusters will be tighter and focus on local relationships (smaller values) or create clusters with wider spread, focusing on global separation (higher values). The metric parameter determines how distances between data points are calculated. We also ran our data set using UMAP with default settings (n_neighbors=15, min_dist= 0.1, metric = Euclidean) and then again with optimization as follows: n_neighbors 15, min_dist 0.5, and reduced to 2 and 3 components for later comparison. For a more detailed look at UMAP hyperparameters see McInnes et al. 2018.

In terms of the clustering algorithm itself, k-means tends toward identification of clusters of similar size and density; this tendency may limit cluster performance in seismic data as the data distributions are often of variable sizes and density (Jain 2010). We note better clustering performance when K-PCA and UMAP are applied to k-means as these methods preserve the local and global structures of the data (maintains accurate geologic feature representation) when changing the dimension space. PCA may perform less optimally in preserving cluster uniqueness by oversimplifying the data, while it preserves variance, it may sacrifice crucial geologic information affecting overall performance. When comparing the use of 2 versus 3 components, 2 components may outperform 3 components if the most meaningful information is better preserved with fewer components. Introducing a third component can bring extraneous data or noise, leading to lower cluster performance by complicating the data structure unnecessarily.

Selecting the appropriate DR technique for the dataset—linear methods for linear data and non-linear methods for non-linear data—is important for cluster performance enhancement and combining them with advanced clustering methods, such as k-means++ for centroid initialization can prevent converging to suboptimal clustering solutions (Arthur and Vassilvitskii 2007). These rationales for algorithm performance are important to consider for the effective application of DR in seismic data analysis, ensuring that both statistical integrity and geological relevance are maintained.

CONCLUSION

We have observed that performing DR prior to unsupervised clustering does improve cluster performance as determined through the analysis of statistical model results. This improved performance extends beyond clustering directly to the geologic interpretability of seismic data. K-PCA was observed to improve geologic resolution for intra-channel delineation, and UMAP improved lateral resolution for identification of channel-related features (observations made in horizon view or what some may refer to as map view). In our statistical evaluation, we find the top-performing models to be K-PCA followed by UMAP, with our top performer outperforming our highest-ranking PCA by almost double (using the silhouette score) and outperforming our highest-performing pre-DR k-means by almost triple (using the silhouette score).

While DR demonstrates the important role it has in improving unsupervised clustering, this may not be true across all datasets. Findings and workflows, such as those discussed in this paper, can serve as a good starting point when attempting to determine an optimal workflow. However, the DR techniques and seismic attributes used in our dataset may not be optimal for every users' situation. Interpreters must take the necessary time to evaluate the optimal attributes for their specific dataset and study, prepare the dataset for multi-attribute analysis, create multiple models for comparison, and compare those models using both QLE and QE evaluation. As geoscientists become more versed in the application of ML, additional evaluations will be necessary to determine the application of novel ML/DR methods for the continued improvement of our workflows. With the abundance of algorithms available, users must also avoid the temptation to try everything; we need to start with an understanding of what our geologic question is and then look for algorithms and DR that are designed to evaluate data to best tease out that information. Maintaining a balance between weighting statistical performance and geologic interpretability is imperative to validate our models without becoming overly dependent on QLE or QE alone. We have helped move this process forward through a detailed study of how K-PCA and UMAP can fit into seismic data analysis. Overall, the hope for this deep dive into the seismic workflow is a broader understanding of how to effectively incorporate these advanced ML with a balanced approach that does not lose sight of the geologic problem.

The geologic interpretation of these clustered models, presented in our next paper, will explore how to best leverage both statistical and visual evaluations, beyond the graphical representation.

ACKNOWLEDGEMENTS

The authors would like to thank the AASPI consortium for providing the software used for the calculation of attributes. We also thank the New Zealand government for making the Romney data available for research.

REFERENCES

Allaoui, M., M. L. Kherfi, and A. Cheriet, 2020, Considerably improving clustering algorithms using UMAP dimensionality reduction technique: A comparative study: International Conference on Image and Signal Processing, Springer International Publishing, 317-325.

Arthur, D., and S. Vassilvitskii, 2007, k-means++: The advantages of careful seeding: SODA, **7**, 1027-1035.

Barnes, A. E. (1997). Genetic classification of complex seismic trace attributes. In *SEG Technical Program Expanded Abstracts 1997* (pp. 1151-1154). Society of Exploration Geophysicists.
Bellman, R., 1956, Dynamic Programming: RAND Corporation.

Bellman, R., 1964, Control theory: Scientific American, **211**, no. 3, 186-201.

Bellman, R., 1969, A new type of approximation leading to reduction of dimensionality in control processes: Journal of Mathematical Analysis and Applications, **27**, no. 2, 454-459.

Bengio, Y., O. Delalleau, and N. Le Roux, 2005, The curse of dimensionality for local kernel machines: Technical Report, **1258**, no. 12, 1.

Bull, S., G. Browne, M. Arnot, and L. Strachan, 2020, Influence of Mass Transport Deposit (MTD) surface topography on deep-water deposition: an example from a predominantly fine-grained continental margin, New Zealand: Geological Society, London, Special Publications, **500**, SP500-2019-192.

Carter, D. C., 1995, Use of Windowed Seismic Attributes in 3D Seismic Facies Analysis and Pattern Recognition: International Symposium on Sequence Stratigraphy, 73-87.

Chopra, S., R. K. Sharma, K. J. Marfurt, H. Bedle, and S. Verma, 2022, Attempts at seismic characterization of a deepwater turbidite channel in Taranaki Basin, New Zealand: First Break, **40**, no. 10, 27-39.

Chopra, S., and K. J. Marfurt, 2019, Unsupervised machine learning applications for seismic facies classification: Unconventional Resources Technology Conference, URTeC, 3135-3142.

Chopra, S., & Marfurt, K. J. (2007). Seismic attributes for prospect identification and reservoir characterization. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.

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.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.

Di, H., L. Truelove, C. Li, and A. Abubakar, 2020, Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand: The Leading Edge, **39**, no. 10, 727-733.

Grahame, J. L., 2015, Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity.

Haibin Di, Muhammad Shafiq, Ghassan AlRegib, Multi-attribute *k*-means clustering for salt-boundary delineation from three-dimensional seismic data, *Geophysical Journal International*, Volume 215, Issue 3, December 2018, Pages 1999–2007, https://doi.org/10.1093/gji/ggy376

Hart, B. S. (2008). Channel detection in 3-D seismic data using sweetness. AAPG bulletin, 92(6), 733-742.

Infante-Paez, L., and Marfurt, K. J. (2018). In-context interpretation: Avoiding pitfalls in misidentification of igneous bodies in seismic data. *Interpretation*, *6*(4), SL29-SL42.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.

Kamp, P. J., A. J. Vonk, K. J. Bland, R. J. Hansen, A. J. Hendy, A. P. McIntyre, ..., and C. S. Nelson, 2004, Neogene stratigraphic architecture and tectonic evolution of Wanganui, King Country, and eastern Taranaki Basins, New Zealand: New Zealand Journal of Geology and Geophysics, **47**, no. 4, 625-644.

King, P. R., & Thrasher, G. P. (1992). Post-Eocene development of the Taranaki Basin, New Zealand: Convergent overprint of a passive margin: Chapter 7: Southwest Pacific and Eastern Indian Ocean Margins.

King, P. R., and P. H. Robinson, 1988, An Overview of Taranaki Region Geology, New Zealand: Energy Exploration & Exploitation, **6**, no. 3, 213-232.

Kroeger, K. F., Thrasher, G. P., & Sarma, M. (2019). The evolution of a Middle Miocene deep-water sedimentary system in northwestern New Zealand (Taranaki basin):

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**, 105427.

La Marca, K., 2020, Seismic attribute optimization with unsupervised machine learning techniques for deepwater seismic facies interpretation: Users vs machines.

Li, J., Wu, X., Ye, Y., Yang, C., Hu, Z., Sun, X., & Zhao, T. (2023). Unsupervised contrastive learning for seismic facies characterization. Geophysics, 88(1), WA81-WA89.

Lubo-Robles, D., and K. J. Marfurt, 2019, Independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: Interpretation, **7**, no. 3, SE19-SE42.

McInnes, L., J. Healy, and J. Melville, 2018, Umap: Uniform manifold approximation and projection for dimension reduction: arXiv preprint arXiv:1802.03426.

Moreno-Ward, A., Vera-Arroyo, A., & Bedle, H. (2025). Improving interpretability: The role of dimension reduction in multiattribute seismic facies identification. *Interpretation*, *13*(3), T475-T496.

Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. In *IOP Conference Series: Materials Science and Engineering* (Vol. 725, No. 1, p. 012128). IOP Publishing.

Naewboonnien, K., 2021, The Use of Seismic Attributes for Submarine Channel Interpretation in Deepwater Taranaki Basin, New Zealand: Bulletin of Earth Sciences of Thailand, **11**, no. 2, 134-143.

Posamentier, H. W., V. Paumard, and S. C. Lang, 2022, Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data: Earth-Science Reviews, **228**, 103963.

Qi, J., Lin, T., Zhao, T., Li, F., & Marfurt, K. (2016). Semisupervised multiattribute seismic facies analysis. Interpretation, 4(1), SB91-SB106.

Reilly, C., A. Nicol, J. J. Walsh, and H. Seebeck, 2015, Evolution of faulting and plate boundary deformation in the Southern Taranaki Basin, New Zealand: Tectonophysics, **651**, 1-18.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Strogen, D. P., H. Seebeck, A. Nicol, and P. R. King, 2017, Two-phase Cretaceous–Paleocene rifting in the Taranaki Basin region, New Zealand; implications for Gondwana break-up: Journal of the Geological Society, **174**, no. 5, 929-946.

Strogen, D. P., K. E. Higgs, A. G. Griffin, and H. E. Morgans, 2019, Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression: Geological Magazine, **156**, no. 10, 1751-1770.

Subrahmanyam, D., & Rao, P. H. (2008, January). Seismic attributes- a review. In 7th International Conference and Exposition on Petroleum Geophysics, Hyderabad, India.

Talebi, H., L. J. M. Peeters, U. Mueller, R. Tolosana-Delgado, and K. G. van den Boogaart, 2020, Towards geostatistical learning for the geosciences: A case study in improving the spatial awareness of spectral clustering: Mathematical Geosciences, **52**, no. 8, 1035-1048.

Troccoli, E. B., A. G. Cerqueira, J. B. Lemos, 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**, 104555.

Verma, S., and S. Bhattacharya, 2021, Unsupervised Machine Learning for Interpreting Shelf-to-Basin Seismic Geomorphology and Paleoclimate: 82nd EAGE Annual Conference & Exhibition, EAGE, Extended Abstracts, 1-5.

Yenugu, M., Marfurt, K. J., & Matson, S. (2010). Seismic texture analysis for reservoir prediction and characterization. The Leading Edge, 29(9), 1116-1121.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29-SAE58.

APPENDIX A: Data Tables

Seismic Survey		Well	
Name	Romney 3D	**Name**	Romney-1
Survey type	3D	**Well Type**	Vertical wildcat exploration
Environment	Marine	**Spud date**	26-Nov-13
Acquisition	2011	**Completion date**	20-Jan-14
Area	35x55 km	**Rig type**	Drillship
Bin size	12.5x25 m	**Latitude**	37° 53' 39.337" S
Sample rate	4 ms	**Longitude**	172° 43' 52.719" E
Record length	7.48 s	**Drilling floor**	25 m above sea level
Dominant frequency	38 Hz	**Water depth**	1546.6 m lowest astronomical tide
Dominant Wavelength	76 m	**Total depth**	4619 m measured depth
Vertical Resolution (l/4)	19 m	**Result**	Dry hole
Lateral resolution	24 m	**Status**	Plugged and abandoned

Appendix A Table A: Romney 3D seismic survey data and Romney-1 well data

Feature Importance—Scaled 0-1 (Loadings)						
	COSPHS	ENV	GLCM	K2-CURV	K1-CURV	SWT
2 COMP						
PCA_1	-0.86	0.32	-0.03	-0.17	0.186	0.29
PCA_2	0.49	0.55	-0.09	-0.14	0.41	0.51
3 COMP						
PCA_1	-0.86	0.32	-0.03	-0.17	0.19	0.29
PCA_2	0.49	0.55	-0.09	-0.13	0.41	0.51
PCA_3	-0.18	0.29	0.2	0.59	-0.6	0.38

Explained Variance %			
2 Comp		3 Comp	
PCA_1	0.38	PCA_1	0.38
PCA_2	0.29	PCA_2	0.29
		PCA_3	0.25

Percent Contribution to Component						
	COSPHS	ENV	GLCM	K2-CURV	K1-CURV	SWT
2 COMP						
KPCA_1	0.213	0.09	0.28	0.06	0.26	0.098
KPCA_2	0.18	0.16	.022	0.033	0.254	0.16
3 COMP						
KPCA_1	0.19	0.18	0.18	0.024	0.25	0.177
KPCA_2	0.21	0.1	0.264	0.234	0.234	0.11
KPCA_3	0.264	0.044	0.3	0.24	0.08	0.071

Comp Magnitude			
2 Comp		3 Comp	
KPCA_1	0.16	KPCA_1	0.16
KPCA_2	0.13	KPCA_2	0.13
		KPCA_3	0.12

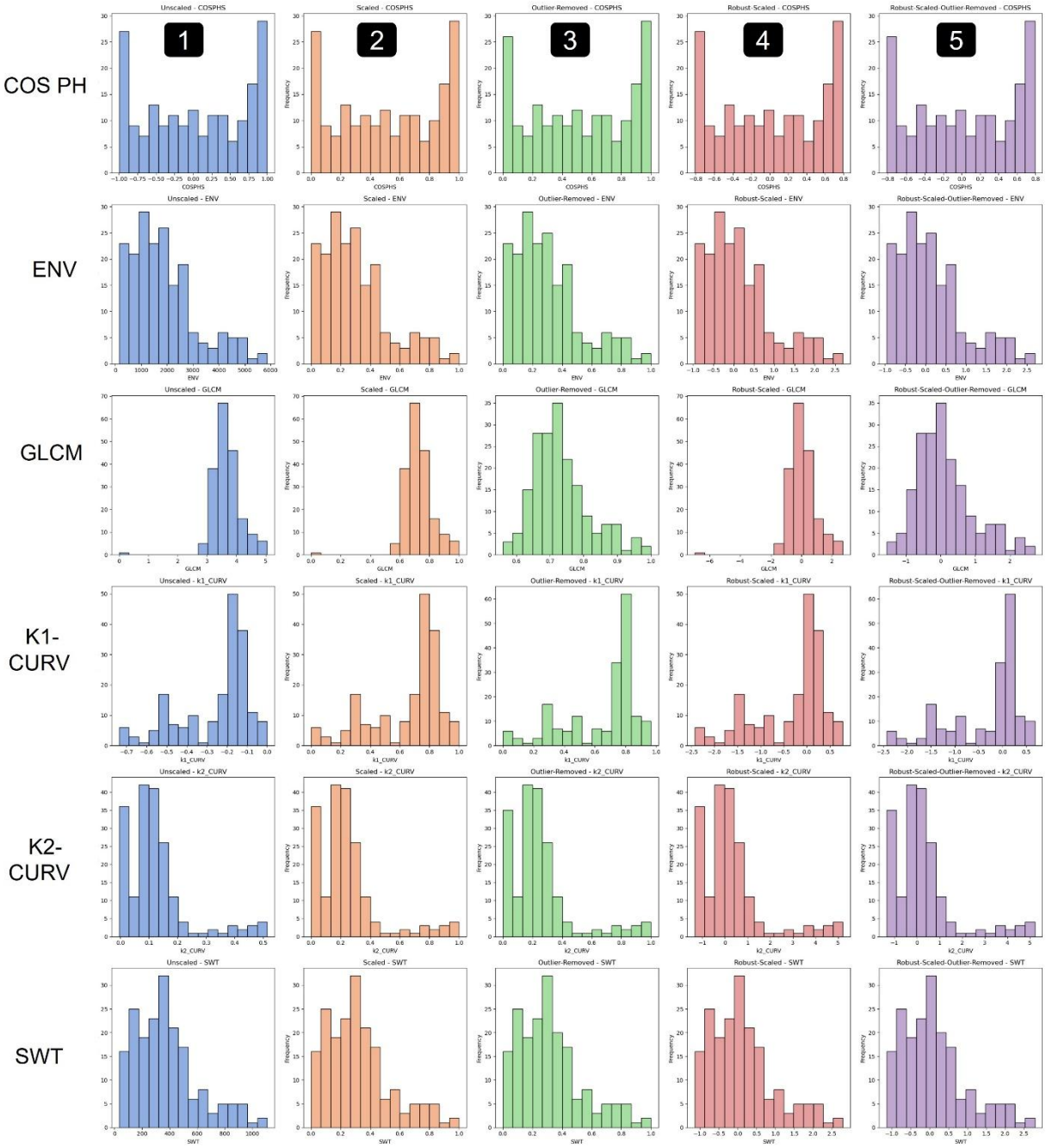
Feature Importance (FI) —Scaled 0-1						
	COSPHS	ENV	GLCM	K2-CURV	K1-CURV	SWT
2 COMP						
UMAP_1	0.013	0.099	1	0.12	0.08	0.11
UMAP_2	0	0.07	0.8	0.08	0.05	0.08
3 COMP						
UMAP_1	0.05	0.133	1	0.15	0.11	0.14
UMAP_2	0.007	0.034	.0312	0.04	0.03	0.04
UMAP_3	0	0.02	0.2	0.02	0.01	0.02

Permutation –Base Feature Importance (PI)						
	COSPHS	ENV	GLCM	K2-CURV	K1-CURV	SWT
2 COMP						
UMAP_1	1.21	0.55	0.23	0.51	1.12	0.45
UMAP_2	1.2	0.6	0.24	0.5	0.9	0.5
3 COMP						
UMAP_1	0.95	0.3	0.189	0.37	0.75	0.34
UMAP_2	0.91	0.41	0.19	0.33	0.82	0.36
UMAP_3	0.93	0.4	0.19	0.36	0.67	0.36

Feature Importance—Scaled 0-1				
Permutation –Base Feature Importance				
2 COMP				
	UMAP1	UMAP2	UMAP1	UMAP2
COSPHS	0.013	0	1.21	1.2
K2-CURV	0.12	0.08	0.51	0.5
K1-CURV	0.08	0.05	1.12	0.9
ENV	0.099	0.07	0.55	0.6
SWT	0.11	0.08	0.45	0.5
GLCM	1	0.8	0.23	0.24

3 COMP						
	UMAP1	UMAP2	UMAP3	UMAP1	UMAP2	UMAP3
COSPHS	0.05	0.007	0	0.95	0.91	0.93
K2-CURV	0.15	0.04	0.02	0.37	0.33	0.36
K1-CURV	0.11	0.03	0.01	0.75	0.82	0.67
ENV	0.133	0.034	0.02	0.3	0.41	0.4
SWT	0.14	0.04	0.02	0.34	0.36	0.36
GLCM	1	0.0312	0.2	0.189	0.19	0.19

Appendix A Figure A: DR algorithms with 2- and 3-component data input weights for evaluation. PCA evaluates input data with feature importance and explained variance %. The first 3 principal components capture 92% of the variance. K-PCA evaluates percent contribution and component magnitude. The component(s) magnitude is relative to each other with higher values indicating more importance. UMAP evaluates the feature importance (FI), and the authors have included the permutation feature importance (PI). While the FI may indicate one attribute has high importance to the projection magnitude, the PI indicates which data have a significant impact on the structure of the projection.



Appendix A Figure B: Data distribution comparison for seismic attributes used for ML workflow. Columns are as follows: 1. original attribute data, 2. data after min-max scaling, 3. data after min-max scaling and outlier removal from GLCM, 4. data after robust scaling, and 5. data after robust scaling and outlier removal from GLM. Data distribution remains unchanged across columns, except for a minimal change for GLCM with outlier removal.

Based on Silhouette Score

just kmeans	# Clusters	Silhouette Score:	DBI:
kMEANS	4	0.352	0.965
kMEANS	5	0.349	0.906
kMEANS	8	0.332	1.002
kMEANS	3	0.330	1.115
kMEANS	2	0.281	1.432
kmeans with	PCA 2		
PCA 2	3	0.451	0.832
PCA 2	4	0.451	0.728
PCA 2	5	0.420	0.772
PCA 2	8	0.420	0.764
PCA 2	2	0.409	0.994
kmeans with	PCA 3		
PCA 3	5	0.398	0.786
PCA 3	4	0.391	0.858
PCA 3	3	0.363	1.020
PCA 3	8	0.363	0.867
PCA 3	2	0.312	1.317
kmeans with	kPCA 2		
kPCA 2	8	0.904	0.145
kPCA 2	5	0.874	0.191
kPCA 2	4	0.855	0.223
kPCA 2	3	0.835	0.268
kPCA 2	2	0.792	0.308
kmeans with	kPCA 3		
kPCA 3	5	0.409	0.769
kPCA 3	4	0.403	0.842
kPCA 3	8	0.374	0.858
kPCA 3	3	0.370	1.002
kPCA 3	2	0.318	1.308
kmeans with	op_kPCA 2		
op_kPCA 2	8	0.894	0.180
op_kPCA 2	5	0.862	0.211
op_kPCA 2	4	0.851	0.243
op_kPCA 2	3	0.817	0.273
op_kPCA 2	2	0.728	0.438
kmeans with	op_kPCA 3		
op_kPCA 3	5	0.604	0.662
op_kPCA 3	4	0.603	0.613
op_kPCA 3	3	0.531	0.848
op_kPCA 3	2	0.474	1.066
op_kPCA 3	8	0.441	0.754
kmeans with	UMAP 2		
UMAP 2	8	0.736	0.370
UMAP 2	5	0.639	0.483
UMAP 2	4	0.555	0.702
UMAP 2	3	0.537	0.618
UMAP 2	2	0.506	0.798
kmeans with	UMAP 3		
UMAP 3	8	0.517	0.647
UMAP 3	4	0.514	0.707
UMAP 3	5	0.500	0.696
UMAP 3	2	0.478	0.772
UMAP 3	3	0.437	0.935
kmeans with	op_UMAP 2		
op_UMAP 2	5	0.637	0.500
op_UMAP 2	4	0.636	0.515
op_UMAP 2	2	0.625	0.586
op_UMAP 2	3	0.621	0.529
op_UMAP 2	8	0.574	0.589
kmeans with	op_UMAP 3		
op_UMAP 3	8	0.739	0.331
op_UMAP 3	5	0.695	0.425
op_UMAP 3	4	0.630	0.513
op_UMAP 3	2	0.585	0.610
op_UMAP 3	3	0.542	0.832

Based on DBI

just kmeans	# Clusters	Silhouette Score:	DBI:
kMEANS	5	0.349	0.906
kMEANS	4	0.352	0.965
kMEANS	8	0.332	1.002
kMEANS	3	0.330	1.115
kMEANS	2	0.281	1.432
kmeans with	PCA 2		
PCA 2	4	0.451	0.728
PCA 2	8	0.420	0.764
PCA 2	5	0.420	0.772
PCA 2	3	0.451	0.832
PCA 2	2	0.409	0.994
kmeans with	PCA 3		
PCA 3	5	0.398	0.786
PCA 3	4	0.391	0.858
PCA 3	8	0.363	0.867
PCA 3	3	0.363	1.020
PCA 3	2	0.312	1.317
kmeans with	kPCA 2		
kPCA 2	8	0.904	0.145
kPCA 2	5	0.874	0.191
kPCA 2	4	0.855	0.223
kPCA 2	3	0.835	0.268
kPCA 2	2	0.792	0.308
kmeans with	kPCA 3		
kPCA 3	5	0.409	0.769
kPCA 3	4	0.403	0.842
kPCA 3	8	0.374	0.858
kPCA 3	3	0.370	1.002
kPCA 3	2	0.318	1.308
kmeans with	op_kPCA 2		
op_kPCA 2	8	0.894	0.180
op_kPCA 2	5	0.862	0.211
op_kPCA 2	4	0.851	0.243
op_kPCA 2	3	0.817	0.273
op_kPCA 2	2	0.728	0.438
kmeans with	op_kPCA 3		
op_kPCA 3	4	0.603	0.613
op_kPCA 3	5	0.604	0.662
op_kPCA 3	8	0.441	0.754
op_kPCA 3	3	0.531	0.848
op_kPCA 3	2	0.474	1.066
kmeans with	UMAP 2		
UMAP 2	8	0.736	0.370
UMAP 2	5	0.639	0.483
UMAP 2	3	0.537	0.618
UMAP 2	4	0.555	0.702
UMAP 2	2	0.506	0.798
kmeans with	UMAP 3		
UMAP 3	8	0.517	0.647
UMAP 3	5	0.500	0.696
UMAP 3	4	0.514	0.707
UMAP 3	2	0.478	0.772
UMAP 3	3	0.437	0.935
kmeans with	op_UMAP 2		
op_UMAP 2	5	0.637	0.500
op_UMAP 2	4	0.636	0.515
op_UMAP 2	3	0.621	0.529
op_UMAP 2	2	0.625	0.586
op_UMAP 2	8	0.574	0.589
kmeans with	op_UMAP 3		
op_UMAP 3	8	0.739	0.331
op_UMAP 3	5	0.695	0.425
op_UMAP 3	4	0.630	0.513
op_UMAP 3	2	0.585	0.610
op_UMAP 3	3	0.542	0.832

Appendix A Table B: All models ranked (grouped by clustering method), highest to lowest, by silhouette score (left) and DBI (right). A pink cell indicates the highest performer per clustering method, where silhouette and DBI agree. A light orange cell indicates a high silhouette performer, and a green cell

Based on Silhouette Score

Method	# Clusters	Silhouette Score:	DBI:
kPCA 2	8	0.904	0.145
op_kPCA 2	8	0.894	0.180
kPCA 2	5	0.874	0.191
op_kPCA 2	5	0.862	0.211
kPCA 2	4	0.855	0.223
op_kPCA 2	4	0.851	0.243
kPCA 2	3	0.835	0.268
op_kPCA 2	3	0.817	0.273
kPCA 2	2	0.792	0.308
op_UMAP 3	8	0.739	0.331
UMAP 2	8	0.736	0.370
op_kPCA 2	2	0.728	0.438
op_UMAP 3	5	0.695	0.425
UMAP 2	5	0.639	0.483
op_UMAP 2	5	0.637	0.500
op_UMAP 2	4	0.636	0.515
op_UMAP 3	4	0.630	0.513
op_UMAP 2	2	0.625	0.586
op_UMAP 2	3	0.621	0.529
op_kPCA 3	5	0.604	0.662
op_kPCA 3	4	0.603	0.613
op_UMAP 3	2	0.585	0.610
op_UMAP 2	8	0.574	0.589
UMAP 2	4	0.555	0.702
op_UMAP 3	3	0.542	0.832
UMAP 2	3	0.537	0.618
op_kPCA 3	3	0.531	0.848
UMAP 3	8	0.517	0.647
UMAP 3	4	0.514	0.707
UMAP 2	2	0.506	0.798
UMAP 3	5	0.500	0.696
UMAP 3	2	0.478	0.772
op_kPCA 3	2	0.474	1.066
PCA 2	3	0.451	0.832
PCA 2	4	0.451	0.728
op_kPCA 3	8	0.441	0.754
UMAP 3	3	0.437	0.935
PCA 2	5	0.420	0.772
PCA 2	8	0.420	0.764
PCA 2	2	0.409	0.994
kPCA 3	5	0.409	0.769
kPCA 3	4	0.403	0.842
PCA 3	5	0.398	0.786
PCA 3	4	0.391	0.858
kPCA 3	8	0.374	0.858
kPCA 3	3	0.370	1.002
PCA 3	3	0.363	1.020
PCA 3	8	0.363	0.867
kMEANS	4	0.352	0.965
kMEANS	5	0.349	0.906
kMEANS	8	0.332	1.002
kMEANS	3	0.330	1.115
kPCA 3	2	0.318	1.308
PCA 3	2	0.312	1.317
kMEANS	2	0.281	1.432

Based on DBI

Method	# Clusters	Silhouette Score:	DBI:
kPCA 2	8	0.904	0.145
op_kPCA 2	8	0.894	0.180
kPCA 2	5	0.874	0.191
op_kPCA 2	5	0.862	0.211
kPCA 2	4	0.855	0.223
op_kPCA 2	4	0.851	0.243
kPCA 2	3	0.835	0.268
op_kPCA 2	3	0.817	0.273
kPCA 2	2	0.792	0.308
op_UMAP 3	8	0.739	0.331
UMAP 2	8	0.736	0.370
op_UMAP 3	5	0.695	0.425
op_kPCA 2	2	0.728	0.438
UMAP 2	5	0.639	0.483
op_UMAP 2	5	0.637	0.500
op_UMAP 3	4	0.630	0.513
op_UMAP 2	4	0.636	0.515
op_UMAP 2	3	0.621	0.529
op_UMAP 2	2	0.625	0.586
op_UMAP 2	8	0.574	0.589
op_UMAP 3	2	0.585	0.610
op_kPCA 3	4	0.603	0.613
UMAP 2	3	0.537	0.618
UMAP 3	8	0.517	0.647
op_kPCA 3	5	0.604	0.662
UMAP 3	5	0.500	0.696
UMAP 2	4	0.555	0.702
UMAP 3	4	0.514	0.707
PCA 2	4	0.451	0.728
op_kPCA 3	8	0.441	0.754
PCA 2	8	0.420	0.764
kPCA 3	5	0.409	0.769
PCA 2	5	0.420	0.772
UMAP 3	2	0.478	0.772
PCA 3	5	0.398	0.786
UMAP 2	2	0.506	0.798
op_UMAP 3	3	0.542	0.832
PCA 2	3	0.451	0.832
kPCA 3	4	0.403	0.842
op_kPCA 3	3	0.531	0.848
PCA 3	4	0.391	0.858
kPCA 3	8	0.374	0.858
PCA 3	8	0.363	0.867
kMEANS	5	0.349	0.906
UMAP 3	3	0.437	0.935
kMEANS	4	0.352	0.965
PCA 2	2	0.409	0.994
kMEANS	8	0.332	1.002
kPCA 3	3	0.370	1.002
PCA 3	3	0.363	1.020
op_kPCA 3	2	0.474	1.066
kMEANS	3	0.330	1.115
kPCA 3	2	0.318	1.308
PCA 3	2	0.312	1.317
kMEANS	2	0.281	1.432

Appendix A Table C: All models ranked, highest to lowest, by silhouette score (left) and DBI (right). A pink cell indicates the highest performer per clustering method, where silhouette and DBI agree. A light orange cell indicates a high silhouette performer, and a green cell indicates a high DBI performer, per each clustering method. Color coding is held from the previous figure.

Compared Statistics Before/After Scaling and Post Outlier Removal

	count	mean	std	min	0.25	0.5	0.75	max
U_COSPHS	188	0.0383	0.6795	-0.9999	-0.5505	0.0127	0.6959	1.0000
S_COSPHS	188	0.5191	0.3398	0.0000	0.2247	0.5063	0.8480	1.0000
O_COSPHS	187	0.5217	0.3388	0.0000	0.2288	0.5080	0.8518	1.0000
U_ENV	188	1859.5115	1268.0259	145.7260	919.9796	1647.0233	2486.2750	5865.0562
S_ENV	188	0.2996	0.2217	0.0000	0.1354	0.2625	0.4092	1.0000
O_ENV	187	0.2996	0.2223	0.0000	0.1354	0.2603	0.4101	1.0000
U_GLCM	188	3.6820	0.5045	0.0000	3.3922	3.6371	3.9137	5.0517
S_GLCM	188	0.7289	0.0999	0.0000	0.6715	0.7200	0.7747	1.0000
O_GLCM	187	0.7328	0.0846	0.5596	0.6731	0.7202	0.7748	1.0000
U_k2_CURV	188	0.1213	0.1098	-0.0047	0.0649	0.1015	0.1441	0.5140
S_k2_CURV	188	0.2429	0.2117	0.0000	0.1341	0.2047	0.2868	1.0000
O_k2_CURV	187	0.2442	0.2115	0.0000	0.1380	0.2053	0.2877	1.0000
U_k1_CURV	188	-0.2593	0.1773	-0.7374	-0.3735	-0.1801	-0.1439	0.0000
S_k1_CURV	188	0.6484	0.2404	0.0000	0.4935	0.7557	0.8048	1.0000
O_k1_CURV	187	0.6465	0.2396	0.0000	0.4909	0.7551	0.8045	0.9663
U_SWT	188	375.7006	231.6662	32.5187	205.7186	344.7879	479.5458	1108.5056
S_SWT	188	0.3189	0.2153	0.0000	0.1610	0.2902	0.4155	1.0000
O_SWT	187	0.3187	0.2159	0.0000	0.1581	0.2898	0.4158	1.0000

U is unscaled
S is scaled
O is outlier removal

Appendix A Table D: Statistics for chosen attributes before scaling, after scaling, after scaling, and outlier removal. Scaling conducting using min-max.

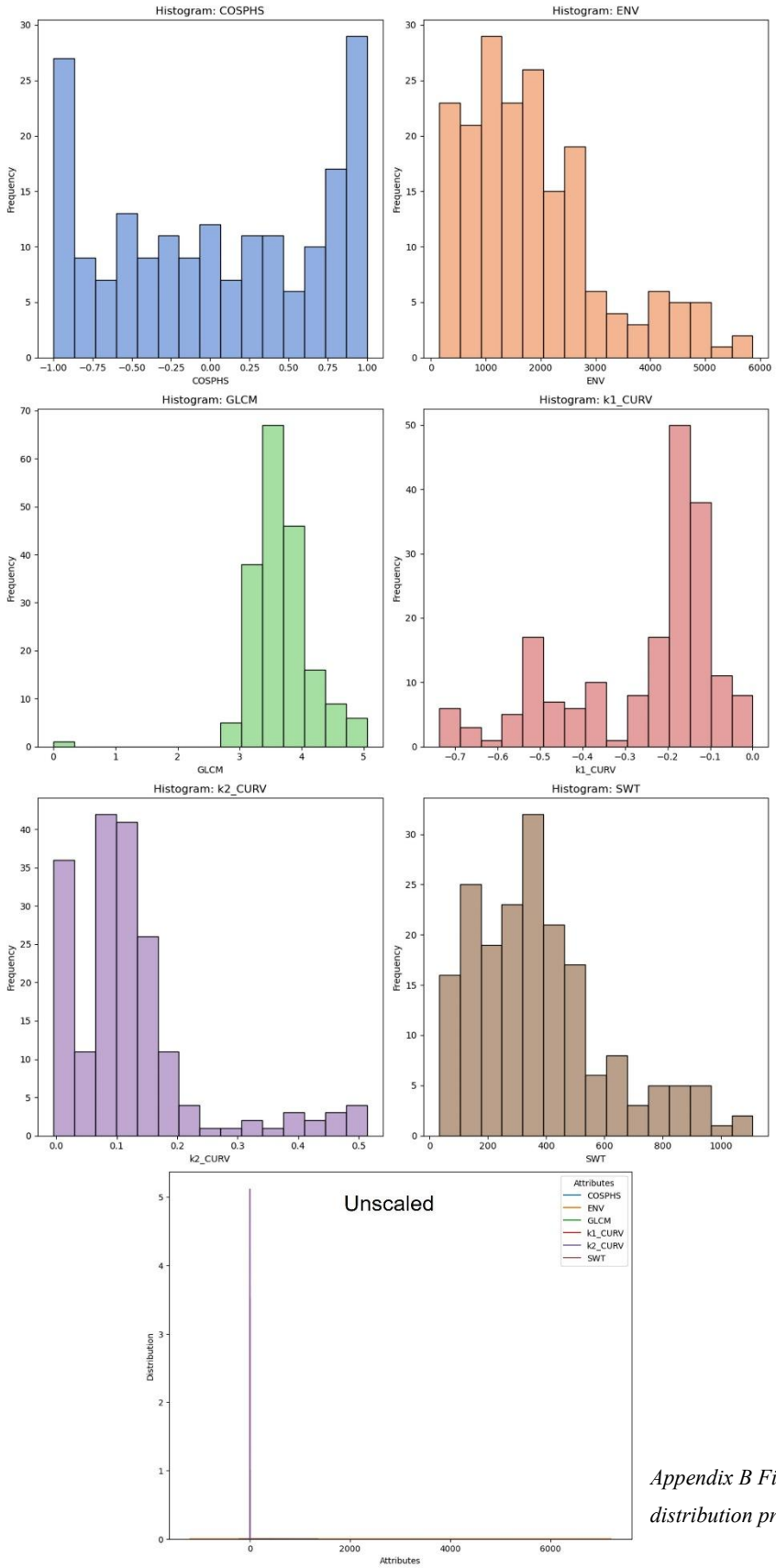
Compared Statistics Before/After Scaling and Post Outlier Removal

	count	mean	std	min	25%	50%	75%	max	range
U_COSPHS	188	0.0383	0.6795	-0.9999	-0.5505	0.0127	0.6959	1.0000	1.9999
S_COSPHS	188	0.5191	0.3398	0.0000	0.2247	0.5063	0.8480	1.0000	1.0000
O_COSPHS	187	0.5217	0.3388	0.0000	0.2288	0.5080	0.8518	1.0000	1.0000
Robust_S_COSPHS	188	0.0205	0.5451	-0.8124	-0.4519	0.0000	0.5481	0.7921	1.6044
Robust_O_COSPHS	187	0.0246	0.5436	-0.8124	-0.4452	0.0027	0.5543	0.7921	1.6044
U_ENV	188	1859.5115	1268.0259	145.7260	919.9796	1647.0233	2486.2750	5865.0562	5719.3302
S_ENV	188	0.2996	0.2217	0.0000	0.1354	0.2625	0.4092	1.0000	1.0000
O_ENV	187	0.2996	0.2223	0.0000	0.1354	0.2603	0.4101	1.0000	1.0000
Robust_S_ENV	188	0.1357	0.8096	-0.9585	-0.4642	0.0000	0.5358	2.6930	3.6515
Robust_O_ENV	187	0.1356	0.8117	-0.9585	-0.4642	-0.0082	0.5391	2.6930	3.6515
U_GLCM	188	3.6820	0.5045	0.0000	3.3922	3.6371	3.9137	5.0517	5.0517
S_GLCM	188	0.7289	0.0999	0.0000	0.6715	0.7200	0.7747	1.0000	1.0000
O_GLCM	187	0.7328	0.0846	0.5596	0.6731	0.7202	0.7748	1.0000	0.4404
Robust_S_GLCM	188	0.0860	0.9673	-6.9737	-0.4697	0.0000	0.5303	2.7123	9.6860
Robust_O_GLCM	187	0.1237	0.8193	-1.5537	-0.4539	0.0017	0.5310	2.7123	4.2660
U_K2_CURV	188	0.1213	0.1098	-0.0047	0.0649	0.1015	0.1441	0.5140	0.5187
S_K2_CURV	188	0.2429	0.2117	0.0000	0.1341	0.2047	0.2868	1.0000	1.0000
O_K2_CURV	187	0.2442	0.2115	0.0000	0.1380	0.2053	0.2877	1.0000	1.0000
Robust_S_K2_CURV	188	0.2502	1.3864	-1.3409	-0.4626	0.0000	0.5374	5.2090	6.5499
Robust_O_K2_CURV	187	0.2584	1.3855	-1.3409	-0.4369	0.0041	0.5434	5.2090	6.5499
U_K1_CURV	188	-0.2593	0.1773	-0.7374	-0.3735	-0.1801	-0.1439	0.0000	0.7374
S_K1_CURV	188	0.6484	0.2404	0.0000	0.4935	0.7557	0.8048	1.0000	1.0000
O_K1_CURV	187	0.6465	0.2396	0.0000	0.4909	0.7551	0.8045	0.9663	0.9663
Robust_S_K1_CURV	188	-0.3447	0.7722	-2.4277	-0.8423	0.0000	0.1577	0.7848	3.2125
Robust_O_K1_CURV	187	-0.3508	0.7698	-2.4277	-0.8506	-0.0018	0.1566	0.6766	3.1043
U_SWT	188	375.7006	231.6662	32.5187	205.7186	344.7879	479.5458	1108.5056	1075.9869
S_SWT	188	0.3189	0.2153	0.0000	0.1610	0.2902	0.4155	1.0000	1.0000
O_SWT	187	0.3187	0.2159	0.0000	0.1581	0.2898	0.4158	1.0000	1.0000
Robust_S_SWT	188	0.1129	0.8460	-1.1404	-0.5079	0.0000	0.4921	2.7891	3.9294
Robust_O_SWT	187	0.1119	0.8482	-1.1404	-0.5190	-0.0016	0.4935	2.7891	3.9294

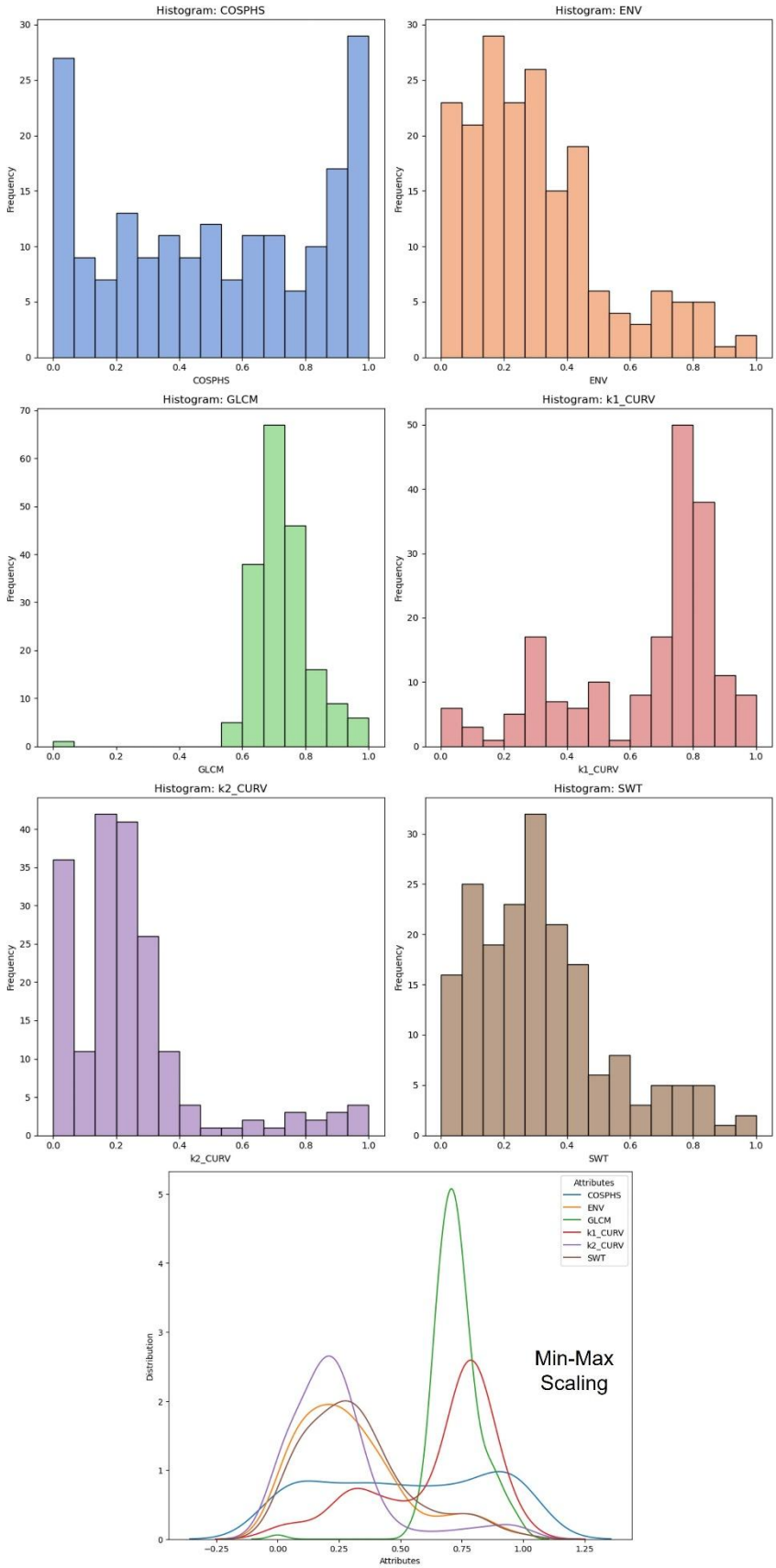
U is unscaled
S is scaled
O is outlier removal

Appendix A Table E: Statistics for chosen attributes before scaling, after scaling, after scaling, and outlier removal. Scaling was conducted using both min-max and robust scaling for comparison.

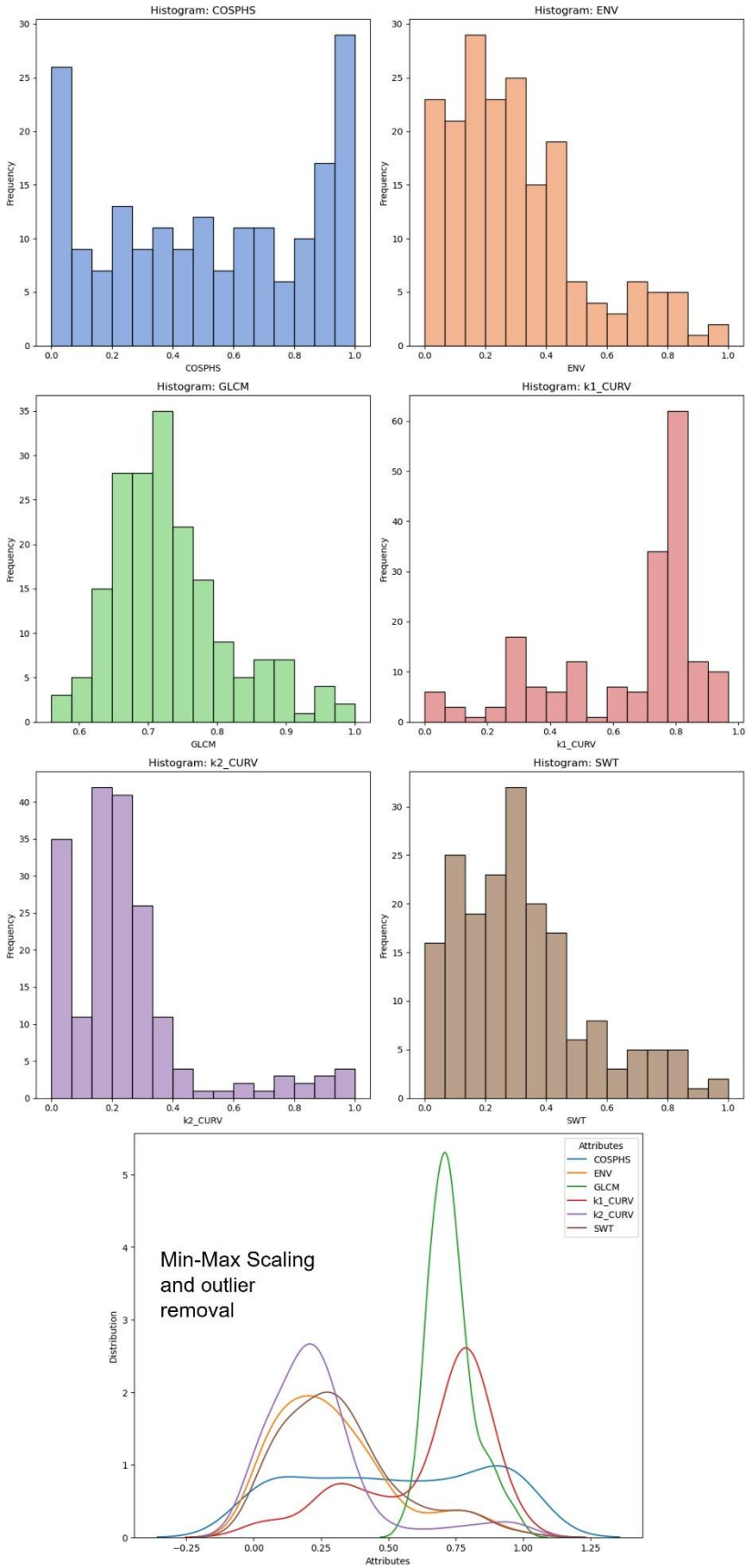
APPENDIX B: Histograms



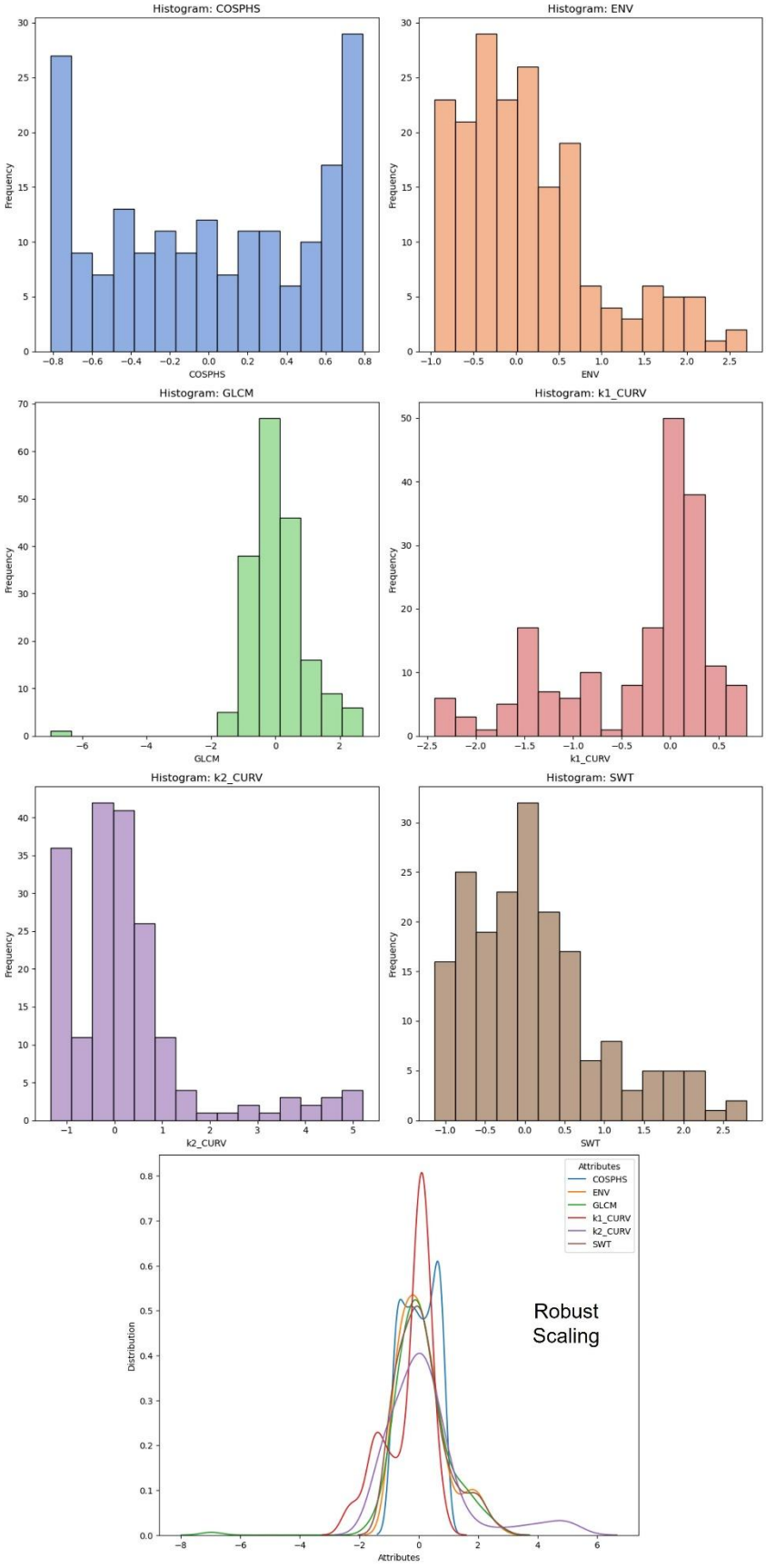
Appendix B Figure A: Data distribution prior to scaling



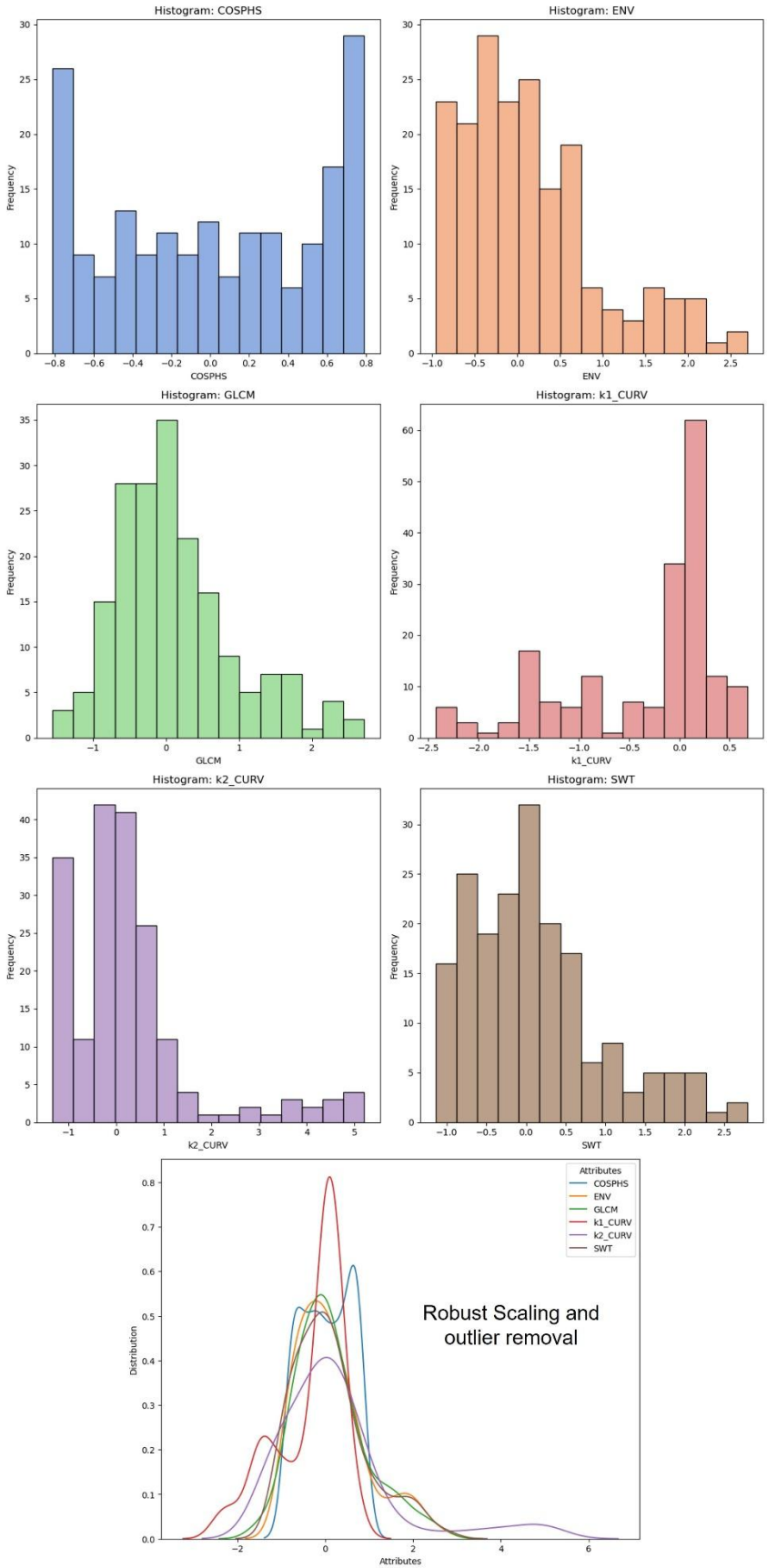
Appendix B Figure B: Data distribution after min-max scaling



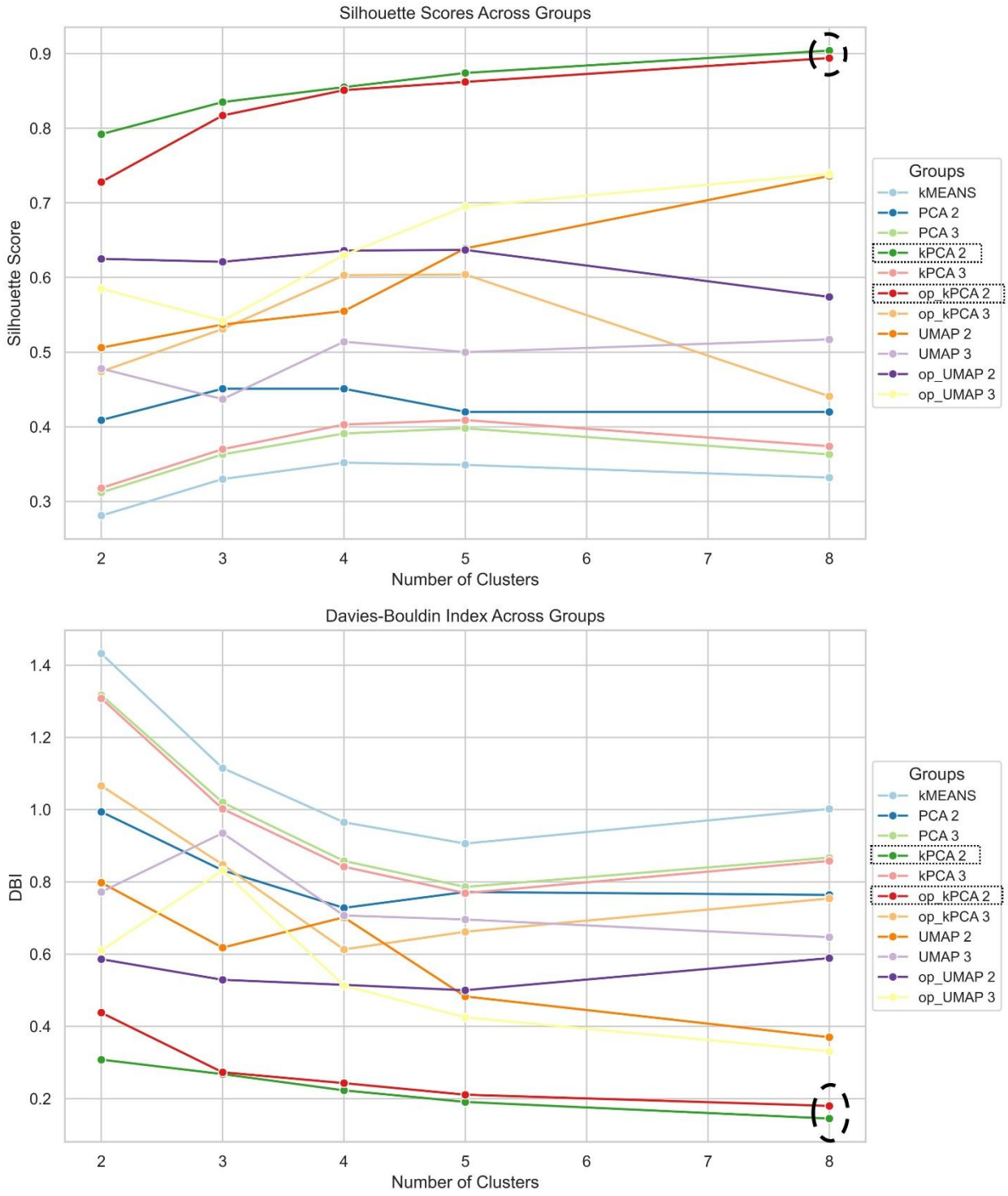
Appendix B Figure C: Data distribution after min-max scaling and outlier removal, from GLCM.



Appendix B Figure D: Data distribution after robust scaling

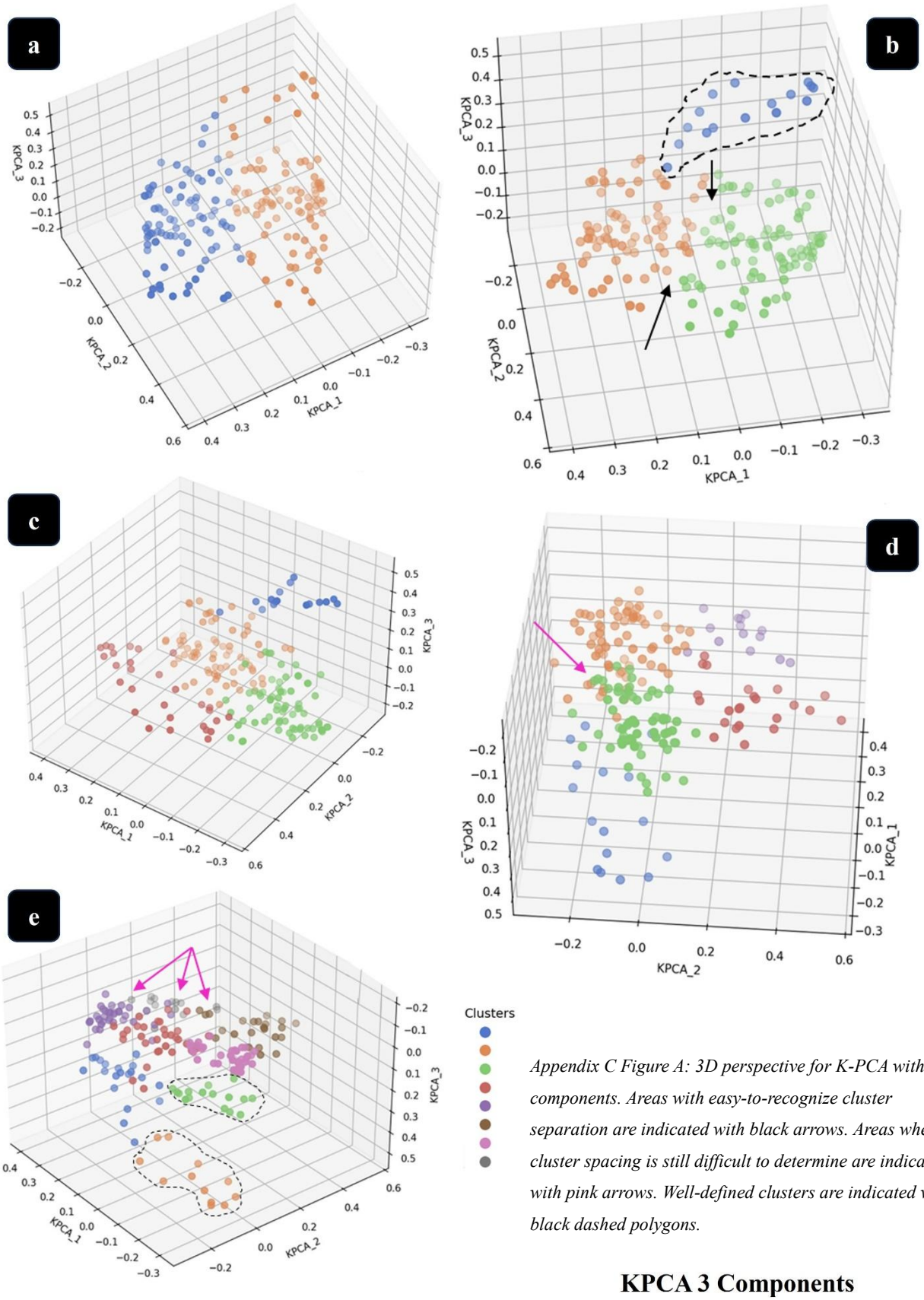


Appendix B Figure E: Data distribution after robust scaling and outlier removal, from GLCM



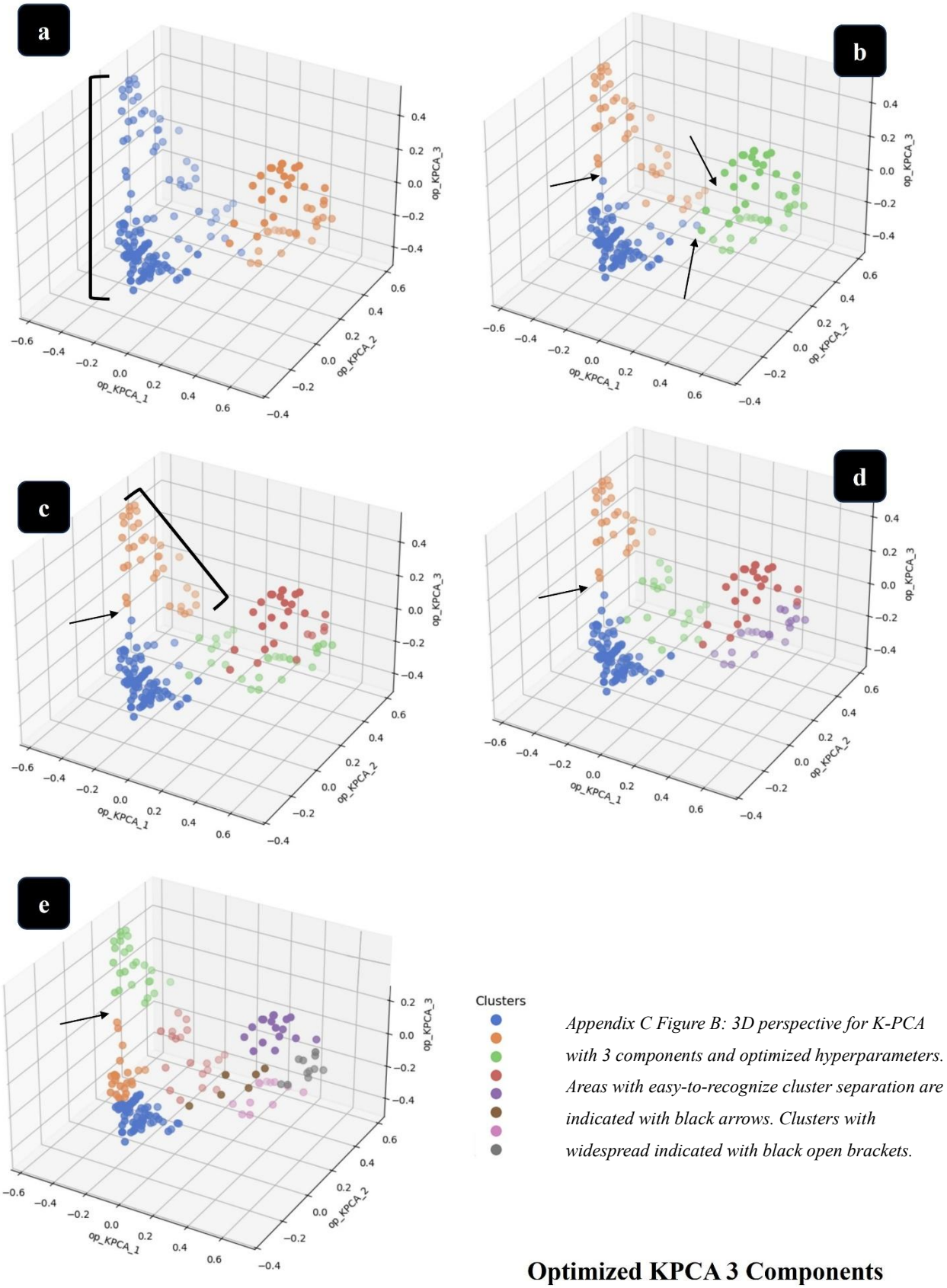
Appendix B Figure F: Top- shows all models and the number of clusters graphed using the silhouette score to provide a visual for which DR technique performs better for cluster performance. Bottom- shows all models and the number of clusters graphed using the DBI to provide a visual for which DR technique performs better for cluster performance.

APPENDIX C: Clustered Models

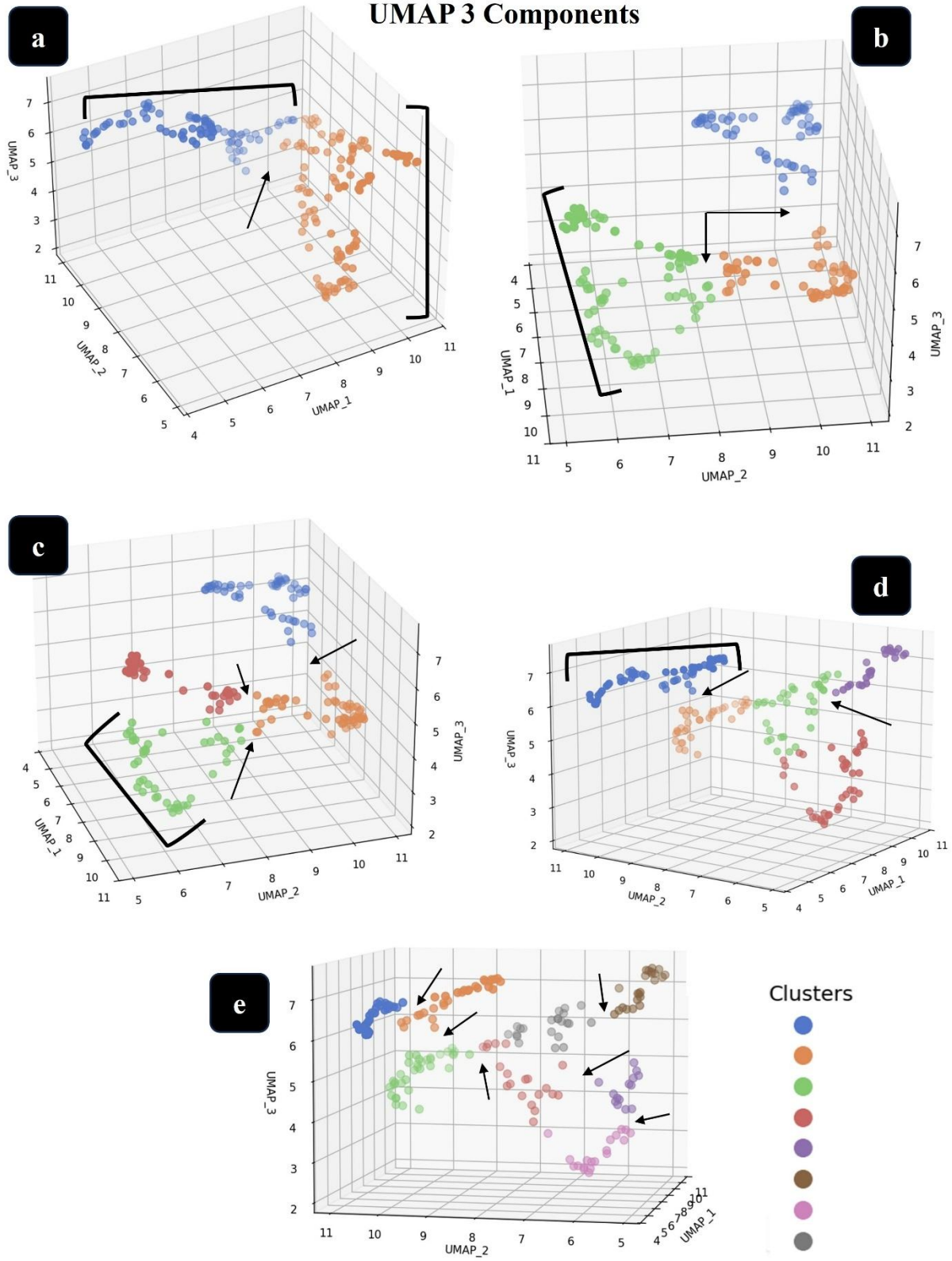


Appendix C Figure A: 3D perspective for K-PCA with 3 components. Areas with easy-to-recognize cluster separation are indicated with black arrows. Areas where cluster spacing is still difficult to determine are indicated with pink arrows. Well-defined clusters are indicated with black dashed polygons.

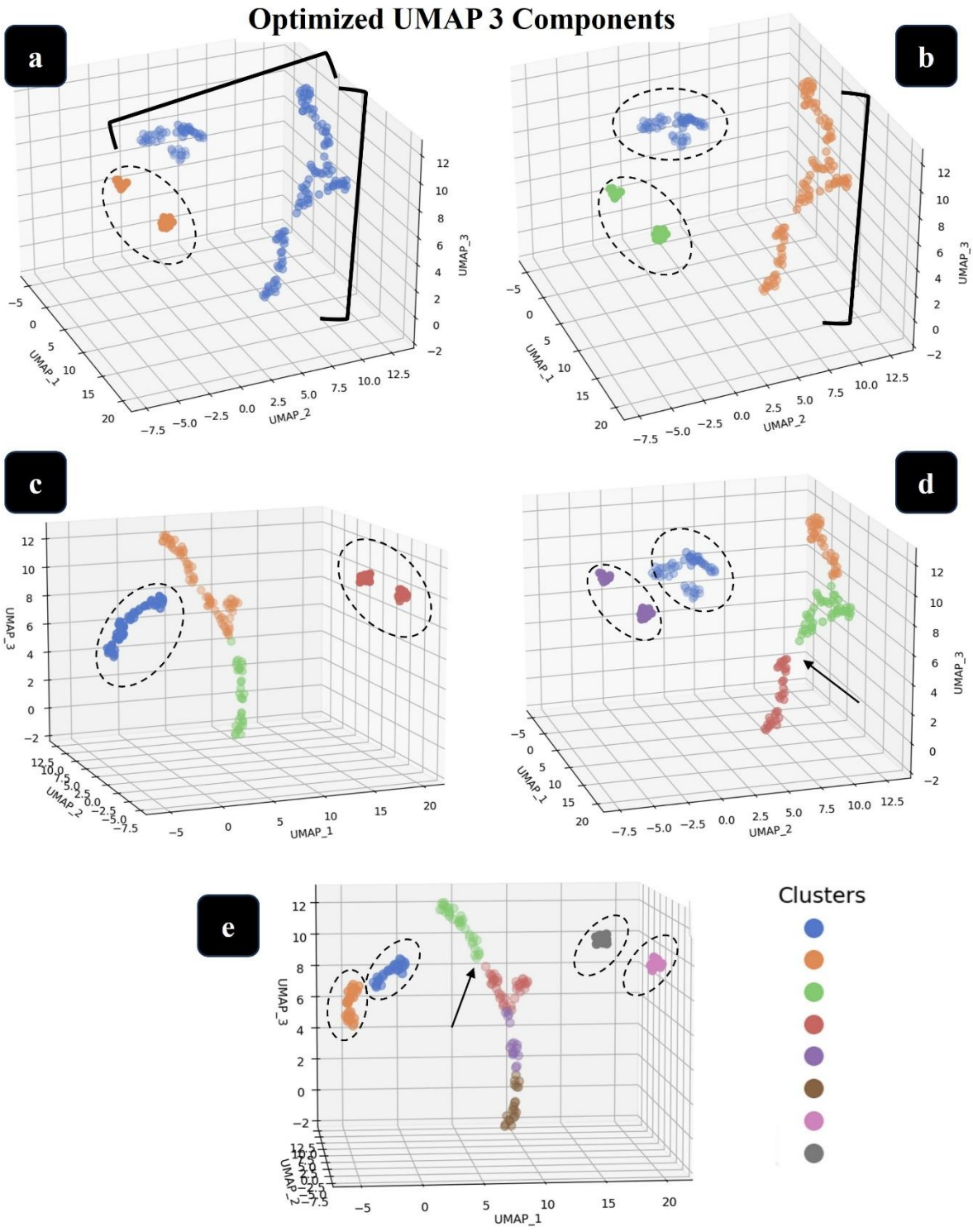
KPCA 3 Components



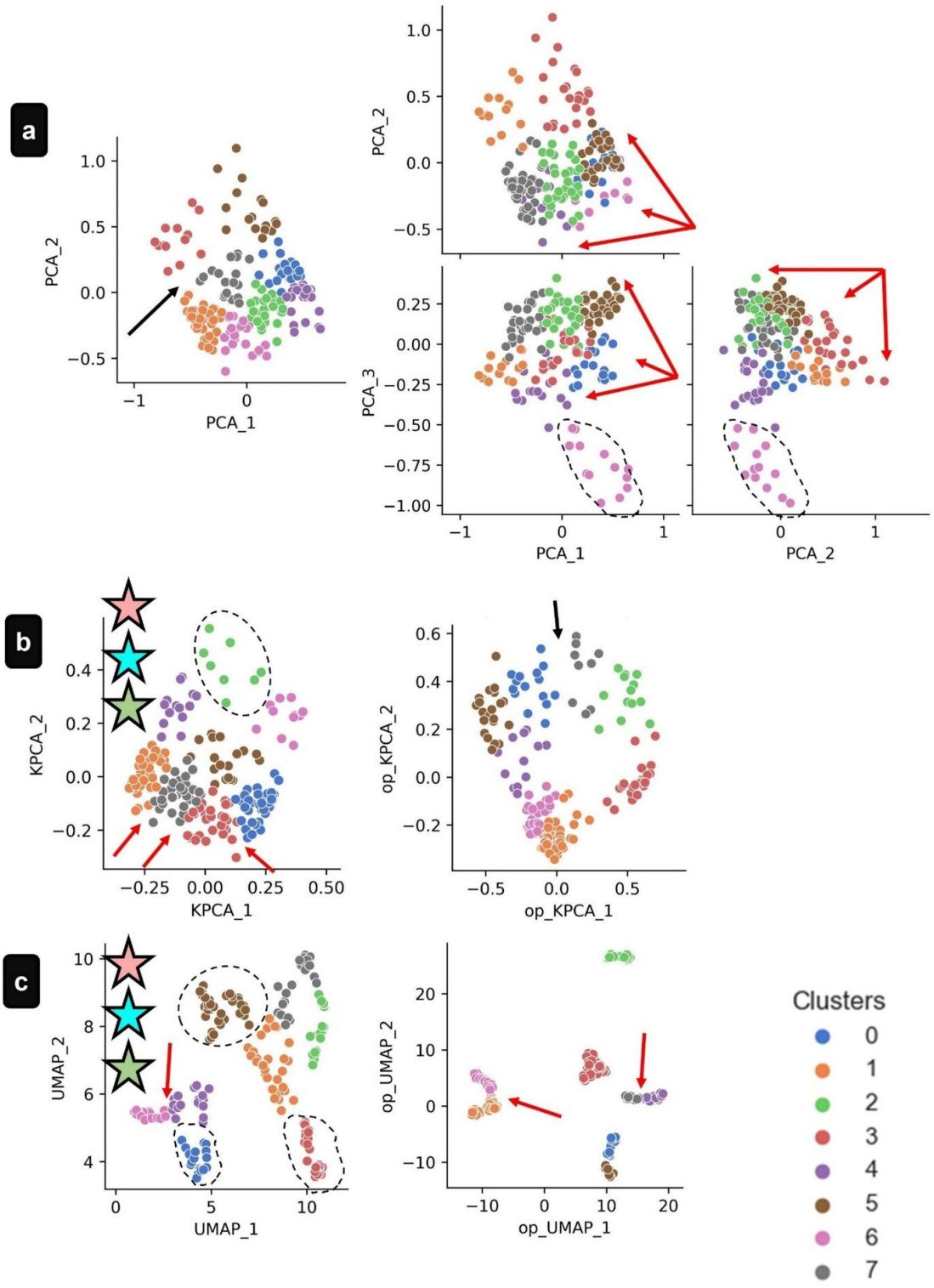
Appendix C Figure B: 3D perspective for K-PCA with 3 components and optimized hyperparameters. Areas with easy-to-recognize cluster separation are indicated with black arrows. Clusters with widespread indicated with black open brackets.



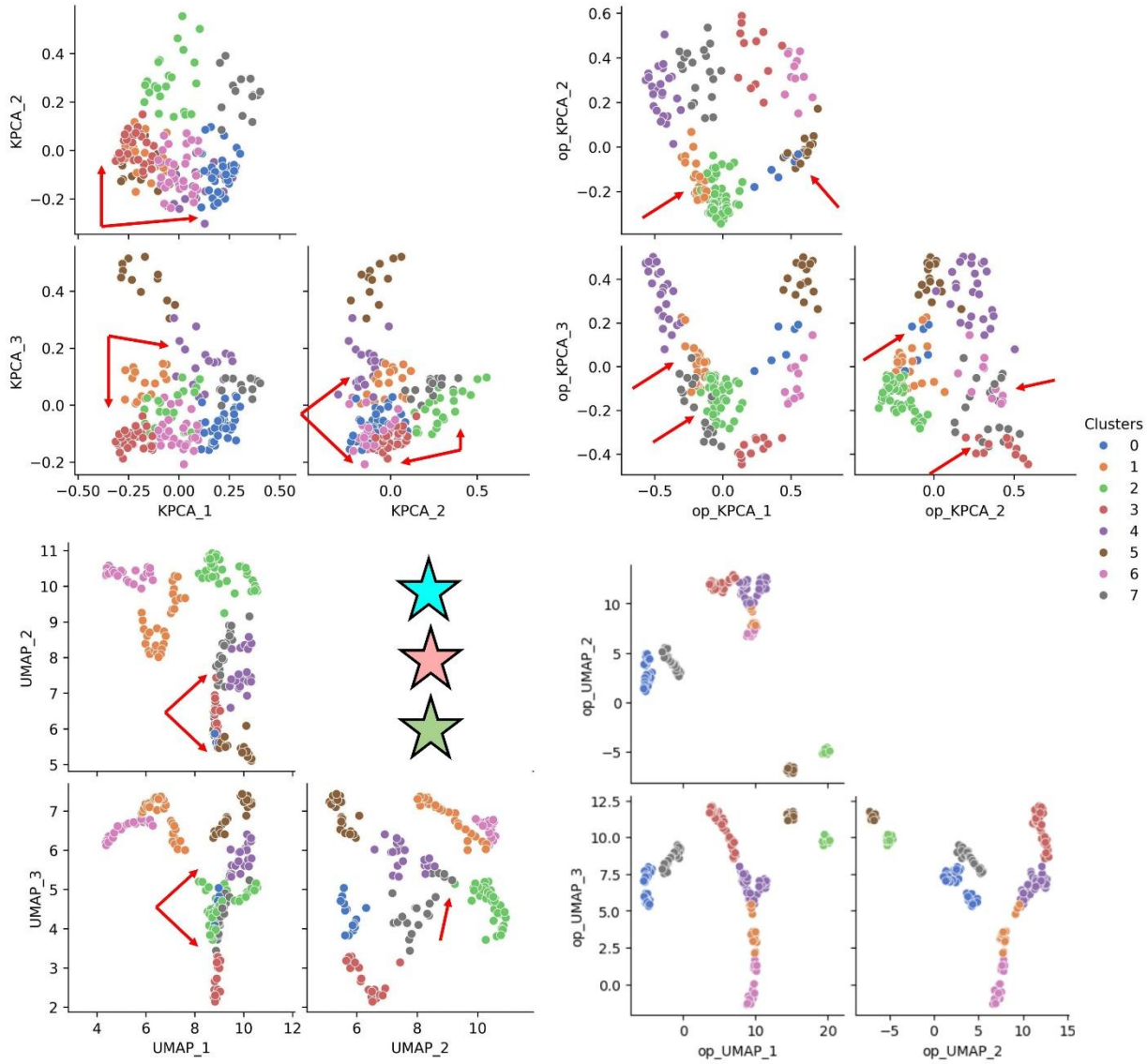
Appendix C Figure C: 3D perspective for UMAP with 3 components. Areas with easy-to-recognize cluster separation are indicated with black arrows. Clusters with widespread indicated with black open brackets



Appendix C Figure D: 3D perspective for UMAP with 3 components and optimized hyperparameters. Areas with easy-to-recognize cluster separation are indicated with black arrows. Clusters with widespread indicated with black open brackets. Well-defined clusters are indicated with black dashed polygons.



Appendix C Figure E: All post-DR models, 2 components, and 8 clusters are shown here. a) PCA b) K-PCA c) UMAP. The model(s) with the more unique clusters are marked with a blue star, Silhouette high performer(s) are noted with a pink star, and DBI high performer(s) are noted with a green star (as ranked within the post-DR group). Areas with more data overlap are indicated by a red arrow. Clusters with good separation are indicated by a black arrow and well-separated clusters are noted with black-dashed polygon. Well-defined clusters are indicated with black dashed polygons.



Appendix C Figure F: Cluster distribution for post-DR K-PCA and UMAP 3 components with and without optimization. Shown are 8 clusters only. A pink star indicates a high performer per silhouette score, and a green star indicates a high performer based on DBI score (as ranked within the post-DR group). Areas with more data overlap are indicated by a red arrow. Clusters with good separation are indicated by a black arrow and well-separated clusters are noted with black-dashed polygon.

Chapter 4: Beyond Amplitude Alone: A Corendering Technique for Enhanced Deepwater Channel Visualization – a Tutorial Part 1: Systematic Optimization Using Cosine Instantaneous Phase and Relative Acoustic Impedance*

*This study is in review in SEG Interpretation

ABSTRACT

Seismic interpretation of deepwater channel systems requires delineation of architectural elements (AEs) whose acoustic impedance contrasts are often subtle and whose amplitude responses can be non-unique. Standard amplitude displays conflate reflector geometry with lithologic properties, complicating facies predictions and architectural boundary mapping.

This tutorial demonstrates a systematic corendering technique using two widely available attributes: cosine of instantaneous phase (cosφ) and relative acoustic impedance (RAI). These attributes address amplitude interpretation limitations by separating complimentary tasks: cosφ provides amplitude-independent mapping of reflector terminations and (dis)continuous reflector patterns that define channel stratigraphy, whereas RAI enables property-based characterization of lateral and vertical facies variations. We present a five-step optimization workflow (attribute extraction, corendering methodology, colormap selection, sampling rate adjustments, and spectral balancing) applied to a deepwater channel complex in New Zealand's Taranaki Basin.

Results demonstrate significant improvement in architectural element visibility compared to conventional amplitude interpretation, enabling confident identification of erosion surfaces, channel stacking patterns, and stratigraphic terminations. This optimized visualization approach provides the foundation for systematic interpretation workflows applicable to exploration settings where limited well control and complexity of deepwater channel settings require transparent evaluation of interpretation confidence.

INTRODUCTION

Geologic interpretation of seismic data depends on identifying geologically significant features. Our channel-focused work emphasizes the need to delineate channel boundaries and

internal architecture. This delineation enables recognition of key stratigraphic termination patterns, including onlap, offlap, and downlap, that are diagnostic of unconformities and stratigraphic/ sequence boundaries.

Once these key markers and patterns have been identified, the interpreter can begin the process of defining the relative timing of depositional and erosional events, which allows for the reconstruction of a depositional environment. When adequate well control is available, these relationships support the development of detailed chronostratigraphic frameworks, as established by Vail et al. (1984). However, in areas with little to no well control, seismic data are often the only tool available. For more information on deepwater channel visualization, see Slatt 2013; Weimer et al., 2006; and Posamentier 1993.

Contemporary geochronostratigraphic workflows are enhanced by specialized software that, under interpreter guidance, automatically tracks surfaces across study areas to establish "relative geologic time" relationships (e.g. Wharton, 2015; Qayyum et al., 2015). While ongoing research addresses automated termination identification, current interpretation practice relies primarily on interactive recognition of these critical horizon relationships. Our analysis focuses on distinguishing individual architectural components within the channel system. This detailed characterization enables application of depositional process understanding to reconstruct system evolution and identify elements exhibiting different lithologic properties, specifically discriminating between sand-dominated and shale-rich architectural elements (AEs).

Accurate delineation of AEs remains challenging due to the subtle acoustic impedance contrasts from one layer to the next. Conventional amplitude analysis may fail to resolve these critical boundaries and internal geometric indicators necessary for robust depositional environment reconstructions, as amplitude primarily focuses on interface properties rather than intra-layer characteristics necessary to provide a robust interpretation.

To address these challenges, we present a five-step workflow for optimizing seismic displays: 1) attribute selection to highlight channel boundaries and internal variations, 2) corendering to delineate key features, 3) colormap selection to maximize feature discrimination, 4) sampling rate optimization, and 5) spectral balancing considerations. This systematic approach to visualization enhancement enables clearer delineation of important stratigraphic features and architectural elements critical for interpretation.

While advances in machine learning are successful in mapping seismic facies that exhibit a similar seismic expression across a given survey (e.g., salt or karst collapse; see Wang et al., 2018; Maas et al., 2023; Lubo-Robles et al., 2022; Ebuna et al., 2018; Gui et al., 2025), channel systems are more challenging, where each architectural element that comprises a channel system may appear as a distinct seismic facies. These individual architectural elements are then spatially linked to each other through geologic processes. Mastery of such processes is essential for successful human interpretation. At present, it is unclear how to best "teach" a machine how to use such geologic models. Our workflow demonstrates how by properly corendering two simple, widely available attributes, the cosine of phase, cosφ, and relative impedance inversion, RAI, can better delineate stratigraphic boundaries than using amplitude. This integrated visualization enhances deepwater channel analysis without requiring specialized algorithms or extensive computational resources. By better delineating these architectural elements, skilled interpreters can use their understanding of deepwater channel depositional processes to better map stratigraphic elements that may form reservoirs and possible flow barriers.

GEOLOGIC SETTING

The Taranaki Basin provides an ideal case study for testing interpretation methodologies due to its extensive research coverage. Located offshore northwestern New Zealand (Figure 1), this Mesozoic-age rift system has been thoroughly characterized through comprehensive geologic investigations (Bull et al. 2020; Di et al. 2020; Grahame 2015; Kamp et al. 2004; Thrasher 1992, King 1988; Kroeger et al. 2019; Naewboonien 2021; Li et al 2017; Reilley et al. 2015; Strogen et al. 2017; Strogen et al. 2019) and advanced machine learning applications to geophysical processes (Chopra et al. 2022; Lubo-Robles 2019; Li et al. 2023; La Marca 2022 and 2020; Chopra et al. 2019; Verma 2021).

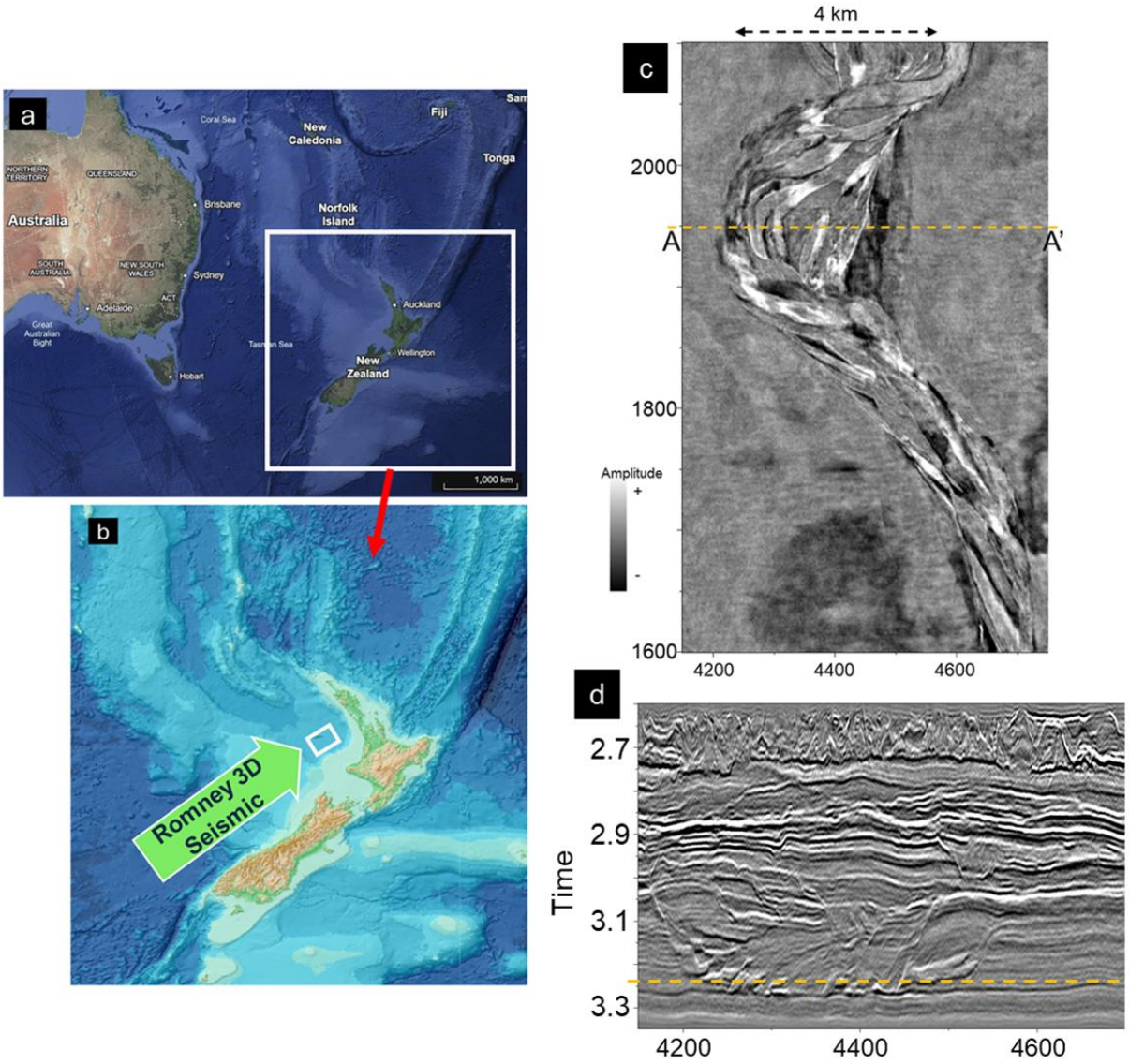


Figure 1. (a) Area of study is off the northwest side of the north island of New Zealand, courtesy Google Earth. (b) Small white polygon indicates the location of the Romney 3D seismic survey, courtesy of The GEBCO grid. (c) Romney timeslice at -3.24 and (d) inline at 1950.

The Romney 3D seismic survey, acquired in 2011, provides the opportunity to study a well-defined deepwater channel complex. The survey covers a 35 km x 55 km area with a 12.5 m x 25 m bin size and 4 ms sampling interval. While the seismic data quality allows for structural and stratigraphic analysis, the channel architecture elements benefit significantly from the application of our integrated visualization workflow. The Romney-1 well provides the sole direct geologic control within the volume, though positioned outside the primary channel target area. This limited well control scenario makes the Romney seismic dataset suitable for demonstrating

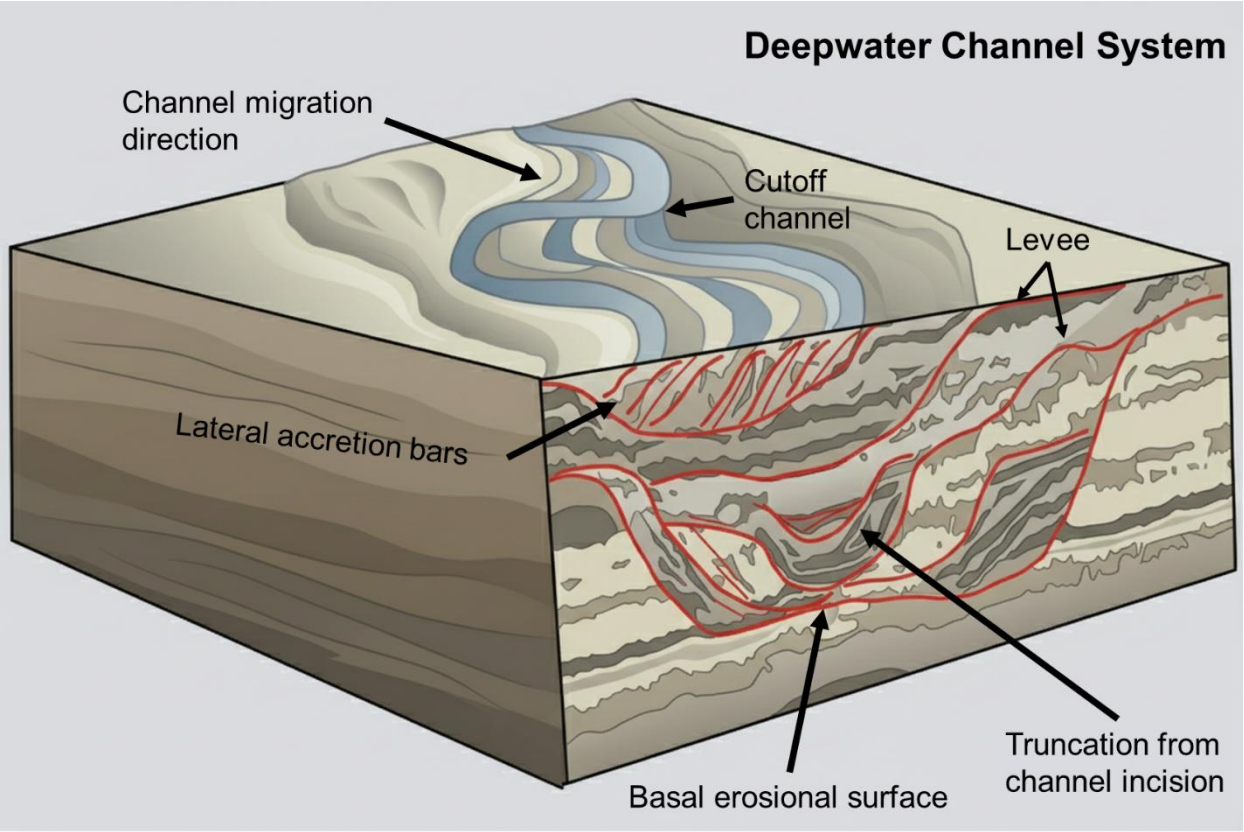


Figure 2. Cartoon illustrating some of the key architecture elements of the turbidite channel addressed in this paper

methodologies applicable to exploration settings. In this tutorial, we focus on identifying key elements, including channel incisions (erosional basal surfaces), progradational features such as lateral accretion bars, confining levees, and abandoned cutoff channels. Each element exhibits characteristic seismic expressions related to its geometry and internal fill characteristics, making accurate delineation critical for understanding channel evolution and reservoir distribution.

We apply the term "channel stratigraphy" to describe the geometric relationships between reflectors within and bounding these architectural elements, analogous to sequence stratigraphic principles but applied at the channel scale. Specifically, we use stratigraphic termination patterns (truncation, onlap, downlap, etc.) to interpret channel morphology, fill chronology, and migration history. For example, basal truncations indicate erosional channel incision, onlap geometries may reveal progressive fill or lateral migration, and draping patterns suggest abandonment phases. Figure 2 illustrates a few of the key architectural elements associated with a turbidite deepwater channel complex, showing basal erosional surfaces, truncation from channel incision, and levees. While these stratigraphic relationships use larger-scale sequence stratigraphy concepts, their interpretation at the channel scale provides insights into localized flow dynamics, sediment

bypass versus depositional accumulation, and temporal evolution within the deepwater system Clear identification of these geometric patterns form the foundation for the interpretation framework presented in Moreno-Ward et al., in review.

METHODS: HOW WE OPTIMIZED THE DISPLAY

Our workflow consists of five steps focused on optimizing two fundamental seismic attributes (cosine of phase and relative acoustic impedance) that are readily available in most commercial software to enhance the visibility and identification of deepwater channel stratigraphic features (Figure 3). **Step one** is attribute selection; we focus on cosφ and relative acoustic impedance for their complementary capabilities in boundary detection and layer characterization. Other combinations of attributes may be used based on the geologic question and targeted features. **Step two** implements corendering to integrate both attributes in a single display, allowing simultaneous visualization of reflector terminations from cosφ and acoustic property variability from RAI. **Step three** addresses colormap optimization. Brown (2011) states that monochrome color bars are best in enhancing edges in seismic and attribute data. We therefore use gray scale and black/white for cosφ to maximize human visual sensitivity to spatial discontinuities, while modifying the RAI initial dual polarity colormap to reduce visual fatigue from high-contrast colors. The **third step** addresses colormap optimization. We implemented the stark black/white for cosφ to maximize human visual sensitivity to spatial discontinuities, while modifying the RAI initial colormap to reduce visual fatigue from high-contrast colors.

Step four addresses the sampling interval, which is critical for cosφ fidelity. The original 4 ms sampling interval only captures approximately 6 samples per wavelength at the 50 Hz dominant frequency; this can create graphical aliasing artifacts, which can appear as discontinuities. We addressed this limitation by interpolating to a 1 ms sample interval before attribute computation, increasing the number of samples per second, which better represent phase

variations, thereby reducing feature smearing that degrades the resolution of thinner beds and can mask stratigraphic boundaries. Finally, **step five** applies spectral balancing to enhance the vertical resolution, critical to stratigraphic analysis.

Our workflow is based on the following observations:

- Zero-phase wavelets and well-defined horizons produce reflector peaks and troughs that correspond to geological layer interfaces
- Cosφ maps these features with +1 values for peaks and -1 values for troughs
- Relative acoustic impedance applies a -90° amplitude rotation while enhancing low-frequency content, better representing layer properties between the interfaces defined by amplitude peaks and troughs.

This analysis supports our hypothesis that integrating relative acoustic impedance with cosφ in vertical section displays will enhance geologic unit delineation. The workflow sequence we applied to the Romney seismic dataset provides one systematic approach that readers can replicate, adapt, or reorder based on their specific data quality, geologic objectives, and available software capabilities. While we've demonstrated this framework using a deepwater channel system, these steps can be followed for any stratigraphic reservoir characterization problem. The fundamental principle remains constant: leverage the complementary information from amplitude-based and phase-based attributes to better resolve geologic stratigraphic features. Interpreters should view this workflow as a flexible template rather than a rigid order of processes, adjusting the sequence of operations, parameter selections, and attribute combinations

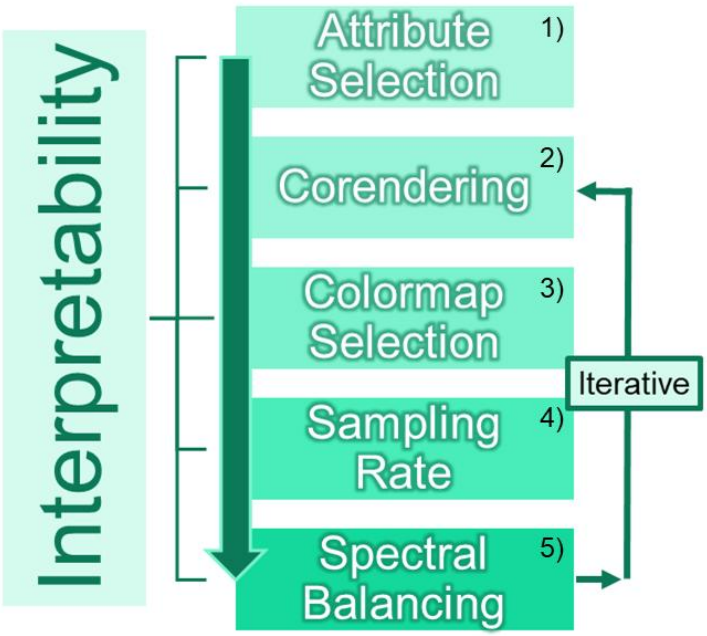


Figure 3: Workflow for improved interpretability, which consists of five major steps through an iterative process.

to suit the specific geologic objectives. The key for channel stratigraphic interpretation is to integrate attributes that highlight geologic boundaries and give information regarding rock properties, allowing for more confident interpretation based on improved visualization.

STEP ONE: ATTRIBUTE SELECTION

Channel stratigraphy interpretation and identification are dependent on the interpreter's ability to identify key geologic and geometric features. While there are many attributes associated with channel features, such as coherence, curvature, and sweetness, to list a few, we chose to focus our attention on two attributes that are commonly available in seismic software. Relative acoustic impedance provides information about potential changes in lithologies and rock properties (beyond seismic resolution), while cosφ helps identify reflector terminations associated with the AEs of interest. For example, Figure 4a shows a time slice at t=3.21 s; while we can use coherence to identify some of the channel boundaries in plan view, if we take a cross-section from A-A' (Figure 4b), it is inferior to the original seismic amplitude and captures little of the variability within this deepwater channel complex. Looking at the same crossline, but this time in amplitude (4d), we see substantially more lateral and vertical boundaries. Yellow arrows indicate the basal channel boundary, which can easily be traced to the right and left, with the white arrow marking the left-most bounding edge of the channel complex. Orange arrows indicate possible levee development; green is just left of a younger channel that shows internal patterns associated with lateral progradation. The pink arrows show a break in the lateral continuity of the reflector; we can follow this reflector dip and map out another iteration of channel incision here.

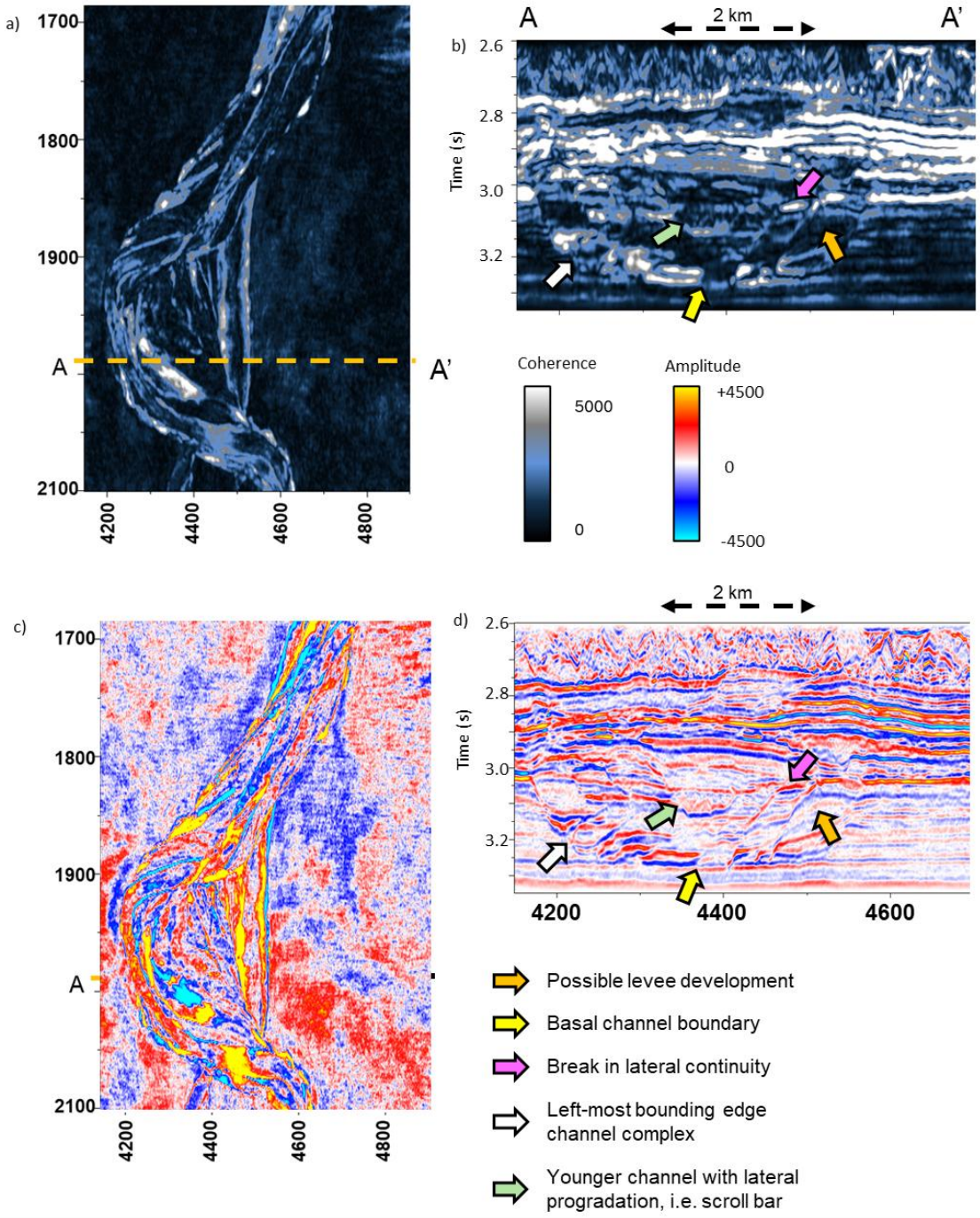


Figure 4: (a) Time slice at t=3.21 s through a coherence volume showing the edges of individual channels that comprise a turbidite system. Line AA' showing a complex channel through the (b) coherence, and the original seismic amplitude volumes in (c) plan view and (d) cross-sectional view. Although coherence does a good job of delineating channel edges when seen on the time slice in (a), it does a poor job on the vertical slices, providing less detail than the original seismic amplitude data in (d). For this example, the seismic data have a 4 ms sample increment and a 12.5 m by 25 m bin size. The coherence was computed using an analysis window containing 9 traces and 3 samples along the structural dip. Seismic data courtesy of NZP&M.

PHASE

Seismic trace analysis fundamentally involves two components: the directly recorded hydrophone measurements (representing kinetic energy) and a 90° phase-rotated variant of the Hilbert transform calculations of adjacent samples (representing kinetic energy density). The instantaneous attribute suite, established by Taner et al. (1979), provides quantitative analysis of seismic reflection characteristics at individual pixels (2D) or voxels (3D), with instantaneous phase being a key component. Since these attributes derive from both the original measurements and their Hilbert transforms, they represent localized rather than truly instantaneous calculations.

This section examines three key aspects of phase-based interpretation:

- Standard interpretation practice involves direct mapping of structural and stratigraphic discontinuities from cosφ displays. However, phase calculations inherently ignore reflection strength, allowing noise from weak amplitude vents to contaminate the interpretation. Hart (2008b) mitigated this through the envelope weighting of cosφ, while alternative approaches involve corendering with seismic amplitude.
- Figure 5 shows line AA' viewed through (a) corendered amplitude and the cosine of instantaneous phase, cosφ, (Figure 5b) instantaneous phase, φ, (Figure 5c) the cosine of instantaneous phase, cosφ, and (Figure 5d) RAI resampled to 1 ms corendered with cosφ. We have overlain the same arrows from Figure 4 to observe how each attribute or corendering of attributes delineates these features.
- Figure 5a demonstrates a common visualization challenge: the default amplitude colormap (yellow, orange, red, brown, white, gray, blue, cyan) corendered with cosφ overlay (white for cosφ > +0.8 and black for cosφ < -0.8, with maximum opacity), produces excessive white throughout the panel that obscures rather than clarifies reflector continuity. The juxtaposition of higher amplitude reds adjacent to white cosφ values creates visual fatigue and makes sustained interpretation difficult, while the limited presence of yellow and cyan against the whites and reds further reduces interpretability.

This approach fails for a couple of reasons. Visually, amplitude variability using this colormap creates display clutter that obscures the geometric information provided by cosφ. The other issue with using amplitude is that these responses are non-unique: multiple effects become conflated, including bed thickness, lithologic variations, cementation, limited variability between layers, and processing artifacts. Amplitude provides information about acoustic impedance contrast strength, but it cannot distinguish whether amplitude changes reflect changes in bed thickness, for example, or lithologic variation (facies changes, or changes in cementation), limiting its effectiveness for delineating channel architectural elements. Figure 5b shows φ displayed with the roma colormap, which reveals some stratigraphic features and improves visual comfort compared with Figure 5b, but lateral reflector continuity remains incompletely resolved. The φ display provides amplitude-independent geometry but cannot fully delineate continuous reflector surfaces without additional property-based context. Figure 5c displays cosφ alone, which excels at revealing lateral reflector continuity and stratigraphic termination patterns that define channel architecture. However, this amplitude-independent geometric display lacks information about impedance variations within and between layers, preventing interpretation of lithologic properties, facies distributions, or depositional processes occurring within the stratigraphic units.

Figure 5d demonstrates the optimized workflow result: co-rendering cosφ with RAI integrates the superior lateral reflector continuity tracking capability of cosφ with the impedance-based lithologic information provided by RAI. This combination enables interpreters to simultaneously identify stratigraphic boundary geometry and assess relative impedance variations within layers, providing both the geometric framework and the property-based context necessary for robust stratigraphic interpretation. Orange arrows indicate locations with possible levee development, and pink shows areas with a break in lateral continuity. Yellow, white, and green arrows indicate channel boundaries.

Standard instantaneous phase displays utilize cyclical color schemes that cycle through the color spectrum (Figure 5b), assigning identical colors to ±180° values. However, Brown (2011) demonstrated enhanced human visual sensitivity to spatial boundaries using monochromatic displays. While constructing monochromatic color bars with equivalent ±180° presentations is possible, computing the cosφ proves more practical and offers additional advantages, slower

variation, and a 0° and ±180° that accentuate peak and trough contribution in zero-phase amplitude data. Our grayscale implementation assigns cos (0°) peaks to white and cos (±180°) troughs to black (Figure 5c).

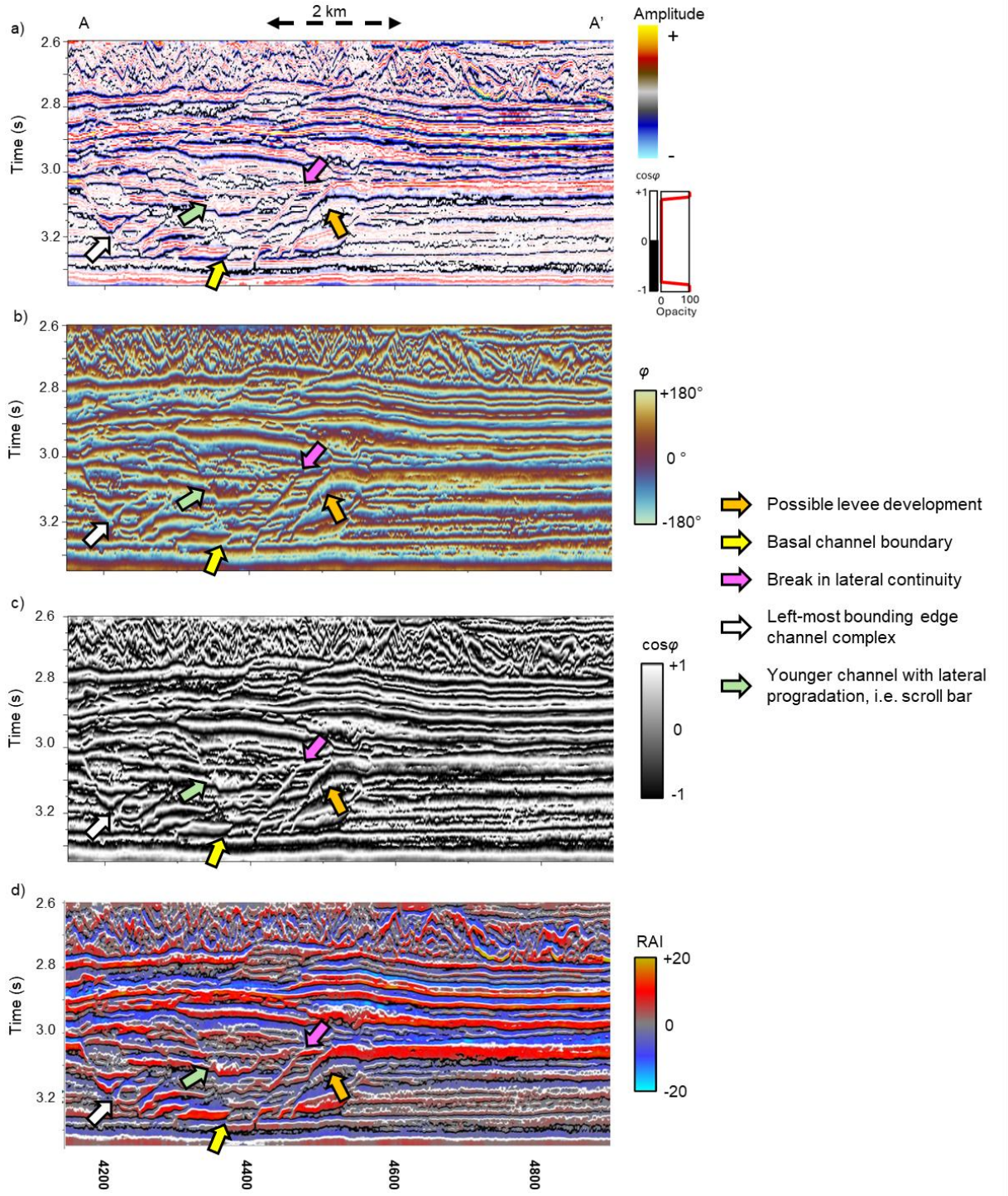


Figure 5: Line AA' through (a)) corendered amplitude and the cosine of instantaneous phase, cosφ (b) instantaneous phase, φ, (c) the cosine of instantaneous phase, cosφ,, and (d) relative acoustic impedance corendered with the cosine of instantaneous phase, cosφ. Corendering the cosine of instantaneous phase with the original seismic amplitude failed to better delineate stratigraphic boundaries. The color bar used for the cosφ overlay in (c) displays only the strongest values with white for cosφ > +0.8 and black for cosφ < -0.8. (d) In contrast, corendering cosφ with RAI resampled to 1 ms after spectral balancing, better delineates stratigraphic boundaries.

STEP 2: CORENDERING THE COSINE OF INSTANTANEOUS PHASE AND RELATIVE ACOUSTIC IMPEDANCE

Integrated seismic attribute analysis seeks to enhance interpretational capabilities by systematically combining information from multiple seismic attribute volumes, creating corendered displays that exceed the interpretational value of the individual attributes. This approach addresses a fundamental limitation in traditional single-attribute interpretation, the inability to simultaneously capture both geometric and amplitude characteristics of deepwater channel features within a unified analytical framework. Beyond the technical advantages, integrated attribute workflows facilitate communication across multidisciplinary teams by generating intuitive visual representations that convey complex geologic relationships more effectively than single attribute displays alone. These enhanced visualizations prove particularly helpful when presenting seismic interpretations to diverse audiences who may lack extensive seismic or geologic interpretation experience but need to have a clear understanding of the subsurface for decision-making processes. With our attributes corendered, **Step** 3 optimizes the colormaps to ensure both attributes remain visually distinct and the features of interest can be identified.

STEP 3: THE CHOICE OF COLOR BARS

The first issue with Figure 5a is that it is difficult to see the black and white extrema of the cosφ *volume* on top of the RAI volume using a yellow-red-white-blue-cyan amplitude color bar. Clearly, we don't wish to use white for the zero crossing of an amplitude or relative acoustic impedance volume. Furthermore, the white peak of cosφ is difficult to see on yellow, as is the black trough of cosφ on blue (Figure 6a). We therefore modify our relative impedance color bar as indicated by Figure 6b to generate the vertical slices of RAI shown in Figure 7a and RAI corendered with cosφ in Figure 7b where we can now more easily follow the peak and trough values of the cosφ attribute. For zero-phase data, the peaks and troughs of seismic amplitude

represent the interfaces between geologic layers. In contrast, the RAI better represents the relative acoustic impedance of the layers themselves, at least within the limits of seismic resolution.

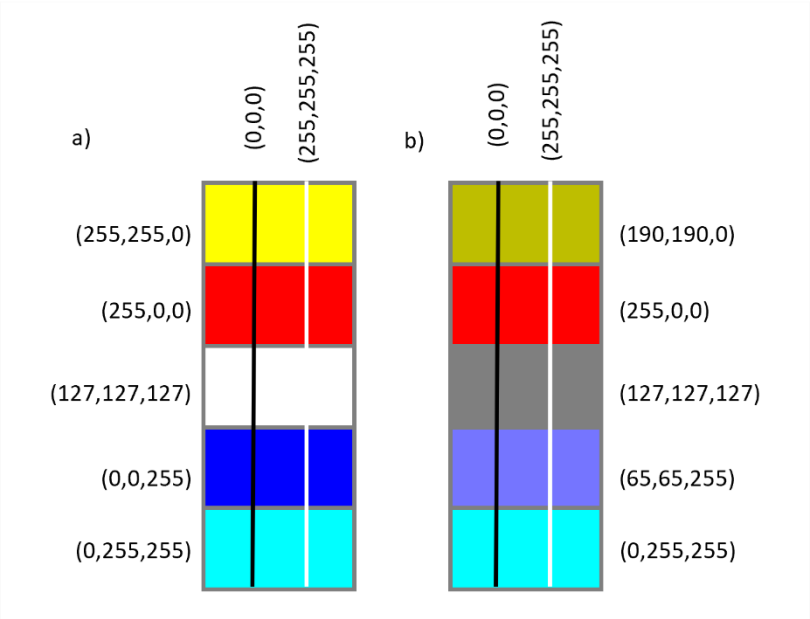


Figure 6. (a) The major components of the color bar initially considered. Clearly, the white value of cosφ=1 cannot be seen on white, but it also shows up poorly on yellow. In contrast, the black value of cosφ=-1 shows up well on white and yellow but poorly on blue. (b) A modified color bar, which we will use in the next figures to display relative acoustic impedance that allows corendering of cosφ. The triplets indicate the RGB values of each color.

STEP 4: INTERPOLATION TO A FINER SAMPLE INCREMENT

Close inspection of Figure 7b reveals discontinuous cosφ anomalies that limit their value in delineating the seismic facies indicated by the RAI. The spectrum of these data range between 5 and 85 Hz, such that reflectors with a dominant frequency of 50 Hz (or period of 20 ms) will be sampled only six times, potentially missing the extrema of $|\cos\varphi|>0.8$. To address this display issue, Figure 6c shows the same corendered attributes computed from the data resampled to a 1 ms sample increment. Figure 7d and 7e show zoomed images of the data that fall in the green rectangles in Figures 7b and 7c. Note the improved continuity of the cosφ attribute indicated by

the yellow, green, and orange arrows achieved by simple interpolation. This improvement is solely due to the display of data at discrete samples, with no increase in seismic resolution.

Figure 7 (a-c) shows our same A-A' line: (a) is RAI alone and (b) shows RAI corendered with cosφ in the original 4 ms sample increment data. (c) Is the same line AA' but now with the data having been upsampled to 1 ms. Focusing on the section highlighted in the green rectangle, we can compare how sample increment can affect reflector response. We see more continuous stratigraphic delineations in (e) data interpolated to 1ms vs. (d) data in the original 4 ms. Yellow arrows show a potential levee development in two different phases of channel migration, green arrows show a break in the lateral continuity of the reflector at the bounding edge of the channel complex, and the orange arrows indicate a broken delineation in (d) and more continuous stratigraphic delineations in (e). The improved continuity achieved through resampling in **step** 4 revealed that our visualization was still limited by the narrow frequency content of the original data. Let's address this now in **step** 5.

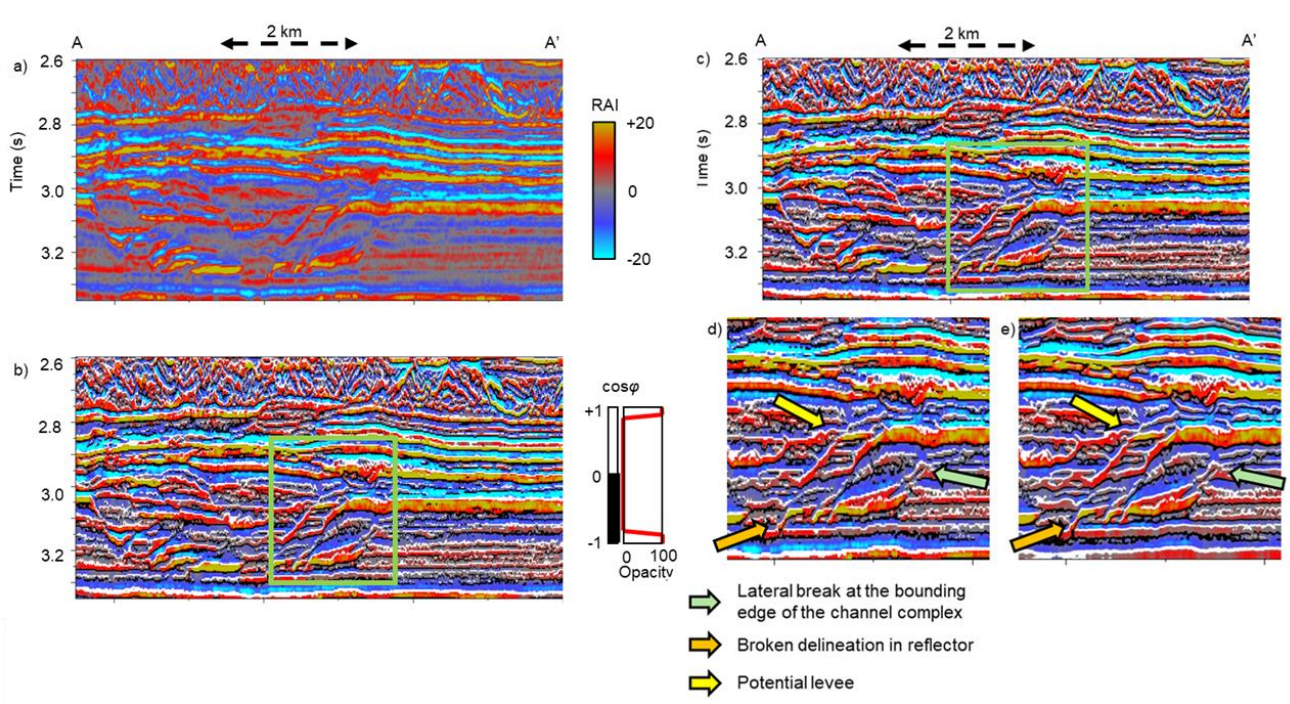


Figure 7. (a) Line AA' through the relative acoustic impedance (RAI) volume, which rotates the seismic amplitude data by -90° and increases the low-frequency component of the data. (b) RAI corendered with cosφ. This image shows that the white cosφ≈+1 anomalies show up well against the dirty yellow high positive RAI response, and the black cosφ≈-1 anomalies show up well against the light blue moderately negative RAI anomalies. These two attributes were computed from the original 4 ms sample increment data. (c) Line AA' but now through RAI and cosφ volumes computed from data that have been interpolated to a sample increment of 1 ms. For a period of 50 Hz, there are only six samples per cycle, such that cosφ is under-sampled at the 4 ms sample increment in (b). Zoomed images corresponding to the green rectangles of (d) the 4 ms volume shown in (b), and (e) the 1 ms volume shown in (c).

STEP 5: SPECTRAL BALANCING

Initial analysis of the seismic data revealed frequency content concentrated primarily within the 10-55 Hz range (Figure 8a), exhibiting the time-variant spectral characteristics typical of marine seismic data. To optimize resolution for stratigraphic interpretation, we applied the spectral balancing methodology described by Chopra and Marfurt (2015), which broadens and flattens the frequency spectrum to achieve more uniform amplitude distribution across the available bandwidth (Figure 8b). This processing extended the effective frequency range to 5-80 Hz, significantly enhancing the volume's resolving power for thin-bed analysis.

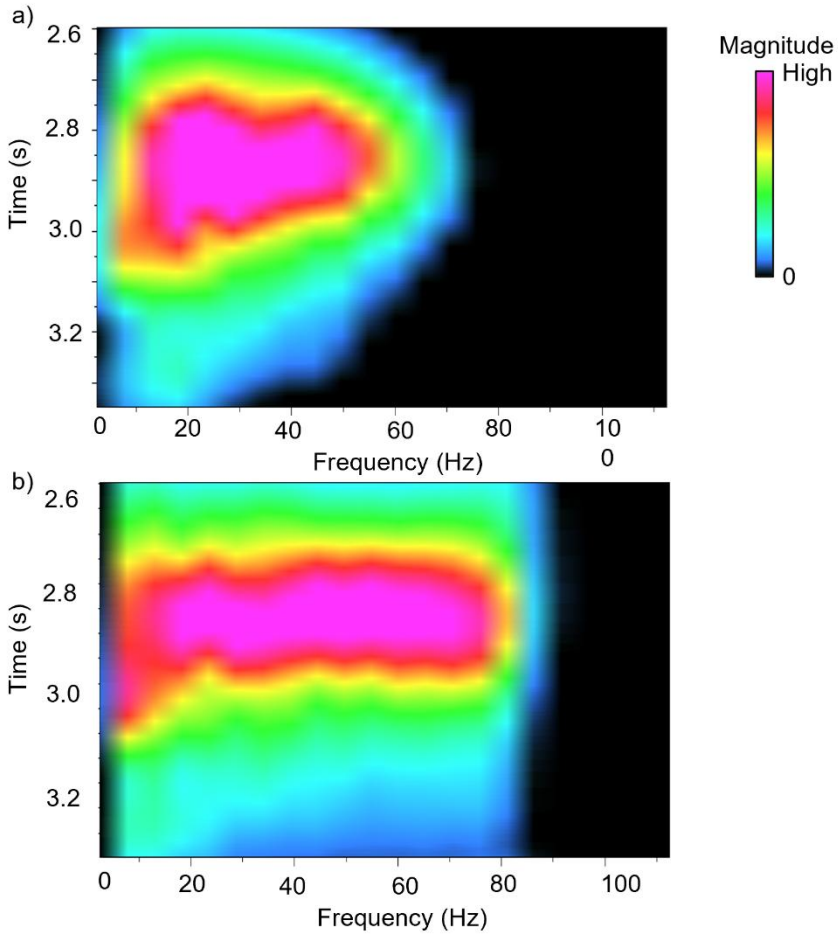


Figure 8. Time-variant spectra of the seismic amplitude averaged over the entire survey (a) before and (b) after spectral balancing. The spectrum in (b) was subjected to a 0-4-80-100 Hz Ormsby filter to suppress high and low frequency artifacts. Notice the spectra in (b) are nearly constant (appearing as the same color) for a given time level. A single time-variant operator is applied to the entire survey, thereby preserving relative amplitude.

RESULTS: WHAT THE OPTIMIZED DISPLAY HELPS REVEAL

Application of our five-step workflow to the Romney deepwater channel complex produced marked improvements in feature visibility and interpretability. The optimized corendering of cosφ with RAI, combined with tailored colormaps, 1 ms resampling, and spectral balancing, revealed previously obscured stratigraphic details critical for channel architecture analysis. How does this balancing affect the interpreter's ability to identify important geologic features? We start with a comparison of amplitude before and after spectral balancing to see if visual changes are observed.

Figure 9a shows line AA' through the original seismic amplitude volume, whereas Figure 9b shows the same line after spectral balancing. Arrows indicate areas within the channel of improved vertical resolution. Yellow, green, and white arrows indicate areas where the lateral continuity of the seismic reflector has been broken, while the pink and orange arrows show more lateral continuity across the reflector. This becomes important when determining not only boundary edges but also channel timing. For example, we can infer that the yellow and green arrows indicate older channels that have been cut by channel migration. The orange arrow, on the other hand, shows a bounding edge that preserves the shape of the channel, indicating that this channel is much younger, as it has not been crosscut by another channel iteration at this location. Figure 8c and 8d shows the same line, after spectral balancing, but this time with RAI alone (Figure 9c) and (Figure 9d), corendered with cosφ. Within the channel, areas with higher RAI values correspond to where we expect to have coarser-grained deposits, while areas with lower RAI where we expect to have finer-grained deposits. With improved feature delineation from spectral balancing done, do we see an improvement in interpreting those bounding edges? Interpreting channel boundaries from the spectral balanced amplitude from (Figure 9e), we can pick out the major features but lose some detail that prevents us from confidently continuing some of those interpreted reflectors, as indicated by the purple arrow. However, using the spectrally balanced data to recompute RAI corendered with cosφ we can fill in those missing areas (Figure 9f) and pick out even more of the fine details within the channel complex, leading to the identification of those fine deposits (blue arrows), coarse deposits (red arrows), and a levee (green arrow).

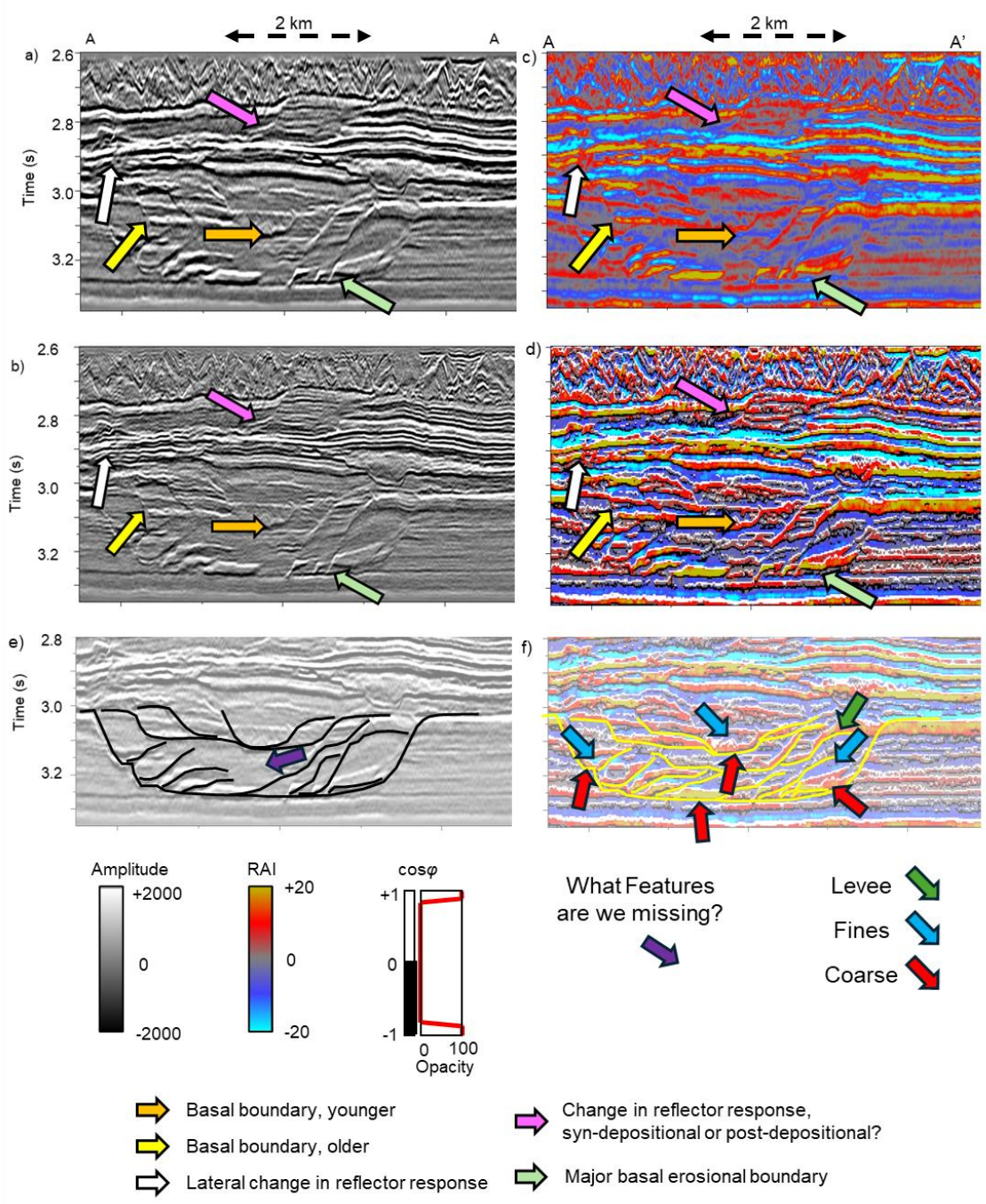


Figure 9. Line AA' through the amplitude volume (a) before and (b) after spectral balancing, where the colored arrows indicate improved vertical resolution within the turbidite system. The spectra of the two images are shown in the previous figure. Line AA' through volumes of (a) amplitude, (b) h cosφ, (c) relative acoustic impedance, and (d) relative acoustic impedance corendered with cosφ. Stratigraphic boundaries mapped guided by (e) cosφ alone, and (f) relative acoustic impedance corendered with cosφ. Using our understanding of geologic processes and the environment of deposition, we can push our interpretation further and hypothesize areas of finer- (blue arrows) and coarser-grained material (red arrows).

This improved vertical resolution in the seismic amplitude volume directly translates to better resolution in our derived attributes. Figure 10a shows line AA' through cosφ and Figure 10b displays the same line through our corendered RAI and cosφ volumes. Arrows indicate the same locations shown in the previous figure, demonstrating significant resolution improvement over results computed from the original, unbalanced data volumes, as seen in Figure 9a and 9c.

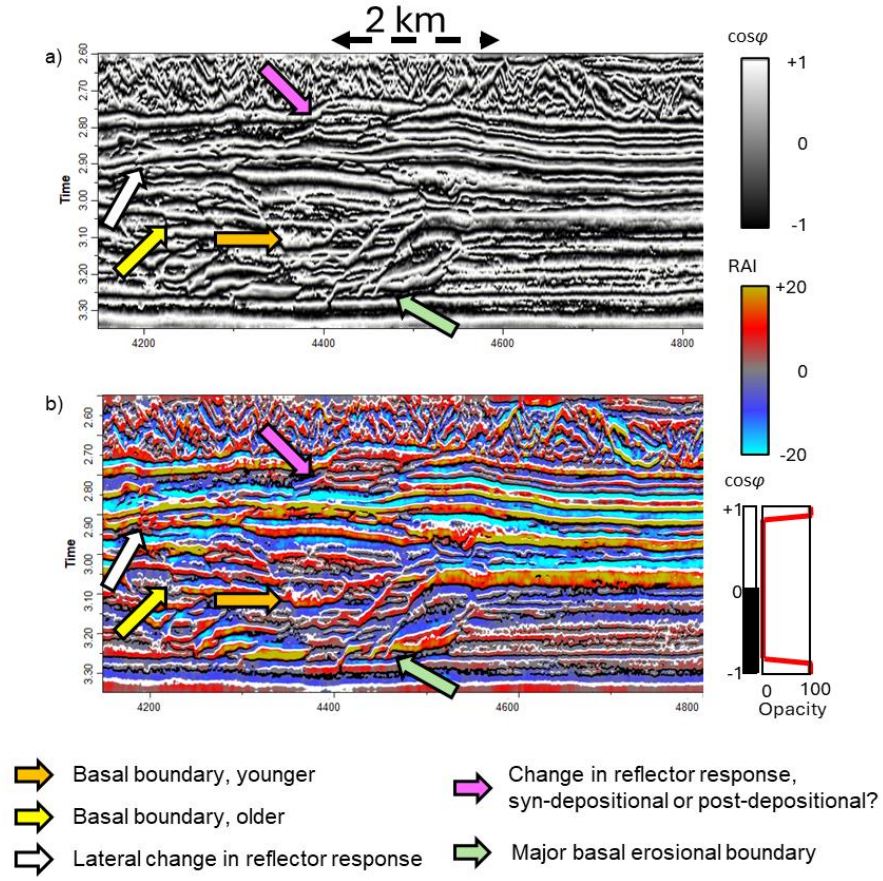


Figure 10. Line AA' through the (a) cosφ volume and the (b) corendered RAI and cosφ after spectral balancing of the 1 ms sample increment data. Arrows indicate areas with the channel that exhibit higher resolution than the comparable images in Figure 9a and 9c.

This enhanced vertical resolution from our workflow to optimize feature delineation provides improved attribute quality, as both cosφ and relative acoustic impedance calculations benefit from the increased frequency content and improved signal-to-noise characteristics. The processing proves particularly valuable for this particular deepwater channel interpretation, where subtle stratigraphic relationships require maximum resolution to distinguish between AE with similar acoustic properties. We have seen several examples of improved delineation in the cross-sectional view. Do these improvements extend into the time-slice perspective? Figure 11

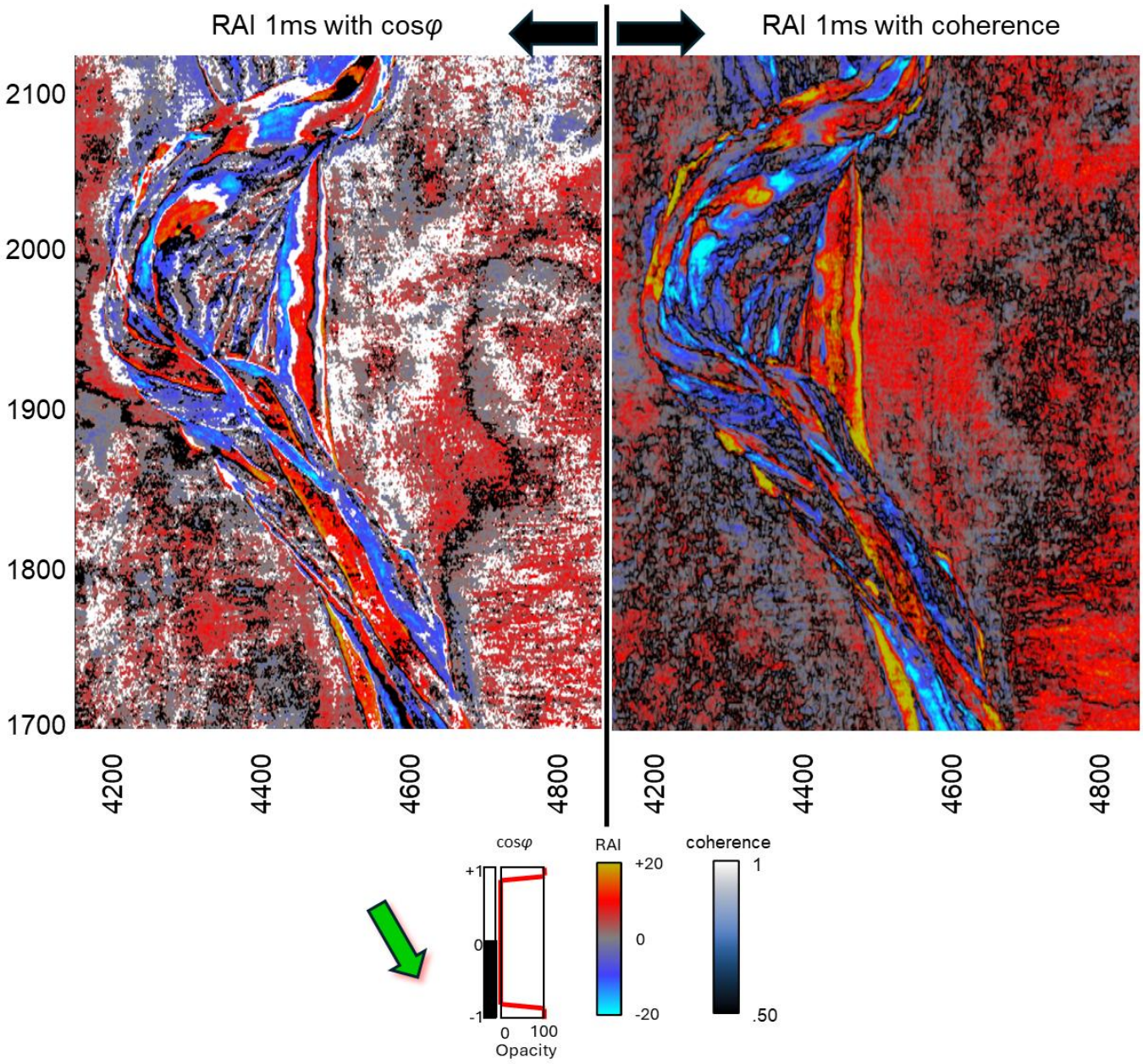


Figure 11: Time slice at t=3.21 s through the corendered RAI and cosφ volumes (left) sampled at a 1 ms increment after spectral balancing and through the corendered RAI and coherence volumes (right). Because the reflectors seen in Figure 9 are relatively flat, the time slice is parallel to the extrema of cosφ in many locations, resulting in the broad white and black smears between the positive and negative RAI anomalies compared to the sharp edges in coherence shown in Figure 4a.

presents a time-slice through the corendered volumes after spectral balancing. We find these horizontal plan-view images significantly less useful than the vertical sections. The white and black smears that are observed, rather than the expected sharp boundaries, are likely a result of stratigraphic layers that are nearly parallel to the horizontal time slices. We still need to keep in mind that interpretation should not only be created in two-dimensional space.

The next consideration should be how the interpreter can extend this interpretation from a cross-sectional perspective to include the three-dimensional reality. We will look at two examples next: one follows an area of high RAI, and the second with an area of low RAI. Figure 12a focuses on a clearly defined reflector, indicating lower RAI (pink arrow) in cross-sectional perspective. Moving to the plan view (Figure 12b), we observe indications of lateral smearing associated with cosφ; as previously mentioned, cosφ is very effective for noting lateral and vertical changes, however, it is not as effective when viewed in the plan view image. Keeping this limitation in mind, we examine Figure 12c for the combined inline/timeslice (chair) view. Despite the smearing, we are better able to get a sense of channel movement using the combined view and thinking about channel migration in the three-dimensional world.

The second example (Figures 12d-f) follows a similar progression but focuses on an area with higher RAI (yellow arrow) shown in cross-section (Figure 12d). The plan view (Figure 12e) again shows the lateral smearing limitations of cosφ, while the combined chair view (Figure 12f) provides the three-dimensional context that is key for understanding channel architecture and migration patterns. Now that we have shown how to optimize the display, what geologic features are we better able to delineate and incorporate into a defendable interpretation?

The optimized display significantly improves the visibility of key channel architectural features compared to conventional amplitude displays. Sharp amplitude-independent reflector discontinuities become traceable laterally regardless of impedance contrast variations, enabling clearer identification of potential erosional surfaces and allowing for chronological interpretation of channel movement. Internal patterns within channel fills become more apparent where subtle inclined geometries indicate possible lateral accretion consistent with lateral channel migration. Along channel margins, reflectors build upward at the channel edges, then thin systematically away from the channel, creating wedge-shaped packages with elevated crests. This geometric pattern is consistent with levee development through overspill deposition. Draping reflectors that may represent abandonment surfaces becomes easier to trace. This enhanced feature delineation establishes the foundation for a systematic interpretation workflow, with detailed geologic analysis of the Romney deepwater channel complex present in the companion paper (Moreno-Ward et al., in review).

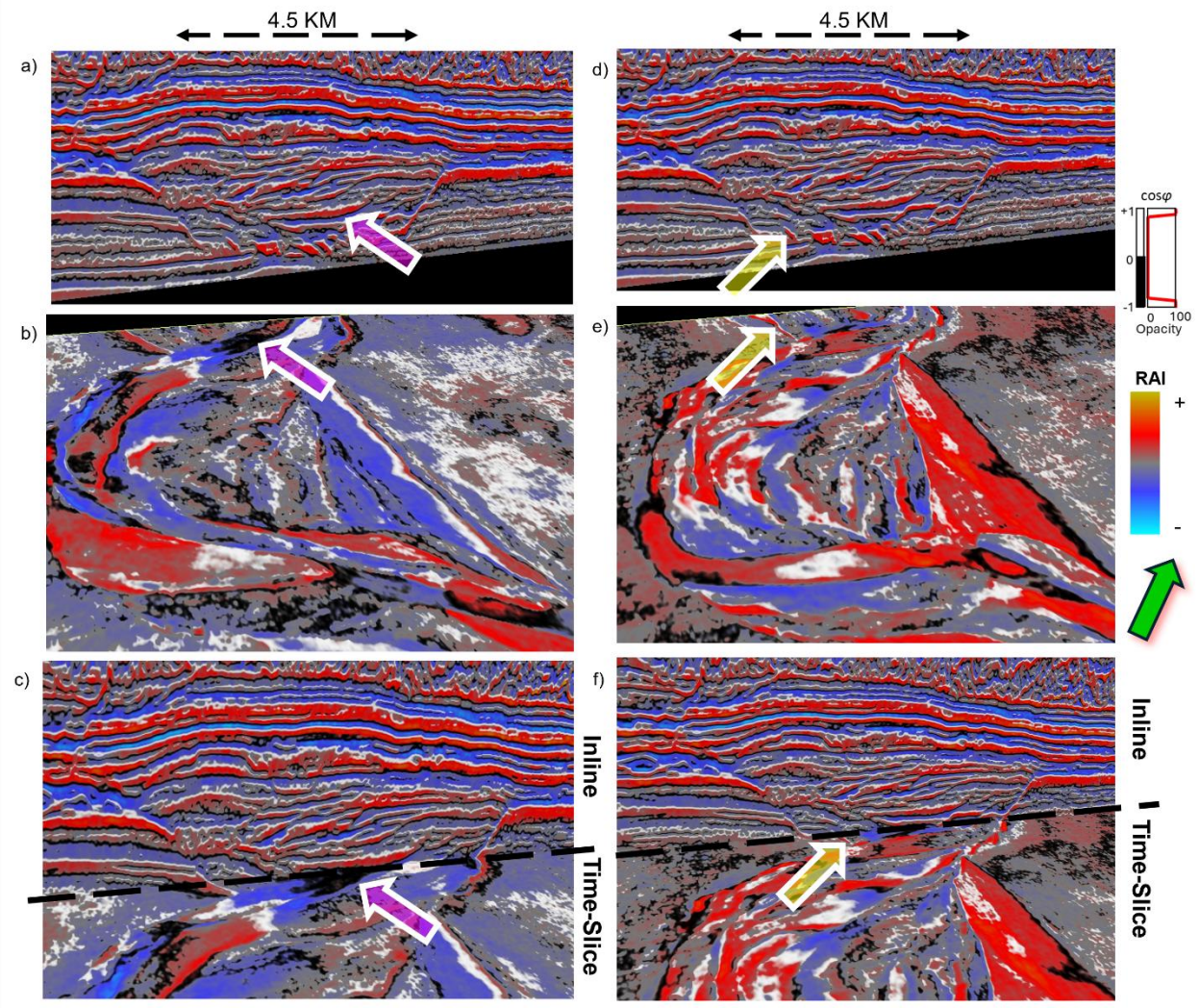


Figure 12: Vertical slice through Line AA'n (a) and (d) with areas of interest marked with arrows. The pink arrow indicates an area of lower RAI, whereas the yellow arrow indicates an area of higher RAI. Because the stratigraphy is relatively flat, timeslices (b) and (c) through these same features exhibit broad cosφ anomalies. Displaying these images a chair diagrams in (c) and (f) allow us to better link the stratigraphic edges seen in the vertical section to those seen on the time slices.

CONCLUSIONS

Accurate interpretation of depositional environments and architectural elements in deepwater channel systems relies heavily on seismic stratigraphic analysis, regardless of basin maturity or available well control. Our approach integrates two widely available seismic attributes, cosine instantaneous phase and relative acoustic impedance, through corendering techniques that significantly improve stratigraphic package identification and interpretation. The

systematic five-step workflow presented here addresses each component that influences display quality: attribute extraction with appropriate parameter selection, corendering methodology that preserves both geometric and lithologic information, perceptually uniform colormap selection that avoids visual artifacts and fatigue, interpolation to finer sample intervals that improves phase attribute fidelity by better capturing wavelet characteristics, and spectral balancing that enhances boundary resolution by compensating for frequency-dependent attenuation. Each optimization step contributes independently to improve feature visibility, though maximum benefit derives from their systematic integration.

Critically, these techniques reduce interpretation uncertainty by delineating subtle features and geometric patterns that may remain obscured in conventional amplitude displays. Amplitude-independent phase information enables confident mapping of erosional boundaries, reflector terminations, and lateral continuity patterns, which are the fundamental geometric relationships that geoscientists use to reconstruct depositional processes, establish stratigraphic architecture, and constrain channel evolution models. The addition of RAI provides impedance-based lithologic context independent of tuning effects, allowing interpreters to distinguish facies variations from geometric complexity. This separation of geometric information and property information addresses the conflation inherent in amplitude responses, enabling more defensible interpretations even where well control is limited or diagenetic alteration has modified acoustic signatures.

Implementation success depends on understanding how each optimization component influences the final display. Proper colormap selection prevents visual artifacts that can obscure information, create false patterns, or lead to eye fatigue. Strategic application of spectral balancing enhances boundary resolution that helps sharpen reflector terminations and improve delineation of subtle features. Although automated horizon picking offers efficiency advantages, our method demonstrates substantial improvements in stratigraphic interpretation on vertical sections through systematic optimizations of attribute corendering and display parameters.

We encourage readers to experiment with these optimization techniques on their own datasets, recognizing that specific sequences and parameters may require adaptation based on data quality, dominant frequency content, and interpretation objectives. This systematic approach is transferable across deepwater settings and provides particular value in exploration contexts

where limited well control requires maximum extraction of geometric and lithologic information from seismic data. By reducing interpretation ambiguity and revealing features critical to understanding depositional processes, these techniques enable more confident characterization of channel architecture, more realistic assessment of exploration risk, and more effective communication of interpretive confidence to decision-makers. The companion paper (Moreno-Ward et al., in review) demonstrates application of this enhanced visualization within a checkpoint-based interpretation framework, illustrating how optimized displays support systematic geologic reasoning for an improved interpretation in deepwater channel characterization.

ACKNOWLEDGEMENTS

Thanks to New Zealand Petroleum and Minerals for providing access to their Romney 3D data volume. All attributes were computed and displayed using software provided by the OU AASPI consortium.

REFERENCES

Brown, A. R., 2011, Interpretation of Three-Dimensional Seismic Data: Investigations in Geophysics AAPG Memoir **42** and SEG Investigations in Geophysics No. 9; 7th ed.

Bull, Suzanne AU - Browne, Greg AU - Arnot, Malcolm AU - Strachan, Lorna PY - 2020/01/17 SP - SP500 EP - 2019 T1 - Influence of Mass Transport Deposit (MTD) surface topography on deep-water deposition: an example from a predominantly fine-grained continental margin, New Zealand VL - 500 DO - 10.1144/SP500-2019-192 JO - Geological Society, London, Special Publications ER -

Chopra, S., R. K. Sharma, K. J. Marfurt, H. Bedle, H., and S. Verma, 2022, Attempts at seismic characterization of a deepwater turbidite channel in Taranaki Basin, New Zealand. First Break, **40** (10), 27-39.

Chopra, S., and K. J. Marfurt, .2019, Unsupervised machine learning applications for seismic facies classification Unconventional Resources Technology Conference, Paper no. URTEC-2019-557-MS, https://doi.org/10.15530/urtec-2019-557

Chopra, S., and K. J. Marfurt, 2015, Spectral decomposition and spectral balancing of seismic data: TLE, **35**, 176-179.

Di, H., L. Truelove, C. Li, and A. Abubakar, 2020, Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand: The Leading Edge, **39**, no. 10, 727-733.

Ebuna, D. R., Kluesner, J. W., Cunningham, K. J., & Edwards, J. H. (2018). Statistical approach to neural network imaging of karst systems in 3D seismic reflection data. *Interpretation*, *6*(3), B15-B35.

Grahame, J. L., 2015, Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity.

Gui, Z., Zhang, J., Zhang, Y., & Sun, C. (2025). Characterization of fault-karst reservoirs based on deep learning and attribute fusion. *Acta Geophysica*, *73*(2), 1335-1347.

Hart, B., 2008, Stratigraphically significant attributes: The Leading Edge, **27**, 320-324.

Hart, B. S. (2013). Whither seismic stratigraphy?. Interpretation, **1**, SA3-SA20.

Kamp, P. J., A. J. Vonk, K. J. Bland, R. J. Hansen, A. J. Hendy, A. P. McIntyre, ..., and C. S. Nelson, 2004, Neogene stratigraphic architecture and tectonic evolution of Wanganui, King Country, and eastern Taranaki Basins, New Zealand: New Zealand Journal of Geology and Geophysics, **47**, no. 4, 625-644.

King, P. R., and P. H. Robinson, 1988, An Overview of Taranaki Region Geology, New Zealand: Energy Exploration & Exploitation, **6**, no. 3, 213-232.

Kroeger, K. F., Thrasher, G. P., & Sarma, M. (2019). The evolution of a Middle Miocene deep-water sedimentary system in northwestern New Zealand (Taranaki basin): Depositional controls and mechanisms. *Marine and Petroleum Geology*, *101*, 355-372.

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**, 105427.

La Marca, K. (2020). Seismic attribute optimization with unsupervised machine learning techniques for deepwater seismic facies interpretation: Users vs machines.

Li, J., Wu, X., Ye, Y., Yang, C., Hu, Z., Sun, X., & Zhao, T. (2023). Unsupervised contrastive learning for seismic facies characterization. Geophysics, 88(1), WA81-WA89.

Li, Q., Yu, S., Wu, W., Tong, L., & Kang, H. (2017). Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: a case study from the Taranaki Basin, New Zealand. *New Zealand Journal of Geology and Geophysics*, *60*(3), 199-208.

Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K. J., & Pranter, M. J. (2022). Quantifying the sensitivity of seismic facies classification to seismic attribute selection: An explainable machine-learning study. *Interpretation*, *10*(3), SE41-SE69.

Lubo-Robles, D., and Marfurt, K. J., 2019, Independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: Interpretation, **7**, no. 3, SE19-SE42.

Maas, M. V. R., Bedle, H., & de Matos, M. C. (2023). Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data. *Marine and Petroleum Geology*, *151*, 106199.

Moreno-Ward, A., Bedle, H., Lubo-Robles, D., and Marfurt, K.J., Beyond visualization alone: a systematic interpretation tutorial for deepwater channel stratigraphy identification. Interpretation. In Review

Naewboonnien, K., 2021, The Use of Seismic Attributes for Submarine Channel Interpretation in Deepwater Taranaki Basin, New Zealand: Bulletin of Earth Sciences of Thailand, **11**, no. 2, 134-143.

Posamentier, H.W., ed., 1993, Sequence stratigraphy and facies associations: International Association of Sedimentologists Special Publication 18, Wiley, 664 p.

Qayyum, F., O. Catuneanu, and C. E. Bouanga, 2015, Sequence stratigraphy of a mixed siliciclastic-carbonate setting, Scotian Shelf, Canada: Interpretation, **3**, SN21–SN37.

Reilly, C., A. Nicol, J. J. Walsh, and H. Seebeck, 2015, Evolution of faulting and plate boundary deformation in the Southern Taranaki Basin, New Zealand: Tectonophysics, **651**, 1-18.

Slatt, R. M. (2013). *Stratigraphic reservoir characterization for petroleum geologists, geophysicists, and engineers* (Vol. 61). Newnes.

Strogen, D. P., H. Seebeck, A. Nicol, and P. R. King, 2017, Two-phase Cretaceous–Paleocene rifting in the Taranaki Basin region, New Zealand; implications for Gondwana break-up: Journal of the Geological Society, **174**, no. 5, 929-946.

Strogen, D. P., K. E. Higgs, A. G. Griffin, and H. E. Morgans, 2019, Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression: Geological Magazine, **156**, no. 10, 1751-1770.

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

Thrasher, Glenn P. (1992). Late Cretaceous Geology of Taranaki Basin, New Zealand. Thesis, Victoria University of Wellington. http://hdl.handle.net/10063/525

Vail, P. R., J. Hardenbol, and R. G. Todd, 1984, Jurassic unconformities, chronostratigraphy and sea-level changes from seismic stratigraphy and biostratigraphy, *in* J. S. Schlee, ed., Interregional unconformities and hydrocarbon accumulation, AAPG Memoir **36**, 129–144.

Verma, S., and S. Bhattacharya, 2021, Unsupervised Machine Learning for Interpreting Shelf-to-Basin Seismic Geomorphology and Paleoclimate: 82nd EAGE Annual Conference & Exhibition, EAGE, Extended Abstracts, 1-5.

Wang, Z., Di, H., Shafiq, M. A., Alaudah, Y., & AlRegib, G. (2018). Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. *The Leading Edge*, *37*(6), 451-461.

Weimer, P., Slatt, R. M., Bouroullec, R., Fillon, R., Pettingill, H., Pranter, M., & Tari, G. (2006). Introduction to the Petroleum Geology of Deepwater Setting.

Wharton, S. R., 2015, Seismic chronostratigraphy and basin development at a Mid-Cretaceous intrashelf basin margin: Interpretation, 3, SN1-SN20.

Zhang, W. B., T. Z. Duan, Z, Q, Liu, Y. F. Liu, L. Zhao, and R. Xu, 2017, Architecture mode, sedimentary evolution and controlling factors of deepwater turbidity channels: A case study of the M Oilfield in West Africa: Petroleum Science, **14, No. 3**, 493-506.

Chapter 5: Beyond Visualization Alone: A Systematic Interpretation Tutorial for Deepwater Channel Stratigraphy Identification Part 2: Framework application and Romney 3D Case Study*

*This study is in review in SEG Interpretation

ABSTRACT

Effective deepwater channel interpretation and identification of stratigraphic features require more than enhanced seismic visualization. The interpreter must create and follow a systematic approach to translate observations into a defensible depositional model utilizing process-based geologic thinking. It requires the identification of key stratigraphic channel patterns (erosional boundaries and reflector termination characteristics such as onlap and downlap) that reveal depositional setting complexities while assessing confidence levels for each observation. Validation of interpretations across multiple display perspectives is a must, and critical evaluation of alternative scenarios should be considered to avoid blindly following meaningless patterns. This tutorial presents an interpretation framework incorporating critical thinking checkpoint questions at each workflow stage, applied to a deepwater channel complex in New Zealand's Taranaki Basin. Building upon the optimization of corendered displays of cosine instantaneous phase ($\cos\varphi$) and relative acoustic impedance (RAI) shown in the companion paper, part 1, we systematically delineate architectural elements (AEs) and reconstruct channel system evolution through deliberate observation protocols rather than just pattern matching. The Romney-1 well control provides critical calibration, enabling seismic facies characterization and interpolation from the well-bore into the channel fairway. These checkpoints and systematic framework produce more confidence in the interpretation of channel architecture in exploration settings where well control is limited. This process-based interpretation approach, which integrates depositional mechanics with pattern recognition, captures the full complexity of multi-phase channel evolution through critical thinking. This method requires interpreters to be transparent about what can be confidently interpreted from seismic data (geometry and architectural relationships), what can be inferred with lower certainty (detailed facies predictions), and what remains beyond the resolution of available data.

INTRODUCTION

Deepwater channel interpretation requires visualization approaches that delineate subtle architectural features and boundaries, particularly reflector termination and lateral discontinuities. Part 1 (Moreno-Ward et al., in review) addresses this challenge by demonstrating that corendering of relative acoustic impedance (RAI) with cosine instantaneous phase (cos*φ)* addresses amplitude interpretation limitations by resolving geometric boundaries and lithologic variations that single attribute displays obscure. The approach utilized an accessible five-step workflow to improve overall interpretability, including:

- Proper attribute selection: we combine cosφ for geometric feature identification and relative acoustic impedance for layer characterization.
- Corendering: Integration of both attributes into a single display helps us to simultaneously visualize reflector terminations and acoustic property variations.
- Colormap optimization: We used a black/white color scale for cosφ to maximize discontinuity detection and a modified relative acoustic impedance colormap to reduce visual fatigue.
- Resampling: Interpolation from a 4ms to a 1ms sample interval helped to eliminate graphical aliasing artifacts and better preserve phase variations.
- Spectral balancing: The final step enhanced vertical resolution by applying spectral balancing to resolve thin beds and stratigraphic boundaries.

Building on this foundation, part 2 applies this systematic data preparation framework to help characterize features within the Romney channel complex, located offshore northwest of New Zealand's North Island, demonstrating how process-based geologic understanding combined with improved seismic visualization can improve confidence in the interpretation model. This tutorial is designed for geoscientists with basic seismic interpretation experience who want a more systematic approach to deepwater channel analysis. Readers should be familiar with fundamental seismic attributes and sequence stratigraphic concepts.

Before moving into the interpretation process, let's discuss a modern analogue. Modern systems provide a direct observable and mappable system with channel dimensions, migration patterns, facies distributions, channel cutoff, and scroll bars that can inform the seismic-scale

interpretation and help build confidence when observing ambiguous features in the deepwater setting.

To help guide seismic interpretation of our deepwater channel complex, we first established a conceptual framework using a modern analogue. We have picked a portion of the Mississippi River near the Ten Mile Pond Conservation area (Figure 1a) and zoomed into a smaller portion (Figure 1b) to begin to sketch out abandoned surficial features. The open blue arrow indicates the easternmost extent of the current floodplain in this section, while the solid blue arrow indicates an abandoned channel loop. Prograding scroll bars are marked with yellow lines and arrows, while the direction of channel migration is noted with a green arrow. Figure 1 c) and d) show pictures of channel deposits collected with hand augers from a portion of the Missouri River near Yankton, SD. Understanding how channel morphology impacts deposition of sediments allows the interpreter to consider where, within various channels, we could expect to have finer-grained muds vs. silts and sands. We have marked with dashed black arrows the expected location within the Mississippi River to find similar sediments. Figure 1e) Illustrates a portion of the 3D seismic data, amplitude, across our channel of interest, with Figure 1f) showing a rough sketch of where we may expect to find various grain sizes of deposition. While these changes in grain size will affect compaction, we have additional considerations regarding diagenesis and potential fluids within any available pore space to consider when thinking about seismic responses. With this in mind, we proceed through our workflow to determine how we can use seismic attributes to help us interpret channel stratigraphic indicators and improve confidence in our models.

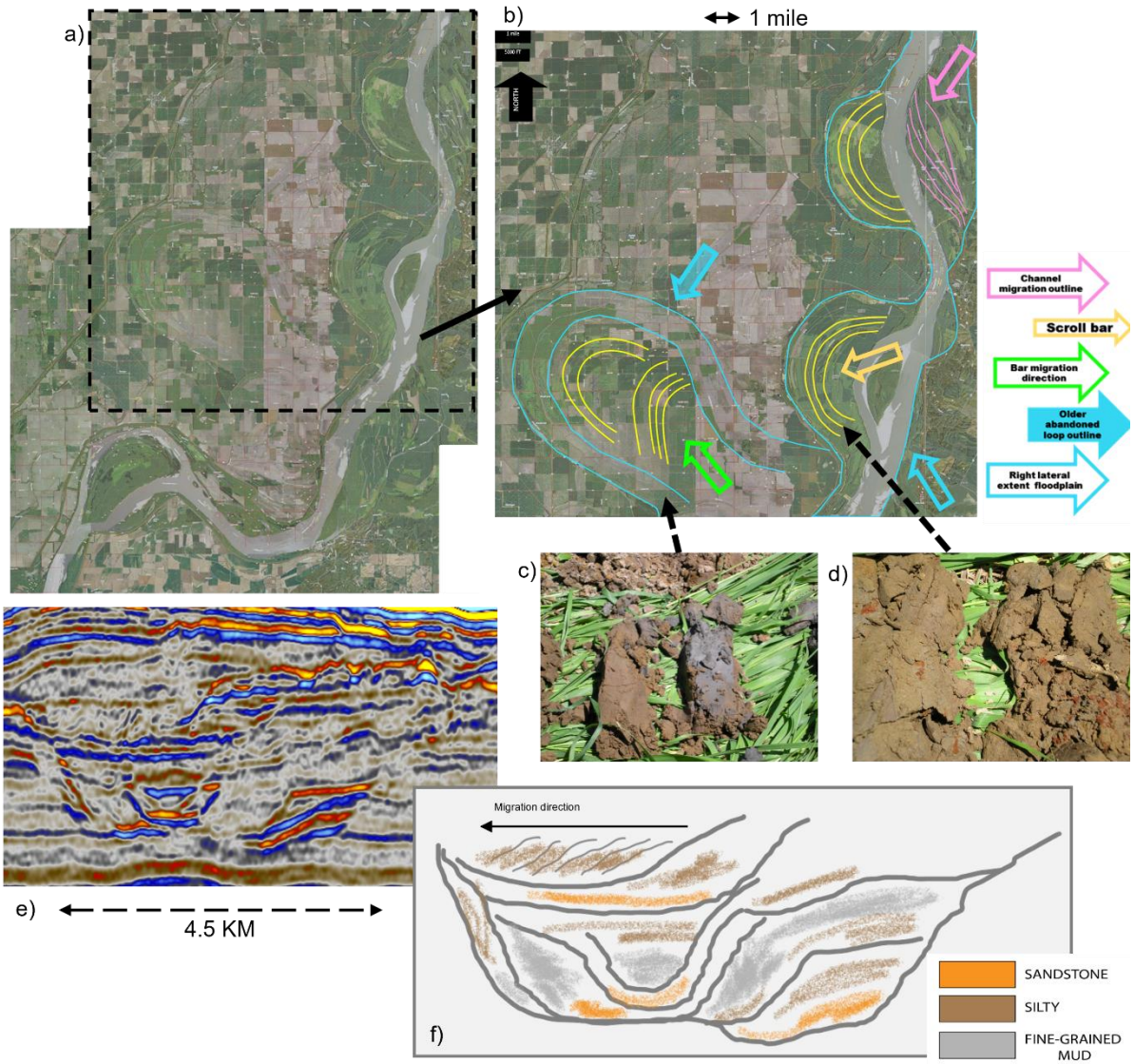


Figure 1: Considerations of expected morphology based on more modern analogues. a) is a portion of the Mississippi River (near Ten Mile Pond Conservation Area), zoomed in to focus (b) on abandoned meander loops and subsequent architectural elements. c) and d) are samples taken from the Missouri River near Yankton, SD. These samples were taken using hand augers across the river valley to note changes in sediment deposition both vertically and laterally in response to river migration. The dashed black arrows indicate the equivalent location these sediment samples would have been found in the abandoned channel features. e) Shows a cross-section of the Romney channel, with f) showing a quick sketch of the expected grain size.

METHODS: HOW TO APPROACH INTERPRETATION OF CHANNEL STRATIGRAPHY

We present our critical thinking framework for interpreting deepwater channels to demonstrate the value of establishing deliberate checkpoints throughout the interpretation process. While other depositional environments may require different questions and emphasis, the key is to develop a systematic approach to ensure geologic reasoning is guiding the interpretive decisions. Figure 2 shows our 14-step process for interpretation. The following sections expand this framework, explaining both the technical foundation and the critical thinking required at each step.

Foundation: Understanding Key Seismic Attributes

Before diving into the interpretation workflow, let me explain why we selected these specific seismic attributes for this analysis. Two primary attributes form the foundation of our approach: RAI and cosφ. We chose RAI over amplitude as the property-based attribute for several key reasons. RAI is less sensitive to tuning effects, provides impedance information that directly relates to lithology when calibrated to well control, and maintains interpretability where diagenesis may reduce amplitude contrasts. As an integrated attribute representing cumulative impedance trends rather than individual reflections, RAI provides greater stability for facies interpretation while avoiding the polarity ambiguities and thickness-dependent variations that complicate amplitude analysis.

The selection of cosφ over the amplitude or raw phase addresses fundamental limitations in boundary detection. Phase-based attributes track wavelet position within the seismic cycle independent of reflection strength, enabling identification of stratigraphic boundaries even where impedance contrasts are subtle. These attributes reveal reflector (dis)continuity patterns, for example, a continuous phase indicates coherent stratigraphy, while phase breaks indicate truncations, erosional surfaces, or facies transitions, information that remains interpretable regardless of amplitude variability (Figure 3).

Cosine of phase is employed rather than the raw instantaneous phase because the cosine transformation eliminates wraparound artifacts in the -180° to +180° range and provides a bounded, continuous scale (-1 to +1), while preserving amplitude-independent geometric information. Furthermore, cosφ values carry direct physical meaning: values near zero correspond to wavelet peaks and troughs, while values of ±1 indicate zero-crossings where actual impedance contrasts occur. These zero-crossing positions represent the most stable portion of the seismic wavelet for stratigraphic interpretation, enabling systematic, repeatable boundary delineation independent of amplitude behavior.

The Interpretation Workflow: Integrating Geological and Geophysical Perspectives

As a geoscientist integrating geological and geophysical perspectives, systematic seismic interpretation ensures pattern recognition is grounded in geologic understanding rather than arbitrary mapping. The following 14-step workflow demonstrates how to integrate geologic reasoning with geophysical analysis, ensuring the interpreter maintains geologic feasibility throughout the interpretation process. The Romney dataset exemplifies a common exploration challenge: interpreting reservoir architecture with distant well control. While Romney-1 provides lithologic constraints, its location 8 km from the channel forces reliance on seismic character and geologic reasoning, precisely the situation where systematic workflows become essential. Figure 4 shows the key questions that should be considered at each step.

Initial Target Definition and Attribute Application

Step 1: Defining the Geologic Target: The first question to consider is: "Why this particular target?" For the Romney deepwater channel complex, we need to consider: What type of reservoir system does this represent? Could this be a viable hydrocarbon reservoir, or might it serve better for carbon sequestration? Without direct penetration of the channel system (remember, the Romney-1 well missed the channel entirely), the interpreter has to proceed using an exploration mindset, acknowledging that there is no direct calibration within the target.

Steps 2-3: Strategic Attribute Selection and Boundary Mapping. Here's where we leverage the complementary nature of my selected attributes (**step** 2). Cosφ is used as the primary tool for geometric interpretation because its amplitude-independent nature prevents the misidentification of boundaries due to fluid effects or lithology changes that don't represent actual stratigraphic surfaces. When mapping boundaries (**step** 3), it's not just drawing lines; we should be thinking about hierarchical channel architecture. What are the major bounding surfaces? What internal subdivisions can I identify? This is the larger-scale geologic system thinking: recognizing that deepwater channel systems are complex and dynamic. The interpreter needs to understand the chronological evolution of the channel, and these boundaries are my starting point for piecing that story together.

For facies work, RAI becomes the tool of choice because it reveals those internal variations that tell the depositional story. There should be a constant reminder that RAI values are relative; they won't definitively tell us what lithology is being looked at, but they effectively track changes. Despite the well's position outside the channel, it still provides crucial baseline information that can be carefully extrapolated into the channel system.

Well Calibration and Data Integration

Steps 4-6: Critical Well Calibration. Now ask yourself: "What do we actually know from the well?" Then tie the well (**step** 4) and review the log data carefully, examining how log curve characteristics correlate with actual lithology (**step** 5). During the well tie, pay attention to any formations that prove difficult to correlate; these zones may signal lateral facies changes or structural complexities that will affect your interpretation away from well control.

Look closely at the mudlogs to understand what rock types actually exist in the wellbore. In the Romney-1, we are approximately 8 km outside the channel system, encountering primarily calcareous claystone at the same depth as the channel system. Interbedded siltstone and limestone stringers are also observed. These facies associations tell us something important: we are in a pelagic/hemipelagic setting with periodic higher-energy events (these silt stringers) and carbonate deposits. While the coarsening-upward sequence we observe in well facies B cannot

directly calibrate channel facies given the distance, it gives us regional stratigraphic context and background sedimentation rates. The 8 km separation means we must be cautious about extrapolating well observations into the channel; however, we can use the well to understand the regional backdrop and help us understand seismic response to lithology.

This brings us to **step** 6: systematically evaluating the seismic response at the well location to understand how the seismic attributes behave in these known facies. Notice the relationships: well facies A limestone produces a strong positive amplitude that is easy to follow across the seismic. While facies B presents a more complex signature: nice parallel to sub-parallel reflectors with slightly chaotic sections between them. Specifically, we see a strong positive parallel reflector at the top, then mid-range amplitudes with somewhat chaotic response in the middle, followed by a low-amplitude parallel reflector at the base. What does this tell us? The chaotic middle section likely represents the heterogeneous mixture of clay and silt, while bounding parallel reflectors mar the more continuous deposits. This seismic character becomes our template for recognizing similar facies/seismic responses away from the well.

This isn't just about depth-matching; it's about understanding what validated lithology looks like in the attribute space. Can you correlate specific RAI behaviors to known lithologies? At this point, you should be building a bridge between hard data and seismic response. Keep in mind, however, that this bridge will need modification as you move into different parts of the depositional system, particularly into the channel itself where energy levels and depositional processes differ dramatically.

Step 7: Intelligent Interpolation. When interpolating facies information into the channel, don’t think of it as a simple spreading of absolute values. Instead, we should be interpolating behaviors and patterns, thinking about what geologic characteristics make sense within the depositional environment. This approach respects geological principles while acknowledging the limitations of seismic resolution.

Integration Phase: Where Geological Thinking Takes Center Stage

Step 8: Process-Based Reasoning. This is where geologic background knowledge really drives the interpretation. Before interpreting another line of seismic, ask: "What processes controlled this channel's evolution?" This isn't just academic curiosity, but rather, it's about building a process-based prediction framework for facies distribution. Do we understand how a deepwater channel works? Are we thinking about bottom currents, about how finer-grained material gets winnowed early while coarser sediment continues downstream? Do we know what a change in flow velocity might do to channel morphology? Is the channel confined and unable to migrate laterally? If we understand these processes, then we can infer where different facies should occur, even where seismic resolution is limited.

Steps 9-10: Building Three-Dimensional Understanding Working through cross-sections (**step** 9), we examine channel boundaries slice by slice, but we should always be thinking about the planform geometry. What would this system look like from above? How does cross-sectional architecture relate to map-view patterns (**step** 10)? A channel that shows high lateral mobility should show meander loops that leave a scroll bar geometry in map view, not a single straight channel. In cross-section we should see more amalgamated deposits. If you are, however, seeing less lateral migration in map view and an overall straighter channel system, we should see aggregation stacking patterns in cross-section. In our Romney channel system, we see areas with higher sinuosity and associated scroll bar geometries. In cross-sectional view, we see both lateral migration and vertical aggradation with amalgamation, indicating a channel system that has been highly mobile through time.

This constant mental switch between 2D and 3D, between section and map, is crucial for building a coherent interpretation. Each view provides different perspectives: cross-sections help to identify stacking patterns and internal architecture, while map view shows spatial relationships, migration directions that control sand body connection, and chronological migration (cross-cutting relationships). It's not enough to have a beautiful cross-section if it creates an impossible map-view geometry.

Step 11: Temporal Evolution Analysis Here's a question that separates geologic thinking from pure pattern recognition: "How did this system evolve through time?" We are not looking at a static image but unraveling a story. Which channels cut which? Where are the abandonment surfaces? What do the vertical stacking patterns tell me about the balance between

aggradation and incision? This temporal framework transforms my static seismic data into a dynamic depositional history.

Step 12: Coherence Verification. At this stage, the interpreter should take a step back and ask: "Does my interpretation make geologic sense across all views?" This is where many interpretations fall apart; they might look good on one section but create impossible geometries in 3D, or they might violate basic stratigraphic principles. This comprehensive review ensures we haven't gotten lost in the details and missed the bigger picture.

Step 13: Consider Alternative Interpretations. What could we be wrong about? The interpreter should develop alternative interpretations not from a lack of confidence, but because good science demands it. If an alternative seems plausible, then loop back to **step** 9 and work back through the steps, ensuring that during our validation process, we are confident that the model we are presenting has geologic meaning. Figure 5 illustrates a good example of looping back. (Figure 5a) Observing reflector response using cosine instantaneous phase cosφ, we initially focus on what appears to be inclined strata. Following this laterally, we may be tempted to interpret this as a shallow channel. (Figure 5b) Expanding the view helps to put this into geologic context. As there is clear channel geometry present, we change the visual perspective to planform at 2.83 s (Figure 5c) and 2.86 s (Figure 5d) and begin looking for geometries representative of channel morphology. Finding no clear channel morphologies at these depths, we change the visual perspective and continue moving deeper within the data set. Figure 5e and 5f illustrate a chair view at inline 1980 and time 2.91 s, and then just the planform. At this depth, we do begin to see evidence of a channel; however, this channel lies just to the NNE of the reflector response of focus. Continuing to follow this a bit deeper, we again view in both chairview and planform (Figure 5g and 5h) at time 2.93 s. At this depth, the channel has become more evident. The inclined reflectors result from differential compaction, where coarser channel fill compacts less than surrounding fine-grained sediments, creating the reflector dip that initially drew attention.

It is important to verify this interpretation because differential compaction-induced dips can influence subsequent deposition: younger channels or gravity flows may preferentially follow these structurally controlled lows, creating secondary channel systems. While these later

features may appear related in seismic data, they are chronologically distinct from the original channel that led to the observed compaction.

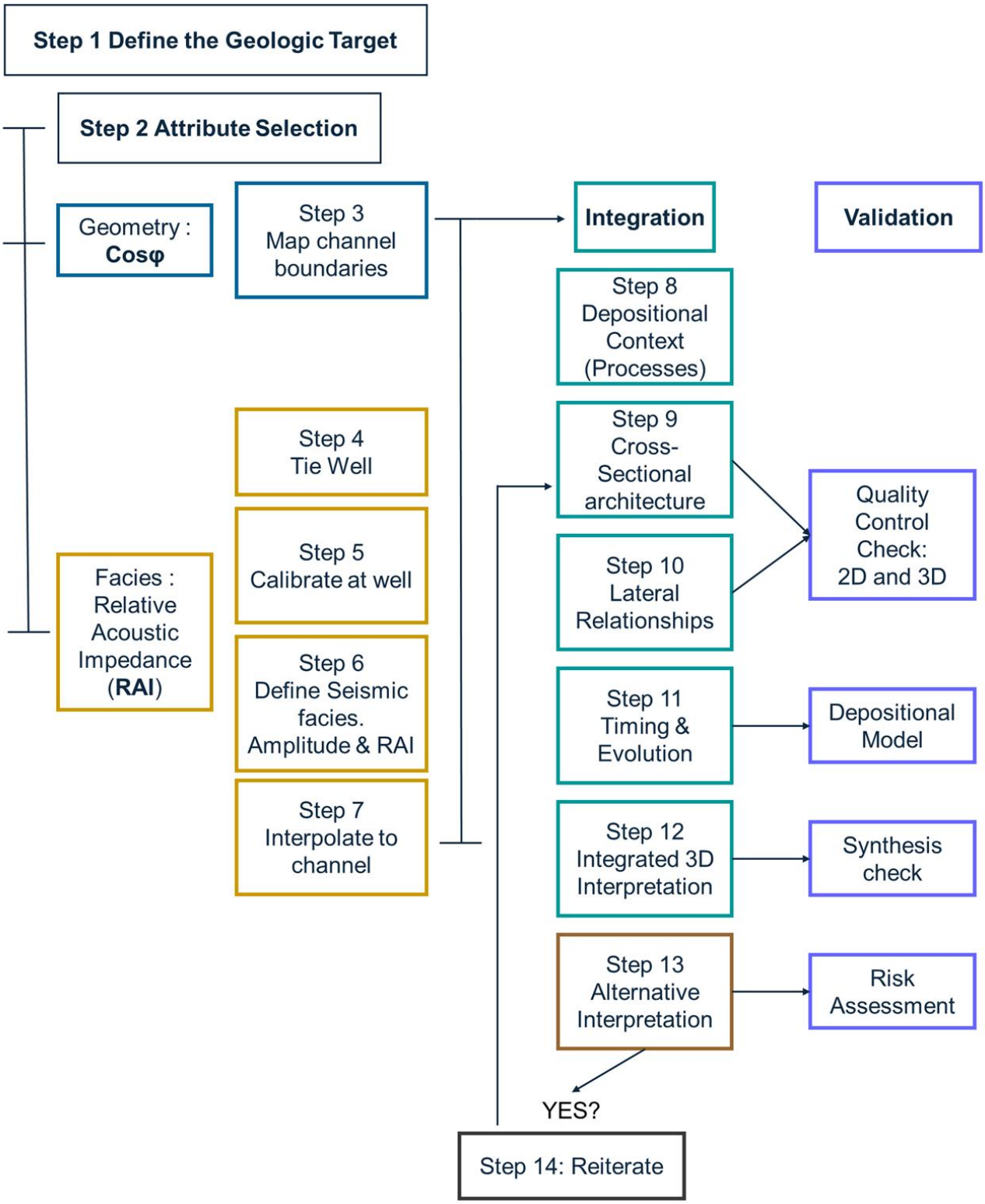


Figure 2: Interpretation workflow for identifying stratigraphic responses using both cosφ and relative acoustic impedance. The 14-step framework separates geometry and facies identification and then helps us combine these as we integrate through steps 8-12. If the overall interpretation is inconsistent or does not represent geologic meaning, reiteration to step 9 should be considered until the interpretation is consistent, considering geometric, lithologic, and depositional constraints.

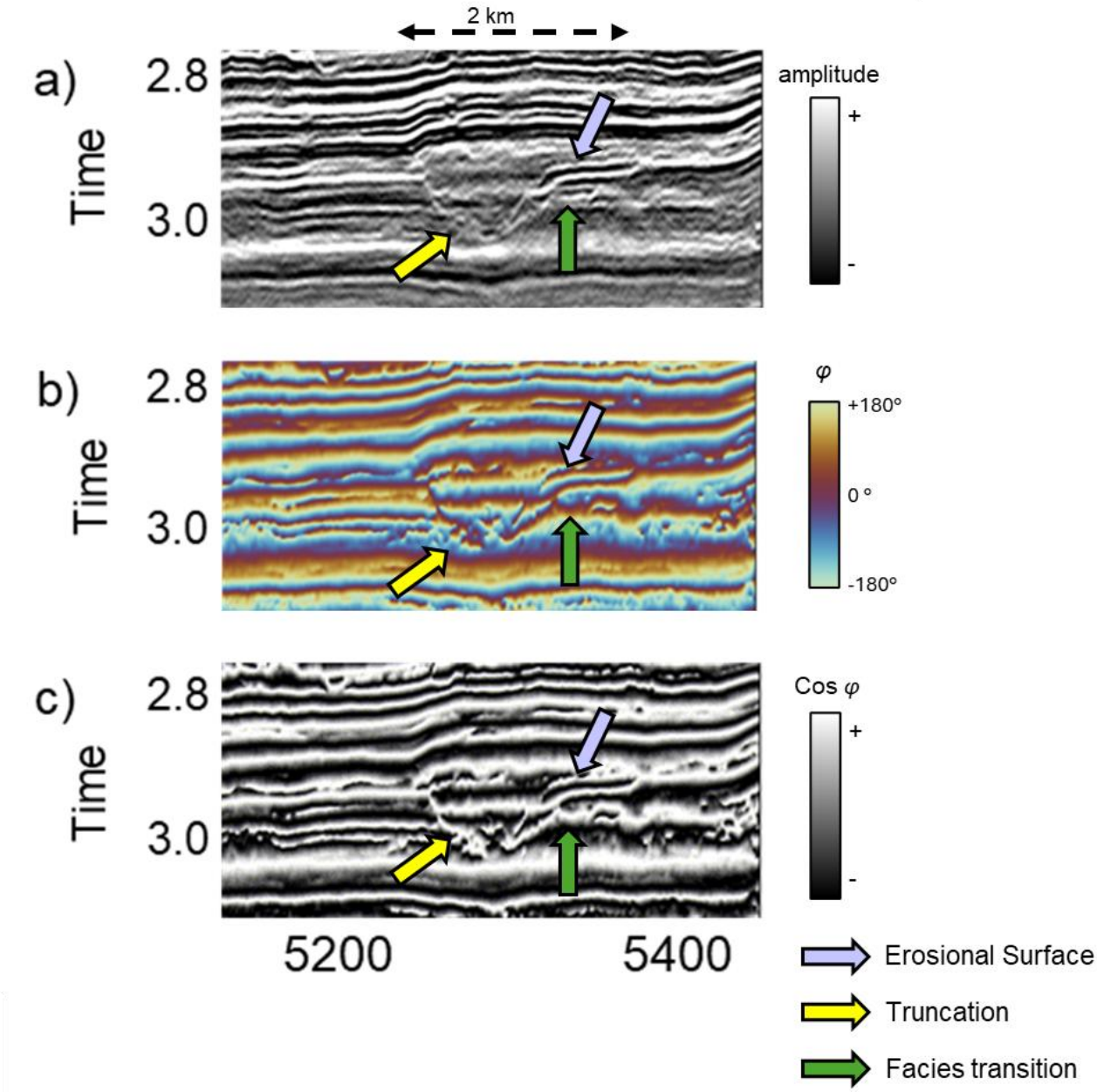


Figure 3: Inline 1980 (a) amplitude, (b) instantaneous phase, and (c) cosine of instantaneous phase. Comparison of how each attribute helps delineate stratigraphic features.

Critical Iteration and Validation

Step 14: The Final Validation Loop. This phase is perhaps the most important part of the workflow, and where geological reasoning really shines. By forcing ourselves through this

critical thinking process, we avoid getting caught in the trap of our own biases and see only what we expect or want to see. Working through iterations and revamping the interpretation models isn't failure; it's refinement and good scientific practice.

The validation phase isn't just checking boxes. It's the systematic verification that the interpretation honors both the data and geological principles:

- **Two- and Three-Dimensional Quality Control**: Do the cross-sections and maps tell the same story?
- **Depositional Model Validation**: Is the temporal evolution model geologically reasonable?
- **Synthesis Verification**: Does everything hold together when viewed from various perspectives?
- **Risk Assessment**: Have we properly weighed the uncertainties?

Throughout validation, we should be constantly checking against geologic principles: cross-cutting relationships tell us relative timing; lateral migration and vertical stacking patterns add geometric complexity; channels rework older deposits in predictable ways. These aren't just observations; instead, they're constraints that guide us toward geologically reasonable interpretations.

The Geological Mindset: Beyond Pattern Recognition

The strength of this workflow lies in combining geological and geophysical perspectives through systematic critical thinking. It’s not just "What does the seismic show?" but "What geologic process created this pattern?" and "Does this make sense given what we know about how these systems form?" It's the difference between pattern recognition and process understanding, between seeing channels and understanding channel systems.

Consider how these two approaches differ when interpreting the Romney channel complex. A pattern recognition approach might observe the sinuous high-amplitude features and

immediately label them as “meandering channel deposits”. However, a process-based approach asks: What flow conditions would create the sinuosity observed in this deepwater channel? How would lateral and down-flow migration, in this setting, sort and deposit sediment? Where would abandonment surfaces form given the observed stacking patterns? Should we see a scroll bar like geometry associated with this resulting channel migration? By working through these process questions, we understand that the sinuosity may actually represent multiple phases of channel migration and avulsion, and not necessarily a single meandering system. This is where thinking through the system in 2-dimensional and 3-dimensional space plays a critical role. Next, we need to think about cross-cutting relationships and how the channel movement can rework material and make it more difficult to put the chronology of the channel migration back into order. What we may end up with is a multi-step process where initial flow deposits coarser-grained material due to higher velocity, followed by deposition of some finer-grained material as the velocity decreases. Thinking through the geologic process in conjunction with pattern recognition is an important critical thinking workflow that can help capture reservoir complexity. Relying on pattern recognition alone means missing the depositional nuances that control facies distribution.

What it really comes down to is thinking about those cross-cutting relationships within channel complex environments to help us get a feel for timing and migration. This iterative, process-driven approach might seem more complex and time-consuming than straightforward seismic mapping, but it leads to more robust, defensible interpretations that honor both the physics of our seismic data and the geology of our depositional systems. In exploration settings like the Romney example, where direct calibration within the target is unavailable, this systematic approach becomes even more critical. We're not just making pretty pictures; we're building defensible geologic models despite inherent uncertainties.

The value of this workflow is that it forces the interpreter to be honest about what they know, what they think they know, and what they are guessing at. And in exploration geology, that honesty is often the difference between success and costly failure.

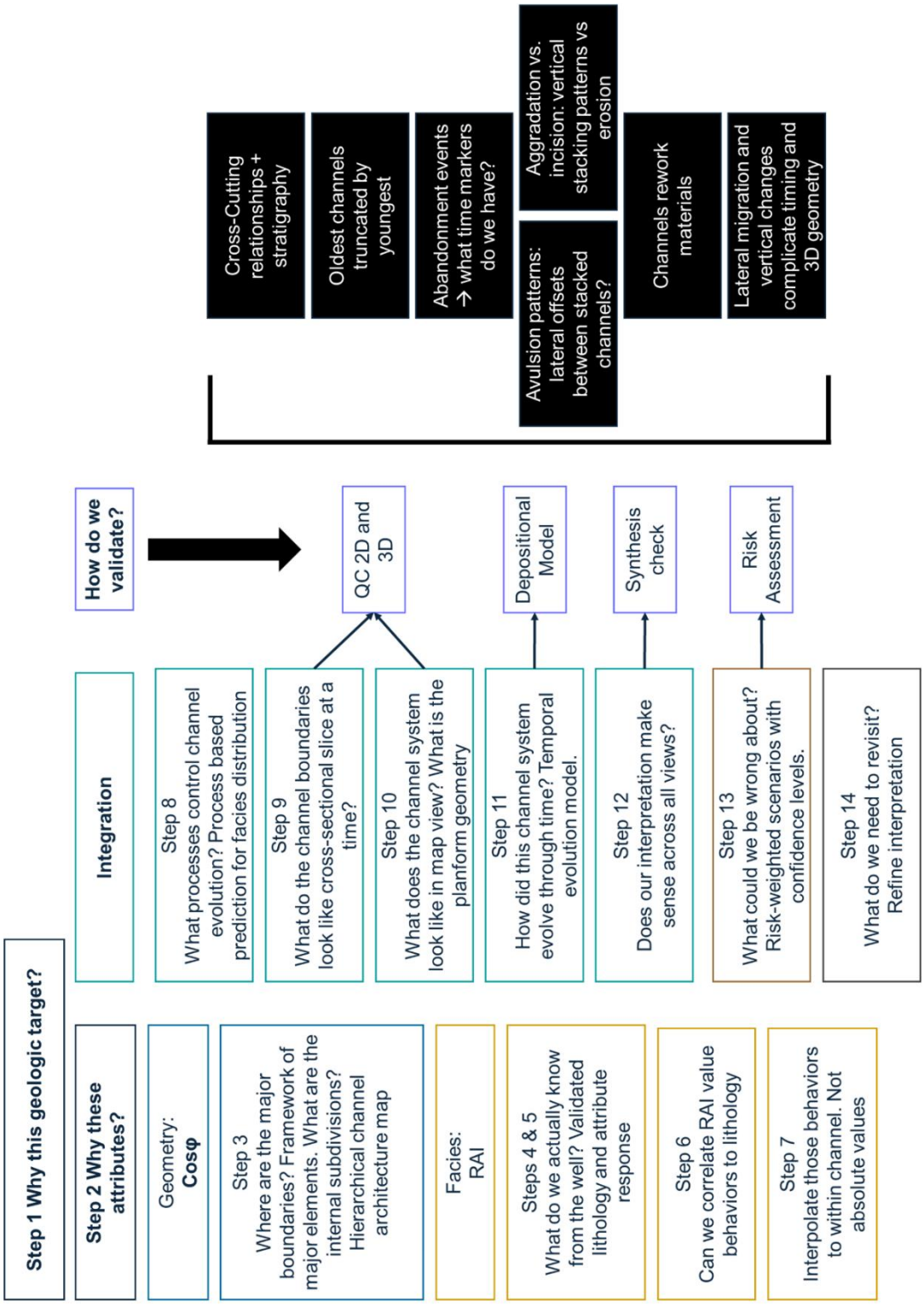


Figure 4: Using the same 14-step workflow, interpreters should assess whether observations are consistent and geologically reasonable before proceeding. There is the risk of moving too quickly without considering alternative reasons for what is being observed.

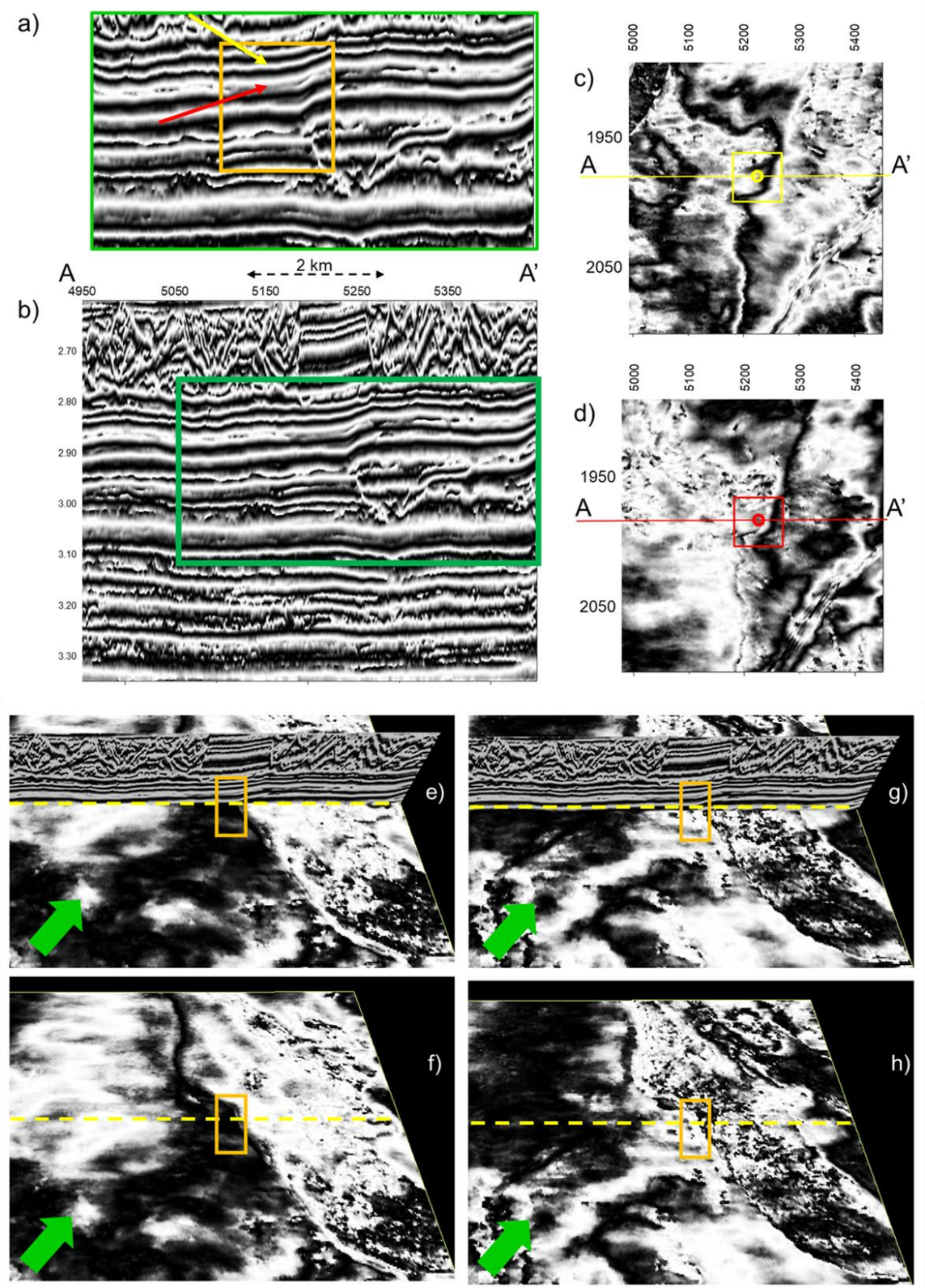


Figure 5: Cosine instantaneous phase showing inclined reflectors from differential compaction near a secondary channel. (a-b) Inline 1980, at two different scales, showing features that could resemble a shallow incised channel. (c-d) Timeslices at 2.83 s and 2.86 s lack typical channel geometries in planform. (e-f) Chair view and timeslice at 2.91s begin to show evidence of a

nearby channel system. (g-h) Deeper view at 2.93 s reveals the actual channel. The inclined strata result from differential compaction rather than channel architectural elements. Arrows in (a) mark the time positions for timeslices (c) yellow arrow and (d) red arrow. Polygons and circles use matching colors to show corresponding locations in (a), (c), and (d). Orange rectangles in (e-h) outline the approximate location for the feature of interest.

IMPROVED FEATURE VISIBILITY AND IDENTIFYING AES

Standard amplitude displays may show some bounding reflectors but provide limited information about intra-formation composition. RAI alone can reveal some internal acoustic variations, the corendered cosφ-RAI display will significantly enhance the visibility of both channel boundaries and AEs while providing direct information about the formation itself. Let's visit Figure 5 to compare how we can improve upon single attributes visualization methods to improve AE identification within the Romney channel complex.

The improvement is more evident when comparing interpreted sections. Using cosφ alone with 50% transparency (Figure 6a), channel complex boundaries can be mapped, but internal architecture remains ambiguous (pink arrow), and lithologic variations are not captured. If we interpret based on this alone, we would essentially be relying on our understanding of where certain materials should likely be deposited rather than on direct evidence from the seismic data. In contrast, the corendered display with 50% transparency (Figure 6b) enables detailed architectural interpretation: channel boundaries are clearly defined, coarse-grained deposits are identifiable (red arrows), fine-grained deposits are distinguishable (blue arrows), and a probable levee structure is visible (purple arrow). This enhanced visualization reveals not only the channel complex geometry but also the internal heterogeneity necessary for understanding depositional processes and timing.

While a cross-sectional view of channel deposition can help us get an understanding of channel bounding edges and some migration timing, it is not necessarily the best view to get a full picture. Let's look at an alternative visualization next.

MEASURED GEOMETRIES AND RELATIONSHIPS

Mapping channel migration patterns requires moving beyond a single cross-section to leverage the full 3D seismic volume. While cross-sectional views like Figure 6 provide initial insights, analyzing individual inlines or crosslines is time-intensive and may not provide a complete picture. Instead, combining cross-sections with map-view slices in chair displays can provide a clearer view of what is happening, not just

However, interpreting from amplitude alone requires assumptions about grain size distribution in deepwater settings, as we are relying heavily on expected relationships between amplitude strength and depositional facies. The question becomes: can the corendered cosφ-RAI visualization provide more direct evidence for channel settings and migration patterns? We will examine the same features using our optimized display in the next section.

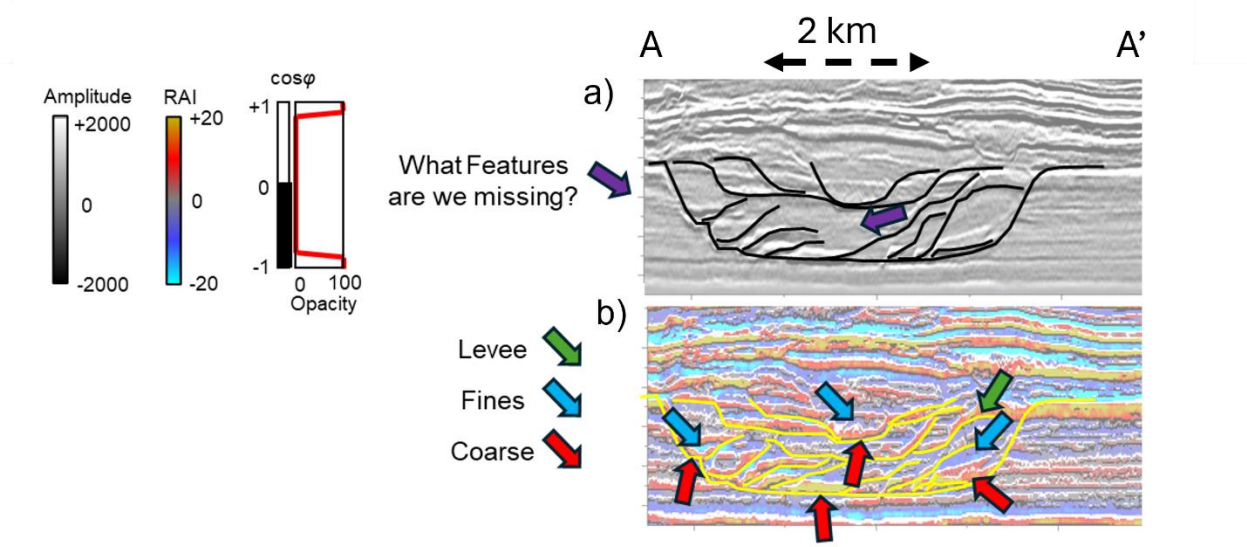


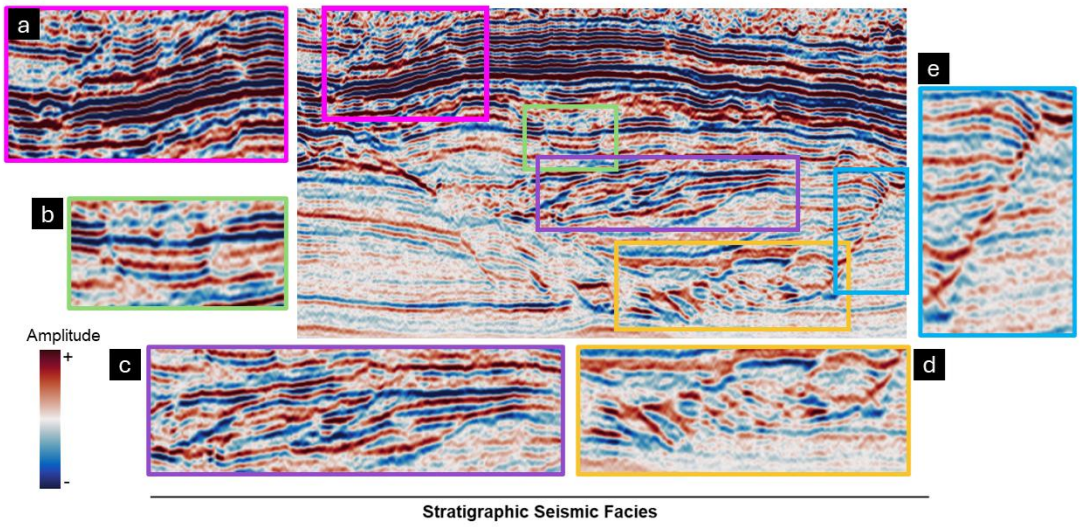
Figure 6: Line AA' through the amplitude volume Yellow, green, and white arrows indicate a broken delineation, and pink and orange arrows indicate more continuous stratigraphic delineations across each reflector. (a) shows how effective amplitude alone is for mapping channel complex features in a cross-sectional view. While we can use the horizontal reflector changes to sketch out some features, we can get a better feel for it using (b) relative acoustic impedance, corendered with cosφ. We can take this a step forward and indicate where finer-grained material could be expected (blue arrow) and coarse-grained material (red arrow).

RESULTS: WHAT IS THE GEOLOGIC MEANING?

IDENTIFICATION OF DEEPWATER CHANNEL ARCHITECTURAL ELEMENTS

Deepwater channel stratigraphy (observed geometric characteristics of channel morphology similar to what is seen in sequence stratigraphy) is fundamentally controlled by both shelf processes and shoreline changes (Catuneanu, 2022); however, there are smaller spatial and temporal scale controls that affect overbank flow and sediment distribution that should not be overlooked (for a brief summary, see Appendix Figure A).

Deepwater channel systems comprise several distinct architectural element components, each with characteristic seismic expressions that can provide critical information for stratigraphic interpretation. While detailed analysis of how these elements affect reservoir connectivity is beyond the scope of this paper (see Funk et al., 2012 for outcrop-based connectivity analysis), accurate identification of these elements in seismic data forms the foundation for such reservoir characterization. We will briefly discuss six stratigraphic seismic facies observed when viewing inline 2080 in Figure 7. Channel fill can affect acoustic properties depending on its lithologic composition; for example, a mud-filled channel may appear as a reflector sag while a sand-filled channel may appear as a structural high due to differential compaction (Figure 7a). We see inclined strata that appear similar to what may be expected in a progradation sequence; however, in this instance, taking a larger view, we see this geometry as a direct result of overburden material leading to compaction.



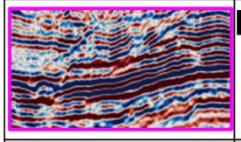
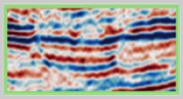
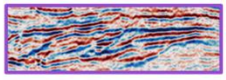
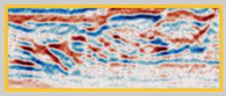
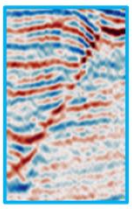
Seismic Response	Appears as	Channel architectural element	Significance
	a Downlap	Differential Compaction	Inclined strata terminating downdip against horizontal surfaces; typical of progradational sequences
	b Truncation	Secondary channel	Minor channel cutting through previous channel fill
	c Onlap	Channel terrace and lateral migration	Younger strata terminating against older surfaces indicate lateral migration of the channel system
	d Truncation	Major unconformity	Strata cut by younger erosional surfaces; indicate active incision
	e Terminations	Major channel unconformity	Channel complex boundary and defining channel unconformity

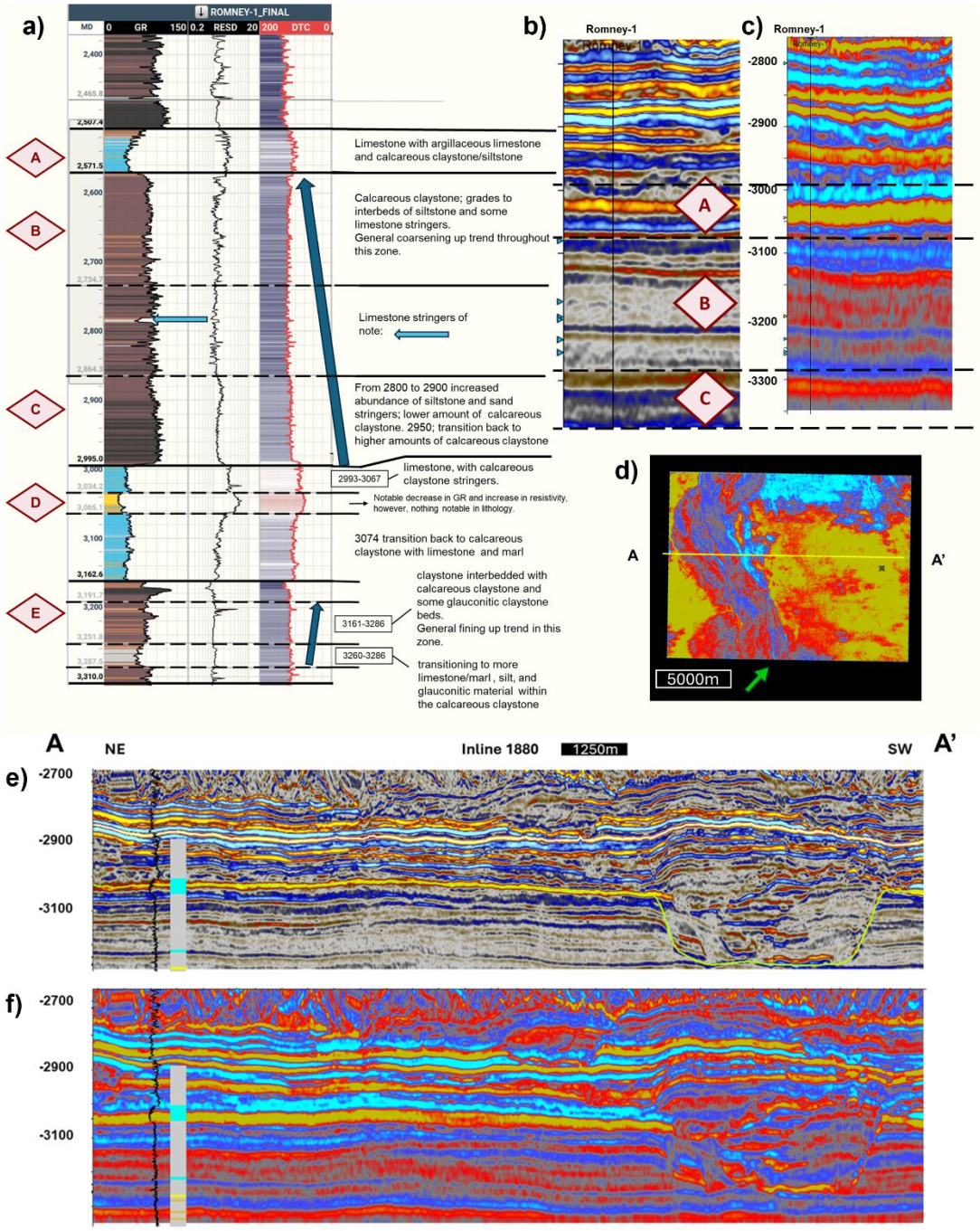
Figure 7: View at Inline 2080 using amplitude. Examining seismic responses in various areas within the channel and considering them collectively will enable us to gain a better understanding of what is happening during the evolution of the depositional environment. However, this is a quick snapshot of just this single location. We need to consider what is happening in both the vertical and horizontal directions. Identifying stratigraphic seismic facies is a multi-part process, beginning here with the seismic response in amplitude. While we use terminology related to stratigraphy to describe similar geometries, the scale we are looking at is different. This figure offers a quick summary of what the seismic response can signify.

Observed truncation may result from a major unconformity and subsequent deposition Figure 7d and 7e, or from secondary channel migration and deposition (Figure 7b). Onlap (Figure 7c), in this example, is interpreted as lateral migration of the channel system and the development of channel terracing. An interpreter must be cautious when beginning the process of using stratigraphic characteristics to recreate the depositional environment.

So, let's begin with our facies development (Figure 8). The Romney-1 had lithology logs, which were used in conjunction with GR and DTC logs to identify specific lithologic packages or well facies (**step** 5). We identified 5 well facies (**step** 6), but are focusing on only 3, which correspond to the vertical limits of the channel system (Figure 8a). Package C is at the base of a coarsening-up sequence. Comparing that to the seismic response in amplitude (Figure 8b), we see sub-parallel low amplitude reflectors that show some chaotic behavior. Moving up the sequence to package B, we continue with low-amplitude responses. Focusing on package A, we see parallel reflectors begin with the low amplitude and are capped with the high amplitude associated with the limestone. (Figure 8c) Extends our facies to consider the RAI response constrained to these same facies. (Figure 8d) Indicates location of A-A' for (Figure 8e) amplitude and (Figure 8f) RAI. Now that we have defined our facies, we can begin interpreting channel stratigraphy and think about depositional history.

We quickly identified six architectural elements (Appendix Figure B); however, we will discuss our interpretation for only two. Figure 9 illustrates a beautiful example of channel cutoff and subsequent abandonment (**steps** 7-12). Figure 9a and 9b show a chair diagram with an inline intersecting a timeslice with amplitude and relative acoustic impedance (corendered with cosφ in inline only) respectively. We've noted channel cutoff/abandonment with red and black arrows in the cross-sectional view and again in plan view to help the interpreter consider what these features look like in both perspectives. Looking at just the vertical changes in (Figure 9c) and (Figure 9d), we can look for similar interruptions in the lateral continuity of the seismic reflector to begin potentially picking out the edges of other iterations of channel migration. The yellow arrow is drawing attention to one of the more recent cutoff channels. When we consider this same feature in the plan view (Figure 9e and 9f), we see that the channel morphology is straighter than in the previous channels. This suggests that an increase in velocity at this time could have contributed to the channel's straightening in this reach. In terms of timing, as this

channel cuts across every other channel iteration we see in this view, it is the youngest channel at this intersection (**step** 11). The other thing of note here is the seismic response within the channel itself. The more sinuous channels have seismic responses that reflect a slower velocity and more parallel deposition, i.e. we can see more nearly parallel seismic responses. When we look at the higher velocity channel, the fill reflects the higher energy, and the seismic response appears to be more chaotic as well (**step** 12).



igure 8: Facies identification is a multi-step process, as shown here. a) Using the well log from the Romney-1, we note well-log response to lithofacies changes. b) and c) show the seismic response at the well location for amplitude and relative acoustic impedance; d) is the cross-section location. e) amplitude with lithology log and gamma ray plotted, and f) relative acoustic impedance with lithology log and gamma ray log plotted. Romney-1 is located approximately 8 km outside of the channel. Seismic response observed at the well bore is then interpolated into the channel complex.

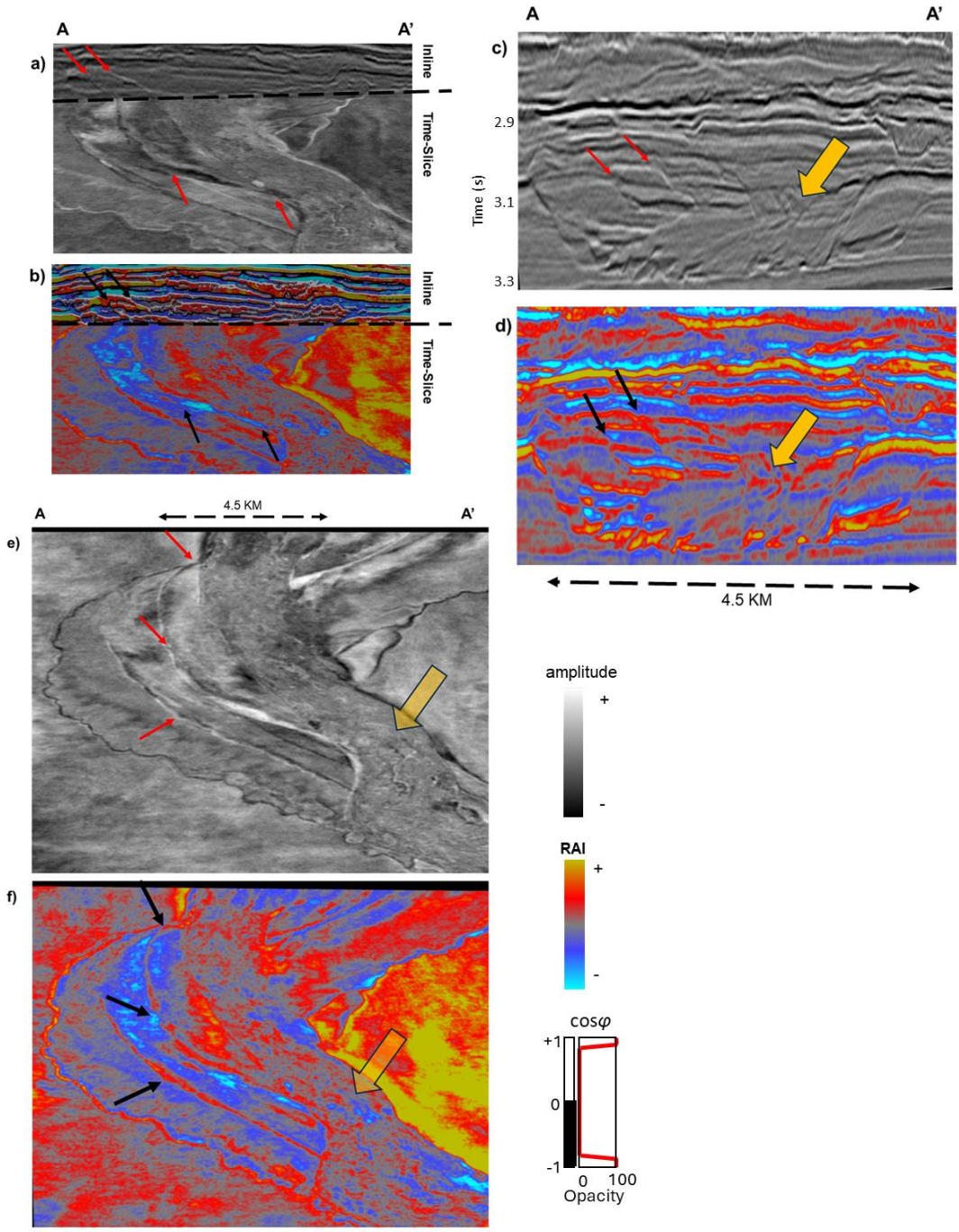


Figure 9: Inline 1951, time-slice 3.068. Chair diagram with a) amplitude and b) RAI, with c) and d) focus on just the cross-sectional view, and e) and f) focusing solely on the time-slice. Red and black arrows indicate channel abandonment, while orange arrows indicate seismic response to increased velocity and straightening of channel geometry. We are better able to see internal variations within the channel complex using RAI (d). The increased flow velocity deposits, as viewed in e) and f), are reflected in a more chaotic seismic response, which begins to resemble what is seen in the MTD responses.

Now that we have identified what channel cutoff and abandonment look like in our 3D seismic data, let's look at point bars and lateral accretion. Moving through the same workflow, we review Figure 10, which focuses a bit deeper on our channel complex at -3.24 seconds, focusing on steps 9-12. Figure 10a and 10b again shows our chair diagram with red and black arrows indicating our focus. In both amplitude and relative acoustic impedance, we see a repetitive pattern of low to high values in the cross-sectional view and can extend that into the plan view as well. Looking with a bit more detail at the vertical changes in Figure 10c and 10d, we can track similar patterns to the right side of the image. Thinking about the point bars as more a result of lateral migration, let’s spend a bit more time focusing on the plan view in Figure 10e and 10f) (steps 9 and 10). We can again see the repetitive pattern of low-high, and the result of this pattern highlights the geometry of a meander loop. Next, we use this pattern recognition with the concept of cross-cutting relationships to understand the timing and evolution of our channel to determine migration direction as indicated by the green arrow (steps 11 and 12).

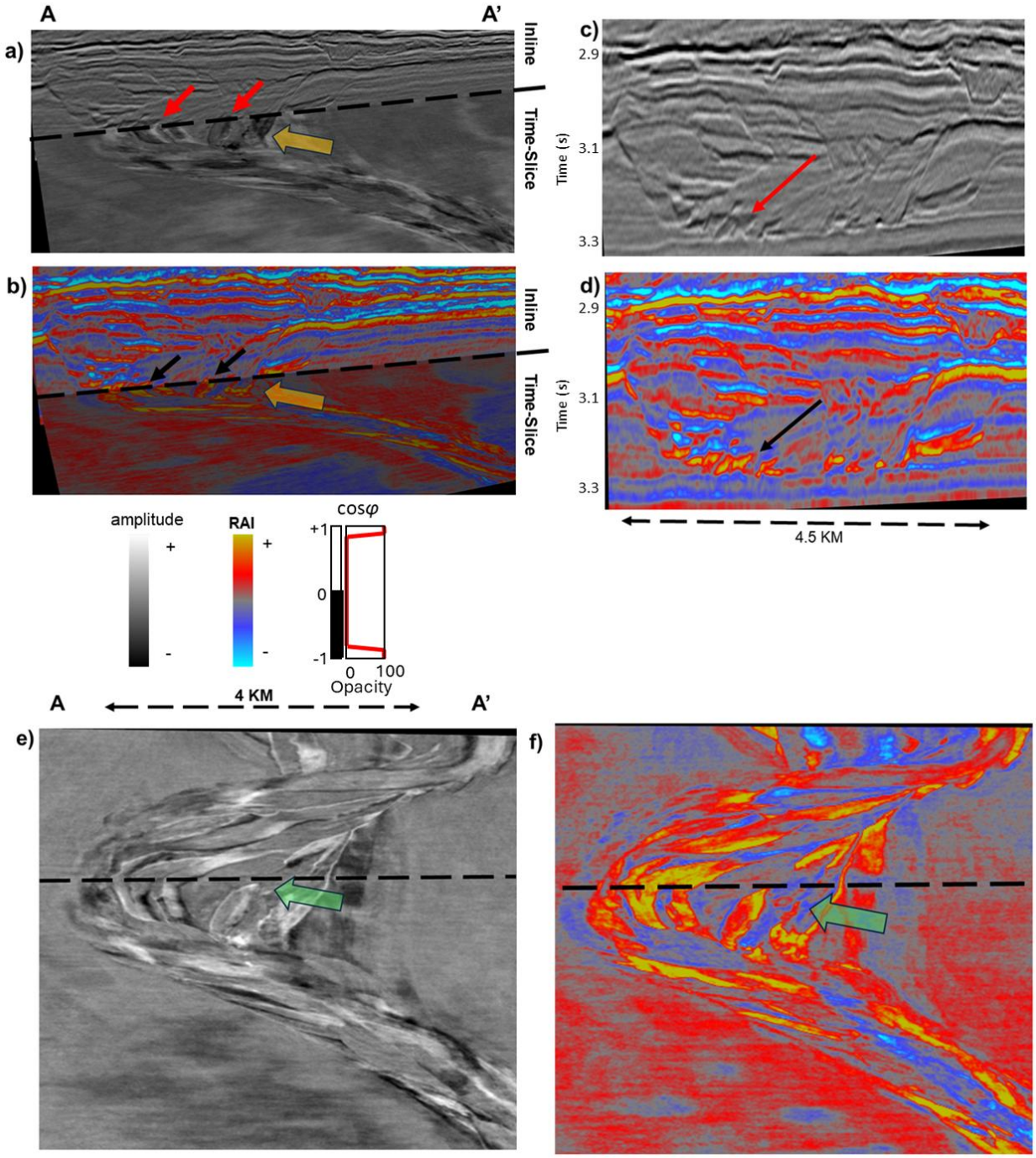


Figure 10: Inline 1951, time-slice 3.244. Chair diagram with a) amplitude and b) RAI, while c) and d) focus on just the cross-sectional view. Orange arrows indicate seismic response to lateral accretion, and the dashed black line indicates the transect for the cross-sectional view. We are better able to see internal variations within the channel complex using RAI (d). e) and f) show just the time slice view with a green arrow indicating the direction of point bar growth.

When evaluating AEs and channel stratigraphy, we use all available information to identify channel components and improve confidence in our depositional interpretation. Other elements important for complete system characterizations include levees, which display asymmetric development. This asymmetry can be seen as heights varying according to channel curvature; outer bends typically develop higher levees than inner bends due to flow dynamics, and these features are predominantly constructed of fine-grained materials that overspill from the main channel during transport events. Subsequent levee collapse and mass wasting can create apparent point bar features composed of reworked levee material rather than traditional lateral accretion deposits (Krastel and Moser, 2022) or can be seen as material splaying out away from the channel.

The architecture of these systems reflects complex flow behavior where the lower, denser portion of flow follows channel-floor topography while upper flow components decouple and spread beyond channel wall confines, creating broader depositional geometries including levees and lobes. In shelf-connected systems, this flow decoupling results in distinct stratigraphic relationships where different elements respond to those shelf processes in predictable patterns that can be traced to stratigraphic frameworks and are dependent on sediment supply variations. However, it may be difficult to distinguish between these elements based on similar acoustic impedance ranges and complex internal geometries. And, tying these features back to those shelf processes becomes more complex as we consider channel movement leading to reworking and redistribution of materials.

For these reasons, we deliberately limit our interpretation scope to channel-scale architecture and focus on what can be reliably inferred from our data: identification of AEs, interpretation of local channel evolution (migration, stacking patterns, fill characteristics), and facies prediction calibrated to Romney-1, explicitly acknowledging that placement within a larger sequence stratigraphic framework would require additional data beyond the scope of what we are discussing here. Let us now think about the evolution of the depositional environment and how proper identification of these AE can help us understand migration and flow within the channel system setting.

MIGRATION AND FLOW OF DEEPWATER CHANNEL SYSTEMS

Channel system evolution is governed by both shared controls common to depositional environments (channels) and deepwater-specific processes that fundamentally distinguish subaqueous from subaerial systems. While sea level, sediment supply, and basin configuration influence all channel systems, deepwater channels operate under additional constraints imposed by bottom water currents (the continuous movement of dense, cold-water masses that can promote mud-rich channel fills), oceanic density currents, and Coriolis effects (for more information on how the Coriolis force effects deepwater channel character see, Davarpanah Jazi et al., 2020, Kolla et al., 2001, Tek 2022, Posamentier and Kolla 2003, and Gong et al., 2021; for more information on the similarities of submarine and fluvial channel systems see Jobe et al., 2016). At higher latitudes, the Coriolis force significantly influences bottom boundary layer dynamics, creating asymmetric flow patterns within channel bends where the density core containing the finest grained sediments consistently deflects toward outer banks, while velocity maximum locations shift based on the interaction between Coriolis and centrifugal forces (Davarpanah Jazi et al., 2020).

Flow velocity controls add another layer of complexity, as density contrasts between turbidity currents and ambient seawater strongly influence both erosional capacity and depositional patterns. Unlike terrestrial systems, where bed substrate affects flow behavior, turbidity currents exhibit hydroplaning characteristics where velocity maxima occur near the seabed, creating fundamentally different erosion and depositional mechanics (Krastel and Mosher, 2022). These complex flow dynamics create predictable but spatially variable depositional patterns that traditional seismic attributes may not resolve, particularly where necessary to distinguish between AE with similar acoustic properties.

The flow character differences between turbidity currents and fluvial systems create distinct architectural signatures. Density-driven turbidity currents exhibit flow stripping behavior where lighter materials separate from the main flow, leading to systematic overbank deposition of fine-grained sediments. As flows progress downslope, progressive mud loss causes levee heights to decrease until eventual fan breakout occurs, at which point deposits become increasingly sand-rich and sheet-like. The density contrast between turbidity currents and

surrounding fluid directly controls flow velocity, with greater contrasts producing faster, more erosive flows that create different channel geometries than their terrestrial counterparts.

Understanding these deepwater-specific processes is essential for interpreting AE and channel stratigraphy patterns from seismic data. Our recommended framework provides a systematic approach to help interpreters identify key features by simultaneously evaluating acoustic impedance contrasts, geometric relationships, and (dis)continuity patterns. This allows us to better distinguish features that may share similar impedance but represent different positions within the depositional model, or in other words, ensures we are not jumping to an easy conclusion without considering alternative interpretations for what we are seeing (**step** 13). We need to ensure we are using the proper geologic context and consider whether our own preconception about what we expect to see is not clouding our judgment when creating our depositional model.

CONCLUSIONS

Interpretation of the Romney deepwater channel complex is improved through the effective combination of optimizing the seismic data, using corendered attributes properly, and a systematic workflow for interpretation. This approach helped in the identification of erosional boundaries, channel abandonment, and lateral accretion while improving confidence throughout our interpretational model. The key distinction between pattern recognition alone and process-based interpretation, demonstrated through our Romney analysis, shows that understanding depositional processes helps constrain facies distributions and architectural relationships even where seismic patterns may be ambiguous. This improved workflow is not for improving visualization quality alone, but additionally moving step-by-step from initial observations through iterative refinement and pausing at each stage to validate consistency across both the 2D and 3D perspectives. We investigated anomalous features, considered alternative scenarios, and assessed confidence levels before advancing. This disciplined approach prevented blind pattern-matching interpretation and ensured each portion of the model represented the most defensible explanation within data constraints. Bottom line, this suggested workflow framework is broadly

applicable to deepwater channel settings where limited well control must be acknowledged to ensure we are honest about what we know, what we must infer, and what we are uncertain about.

ACKNOWLEDGEMENTS

Thanks to New Zealand Petroleum and Minerals for providing access to their Romney3d data volume. All attributes were computed and displayed using software provided by the OU AASPI consortium.

REFERENCES

Catuneanu, O. (2022). Sequence stratigraphy of deepwater systems. In *Deepwater Sedimentary Systems* (pp. 443-470). Elsevier.

Davarpanah Jazi, S., Wells, M. G., Peakall, J., Dorrell, R. M., Thomas, R. E., Keevil, G. M., et al. (2020). Influence of Coriolis force upon bottom boundary layers in a large-scale gravity current experiment: Implications for evolution of sinuous deep-water channel systems. *Journal of Geophysical Research: Oceans*, 125, e2019JC015284. https://doi.org/10.1029/2019JC015284

Funk, J. E., Slatt, R. M., & Pyles, D. R. (2012). Quantification of static connectivity between deep-water channels and stratigraphically adjacent architectural elements using outcrop analogs. *AAPG bulletin*, *96*(2), 277-300.

Gong, C., Steel, R. J., Qi, K., & Wang, Y. (2021). Deep-water channel morphologies, architectures, and population densities in relation to stacking trajectories and climate states. *Bulletin*, *133*(1-2), 287-306.

Hansen, L., Janocko, M., Kane, I., & Kneller, B. (2017). Submarine channel evolution, terrace development, and preservation of intra-channel thin-bedded turbidites: Mahin and Avon channels, offshore Nigeria. *Marine Geology*, *383*, 146-167.

Jobe, Z. R., Howes, N. C., & Auchter, N. C. (2016). Comparing submarine and fluvial channel kinematics: Implications for stratigraphic architecture. *Geology*, *44*(11), 931-934.

Kolla, V., Bourges, P., Urruty, J. M., & Safa, P. (2001). Evolution of deep-water Tertiary sinuous channels offshore Angola (west Africa) and implications for reservoir architecture. *AAPG bulletin*, *85*(8), 1373-1405.

Kolla, V., Posamentier, H. W., & Wood, L. J. (2007). Deep-water and fluvial sinuous channels—Characteristics, similarities and dissimilarities, and modes of formation. *Marine and Petroleum Geology*, *24*(6-9), 388-405.

Krastel, S. and D. C. Mosher (2022), Mapping a river beneath the sea, *Eos, 103,* https://doi.org/10.1029/2022EO220052. Published on 28 January 2022.

Moreno-Ward, A., Bedle, H., Lubo-Robles, D., and Marfurt, K.J., Beyond visualization alone: a corendering technique for enhanced deepwater channel visualization- a tutorial. Interpretation. In Review

Posamentier, H. W., & Kolla, V. (2003). Seismic geomorphology and stratigraphy of depositional elements in deep-water settings. *Journal of sedimentary research*, *73*(3), 367-388.

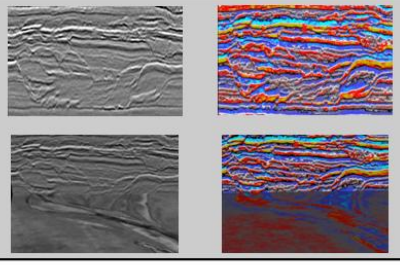
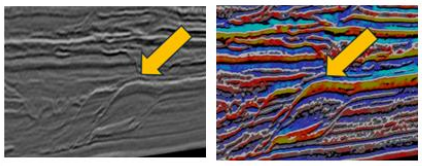
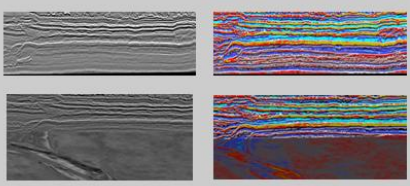
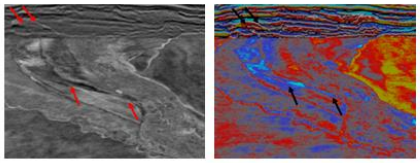
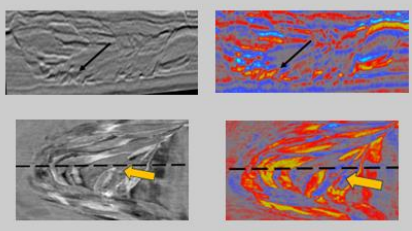
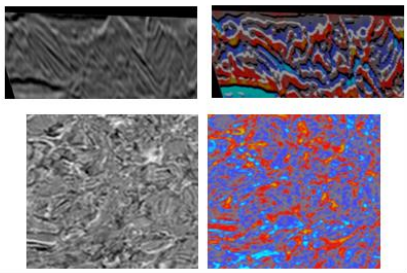
Tek, D. E., McArthur, A. D., Poyatos-Moré, M., Colombera, L., Allen, C., Patacci, M., & McCaffrey, W. D. (2022). Controls on the architectural evolution of deep-water channel overbank sediment wave fields: insights from the Hikurangi Channel, offshore New Zealand. *New Zealand Journal of Geology and Geophysics*, *65*(1), 141-178.

Zhang, W. B., Duan, T. Z., Liu, Z. Q., Liu, Y. F., Zhao, L., & Xu, R. (2017). Architecture mode, sedimentary evolution and controlling factors of deepwater turbidity channels: A case study of the M Oilfield in West Africa. *Petroleum Science*, *14*(3), 493-506.

Appendix 4A: Channel Architectural elements and controls

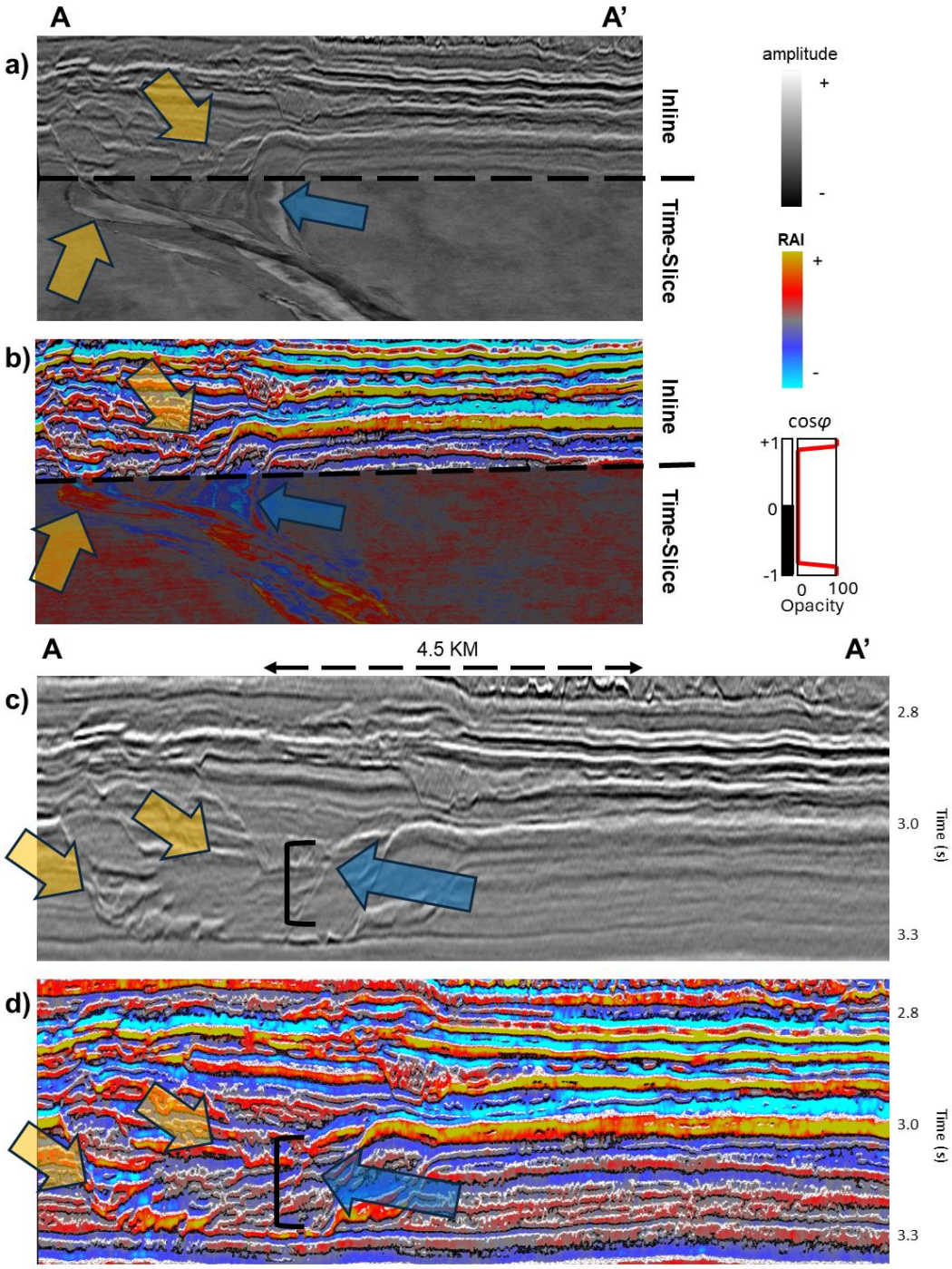
Control Factor	General Mechanism	Effect on Overbank Flow	Applicability & Considerations
Flow-to-Conduit Size Ratio	Thickness of turbidity current relative to channel depth determines overspill magnitude	Thicker flows relative to channel depth generate more vigorous overbank flow and transport coarser sediment	Applicable to all submarine channels
Overbank Gradient	Slope direction and magnitude affects flow acceleration/deceleration	Steep gradients sloping away from channel enhance flow velocity; inward slopes or flat areas hinder flow	Basin geometry and sediment supply rate dependent
Flow Evolution	The longer the flow, the more gradual the loss of sediment from the upper, less sediment-laden portions at each bend or overspill point. Overall flow becomes thinner and less capable of overspilling	Creates systematic downstream decrease in overspill volume and energy, resulting in progressively smaller sediment wave fields along the channel path	Most pronounced in long channel systems (>100s km) with multiple bends where cumulative sediment loss becomes significant
Coriolis Deflection	Earth's rotation deflects flow direction (right in Northern Hemisphere, left in Southern)	Creates systematic cross-channel flow asymmetry, enhancing one margin while hindering the other	Latitude-dependent - strongest at high latitudes, negligible near equator
Contour Currents	Large-scale bottom currents that flow along the seafloor (following bathymetric contours) can either flow with or against the direction of overbank sediment flows	Currents flowing in the same direction as overbank flow enhance sediment transport; opposing currents reduce flow velocity and sediment wave development	Variable by location and time. Current strength and direction can change seasonally, with climate cycles, and over geological time
Centripetal Forces	Channel curvature focuses overspill downstream of bend apex	Creates spatial focusing of thicker, faster overbank flow at outer bends	More important in highly sinuous channels. Southern hemisphere, deflection to west
External Flow Interaction	Flows from other sources (slope failures, tributary channels) interact with channel overspill	Can reduce overbank flow velocity through flow interference or create complex depositional patterns	Common in tectonically active margins and multi-source systems

Appendix Figure A: A brief summary of some control factors that affect deepwater channel morphology. Understanding these variables is important when interpreting the depositional history. For more information on the control factors for deepwater channels, see Tek et al., 2022, Gong et al., 2021, Hansen et al., 2017, and Kolla et al., 2001.

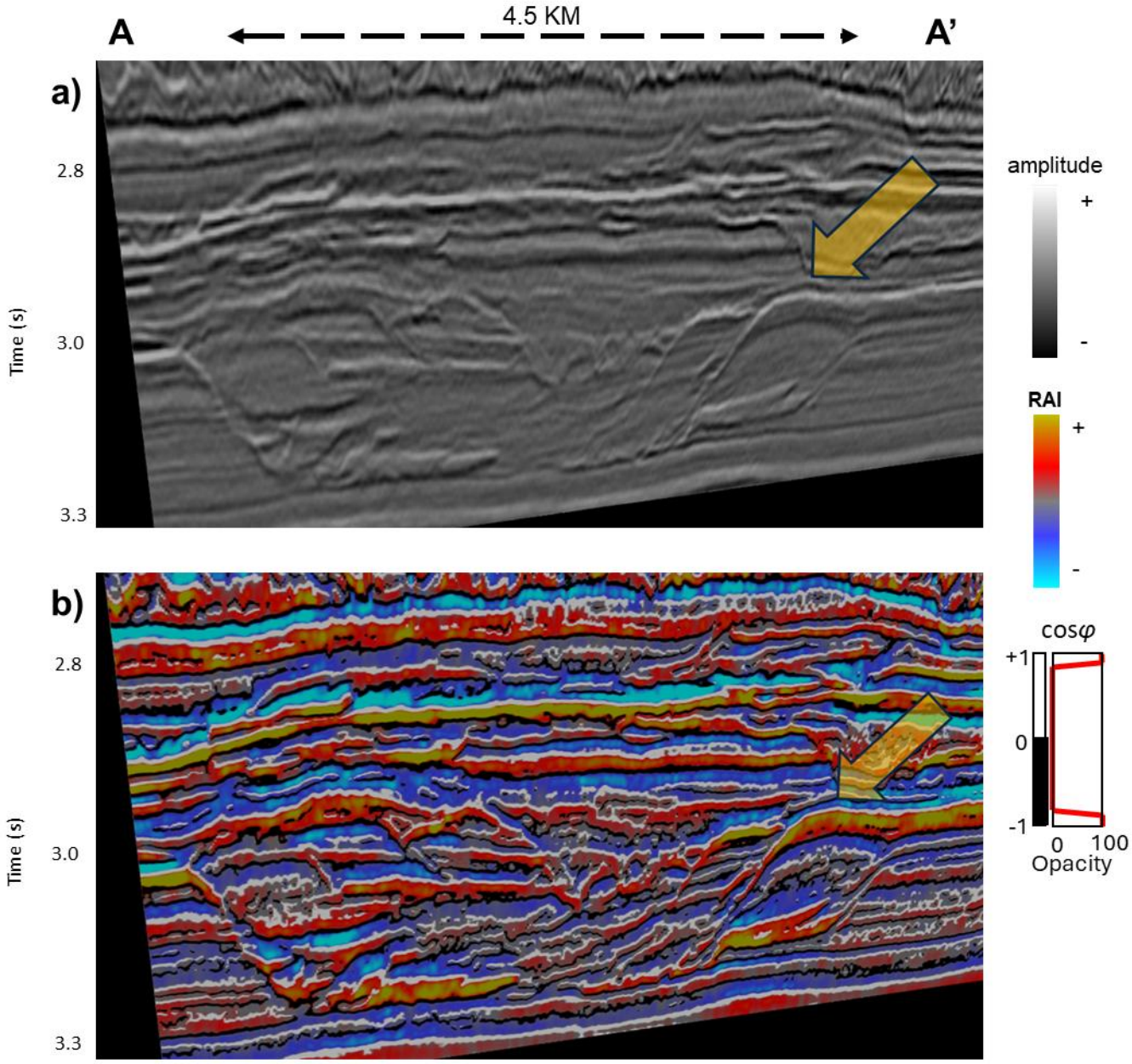
Seismic Expression	Architectural Elements	Geologic Significance	Stratigraphic Interpretation
	Channel Fill	Sediment deposited within the channel axis; composition reflects flow evolution and abandonment processes	Indicates channel activity timing, flow competence, and potential reservoir vs. seal quality.
	Levee	Fine-grained deposits formed by flow overspill; asymmetric due to Coriolis effect and flow dynamics	Defines channel margins, flow direction, and relative flow magnitude; key for compartmentalization
	Over-bank deposit	Widespread fine-grained sediments deposited beyond levee crests during large flow events	Establishes regional flooding surfaces and sequence boundaries; critical for correlation
	Channel Abandonment	Cessation of active flow through channel; often marked by fine-grained fill or draping	Indicates avulsion timing, system evolution, and potential seal development
	Point Bars (Lateral Accretion)	Apparent point bar deposits; in deepwater systems often composed of reworked levee collapse material rather than lateral migration	Indicates channel sinuosity and migration patterns; requires careful analysis to distinguish from fluvial-style deposits
	←INLINE Mass Transport Deposits (MTDs) ←TIMESLICE	Chaotic deposits from submarine landslides, debris flows, or slumps; often triggered by slope instability	Identifies major erosional events, sequence boundaries, and potential drilling hazards

Appendix Figure B: Seismic expression for deepwater channel architectural elements displayed in amplitude (black, grey, white colormap) and RAI corendered with cosφ (colormap defined in figure 9). Amplitude helps identify major channel feature boundaries where reflection strength is strongest but is limited for facies interpretation because 1) high/low amplitude can be

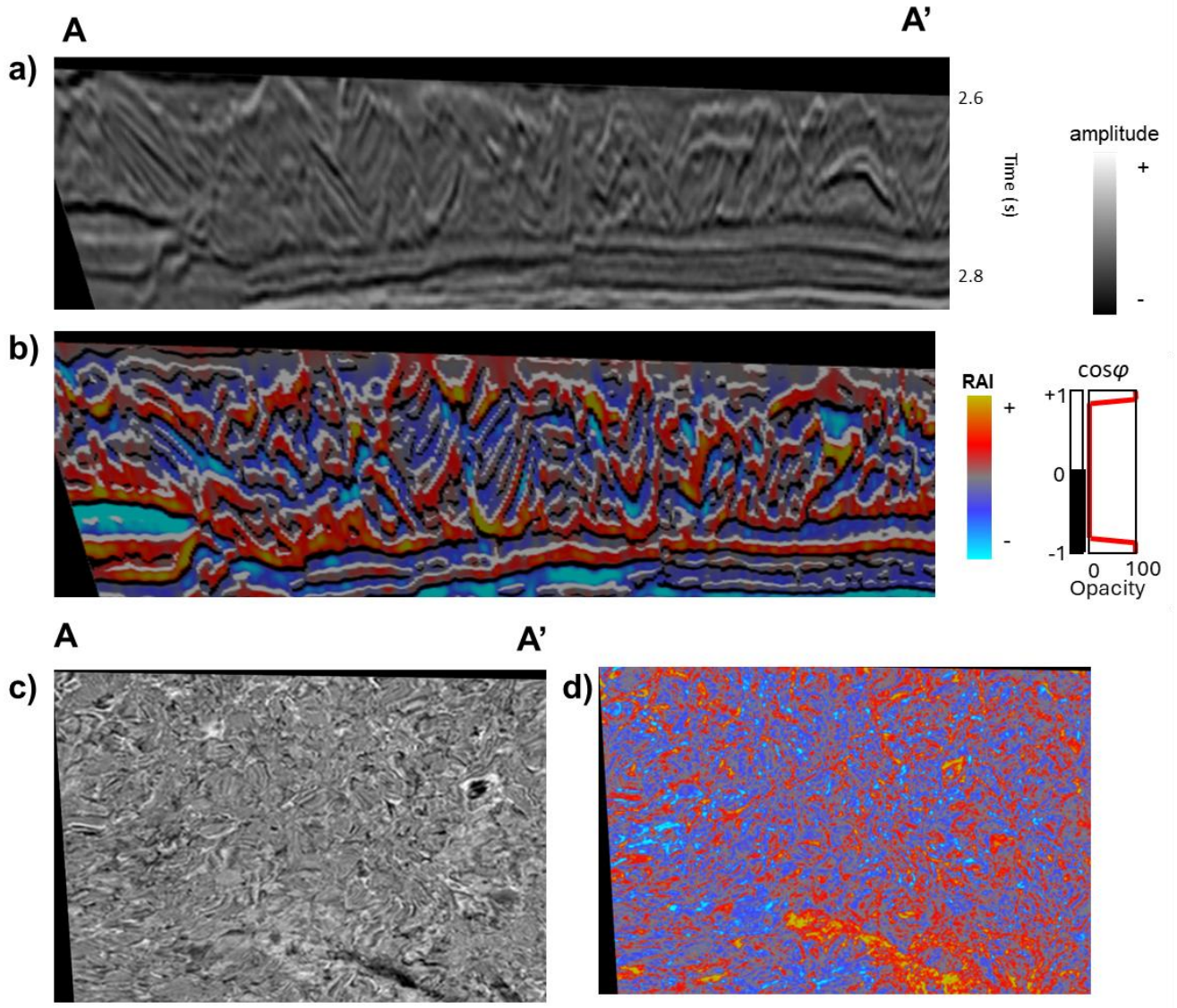
ambiguous, and 2) lack amplitude signal where internal reflectors are absent. RAI can provide more impedance information throughout the strata while cosφ helps to define geometry associated with channel architecture. Both amplitude and RAI are relative measures requiring well control for truthing. RAI is a better estimate for rock properties.



Appendix Figure C: Inline 1960, time-slice 3.18. Chair diagram with a) amplitude and b) RAI while c) and d) focus on just the cross-sectional view. Orange arrows indicate seismic response to channel fill and blue arrows and black bracket focus on overbank deposits. We are better able to see internal variations within the channel complex using RAI (d).



Appendix Figure D: Inline 1965. Cross-sectional view with a) amplitude and b) RAI. Orange arrows indicate seismic response to levee development with more detailed internal variations within the channel complex visible using RAI (b).



Appendix Figure E: Inline 1960, time-slice 3.18. Cross-sectional view with a) amplitude and b) RAI while c) and d) focuses on a time-slice view of MTDS within the seismic volume.

Chapter 6: FROM CLUSTERS TO FACIES: INTEGRATING DOMAIN KNOWLEDGE INTO ML-BASED SEISMIC INTERPRETATION IN DEEPWATER CHANNELS*

This paper is in review in SEG The Leading Edge

ABSTRACT

Machine learning workflows in geoscience are increasingly capable of identifying statistical patterns in complex, multiattribute datasets, but pattern recognition alone does not constitute geologic interpretation. Domain knowledge must be deliberately built into every stage of the workflow, from attribute selection through model evaluation, not applied as an afterthought to algorithm outputs. The final stage of an integrated ML-geologic framework is presented here, demonstrating how unsupervised cluster model outputs are evaluated and interpreted through a two-tier framework that balances statistical performance with geologic interpretability, combining multiattribute seismic analysis, kernel principal component analysis, and unsupervised clustering. Applied to the Romney 3D seismic volume, Taranaki Basin, New Zealand, the workflow identifies nine seismic facies: three response-type facies calibrated at Romney-1 and six stratigraphic interval facies defined in cross-section. Architectural element mapping reveals a multi-phase deepwater channel system with evidence of lateral migration, channel abandonment, and aggradational stacking, with cluster distributions validated across timeslice, chair view, and cross-sectional perspectives. The workflow treats ML as a pattern recognition tool whose outputs must be evaluated against geologic reality, unsupervised in algorithm, geologically supervised in application, and is transferable to any seismic facies identification problem where multiattribute analysis is warranted and well control is limited, provided attribute selection rationale and domain knowledge are rebuilt for each application.

1. INTRODUCTION

Machine learning offers geoscientists an increasingly powerful set of tools for identifying patterns in complex, multiattribute seismic datasets. However, the ability of an algorithm to find statistical structures in data does not guarantee that the resulting groupings carry geologic meaning. This distinction, between mathematical separation and depositional interpretation, is

the central challenge that any ML-based seismic facies workflow must address. Domain knowledge cannot be a final interpretive layer applied after the algorithm has run; it must be embedded at every stage, from attribute selection and dimension reduction through model evaluation and facies labeling. The attribute selection, DR, and clustering components of the larger framework are detailed in companion works (Moreno-Ward et al., 2025; Moreno-Ward and Bedle, In Review): the focus here is the final, most interpretively demanding stage: the integrated evaluation of unsupervised cluster model outputs against geologic reality. Deepwater channel systems make this challenge particularly challenging; their inherent heterogeneity, the scale mismatch between seismic resolution and reservoir architecture, and the sparse well control typical of exploration settings all demand an interpretive framework that is simultaneously rigorous in its use of quantitative metrics and grounded in depositional process understanding. The central question addressed here is therefore not which ML model performs best statistically, but how domain knowledge can be systematically integrated into both the selection and interpretation of unsupervised clustering outputs, using geologic reasoning to identify the most geologically representative model and to translate its outputs into a meaningful and transferable seismic facies framework.

The nearest well is located approximately 8 km outside the channel complex of interest, providing lithofacies calibration at a single interchannel location from which interpretations must be extrapolated into untested channel fairways. Our integration framework bridges quantitative rigor with geologic interpretability by using ML to cluster multiattribute data, observing emergent patterns through objective evaluation, integrating geologic context with statistical results, and maintaining realistic expectations about data resolution limits. The integration strategy combines well calibration through seismic-to-well ties, modern analogues to establish expected facies architectural elements (channel axis, margin, abandonment, lateral migration elements), and seismic geomorphology to map spatial patterns throughout the 3D volume. The underlying logic is that similar seismic responses can indicate similar facies, though we acknowledge this represents one geologically plausible realization rather than a unique solution, and that internal heterogeneity below seismic resolutions remains uncertain.

We have four objectives: (1) apply the integrated workflow to identify geologically interpretable deepwater channel facies; (2) identify mappable architectural elements and stratigraphic relationships at seismic resolution; (3) use this workflow to help target channel fairways for appraisal well targets with reduced uncertainty; and (4) provide a transferable framework applicable to exploration settings with limited well control. Through this process, or any in an exploration setting, we need to ensure we document uncertainties and decisions transparently to support iterative refinement as new data are acquired. This work provides a practical workflow for the common exploration challenge of sparse wells in deepwater channel systems, with transparent methodology demonstrating how to integrate ML-derived results with geologic interpretation. Honest uncertainty characterization clarifies what is known versus unknown, building the foundation for understanding the reservoir and providing a transferable approach for similar exploration settings.

2. Consideration for incorporating ML

2.1 How to approach including ML

For selecting a specific ML workflow, it is worth stepping back to consider what the algorithm is actually being asked to do, and perhaps more importantly, what we are not. A useful framing and one that underlies the decisions made throughout this framework, is that machine learning will answer the question you asked, not the question you meant to ask. This distinction has significant implications for how a workflow is designed and how its outputs are evaluated.

In the context of seismic facies classification, this means that the practitioner must define, as explicitly as possible, what “facies” means within their specific dataset and geologic objective before any model is run. Facies is a term that carries different meanings depending on disciplinary background, scales of observation, and interpretive framework. Without a deliberate working definition, it is easy to either over-constrain the problem, expecting the algorithm to recover a level of geologic detail that the seismic data could not resolve, or to under-constrain it, producing clusters that are statistically valid but geologically ambiguous.

The two-path framework presented in Figure 1 reflects these considerations. Both pathways, direct clustering of the six seismic attributes (path 1) and DR-assisted clustering (path 2), are evaluated not only on statistical grounds but also with geologic interpretability as the primary criterion. When multiple seismic attributes are used simultaneously, the resulting multiattribute feature space can make it difficult for clustering algorithms to resolve meaningful patterns without first reducing that complexity. Path 2 was ultimately found to produce more robust and interpretable results for this reason.

Critically, however, neither path produces a final geologic interpretation on its own. Unsupervised clustering identifies statistical groupings within the data (Hastie et al., 2009); it does not assign geologic meaning to those groupings. That step requires a human interpreter, one who can integrate the quantitative model outputs with regional geologic context, depositional understanding, and an awareness of their own interpretive biases. ML can rapidly identify patterns in large, multiattribute datasets that exceed human perceptual capacity, but it cannot assess whether these patterns are geologically reasonable. The 2 are complementary, not interchangeable.

This section outlines the key considerations to shape our approach, with the intention that it may serve as a practical reference for practitioners working to incorporate similar methods into their own workflows

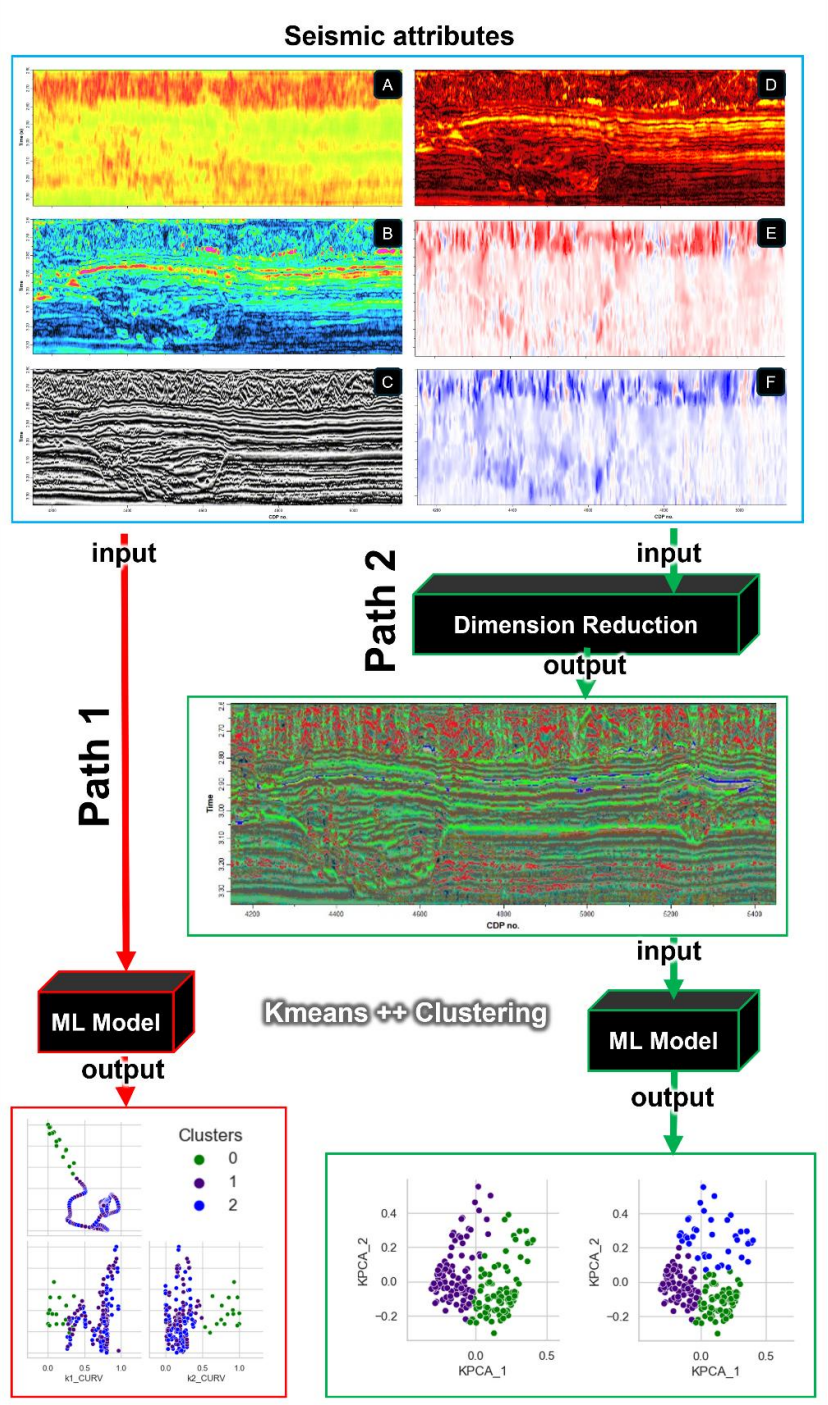


Figure 1: Simplified workflow illustrating the two potential pathways for selecting how to implement ML into the seismic interpretation workflow. Path 1 utilizes all six seismic attributes as direct inputs for unsupervised ML via k-means++ clustering: A) GLCM dissimilarity, B) envelope, C) cosine instantaneous phase, D) sweetness, E) k1 curvature, and F) k2 curvature. Path 2 first employs these same six attributes as inputs for DR, where the resulting output is evaluated using qualitative methods to assess improvements in interpretability before clustering. The DR output then serves as the input for unsupervised k-means++ clustering. The statistical performance of each k-means++ output can be directly compared; however, this comparison alone does not constitute a final determination of geologic interpretability.

2.2 Attributes: single attribute, multiattribute, ML?

Seismic attributes are most informative when selected with a specific geologic question in mind. No single attribute captures the full range of information needed to characterize a complex depositional system, and the six attributes used here illustrate this point directly.

Cosφ is well-suited for resolving channel geometry, boundary identification, termination surfaces, truncations, and scroll bar features, but provides limited information about the facies within the channel itself. Sweetness and envelope complement this by offering sensitivity to internal facies variations, including proxies for relative changes in acoustic character, such as potential grain-size and fluid-content variations. The two curvature attributes, K_1 and K_2, can highlight incisional features and levee geometries. Still, in planview, we cannot distinguish a single full feature, i.e., if we were to map an iteration of a channel meander, we could outline a good portion of it using cosφ. However, we would have difficulties identifying those same full features using one of the two curvature attributes. In cross-sectional view, both curvature attributes and GLCM dissimilarity are susceptible to vertical smearing that limits their ability to resolve internal AEs. Collectively, however, these six attributes were selected because they address two distinct but complementary geologic objectives: characterizing channel geometry and characterizing facies within the channel complex (Figure 2).

Thoughtful attribute selection is a prerequisite for any seismic interpretation approach, and a substantial body of literature exists to guide those decisions (Chopra and Marfurt, 2005 and 2007; Dewett et al., 2021). The intent here is not to reproduce that discussion, but to highlight a practical consequence: when the geologic question requires simultaneous evaluation of more than three attributes, co-rendering is limited to three attributes simultaneously, and even within that limit introduces its own challenges, dominant attributes can visually overwhelm subtler signals, color blending artifacts can obscure geologic features, and the human visual system cannot reliably decompose three independent attribute trends simultaneously. It is at this point that ML offers a meaningful advantage, its ability to evaluate patterns across the full multiattribute feature space in ways that conventional visualization cannot. DR simplifies the

multiattribute feature space for the clustering algorithm, but the resulting cluster outputs must still be evaluated by a human interpreter; the perceptual challenge is therefore transformed, not eliminated.

Figure 2 presents each of the six attributes individually, in both cross-sectional and planview, without interpretation. This is intentional; it allows the reader to assess independently what geologic information each attribute does and does not convey before the combined ML-based evaluation is introduced.

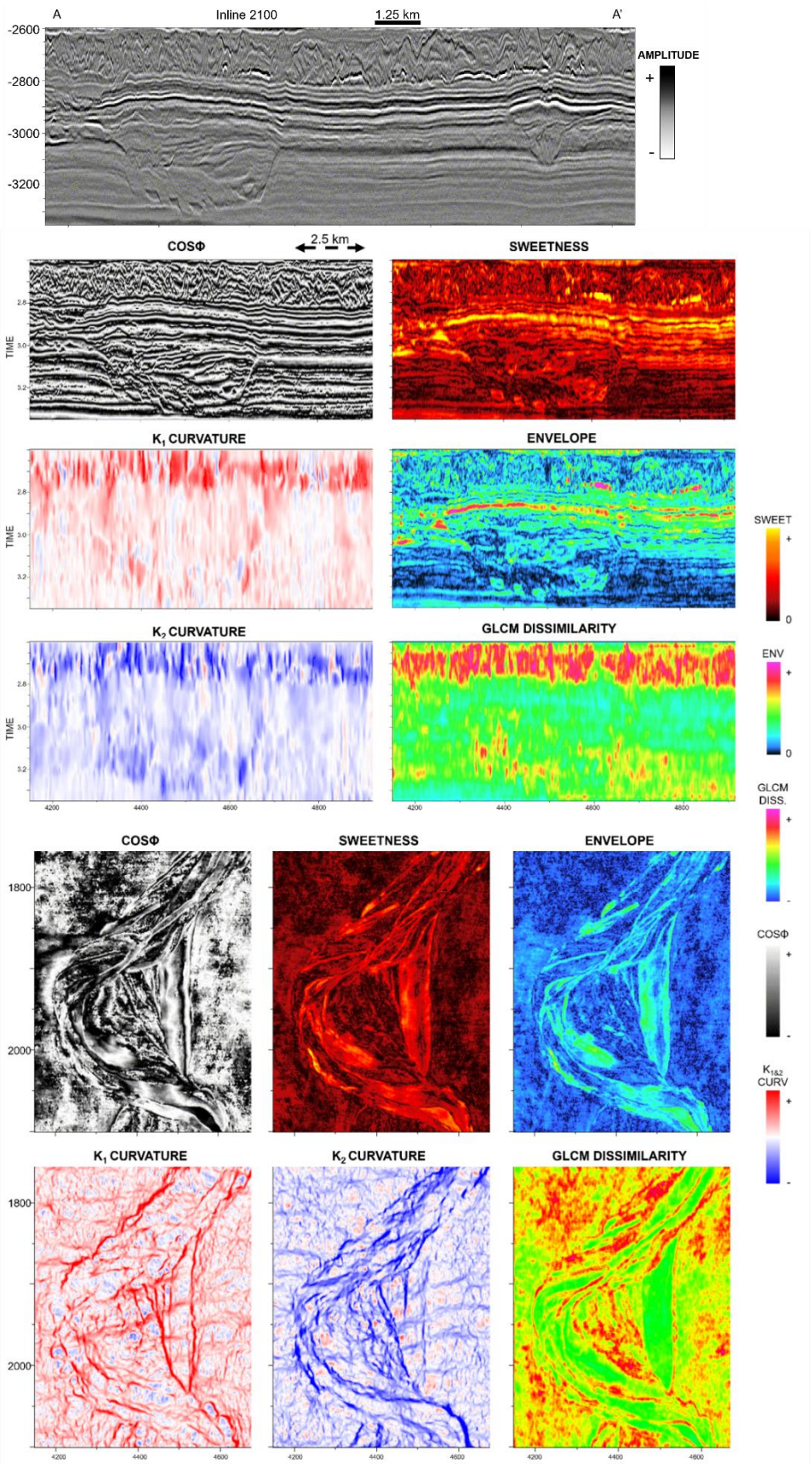

A
Inline 2100
1.25 km
A'
-2600
-2800
-3000
-3200
AMPLITUDE
+
-
COSΦ
2.5 km
SWEETNESS
TIME
K1 CURVATURE
ENVELOPE
K2 CURVATURE
GLCM DISSIMILARITY
SWEET
ENV
GLCM DISS.
COSΦ
K1&2 CURV
1800
2000
4200
4400
4600

Figure 2: *The six input seismic attributes, derived from amplitude, are displayed in both cross-sectional(inline 2100) and planview (timeslice 3.2s). Of the three attributes that show sensitivity to internal architectural elements within the channel complex, cosφ, sweetness, and envelope, cosφ is the only one that consistently resolves those features with sufficient clarity to support the identification and interpretation of these features. The remaining attributes capture laterally variations across the*

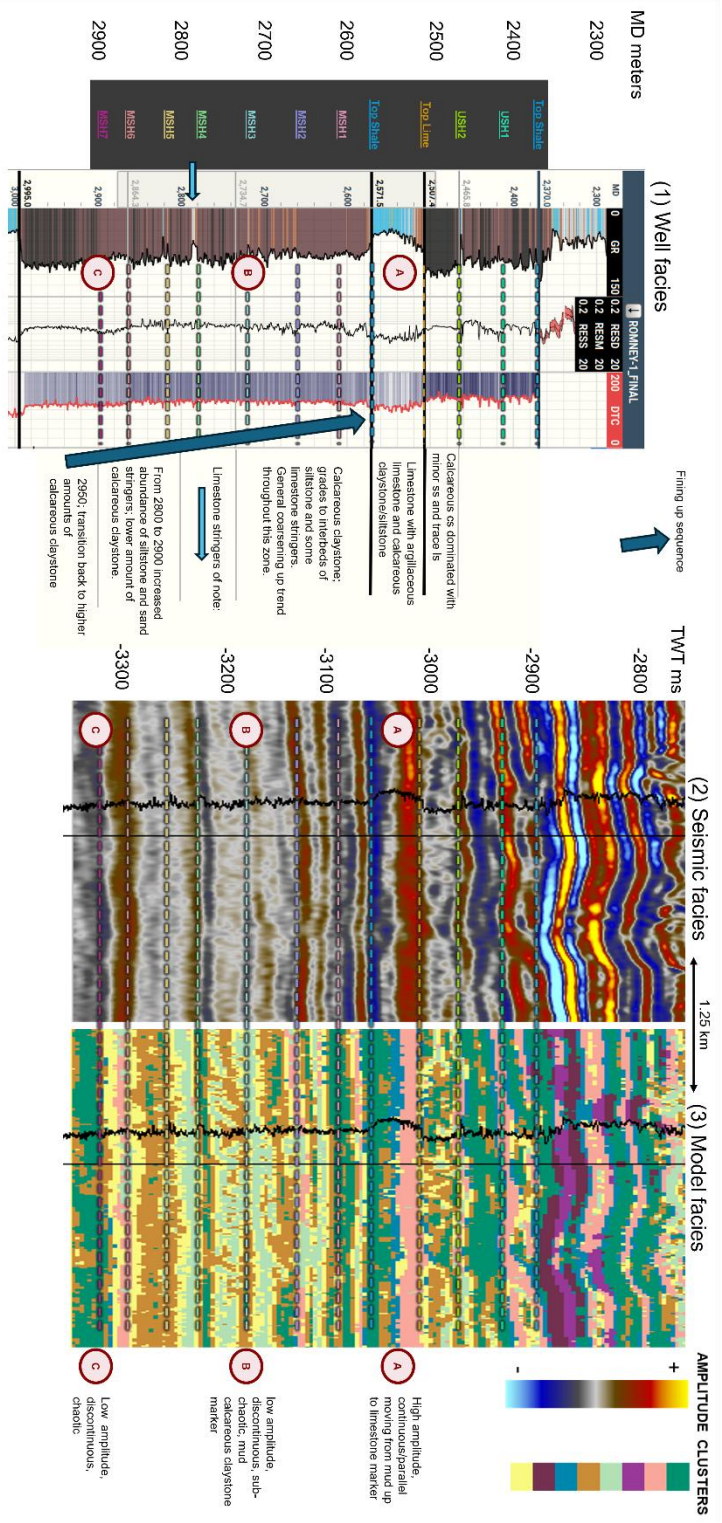
2.3 Facies Calibration Strategy

The facies calibration strategy integrates three complementary lines of evidence to extrapolate from limited well control (Figure 3). At the wellbore, three lithofacies (A-C) were interpreted from wireline and validated against mud log descriptions within the stratigraphic interval (Step 1). These lithofacies were then tied to seismic response through well-to-seismic correlation, establishing the amplitude character associated with each facies at Romney-1 (Step 2). Very high-amplitude reflectors correspond to limestone intervals deposited during transgressive phases, when channels were inactive and pelagic carbonates precipitated in the basin (King and Robinson, 1988; Leila et al., 2022; Posamentier and Kolla, 2003; Posamentier et al., 2022; Strogen et al., 2019). Seismic facies at the wellbore are then compared with model responses from the unsupervised clustering, establishing the facies link between well, seismic, and model (Step 3).

These facies (seismic and model) are then interpolated into the channel of interest. Critically, facies within the channel are **not** lithofacies but rather defined by patterns in seismic response associated with depositional architecture. This distinction reflects a deliberate, process-driven approach that prioritizes AE identification and depositional history first. A practitioner with a different geologic objective would define facies differently, select attributes differently, and arrive at different but equally valid clustering results. The three-step calibration framework outlined here is transferable to any setting where well control exists. What is not transferable is the facies labeling that follows.

Amplitude is the integrated seismic response that geoscientists are trained to interpret for depositional features. While the six input attributes capture different aspects of the seismic signal (e.g. cosine instantaneous phase, k1 k2 curvature, envelope, GLMC dissimilarity, and sweetness), amplitude remains an intuitive and interpretable dataset for identifying channel features. These attributes extract specific mathematical properties from the seismic waveform (texture, continuity, geometry) that enhance pattern recognition for ML, DR, and clustering. However, as amplitude preserves the waveform character that connects directly to impedance contrasts and depositional facies, we are making the conscious decision to cross-reference clustering results directly to the amplitude volume. In doing so, we can determine if the statistically-derived clusters align with geologically significant features visible in the data geoscientists routinely use for interpretation. Additionally, amplitude with bump mapping enhances subtle geometric features that might be less noticeable, making it a good option for interpreting AEs.

This context-dependent interpretation is essential because seismic amplitude response indicates impedance contrast, not a unique lithology (Pemberton et al., 2018). High amplitudes observed within the channel fairway are interpreted differently from high amplitudes at the wellbore, based on geometric position within the depositional system and stratigraphic context.



Within the mapped channel complex, high amplitude reflectors at the channel base and axis are interpreted as coarser clastic material rather than limestone, based on three lines of evidence: (1) geometric association with erosional channel features formed during regressive phases when siliciclastic delivery from the shelf was active, (2) organized parallel reflector character consistent with bedded turbidites, and (3) analogue subaerial systems demonstrating grain size distribution across specific architectural elements. Conversely, low-amplitude, chaotic to sub-parallel seismic character within the channel suggests mud-dominated material, similar in seismic response to the low-amplitude claystone observed at the wellbore.

The well calibrates interchannel facies, while interpretations within the untested channel fairway integrate the following constraints: (1) seismic attributes capture variations in rock properties

Figure 3: Facies calibration workflow. 1) Three well facies (A-C) were identified from GR, RES, and mud log descriptions within the target interval. 2) Well-to-seismic tie, establishes seismic response (amplitude-based) and seismic facies. 3) Model facies are then identified at the well bore using the foundation from steps 1 and 2 and then interpolated into the channel of interest. The model facies are NOT lithofacies.

that can be clustered and mapped spatially; (2) seismic geomorphology identifies depositional positions within the system, informing expected facies based on architectural element context; and (3) modern analogues provide architectural templates for identifying common channel features. Together, these approaches allow facies framework development despite very limited well control, though uncertainty remains highest where well calibration is absent (Mayall et al., 2006).

2.4 GEOLOGICAL AND DATA CONTEXT

2.4.1 Study Area and Depositional Setting

The Romney 3D seismic volume is located in the deepwater Taranaki Basin, offshore New Zealand, approximately 160 km northwest of the Taranaki Peninsula in water depths of ~1450 m (Figure 4). The main channel of interest for this paper is a Late Oligocene to Late Miocene deepwater channel-levee system positioned 60 km from the continental shelf edge (Lutome, 2025; Wang et al., 2025). Seismic evaluation indicates a major complex channel system at -2900 to -3270 ms TWT (channel system A) and a younger, laterally more extensive channel system at -2750 to -2900 ms TWT (channel system B). Channel system A, from a geophysical standpoint, occurs at -2950 to -3300 ms TWT, or, to consider from a more geologic viewpoint, 2438 to 2875 m MDRT (measured depth from rotary table), representing a 437 m thick channel complex built through multiple phases of lateral migration and vertical aggradation (Figure 4 (C)). The channel complex occupies an erosionally incised area ~4.5 km wide when measured in cross-section, though this width varies locally.

These channels formed during regressive phases when lowered sea level triggered sediment delivery via turbidity currents to the basin floor. During transgressive phases, channels became inactive, and the basin received pelagic carbonate deposition, creating distinctive limestone intervals observed in well logs. Modern analogues from subaerial systems provide process-based frameworks for understanding lateral and vertical architectural elements and facies distribution. However, direct application of subaerial analogues to submarine settings does require caution: turbidity flows are episodic rather than continuous, and at 60 km from the shelf edge, maximum

grain size is likely limited to finer sand (Wynn et al., 2007; Posamentier and Kolla, 2003; Kolla et al., 2007).

2.4.2 Available Data

The 3D seismic volume is a pre-stack depth-migrated (PSDM) dataset targeting the Romney structure and surrounding basin. Romney-1 is the only well in the seismic area and was targeting North Cape sands and the Taranaki Delta formations; however, it was drilled approximately 8 km outside of channel system A. Available well data includes a full suite of wireline logs and mud logs with lithologic descriptions. The well provides direct calibration of interchannel facies, with much of the section at the depth of interest composed of 90% calcareous claystone and 10% siltstone, but it does not sample channel axis or proximal levee environments. This combination of 3D seismic and single-well control in an interchannel setting requires an integrated approach that combines seismic attribute analysis, machine-learning-based evaluations, and analogue-informed depositional modeling to map the channel system.

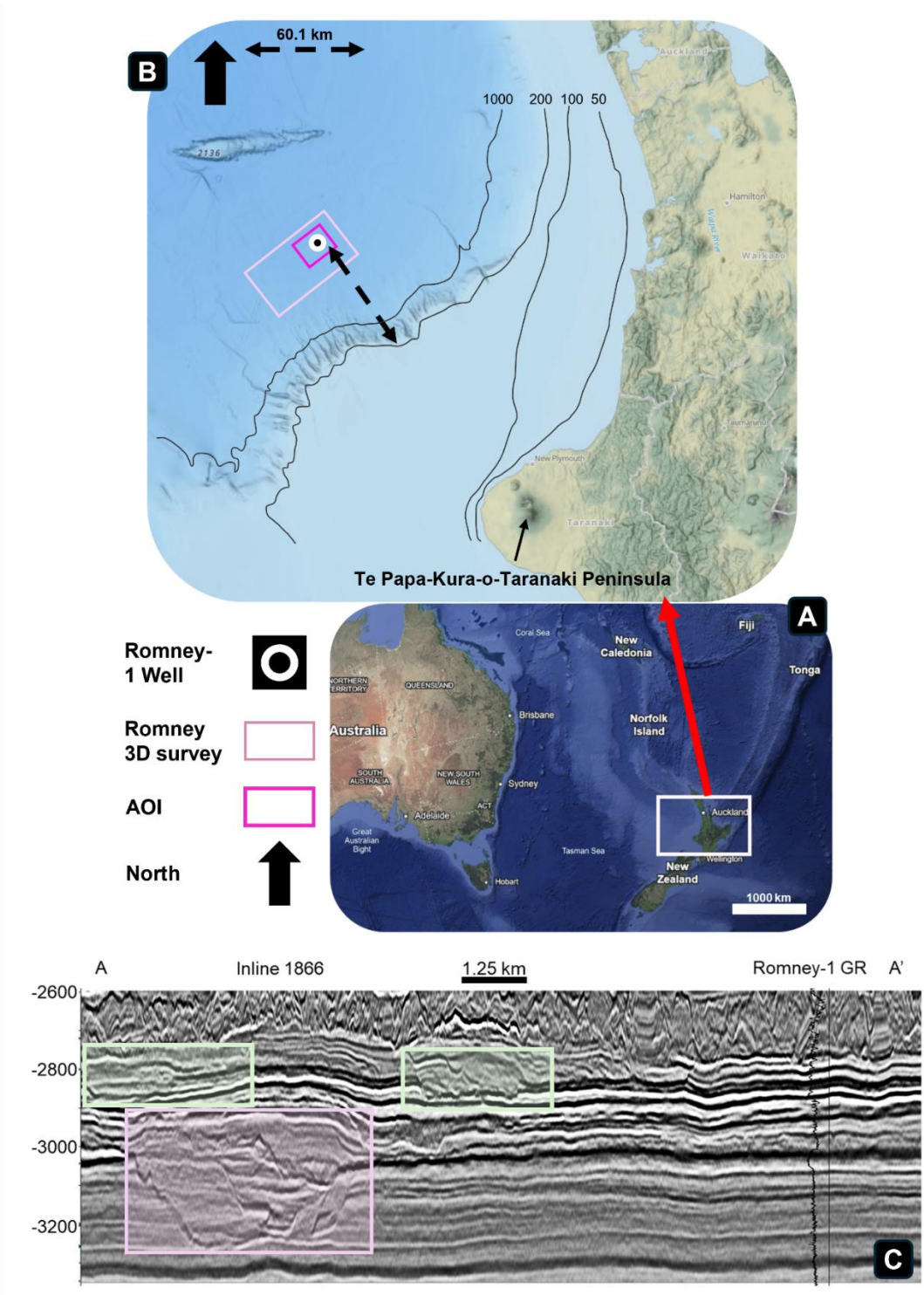


Figure 4: A) Romney dataset location offshore NW of New Zealand, approximately B) 60 km from the continental shelf edge and 160 km NW of the Taranaki Peninsula. C) Inline 1866 with pink polygon noted channel system A, our primary channel of interest, and green polygons indicating channel system B.

3. METHODOLOGY

The following methodology section is organized to summarize the early stages of the larger framework detailed in (Moreno-Ward et al., 2025; Moreno-Ward and Bedle, In Review). The primary focus here begins in section 3.3, where the integrated two-tier evaluation framework, combining quantitative statistical assessment with systematic geologic interpretability criteria, is introduced as a structured and transferable approach to unsupervised model selection for interpretation.

3.1 Seismic Attribute Analysis and Dimension Reduction

Seismic attributes were computed across the 3D volume, and dimensionality reduction (DR) techniques were applied to reduce the multiattribute dataset to a lower-dimensional space suitable for unsupervised clustering. Three DR techniques were tested: principal component analysis (PCA), kernel principal component analysis (K-PCA), and uniform manifold approximation and projection (UMAP). PCA provides linear dimensionality reduction, K-PCA captures non-linear relationships through kernel transformations, and UMAP preserves local neighborhood structure while maintaining approximate global topology. Detailed methodology for attribute selection and DR implementation is provided in our previous work (Moreno-Ward et al., 2025). K-PCA was selected for detailed analysis because it produced the highest qualitative interpretability pre-clustering and demonstrated statistically higher clustering performance post-DR compared to PCA, UMAP, and no DR outputs.

3.2 Unsupervised Clustering and Model Generation

K-means ++ clustering was applied first to the six attributes, with clustering results evaluated for performance (Figure 1). Next, K-means ++ clustering was applied to the K-PCA-reduced attribute space following the methodology established in Papers 1 & 2 (Moreno-Ward et al., 2025; Moreno-Ward and Bedle, In Review) (Figure 1). K-PCA-based models demonstrated superior combined statistical and geologic delineation performance, with the top nine models all derived from K-PCA dimensionality reduction. We will focus on evaluating the top-performing K-PCA models through integrated geologic interpretation, emphasizing how these statistically derived clusters map to depositional architecture and channel-scale stratigraphy.

3.3 Integrated Model Evaluation Framework

Model selection integrated two evaluation tiers. Tier 1 quantitative validation used silhouette scores and the Davies-Bouldin Index to evaluate cluster cohesion and separation. Tier 2 geologic interpretability evaluated each model based on stratigraphic and AE identification and overall facies distribution, providing confidence in the interpretation model. A detailed discussion of statistical metrics and their application is provided in Moreno-Ward and Bedle, In Review.

Critically, model facies are defined not as lithofacies labels, but as geometrically related seismic facies associated with channel AEs. These model facies help constrain channel-scale stratigraphy and geometry (Moreno-Ward et al., In Review), which, when integrated with subaerial analogue frameworks and observed differential compaction, inform interpretations of expected grain size

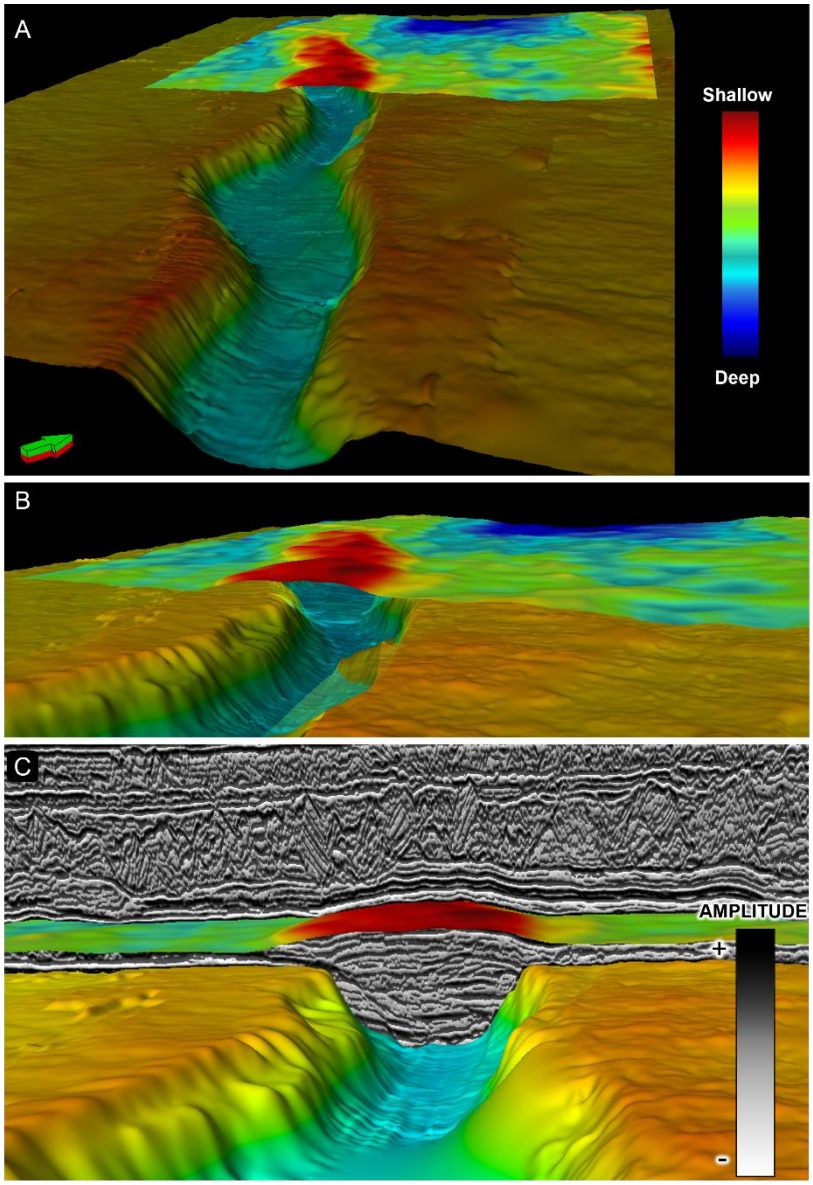


Figure 5: When evaluating this deepwater channel system, we have to acknowledge that while we do not have lithologic truth via a well, we can use other indicators to understand the channel fill. A) The major surface that the channel incised was mapped, and mappable horizon above the channel complex was also mapped. B) Show the same surfaces, but with a slightly different vantage. What we see is a confirmed "structural" high. C) The same two mapped surfaces again, but this time with an inline using amplitude shown. The interpretation for this feature is differential compaction. There is recorded claystone outside of the channel. As this channel is tied to transgressive-regressive cycles we can also infer that the material filling the channel are more coarse material coming from the shelf.

distributions (Figure 5). The optimal model will be balanced between the two evaluation tiers: while higher-performing statistical models are considered for initial evaluation, the best model will ultimately be the one that best preserves geologically significant features. This allows for the confident integration of quantitative clustering results with qualitative geologic interpretations.

Geologic evaluation criteria include facies continuity and AE preservation, consistency with established depositional models, appropriateness of boundary sharpness versus gradation, and channel geometry preservation.

Because KPCA_1 is dominated by cosφ contributions (Moreno-Ward and Bedle, In Review), clusters with strong KPCA_1 influence will preferentially track seismic reflection events regardless of their position relative to the channel complex. Geologic context is therefore essential for distinguishing clusters that map channel margins and basal surfaces from those that reflect unrelated stratigraphic boundaries elsewhere in the volume.

Cluster outputs are not interpreted as discrete classifications but are evaluated in the same manner as any seismic volume, through the spatial patterns that clusters collectively produce. Clusters that are pervasively expressed across the volume tend to reflect broad attribute characteristics common through the dataset, while clusters present in smaller proportions are often more geologically diagnostic, capturing specific seismic responses associated with discrete AE or localized depositional conditions. The spatial relationship between clusters, their relative proportions, and their geometric context within the channel complex also carries geologic meaning. These patterns are evaluated against the amplitude, volume, and known architectural element geometries to ground the interpretation in an observable seismic response that geoscientists routinely use. Additionally, a seismic interpretation model of this channel complex was considered using cosφ corendered with relative acoustic impedance to provide an independent depositional model against which cluster outputs can be assessed (Moreno-Ward et al., In Review) This conventional interpretation, developed independently of the ML workflow, helps to establish a depositional framework, including channel geometry identification and stratigraphic context, that serves as a primary reference for evaluating whether clustering results are geologically meaningful. This three dimension mophological evaluation is therefore treated as a core interpretive step rather than a purely visual one.

3.4 Cluster Normalization and 3D Visualization

Because unsupervised k-means ++ assigns cluster numbers arbitrarily, direct comparison across models requires normalization. Cluster normalization was achieved by creating a discrete colormap and manually matching colors to corresponding clusters across models, ensuring that similar facies maintain consistent visual representation. As cluster count decreases, the colormap is simplified accordingly to reflect the combined facies.

Normalized models were then evaluated through 3D visualization in inline, timeslice, and chair views (where we view inline simultaneously with a timeslice of interest) to assess spatial distribution of seismic facies. Next, geobody extraction was performed to evaluate model performance in delineating channel AEs. Geobody extraction provides an additional and independent line of evidence. A cluster that appears pervasive across the volume may seem geologically non-specific when evaluated as a spatial pattern alone; however, when extracted as a three-dimensional geobody, it can correspond directly to channel scour morphology, revealing the erosional container rather than the fill. This combined visualization and evaluation approach enables the assessment of how well each model preserves and delineates geologically meaningful deepwater channel features, including axis, levee, and margin positions.

4. RESULTS: SEISMIC FACIES ARCHITECTURE

4.1 Response-Type Facies: Seismic Facies Characteristics

Seismic facies characterization operates at two complementary levels in this evaluation: (1) response-type facies defined by seismic response character and calibrated at Romeny-1, which describes recurring seismic signatures applicable across the volume, and (2) stratigraphic interval facies defined in cross-section to describe the vertical organization of the section outside channel package A. The former used to assign geologic meaning to cluster outputs; the latter to provide the stratigraphic context within which those cluster assignments are evaluated. We will begin by evaluating type 1.

Three primary response-type facies were identified through integrated analysis of well-calibration, amplitude response, and clustering model output (Figure 3). These seismic facies

delineate distinct depositional areas within the deepwater channel system, characterized by seismic signature, spatial distribution, seismic attribute information, and architectural context.

Seismic Facies A: High Amplitude, Continuous/Parallel

Seismic Signature: At the wellbore, the seismic response shows high amplitude, continuous to semi-continuous reflectors with parallel to sub-parallel geometry. This response is associated with lithologies that range from calcareous claystone intervals upward to a prominent limestone marker.

Depositional interpretation: High impedance contrast intervals at the wellbore are interpreted as limestone deposition during transgressive phases; within the channel, these are interpreted as coarser clastic material deposited during regressive phases when siliciclastic delivery from the shelf was active. These two interpretations are context-dependent, with the within-channel interpretation based on geometric association and depositional system context.

Reservoir Implications: Context dependent. In interchannel positions, this is non-reservoir; in a channel axis position, it represents a potential reservoir if it is sand-filled.

Seismic Facies B: Low Amplitude, Discontinuous, Sub-Chaotic

Seismic Signature: At the wellbore, the seismic response shows low amplitude, discontinuous reflectors with sub-chaotic behavior. This response is associated with calcareous claystone intervals.

Depositional interpretation: At the wellbore, this is interpreted as a mud-dominated interchannel facies. Within the channel complex, this response may represent a distal levee or abandoned channel fill; depositional position within the system is critical to disambiguation.

Reservoir Implications: Context dependent. In interchannel positions, this is non-reservoir; in a channel position, this facies would represent a baffle of barrier interval.

Seismic Facies C: Low Amplitude, Discontinuous, Chaotic

Seismic Signature: At the wellbore, the seismic response shows low amplitude, discontinuous reflectors with chaotic internal behavior. This response is associated with high proportions of calcareous claystone.

Depositional interpretation: At the wellbore, this is interpreted as a mud-dominated interchannel facies. Within the channel, this response may represent abandoned channel fill. Positioning within the system becomes very important.

Reservoir Implications: Non-reservoir in both interchannel and within-channel in this system.

4.2 Stratigraphic Interval Facies

To place these response-type facies within their broader stratigraphic context, the following section examines the vertical facies organization observed in cross-section along inline 1866, which intersects Romney-1. Six **stratigraphic interval facies**, representing the type (2) facies and labeled as 1*-6, are defined along this cross-section outside channel complex A to provide a stratigraphic context for evaluating cluster behavior in the section surrounding the primary target interval.

Viewing our model in cross-sectional view (Figure 6, line A-A') allows us to get an understanding of the vertical stacking of facies and stratigraphic organization outside of and within the channel. The post-K-PCA clustering model with 8 clusters, both 2 and 3 components, is viewed along inline 1866 in order to extrapolate information directly from Romney-1. This same perspective is viewed with variations in display to aid the interpreter in the selection of stratigraphic facies packages. Figure 6 (A) shows the amplitude volume uninterpreted, with (B) the interpreted equivalent. The post-K-PCA clustering model with 2 components and 8 clusters is shown (C) uninterpreted and (D) interpreted; the same model is shown, corendered with cosφ uninterpreted (E) and interpreted (F). Additionally, post-K-PCA clustering model with 3 components and 8 clusters, shown (G) uninterpreted and (H) interpreted; the same model is shown, corendered with cosφ uninterpreted (I) and interpreted (J). Six primary stratigraphic facies are identified outside channel package A, numbered from oldest (base) to youngest (top).

Stratigraphic Facies 1* (Basal package): The oldest interval is interpreted as basin fill during initial regression, consisting of calcareous claystone with an upward-fining sequence. This

interval forms the base of the studied seismic section and represents interpreted hemipelagic deposition with increasing terrigenous influx.

Stratigraphic Facies 1 (Lower fining package): Transitioning upward, this interval continues the fining-upward trend from siltstone to calcareous claystone-dominated facies. Seismic response shows parallel reflectors bracketing discontinuous, sub-chaotic reflectors. This interval is interpreted as representing continued transgression, with decreasing grain size recorded in the well composite log. The transition to calcareous claystone at the top of the sequence indicates the establishment of hemipelagic sedimentation during advanced transgression.

Stratigraphic Facies 2 (Limestone Marker): A major high-amplitude, continuous limestone interval interpreted as representing maximum flooding, terrigenous sediment delivery cessation, and pelagic carbonate precipitation dominating basin sedimentation. This interval serves as a regional stratigraphic marker and maximum flooding surface correlating to a recorded limestone interval at the wellbore and caps the underlying fining-upward succession, marking the end of this transgressive event.

Stratigraphic Facies 3 (Beginning of regression): Above the limestone marker, this interval shows trace sand at the base transitioning upward to limestone stringers within silty claystone, interpreted as a final pulse of coarser sediment delivery followed by an upward transition to silty claystone. Seismic response is sub-chaotic to sub-parallel.

Stratigraphic Facies 4 (Upper fining package): This interval shows continued upward fining from silty claystone to calcareous claystone-dominated facies, with continuous reflectors in seismic response, and is interpreted as representing continued transgression.

Stratigraphic Facies 5 (gravity flow deposits): Sub-parallel to sub-chaotic seismic character with variable amplitude suggests potential gravity flow deposits, potentially off of the differentially compacted high structure. Unlike the more organized facies below or of the more chaotic MTD above, this interval has some mappable reflectors that are not necessarily continuous, indicating the potential for partially organized downslope flow.

Stratigraphic Facies 6 (Mass transport deposit MTD): The uppermost interval in this view shows highly disrupted, chaotic seismic character interpreted as an MTD. This represents a significant slope failure event, marking a major change in depositional regime at this time.

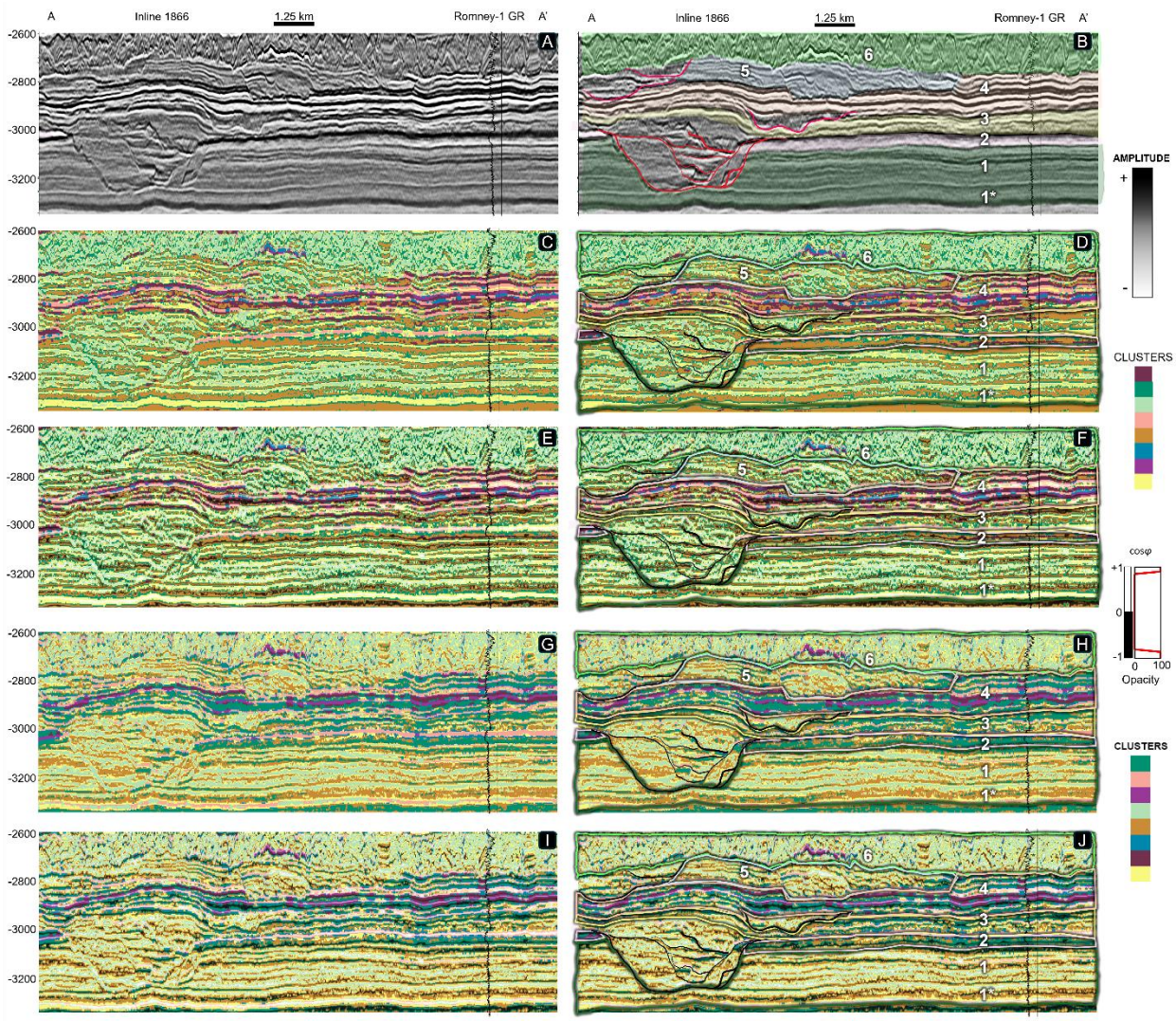


Figure 6: Cross-section A-A' showing seismic facies evaluation. Uninterpreted and interpreted panels are shown for: amplitude(A/B); Post-K-PCA clustering model (8 clusters, 2 components) (C/D); K-PCA 2-component clustering corendered with cosφ (E/F); *Post-K-PCA clustering model (8 clusters, 3 components) (G/H); K-PCA 3-component clustering corendered with cosφ (I/J). Six facies packages are identified from base to top (oldest to youngest). These packages are distinct from the originally picked seismic facies (A-C) and are here interpreted in the context of the broader depositional environment. Facies (1*) marks a basal calcareous claystone that begins an upward fining sequence into Facies (1), characterized by increased siltstone content with sandstone stringers. Facies (2) represents a major limestone marker (transgressive interval), Facies (3) calcareous claystone-dominated with minor sandstone and trace limestone, and Facies (4) silty to calcareous claystone transition. Below the mud log start depth of 2369 m (~ -2887 ms), Facies (5) is interpreted as potential gravity flow deposits, and Facies (6) as a mass transport deposit (MTD). Facies succession reflects transgressive-regressive cyclicity with episodic coarse clastic delivery to the basin floor.*

4.3 Architectural Element Mapping

The practical question this workflow is designed to answer is not which clusters are present, necessarily, but what those clusters mean geologically and how reliably they map depositional architecture. Seismic facies characterization begins with the identification of AEs through amplitude, which is then integrated with clustering model results. Amplitude data (Figure 7 A/A') and bump mapping enhancement (Figure 7 B/B') delineate channel architecture, which is then evaluated against unsupervised cluster distribution from post-K-PCA clustering, with 2 components (Figure 7 C/C') and 3 components (Figure 7 D/D'). Both timeslice and chair view perspectives are provided to ensure a thorough understanding of lateral and vertical facies distribution and channel geometry evolution.

The interpreted panels (Figure 7, E/E'-H/H') reveal multiple channel phases and AEs. Black arrows identify the oldest channel phase in this location, characterized by high amplitude and straight geometry, interpreted as high-energy flow with coarse-grained fill along the channel margins. Orange arrows mark an older meandering channel phase with lower amplitude, interpreted as a mud-filled, lower-energy abandoned channel. Note the distinct cluster association, clusters 3 and 8, in both models. Pink arrows highlight a younger, higher-energy channel phase with a distinct geometric expression from sediment distribution preserved in the clustering models. Yellow arrows mark a feature preserved in the interchannel position, further demonstrating how sediment distribution geometries are preserved in the clustering models. The preservation of these geometries across both attribute and model space demonstrates that the K-PCA clustering workflow maintains geologically significant features rather than introducing artifacts.

Channel complex geometry reflects a multi-phase system with evidence of both lateral migration and channel abandonment, in planview with clear aggradational stacking and terminations viewed in cross-sectional view.

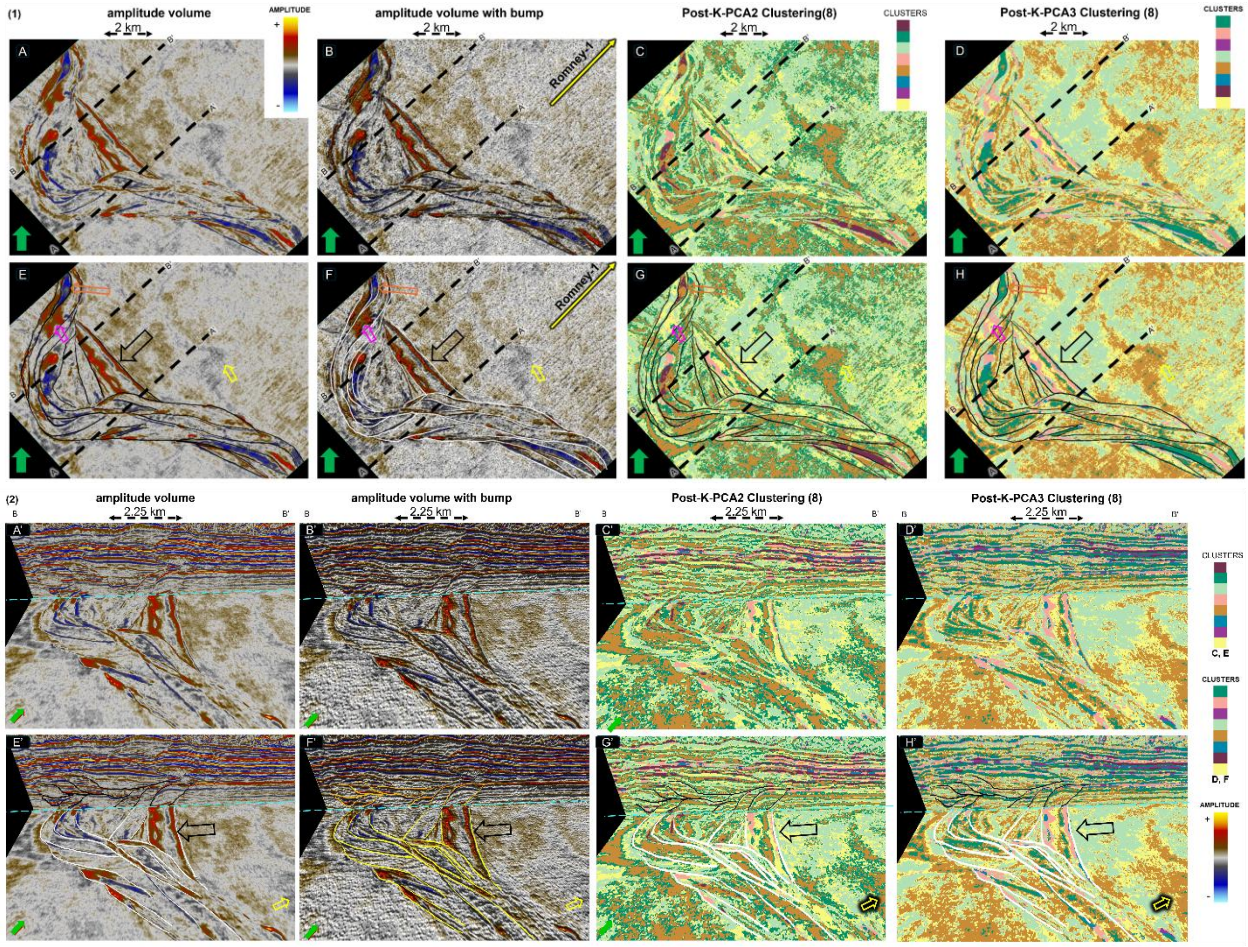


Figure 7: Seismic facies characterization of channel architecture shown in timeslice (1, A-H) and chair view (2, A'-H'). Uninterpreted attributes include amplitude (A/A'), amplitude with bump map (B/B'), K-PCA 2 components with 8 clusters (C/C'), and K-PCA 3 components with 8 clusters (D/D'). Interpreted equivalents are shown in (E/E'), (F/F'), (G/G'), and (H/H'), respectively. The consistency of geometric features across both the attribute space and the model space demonstrates that K-PCA clustering preserves geologic structure rather than introducing artifacts. Black arrows mark the oldest channel phase, characterized by high amplitude and straight geometry, interpreted as high-energy flow with coarse-grained material along channel edges. Orange arrows indicate an older meandering channel with lower amplitude, interpreted as mud-filled with lower-energy flow. Pink arrows highlight a zone of higher-energy flow with preserved sediment distribution shape. Yellow arrows identify preserved sediment distribution within the interchannel area.

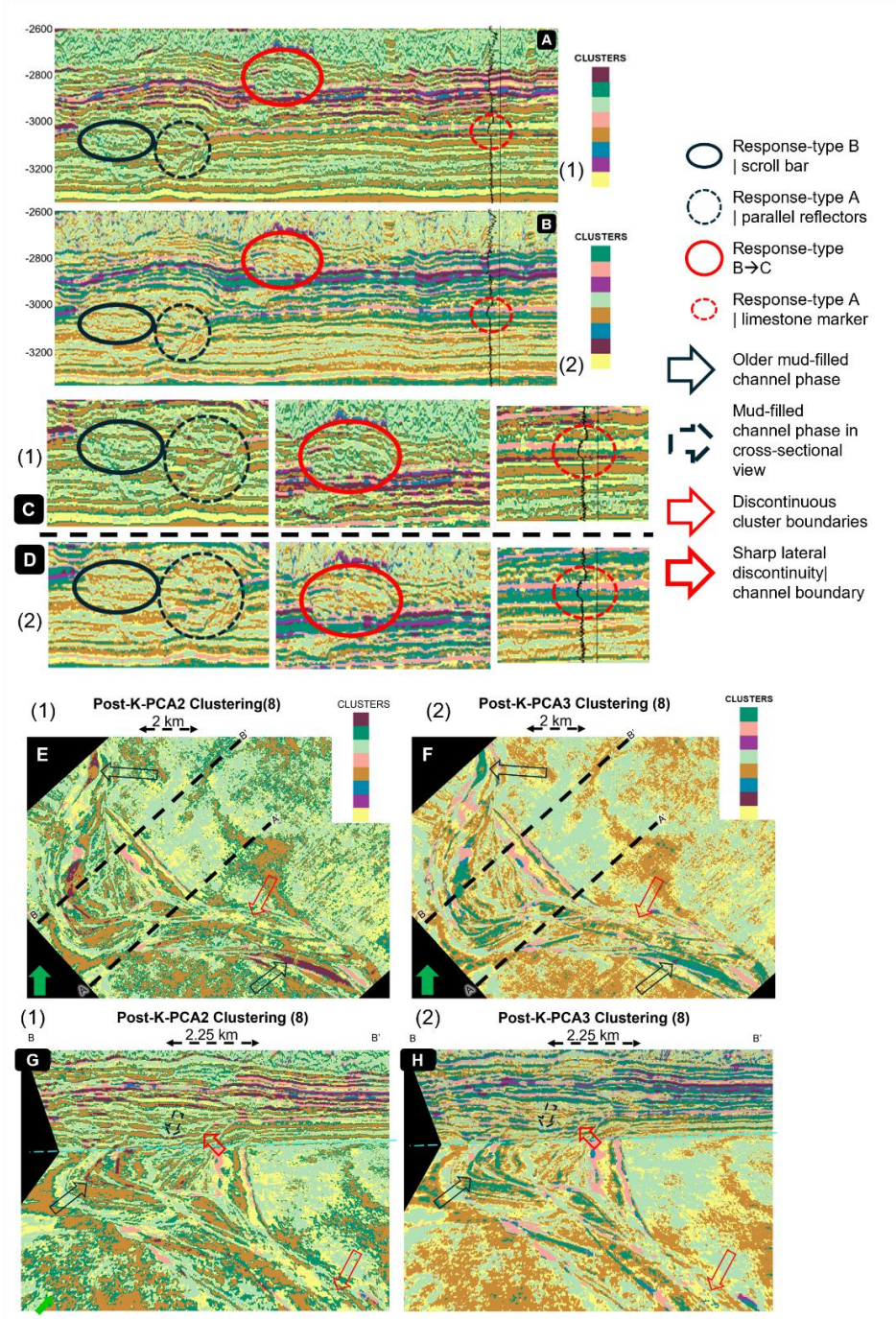


Figure 8: Comparative evaluation of post-K-PCA clustering models with 8 clusters: (1) 2-component model and (2) 3-component model. Panels (A/B) show cross-sectional views of models (1) and (2), respectively, with (C/D) showing zoomed insets of the annotated features in (A/B). Panels (E/F) show timeslice views and (G/H) show chair views of models (1) and (2). Solid black circles highlight scroll bar features associated with response-type B facies; dashed polygons highlight response-type A facies characterized by parallel reflectors. Red circles identify an area above the channel of interest showing response-type B grading to type C at its base. Red dashed polygons highlight the limestone marker at Romney-1, associated with response type A. In Panels (E/F/G/H), black arrows identify an older, low-amplitude, mud-fill channel phase's clustered response; bold red arrows highlight sharp lateral discontinuities that allow for better channel boundary identification; and short red arrows in the cross-section portion of the chair view highlight discontinuous reflectors that define the channel boundary. Overall, model (1) provides

4.4 Direct Model Comparison

Figure 8 presents a direct qualitative comparison of the 2-component (Model 1) and 3-component (Model 2) K-PCA clustering results, both with 8 clusters, evaluated across cross-sectional, time-slice, and chair view perspectives. This comparison is inherently qualitative, relying on the interpreter's domain knowledge to assess which model better preserves and delineates geologically meaningful features.

In the cross-sectional view (Figure 8, A\ B), and their zoomed equivalents (Figure 8, C\ D), four annotated features are evaluated. In the solid black circle, highlighting scroll bar features associated with response-type B facies, Model 1 shows more consistent lateral detail. With crisper cluster boundaries than Model 2. In the dashed polygon, highlighting response-type A facies characterized by parallel reflectors, Model 1 shows a clearer distinction between cluster types and a more continuous response, while Model 2 introduces additional cluster variability that reduces rather than enhances the clarity of the reflector pattern. In the red circle, highlighting an area above the channel of interest where response-type B grades downward into response-type C at its base, Model 1 maintains the same pattern of greater continuity and consistency. In the red-dashed polygon highlighting the limestone marker at Romney-1, associated with the response-type A, both Models show differences in cluster distributions across the limestone package. In Model 1, a pink cluster is largely continuous to the left of the well across the upper portion of the limestone, with an additional cluster (burgundy) emerging toward the base, providing more internal stratigraphic detail despite some lateral discontinuity. In Model 2, a single consistent cluster spans the upper limestone continuously from left to right, but the base of the limestone marker shows limited cluster continuity and less internal differentiation than Model 1.

In the time slice views (Figure 8, E\ F), black arrows identify an older, low-amplitude, mud-filled channel phase’s clustered response, visible in both Models. However, the cluster representing this feature in Model 1 is more spatially focused, distinctly delineating the older channel. Without being distributed broadly across the surrounding area. In Model 2, the equivalent cluster

is more pervasive across the time slice, reducing its diagnostic value for isolating this specific channel element. Red arrows in both panels highlight differences in lateral depositional distributions that affect channel boundary identification; these boundaries are more continuous and easier to map in Model 1 than in Model 2. In the chair views (Figure 8, G\ H), a dashed black arrow on the inline portion highlights a similar older mud-filled channel phase observed in the time slice. In cross-sectional view, Model 1 shows a spatially focused cluster distribution for this feature, while Model 2 shows a higher distribution of the equivalent cluster, again reducing specificity. Bold red arrows highlight sharp lateral discontinuities in the cross-sectional view that delineate an iteration of channel migration, better expressed in Model 1 than Model 2. Short Red Arrows further highlight discontinuous reflectors in the cross-sectional portion that define the channel boundary; this subtle but geologically significant feature is better captured in Model 1, where the cluster response more precisely honors the reflector geometry. On the timeslice portion of the chair view, red arrows again draw attention to the same continuous and discontinuous boundaries, confirming that the channel margins are more consistently mappable in Model 1 than Model 2.

Overall, this qualitative comparative evaluation demonstrates that the 2-component KPCA Model with 8 clusters provides superior delineation of geologically meaningful features across all view perspectives. Model 1 consistently shows greater lateral continuity, crisper cluster boundaries, and a more focused spatial distribution of clusters associated with specific architectural and depositional elements. While Model 2 captures the same broad architectural framework, the additional component introduces cluster variability that in several instances obscures rather than enhances the geologic signal. On the basis of this integrated qualitative assessment, informed by domain knowledge and evaluated across multiple view perspectives, Model 1 is selected as the preferred Model for further interpretation and reservoir characterization. Notably, this geologically preferred model also corresponds to the highest-performing model identified through statistical evaluation, a convergence of quantitative and qualitative assessment that provides additional confidence in the selection and sets the stage for the interpretive discussion that follows.

5 Discussion: Seismic Facies as a foundation for understanding the geologic setting

The seismic facies framework established in section 4, response-type facies calibrated at Romney-1, stratigraphic interval facies defined in cross-section, and AEs mapped in planview, provides the foundation for the following interpretive discussion.

5.1 The Value of Integrated Model Evaluation

Model selection should come from a balanced integration of quantitative statistical metrics and geologic interpretability criteria. While statistical metrics provide a first filter, they cannot guarantee geologic meaningfulness; a statistically well-separated cluster solution may still produce geometrically implausible facies distributions or fail to honor known depositional architecture. This reflects differing optimization objectives rather than a limitation of ML: clustering algorithms optimize for mathematical separation within attribute space (Jain, 2010), whereas geologic interpretation requires process-based reasoning and spatial context that statistical metrics may not capture. The assumption that the highest-performing statistical model will automatically produce the most geologically interpretable results is therefore one that warrants careful scrutiny in any application of unsupervised clustering to seismic facies classification.

The top-performing K-PCA models demonstrated superior statistical performance (silhouette scores and DBI), but final model selection prioritized those that also preserved geologically meaningful features: continuous channel geobodies, stratigraphic consistency, and architecturally plausible facies distributions. In exploration settings with limited control, geologic interpretability becomes essential. A model that clusters effectively but produces geologically implausible geometries provides limited value for appraisal targeting or reservoir framework development.

5.2 The irreplaceable role of domain knowledge and geologic reasoning:

ML offers objective, scalable pattern recognition, but domain knowledge still plays an irreplaceable role in model selection and interpretation. Hard-earned geologic knowledge for understanding the complexity of deepwater channel systems, including the integration of spatial awareness, depositional processes, and stratigraphic context, cannot be quantified statistically but

is critical for evaluating whether clustering results reflect depositional processes or mathematical artifacts. Blind adherence to optimizing metrics risks selecting models that are statistically robust but geologically meaningless. The workflow presented here treats ML as a pattern recognition tool, while geologic reasoning and expertise provide the interpretive framework for model selection, facies labeling, and output evaluation.

Effective qualitative assessment of seismic data requires familiarity with the features being studied, developed through repeated exposure to the data, study of modern analogues, and transfer of knowledge from well-understood subaerial systems. For deepwater channel systems, we apply information from fluvial systems to identify channel AEs, including main channels, splays, levees, slumps, and prograding bars. While we acknowledge that the channel environment may be different between subaerial and submarine conditions, the AE remains largely similar.

Evaluation must also be conducted across multiple view perspectives, inline, crossline, and planview, with interpretations that are consistent regardless of perspective; inconsistencies signal the need for reexamination. For more detail on combining stratigraphy and geomorphology in conjunction with best practices for gaining the most from your seismic data, with proper lighting and color choice, see Posamentier et al. 2022.

Managing interpreter bias

In addition to transfer learning from analogues, we must be cautious of interpreter bias. If we focus too narrowly on a small subset of the data, expecting to see particular features, we will inevitably find patterns that match our preconceived expectations. Infante-Paez and Marfurt (2018) show examples from within the Taranaki Basin illustrating the potential misidentification of a hybrid igneous flow for a channelized turbidite and volcanics for carbonate buildups. Keeping ourselves aware of this risk and broadening our perspective by researching the area's geologic history and examining the features within the larger context can help reduce the potential impact of bias on interpretation.

Practical Recommendations

In practice, statistical metrics such as the silhouette scores and DBI provide a reliable first filter for identifying well-separated, internally consistent cluster solutions. However, practical red flags can emerge at two stages in a workflow like this one. During attribute selection and

clustering, warning signs can include cluster solutions that show no clear separation regardless of parameter tuning, suggesting either inappropriate attribute selection or insufficient geologic variability to resolve distinct facies; clusters dominated by a single attribute response, indicating redundancy rather than meaningful multiattribute discrimination. Attribute dominance can be diagnosed through loading analysis of the DR step, in a companion paper (Moreno-Ward and Bedle, In Review), loading plots revealed that cosφ dominated the first principal component in PCA, which informed the decision to apply kernel transformation to redistribute attribute influence more meaningfully across components before clustering. Where these warning signs emerge, DR before clustering should be considered, as demonstrated in this workflow, K-PCA can resolve structure in attribute space that direct clustering fails to capture. The second and more geologically meaningful stage of evaluation treats the model output as a "seismic attribute": if expected geologic features that should be resolvable at seismic scale are absent, if cluster boundaries reflect artifacts rather than geologic geometry, if features appear in one view perspective but disappear in another, or if the cluster distribution fails to honor well calibration at the borehole, the model warrants reexamination. Throughout this process, the clustering itself remains unsupervised, no labels are provided, and the algorithm determines the data structure independently, but geologic reasoning governs attribute selection on the front end and model evaluation on the back end, ensuring that statistical outputs are assessed against geologic reality rather than accepted uncritically.

5.3 Uncertainties and Path Forward

Several sources of uncertainty constrain this facies framework and require acknowledgement. First, seismic resolution limits (19 m vertical, 24 m lateral) prevent imaging of bed-scale heterogeneity, internal facies variations within AEs, and thin mud drapes that may baffle connectivity. This framework addresses large-scale architectural organization.

Second, distance from well control introduces interpretive uncertainty. Romney-1 calibrates interchannel facies but does not sample channel axis or proximal levee environments for channel system A. Interpretations within the channel fairway integrate seismic geomorphology and analogue frameworks but remain untested by drilling. High amplitude within channels is

interpreted as coarser clastics based on depositional context and geometric association with erosional features; lithology remains unproven.

Third, diagenetic modification of reservoir quality cannot be assessed from seismic data alone. Cementation, compaction, and fluid-rock interaction may significantly alter porosity and permeability distributions. While this workflow cannot definitively address lithofacies, it can help to identify potential reservoir fairways for further testing.

Despite these uncertainties, the seismic data resolve key stratigraphic and architectural relationships that form a robust foundation for reservoir characterization. Above-resolution channel truncations and erosional surfaces are particularly well-expressed in this dataset, driven in part by the effectiveness of cosφ for reflector geometries. In planview, lateral migration and meander loop cutoff geometries are clearly preserved, while cross-sectional views of the same features reveal the vertical relationships between successive channel phases (Figure 7).

5.4 Workflow Transferability

The integrated workflow presented here is transferable in principle to any seismic facies classification problem where multiattribute analysis is warranted, and well control is limited. The core logical framework, deliberate attribute selection, dimension reduction, unsupervised clustering, and integrated two-tier evaluation, balancing statistical performance with geologic interpretability, is not specific to deepwater channel systems. What is specific to this setting is the attribute selection rationale, the facies definitions, the depositional analogue framework, and the domain knowledge required to evaluate cluster outputs meaningfully. These must be rebuilt from the ground up for each new application.

Within deepwater and other channelized systems, the workflow is more directly transferable where similar architectural questions are being asked: identifying channel geometries, mapping lateral migration patterns, and characterizing internal facies distribution at seismic resolution. Above-resolution truncation and erosional surfaces, such as the channel cutoff geometries documented here, represent a class of features most reliably resolved using cosφ or a functionally

equivalent attribute. Transferability is therefore partly resolution-dependent; the workflow cannot recover what the seismic data cannot resolve, and expectations must be calibrated accordingly.

Practical transferability is also constrained by computational resources and project stage. For large seismic volumes with many input attributes, processing time and workstation capacity may limit what is feasible within a given project timeline. More fundamentally, this workflow is best suited to the exploration stage, where the primary objective is developing a systematic understanding of an under-characterized system and where the time investment in a rigorous multiattribute evaluation is justified by the interpretive return. It is not designed, yet, for operational settings where rapid answers are required, where supervised methods with established well control are more appropriate, or where a single well-chosen attribute may answer the immediate question more efficiently. Recognizing this distinction, knowing when ML adds meaningful value over conventional single or multiattribute interpretation, is itself a transferable principle.

The domain knowledge required to apply this workflow does not transfer with it. A practitioner working in a new setting must bring equivalent depth of knowledge about that system's depositional processes, AEs, and expected seismic responses, and must remain cognizant of potential interpreter bias risks outlined in section 5.2.

5.5 Broader Implications

The results presented here reflect a broader shift in geoscience practice toward human-machine collaboration, where ML provides scalable pattern recognition and objectivity, and domain expertise provides process-based reasoning and interpretive judgment. Neither alone is sufficient. As ML adoption in exploration workflows continues to grow, the most transferable skill is not familiarity with any particular algorithm, but the ability to critically evaluate model outputs against geologic reality, a capability that remains fundamentally human.

6.CONCLUSIONS

We present an integrated ML and geological workflow for seismic facies mapping in a deepwater channel system, demonstrating that unsupervised clustering with K-PCA can meaningfully support exploration-stage reservoir characterization when guided by domain knowledge at every stage of the workflow. The methodology combines deliberate multiattribute selection, DR via K-PCA, and a two-tier evaluation framework that balances statistical performance with geologic interpretability, providing a structured and reproducible approach to seismic facies classification in settings with limited well control.

Nine seismic facies were identified through integrated well calibration, amplitude analysis, and clustering model output. Three response-type facies, high amplitude continuous, low amplitude discontinuous sub-chaotic, and low amplitude discontinuous chaotic, delineate distinct depositional environments within the channel system and provide the geological framework for evaluating cluster assignments across the seismic volume. Six stratigraphic interval facies, defined in cross-section along inline 1866, characterize the vertical organization of the section outside channel complex A, documenting a transgressive-regressive cyclicity from basal hemipelagic deposition through a major limestone flooding surface, gravity flow deposition, and a capping mass transport deposit.

AE mapping reveals a multi-phase channel system with evidence of lateral migration, channel abandonment, and aggradational stacking. The clustering model preserves geologically significant geometries across both attribute and model space, including channel axis, channel margin, and interchannel features, demonstrating that the K-PCA workflow maintains depositional structure rather than introducing mathematical artifacts. The consistency of these features across timeslice and chair view perspectives validates the interpretive framework and confirms that the workflow captures real geologic variability rather than mathematical artifacts.

Integrated model selection, combining quantitative metrics with geologic interpretability criteria, was essential to the success of this workflow. Rather than evaluating all candidate models in

equal depth (55 in total), the final stage in the workflow presented here demonstrates a structured approach to model selection: statistical metrics identified the top-performing model as K-PCA with 2 components and 8 clusters, which was then carried forward for full geologic evaluation, with K-PCA with 3 components and 8 clusters included as a comparative evaluation. The proposed workflow treats ML as a pattern recognition tool operating at scale and with objectivity, while geologic reasoning provides the interpretive framework for model selection, facies labeling, and output evaluation. This distinction, unsupervised in algorithm, geologically supervised in application, represents a practical and defensible approach to ML integration in exploration geoscience.

Critically, once this interpretive framework is established, the evaluation of additional candidate models becomes more efficient: the geologist knows what features to look for, what constitutes a meaningful cluster distinction, and what red flags warrant rejection or further evaluation. The workflow, therefore, serves not only as a demonstration of ML-geological integration but also as a reusable framework for systematically and efficiently navigating unsupervised model selection, transferable in principle to any seismic facies identification problem where multiattribute analysis is warranted and well control is limited, provided that attribute selection rationale, facies definitions, and domain knowledge are rebuilt for each new application.

ACKNOWLEDGEMENTS

The authors would like to thank the AASPI consortium for providing the software used for the calculation of attributes. We also thank the New Zealand government for making the Romney data available for research.

REFERENCES

1) Chopra, S., and K. J. Marfurt, 2005, Seismic attributes—A historical perspective: GEOPHYSICS, **70**, no. 5, 3SO-28SO.
2) Chopra, S., and K. J. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization: Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.
3) Dewett, D. T., J. D. Pigott, and K. J. Marfurt, 2021, A review of seismic attribute taxonomies, discussion of their historical use, and presentation of a seismic attribute communication framework using data analysis concepts: Interpretation, **9**, no. 3, B39-B64.
4) Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: Springer.
5) Infante-Paez, L., & Marfurt, K. J. (2018). In-context interpretation: Avoiding pitfalls in misidentification of igneous bodies in seismic data. *Interpretation*, *6*(4), SL29-SL42.
6) Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.
7) King, P. R., and P. H. Robinson, 1988, An Overview of Taranaki Region Geology, New Zealand: Energy Exploration & Exploitation, 6, no. 3, 213-232.
8) Kolla, V., Posamentier, H. W., & Wood, L. J. (2007). Deep-water and fluvial sinuous channels—Characteristics, similarities and dissimilarities, and modes of formation. Marine and Petroleum Geology, 24(6-9), 388-405.
9) Leila, M., I. El-Sheikh, A. Abdelmaksoud, and A. A. Radwan, 2022, Seismic sequence stratigraphy and depositional evolution of the Cretaceous-Paleogene sedimentary successions in the offshore Taranaki Basin, New Zealand: implications for hydrocarbon exploration: Marine Geophysical Research, 43, no. 2, 23.

10) Lutome, Marco Shaban. “Seismic Geomorphology and Evolution of Mid-Late Miocene Deepwater Channels Offshore Taranaki Basin, New Zealand.” *Marine and petroleum geology* 171 (2025): 107202.

11) Mayall, M., Jones, E., & Casey, M. (2006). Turbidite channel reservoirs—Key elements in facies prediction and effective development. *Marine and Petroleum Geology*, *23*(8), 821-841.

12) Moreno-Ward, A., Vera-Arroyo, A., & Bedle, H. (2025). Improving interpretability: The role of dimension reduction in multiattribute seismic facies identification. *Interpretation*, *13*(3), T475-T496.

13) Moreno-Ward, A. & Bedle, H. (In Review). Enhancing multiattribute seismic facies analysis: The effects of dimension reduction on unsupervised clustering. *Interpretation*

14) Moreno-Ward, A., Bedle, H., Lubo-Robles, D., and Marfurt, K., (In Review). Beyond Amplitude Alone: A Corendering Technique for Enhanced Deepwater Channel Visualization – A Tutorial Part 1: Systematic Optimization Using Cosine Instantaneous Phase and Relative Acoustic Impedance. *Interpretation*

15) Moreno-Ward, A., Bedle, H., Lubo-Robles, D., and Marfurt, K., (In Review). Beyond Visualization Alone: A Systematic Interpretation Tutorial for Deepwater Channel Stratigraphy Identification Part 2: Framework Application and Romney 3D Case Study. *Interpretation*

16) Pemberton, E. A., Stright, L., Fletcher, S., & Hubbard, S. M. (2018). The influence of stratigraphic architecture on seismic response: Reflectivity modeling of outcropping deepwater channel units. *Interpretation*, *6*(3), T783-T808.

17) Pettingill, H. S., & Weimer, P. (2001). World-wide deep water exploration and production: Past, present and future.

18) Posamentier, H. W., & Kolla, V. (2003). Seismic geomorphology and stratigraphy of depositional elements in deep-water settings. *Journal of sedimentary research*, *73*(3), 367-388.

19) Posamentier, H. W., Paumard, V., & Lang, S. C. (2022). Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data. *Earth-Science Reviews*, *228*, 103963.

20) Strogen, D. P., K. E. Higgs, A. G. Griffin, and H. E. Morgans, 2019, Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression: Geological Magazine, 156, no. 10, 1751-1770.

21) Wang, G., Wu, W., Li, Q., Liu, W., Zhou, Y., Liang, S., & Sui, Y. (2025). Seismic geomorphology and reservoir conditions of a Middle Miocene submarine channel system in the Taranaki basin, New Zealand. *Marine and Petroleum Geology*, *172*, 107219.

22) Wynn, R. B., Cronin, B. T., & Peakall, J. (2007). Sinuous deep-water channels: Genesis, geometry and architecture. *Marine and Petroleum Geology*, *24*(6-9), 341-387.

Chapter 7: Conclusion

The central research question that guided this work was: How can we better incorporate ML/AI techniques into seismic interpretation workflows to improve interpretation resolution and increase confidence in geologic models? Embedded within that question are deeper, more fundamental thoughts. Why should we implement ML in the first place? What are the appropriate inputs? How do we evaluate whether the outputs are geologically meaningful? And critically, how do we define the facies that are the main outputs from the algorithm itself when the algorithm doesn't know anything about geology?

Answering these questions requires confronting an uncomfortable reality: in many ML applications to seismic interpretation, the wrong problem may be what is solved. Instead of "which algorithm performs best?" maybe the question should be "does this algorithm address my geologic question". Too often what gets optimized are statistical metrics rather than geologic validity. Other times, the model outputs are treated as the answer rather than new data that require interpretation. If we stop our critical evaluation at the algorithm output, accepting cluster models as facies maps without geologic validation, then the gap between a well-performing ML model and a defensible interpretation is exactly where the application of ML in the interpretation workflow fails. As ML techniques are rapidly adopted across geoscience workflows, the more critical question becomes not how to implement ML, but *when, why*, and *whether* it should be implemented at all.

This dissertation addresses that gap by developing and demonstrating a practical framework for integrating unsupervised ML into seismic facies interpretation within a geologically informed context. The research demonstrates that ML can enhance seismic facies interpretation, but only when implemented within a rigorous geologic framework that treats ML as a tool, not a geologist, and that prioritize interpretive integrity over computational convenience.

Key findings:

Dimensionality Reduction Improves Interpretability

Chapter 2 establishes that DR techniques genuinely improve the visual delineation of deepwater channel AE in multiattribute analysis. DR is not simply a computational tool for reducing data complexity, it is also an interpretive tool that can reveal geologic patterns obscured in high-dimensional attribute space. By systematically reducing the dimensionality of the input data, interpreters can more clearly visualize the relationships between seismic attributes and geologic features such as channel axes, margins, levees, and mass transport deposits. This finding is significant because it shows that DR has value beyond its role in preparing data for clustering algorithms; it enhances human interpretation directly.

Dimensionality Reduction Improves Clustering performance:

Chapter 3 evaluated the effects of DR on unsupervised clustering by comparing 55 different model configurations using k-means ++ clustering pre- and post-DR. The results were unequivocal: post-DR clustering consistently outperformed pre-DR clustering across multiple quantitative (statistical) evaluation metrics including silhouette score and Davies-Bouldin Index. This finding confirms that applying DR prior to clustering produces measurably better statistical models.

However, Chapter 3 sets the framework for the next question which is more nuanced and ultimately more important. Does the best-performing statistical model represent the model that is geologically most meaningful? Quantitative metrics can tell us which model is most internally consistent, or which produces the most cohesive and compact clusters, but they cannot tell us directly whether those clusters correspond to meaningful geologic distinctions. A model might achieve excellent silhouette scores by grouping samples with similar attribute characteristics; however, those clusters may span multiple depositional facies and have non-unique lithologies

for example. This distinction between a statistically robust and a geologically robust model is central to the entire dissertation.

Geologic Foundation Must precede ML evaluation:

Chapters 4 and 5 took a deliberate step back from the ML workflow to address a foundational question: before we can evaluate the geologic meaning of a ML model, we must first establish the geologic benchmark. Chapter 4 demonstrated how improving seismic data visualization through the corendering of cosine instantaneous phase and relative acoustic impedance significantly enhanced the interpreter's ability to delineate stratigraphic boundaries and AEs within deepwater systems. This systematic optimization workflow, involving attribute selection, color bar design, interpolation, and spectral balancing, moves interpretation beyond amplitude-based analysis alone and provides a more robust foundation for understanding channel architecture.

Chapter 5 extended this work into a comprehensive interpretation framework for defining and identifying deepwater stratigraphy. The critical insight from this chapter is that "stratigraphy" with a channel system reflects the complex, often chaotic movement and evolution of the channel system itself. Seismic reflection patterns are directly tied to the migration, aggradation, and abandonment phases of channel development. Understanding what these patterns mean in terms of depositional processes, and being able to relate them to specific stratigraphic positions within the channel architecture is essential to the interpretation process.

This critical thinking and checkpoint approach to geologic interpretation is what allows the interpreter to evaluate the ML outputs and provide defensible models. Without it, there is no basis for determining whether a cluster represents a mathematical artifact or a geologically meaningful feature. A firm understanding of the depositional system, what processes shaped it, how those processes are recorded in seismic reflection data, and what AEs should be present, is the interpretive baseline against which all ML models must be measured.

Answering the Research Question:

The findings from these chapters converge on a single, coherent answer: ML can be effectively incorporated into seismic interpretation workflows to improve resolution and confidence, but only when it is implemented within a framework that balances quantitative rigor with qualitative geologic reasoning.

That framework consists of several critical components:

Deliberate attribute selection. Attributes must be chosen for their ability to capture geologically meaningful variations, not simple because they are available or our default preference. The interpreter must understand what each attribute measures and how it relates to the geologic feature of interest.

Strategic application of dimension reduction techniques. DR improves both visual interpretation and clustering performance, but the choice of DR technique matters. Different algorithms make different assumptions about the data structure, and those assumptions must align with the geologic problem being addressed.

Evaluation using both quantitative and qualitative criteria. Statistical metrics provide valuable insight into model performance, but they cannot replace geologic evaluation. The interpreter must assess whether clustered data aligns with geologic meaning, whether facies definitions are geologically defensible, and whether the model enhances or obscures interpretive understanding.

Grounding in depositional process understanding. Before ML can be applied effectively, the interpreter must develop a robust understanding of the depositional system, its process, architecture, and seismic expression. This understanding provides the context necessary to evaluate whether ML outputs are capturing geology or noise.

Treating ML as a tool, not an answer. The algorithm partitions data based on similarity in attribute space. It does not know geology. The interpreter must still provide the geologic

context and meaning, defining facies based on depositional processes rather than cluster membership, and being willing to reject or modify ML outputs when they conflict with geologic constraints.

Implications for Practice

The practical implications of this work extend beyond the specific case study presented here. The framework developed in this dissertation is transferable to other geologic locations. It provides practitioners with a systematic approach to; 1) assess whether ML is appropriate for their specific interpretive challenge; 2) select and prepare appropriate input data; 3) choose and configure ML algorithms with geologic intent; 4) evaluate outputs using defensible geologic criteria; 5) and integrate ML results into traditional interpretation workflows without sacrificing rigor.

In an industry where time pressure often drives rapid adoption of new technologies, this framework offers a counterbalance, a methodology for slowing down enough to implement ML thoughtfully rather than reflexively or to ask the real question, “Is ML necessary right now” understanding that the process will take some time. The framework acknowledges industry realities while maintaining that computational efficiency should never come at the expense of geologic defensibility.

Limitations and future work

This dissertation establishes the framework for evaluating ML-derived facies models but stops short of completing the full geologic interpretation of the Taranaki Basin channel system. Developing and validating a novel ML algorithm for seismic interpretation workflow is iterative, and the systematic interpretive framework established here through traditional seismic analysis serves a deliberate purpose. It defines the geologic benchmark against which ML-derived clustering results will be evaluated. The goal is to demonstrate the full intellectual work required to responsibly vet a novel ML algorithm for the interpretation workflow, making the logic

transparent so that practitioners can apply informed judgment rather than simply accepting algorithmic outputs at face value. The next phase of this research involved applying the evaluation framework discussed in Chapter 5 to systematically compare 2 candidate models, specifically, evaluating whether (model 1) KPCA with 2 components and 8 clusters produces geologically different and more defensible results than (model 2) KPCA with 3 components and 8 clusters. This comparison will test the central premise that the statistically optimal model may not be the geologically optimal model and will demonstrate how the framework guides model selection in practice.

What was found in Chapter 6 is the model ranked as the highest statistical performer, model 1, when compared to model 2 had optimal qualitative result for the identification of channel boundaries and features. However, there are 53 other statistical models that can be reviewed in depth for interpretability. Given the shear number of models, using the statistical metrics provides a practical places to focus attention when looking at moving into the interpretation phase.

While this work focused on unsupervised clustering for facies identification, the framework could be extended to other ML applications in seismic interpretation. The core principles, understanding the geology first, how to critically evaluate the outputs from a geologic perspective, and treating the algorithm as a tool, remain relevant across these applications.

Final reflections

The algorithm doesn't know geology. It never will. But that limitation is not a failing of ML; it is simply the nature of what these tools are designed to do. ML algorithms are extraordinarily powerful pattern recognition engines, capable of identifying relationships in high-dimensional data that would be impossible for humans to discern through visual inspection alone. When properly applied, they can enhance our ability to see and interpret geologic features in seismic data.

The keyword is "properly". Proper application requires more than technical competency with the algorithms. It requires geologic expertise, interpretive discipline, and the intellectual

honesty to recognize when a statistically impressive model is geologically meaningless. It requires building the geologic foundation first, understanding the depositional system, optimizing the seismic data, and establishing context-specific meaning before ever asking an algorithm to partition the data.

This dissertation provides a foundation and framework for how this practitioner has approached this process. The value of this work lies not in proving that ML works for seismic interpretation, that was never really in doubt, but in demonstrating how to make it applicable in a way that geoscientists can trust, defend, and build up. The methodology presented here treats ML neither with uncritical enthusiasm nor with reflexive skepticism, but with the deliberate, critical engagement that any powerful tool deserves.

In the end, the best ML workflow is one that makes the interpreter better, not one that seeks to replace them. This dissertation offers a path toward that goal.

Appendix 1: Enhancing seismic delineation of obscured geologic architectural elements in a deepwater channel complex through multi-attribute analysis: A study from the Taranaki Basin*

*This paper is published in In International Meeting for Applied Geoscience & Energy Expanded Abstracts as

Moreno-Ward, A., Bedle, H., & Vera-Arroyo, A. (2024, August). Enhancing seismic delineation of obscured geologic architectural elements in a deepwater channel complex through multi-attribute analysis: A study from the Taranaki Basin. In SEG International Exposition and Annual Meeting (pp. SEG-2024). SEG.

Summary:

The integration of Machine Learning (ML) into 3D seismic interpretation is advancing and aims to enhance the analysis of complex geological structures through integrated methodologies for multi-attribute analysis and high-resolution interpretation. Our new approach goes beyond previous synthetic data studies building on the work of La Marca and Bedle (2022). The goal is ambitious: to improve the identification of intricate channel features in depositional environments. By focusing on the careful selection of attributes and applying dimensional reduction (DR) techniques, we aim to significantly enhance the visualization of geological features. This approach assesses the practical applicability of ML in real-world seismic interpretation. Beyond the conventional application, this project tackles the challenge of employing DR to provide clearer visualizations that highlight channel signatures. This enhanced clarity facilitates the precise delineation of architectural elements (AE), setting the stage for additional classification techniques such as K-means clustering to categorize channel facies. This methodological advancement underscores ML's potential not just for statistical data enhancement but as a tool for gaining geological insights, thereby expanding its utility in revealing the nuanced features of channel complexes and improving the characterization of depositional environments for better mapping of reservoir extent and connectivity.

Geologic Location:

The Taranaki Basin, a Mesozoic rift basin located off the northwest coast of New Zealand's North Island, has been shaped by complex processes of extension, subsidence, and compression from the Late Cretaceous to the Neogene period. Spanning approximately 100,000 km² with sediment layers up to 10 km thick (Grahame, 2015), it presents unique opportunities and challenges for geological interpretation. The deepwater channel imaged in the Romney 3D seismic volume is a complex, nested channel with well-preserved channel migration architecture. There are several areas where channel movement timing can be inferred; however, because of the active progression of the channel during deposition, AE identification is convoluted.

Motivation:

Interpreting seismic facies in this complex deepwater channel system poses significant challenges; sparse well control and intricate morphology. These difficulties are amplified when interpreting channel architecture and associated lithology indicators, critical for applications like permeability variation studies within reservoirs. Detailed knowledge of reservoir characteristics can significantly impact efficiency and environmental sustainability, making such variations pivotal for optimizing strategies in fields like hydrocarbon extraction and carbon capture. Addressing these challenges, this project employs a combination of seismic attributes and advanced statistical and algorithmic techniques to enhance the geological interpretation process. Central to our method is the use of DR algorithms designed to improve visualization and interpretation of features without losing geological integrity. A systematic evaluation of each algorithm's effectiveness — particularly preservation of geologic meaning and feature enhancement—is a crucial preliminary step. Proper evaluation is critical for selection of appropriate DR techniques and sets the stage for unsupervised clustering. Through this workflow, we aim to better navigate the complexities of seismic data interpretation, directly supporting efforts to accurately understand reservoir characteristics.

Methods:

Data Preparation and Seismic Attribute Computation: Began with the identification of the area of interest (AOI) and calculating key seismic attributes, as shown in Figure 1. These attributes, detailed in Table 1, effectively outline channel edges, AE, and lithology indicators (Figures 1 & 2).

Algorithmic Analysis and Implementation: The project transitioned to statistical and algorithmic analysis, evaluating several algorithms, including ICA, PCA, K-PCA, UMAP, and K-Means.

Data Preprocessing: Generated attributes were concatenated into the project data frame and Min-Max scaling was applied to normalize the data range from 0 to 1. This ensured a balanced analysis across attributes of varying scales.

Dimensionality Reduction and Visualization: Our focus shifted to refining complex, high-dimensional seismic data for better interpretation. Early analysis of unsupervised kmeans clustering showed significant cluster overlap, which was greatly minimized by applying DR techniques before clustering.

Independent component analysis (ICA): is a computational method for separating a multivariate signal into additive, independent non-Gaussian signals (non-orthogonal). ICA is used to identify and separate superimposed geological features and is particularly effective in distinguishing overlapping seismic signals, enabling us to isolate distinct geological phenomena that might be hidden in the raw data.

Principal Component Analysis (PCA): is used to emphasize variation and capture strong patterns through identification of principal components (orthogonal) with maximal variation, thereby identifying the most significant patterns and trends. It simultaneously simplifies the data and retains those characteristics that contribute most to its variance contributing to a clearer understanding of the underlying geological structures. Several studies have been previously conducted focusing in more detail on the nuances of ICA and PCA. For more in depth

discussions, see Lubo-Robles 2019; Chopra 2014; Honorio et al. 2014; Hyvarinen 2000; Roden et al. 2015; and Zhao et al. 2015.

Kernel PCA (K-PCA): While PCA is adept at uncovering linear relationships, K-PCA extends this capability to nonlinear relationships in the data. It uses kernel functions to project the data into a higher-dimensional space where nonlinear relationships can be linearized. This is particularly valuable in seismic analysis, as geological formations often exhibit complex, nonlinear relationships.

Uniform Manifold Approximation and Projection (UMAP): is used for improved visualization and excels in preserving both the local and global structure of high-dimensional data. It can effectively display a broad overview of the data's structure while maintaining detailed relationships between closer data points. In seismic analysis, UMAP provided a comprehensive view of the data's structure, enhancing the ability to interpret complex geological features.

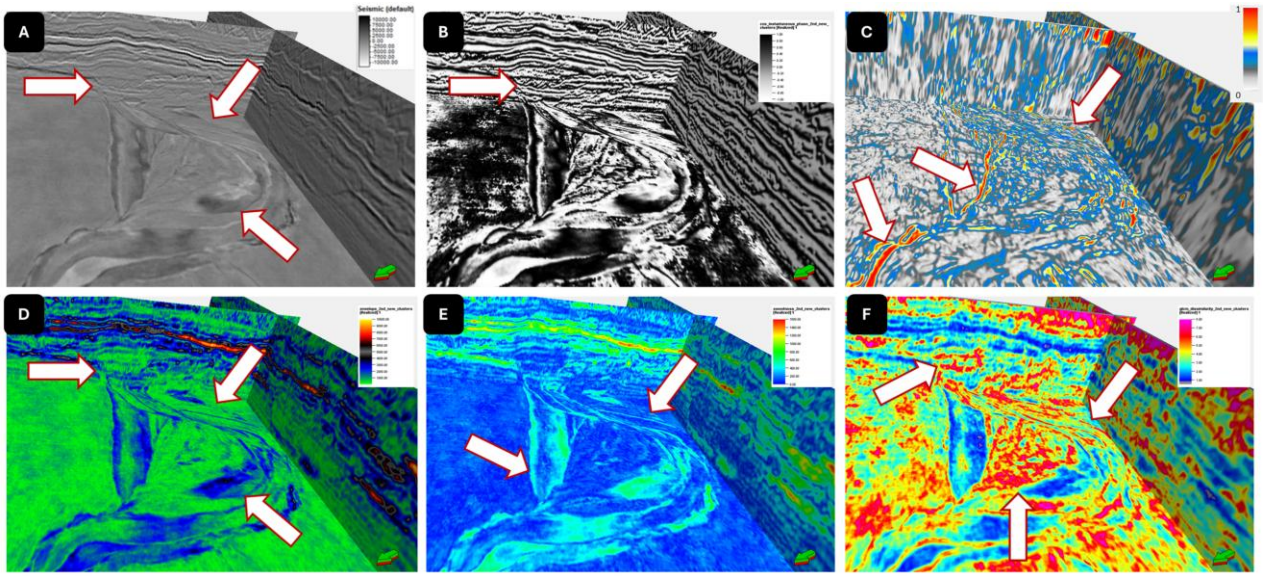


Figure 2: AOI illustrating attribute efficacy at delineating channel features in cross-sectional view. Channel migration delineated in black outlines (red for Cosine Phase); yellow arrows indicating additional in channel structure at smaller scale and pink arrow delineating basal edge features. A amplitude B cosine instantaneous phase C k-curvedness overlain onto COS Phase D Envelope E Sweetness F GLCM Dissimilarity

Model Selection and Visualization:

Two key methods were utilized to evaluate the performance of each iteration of KMeans clustering. The elbow method and the silhouette score are pivotal in comparing overall accuracy of the models, however, final determination of the best number of clusters is up to the interpreter. Each DR model is rendered in seismic software to determine preservation and improvement in visualizing geologic features.

K-PCA and UMAP are designed to handle more complex non -linear data. Given the complexities of channel systems, these algorithms are particularly valuable. They excel in

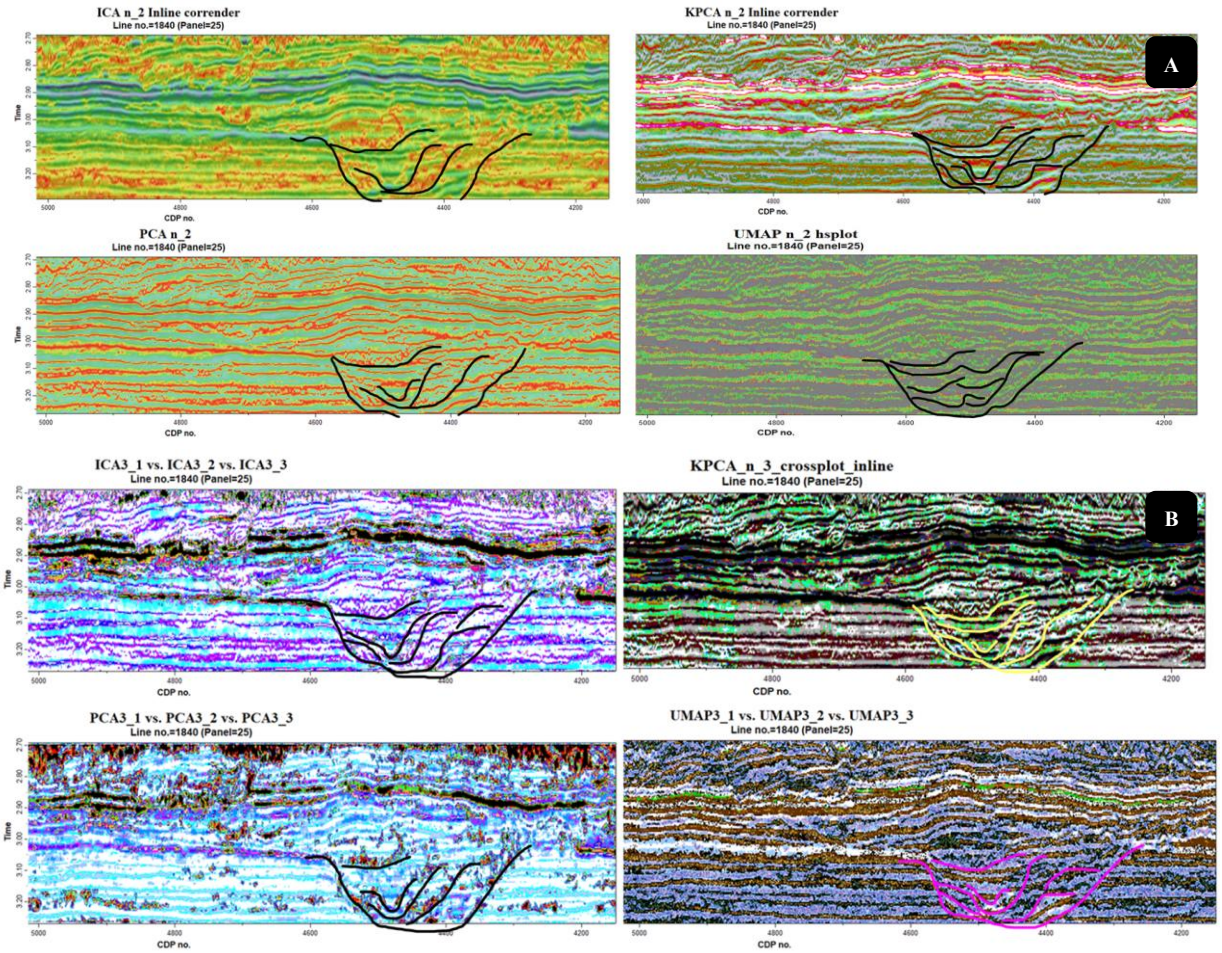


Figure 3: Evaluation of performance for dimensional reduction algorithms, a) 2 components ; K-PCA delineates more internal channel architecture. Visual comparison of architectural features highlighted with selected attributes compared to K-PCA performance shows improvement. b) 3 components; K-PCA again performs well at delineating internal channel architecture, however, UMAP performance is now improved.

transforming complex seismic data into a new feature space that offers a more interpretable representation of the data.

Figure 3 shows the channel inline for each of the DR models. K-PCA and UMAP capture more internal structure within the channel complex across all models. Understanding channel migration, timing, and fill are fundamental to providing a robust geologic interpretation. Calculated seismic attributes aid in highlighting important changes in the depositional environment, however, DR models show even more internal features.

By employing dimensionality reduction and visualization techniques, these data were transformed from the complex, high-dimensional seismic data into manageable and interpretable forms. Each model is then analyzed using KMeans clustering to determine which reduction method leads to a more accurate overall clustering of data.

Overall, UMAP had an improved performance on clustering for both 2 components and 3 components, performing better in reducing cluster size and minimizing data overlap. For more in-depth discussion of these algorithms see, Mika et al. 1998; Van der Maaten 2008; Wattenberg et al. 2016; and McInnes 2018.

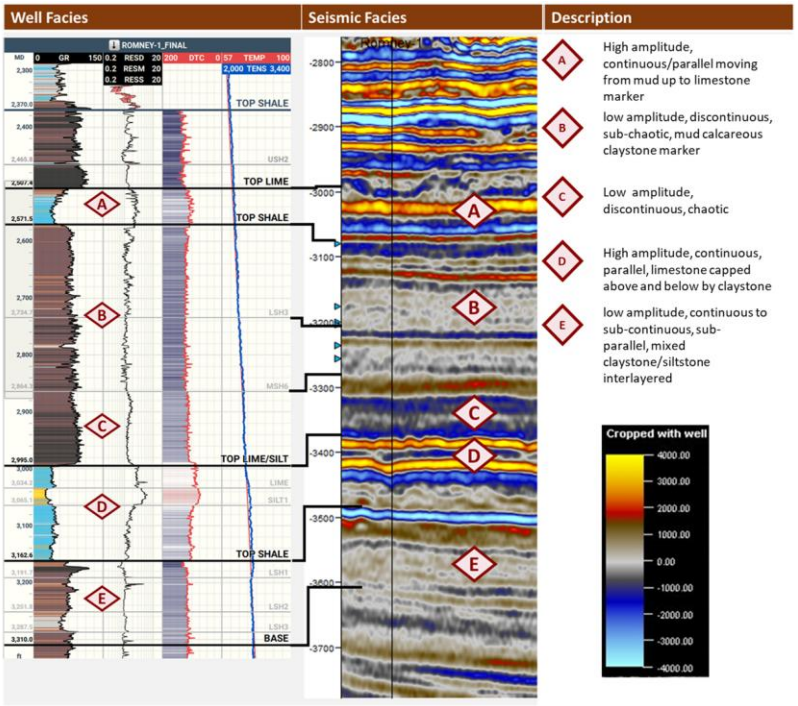


Figure 4, above: Well data and seismic evaluation to determine well-facies and seismic-facies.

Facies Interpretation: Using available well logs for the nearby Romney-1, and seismic data, the interpreter can begin defining facies. Figure 4 illustrates the selection of both well facies and corresponding seismic facies. Well facies were determined by identifying unique well response packages and cross-checked with the available mud logs. The packages are identified on the seismic data and seismic response was recorded. These facies and responses are then interpolated into the channel, and DR facies identified. This same procedure is carried forward into the cluster models.

Figure 5 illustrates interpolated facies into the channel; when cross-checked with amplitude, one can assign the higher amplitudes within the channel to coarser grain material (sands) and the lower amplitudes to finer grain material (silts and muds). Comparing this to the calculated lithology log, there is a general correlation.

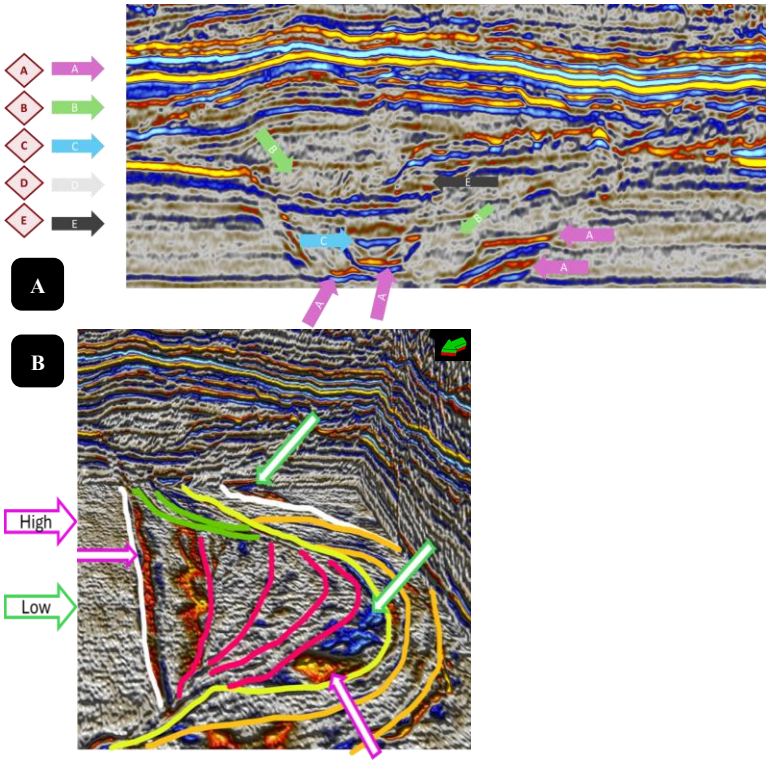


Figure 5: Extrapolating well and seismic facies into the channel for lithology indicators. b) Using a combined x-line, inline, and z slice view, lithology indicators can be applied. Channel migration indicator outlines can begin as a starting stage to overall geologic interpretation.

Conclusion:

Utilizing DR algorithms like K-PCA and UMAP has effectively made complex seismic data more interpretable, enhancing the identification of channel features and AE. This advancement highlights ML's capacity to offer deep geological insights and improve reservoir mapping accuracy—key for understanding reservoir dynamics. Our integrated approach, merging seismic analysis with advanced ML and thorough evaluation, streamlined seismic interpretation, enabling more precise geological insights and informed interpretations**.**

References:

Barnes, A. E., 2007, A tutorial on complex seismic trace analysis: Geophysics, 72, no. 6, W33–W43; https://doi-org.ezproxy.lib.ou.edu/10.1190/1.2785048.

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

Chopra, S., and Marfurt, K. J., 2014, Churning seismic attributes with principal component analysis: SEG Expanded Abstract, 2672-2676; https://doi.org/10.1190/segam2014-0235.1

Grahame, J.L., 2015, Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity.

Hart, B. S., 2008, Channel detection in 3-D seismic data using sweetness: AAPG Bulletin, 92, no. 6, 733-742. https://doi.org/10.1306/02050807127

Honorio, B. C. Z., Crus Sanchetta, A., Pereira Leite, E., and Campane Vidal, A., 2014, Independent component spectral analysis: Interpretation, 2, SA21-SA29; https://doi.org/10.1190/INT-2013-0074.1

Hyvärinen, A., and Oja, E., 2000, Independent component analysis: Algorithms and applications: Neural Networks, 13, 411–430; https://doi.org/10.1016/S0893-6080(00)00026-5

La Marca, Karelia, and Bedle, Heather, 2022, Deepwater Seismic Facies and Architectural Element Interpretation Aided with Unsupervised Machine Learning Techniques; Taranaki Basin, New Zealand: Marine and Petroleum Geology, 136. https://doi.org/10.1016/j.marpetgeo.2021.105427

Li, Q., Yu, S., Wu, W., Tong, L., and Kang, H., 2017, Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: A case study from the Taranaki Basin, New Zealand: New Zealand Journal of Geology and Geophysics, 60, no. 3, 199-208. https://doi.org/10.1080/00288306.2017.1307230

Lubo-Robles, D., and Marfurt, K. J., 2019, Independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: Interpretation, 7, SE19-SE42; https://doi.org/10.1190/INT-2018-0109.1

McInnes, L., Healy, J., and Melville, J., 2018, Umap: Uniform manifold approximation and projection for dimension reduction: arXiv preprint arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426

Mika, S., Schölkopf, B., Smola, A., Müller, K. R., Scholz, M., and Rätsch, G., 1998, Kernel PCA and de-noising in feature spaces: Advances in Neural Information Processing Systems, 11.

Qi, J., Lin, T., Zhao, T., Li, F., and Marfurt, K., 2016, Semisupervised multiattribute seismic facies analysis: Interpretation, 4, no. 1, SB91-SB106. https://doi.org/10.1190/INT-2015-0098.1

Roden, R., Smith, T., and Sacrey, D., 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, 3, no. 4, SAE59–SAE83; https://doi.org/10.1190/INT-2015-0037.1

Subrahmanyam, D., and Rao, P. H., 2008, Seismic attributes-a review: 7th International Conference and Exposition on Petroleum Geophysics, Hyderabad, India.

Van der Maaten, L., and Hinton, G., 2008, Visualizing data using t-SNE: Journal of Machine Learning Research, 9, no. 11.

Wattenberg, M., Viégas, F., and Johnson, I., 2016, How to use t-SNE effectively: Distill, 1, no. 10, e2.

Yenugu, M., Marfurt, K. J., and Matson, S., 2010, Seismic texture analysis for reservoir prediction and characterization: The Leading Edge, 29, no. 9, 1116-1121. https://doi.org/10.1190/1.3485772

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

Appendix 2: Finding the balance: Human insight vs. machine learning techniques to optimize deepwater channel facies identification*

*This paper is published in In *International Meeting for Applied Geoscience & Energy Expanded Abstracts* as

Moreno-Ward, A., & Bedle, H. (2026, February). Finding the balance: Human insight vs. machine learning techniques to optimize deepwater channel facies identification. In International Meeting for Applied Geoscience & Energy Expanded Abstracts (Vol. 44, No. 1, pp. 1420-1424). Society of Exploration Geophysicists.

Summary

To improve the interpretability of deepwater channel seismic facies, we must address several key challenges. These challenges can originate from data constraints – reduced 3D seismic coverage or resolution limitations – limited well control, or simply the inherently complex nature of channel systems. Large-scale channel features may seem relatively straightforward to identify, but delineating subtle, smaller-scale elements requires more sophisticated and time-consuming approaches. By leveraging seismic attributes in conjunction with advanced statistical and algorithmic techniques, we can enhance and refine the geologic interpretation process. However, success lies in maintaining a careful balance between traditional geologic expertise and quantitative methods, ensuring we neither become overly reliant on machine learning (ML) techniques nor allow potential interpreter bias to influence our conclusions. With all these things in mind, we explore how we can better incorporate ML methods to seismic interpretation workflows, and whether these methods can increase confidence in our geologic models.

Rationale for Machine Learning:

Including ML algorithms, such as dimension reduction (DR) and clustering techniques, into our 3D seismic interpretation workflows can improve visual interpretation of deepwater channels (Moreno-Ward et al. 2025). Where there is limited well control available, seismic attributes become increasingly important for creating geologic interpretations.

Interpreters must either view each attribute individually or up to three attributes can be co-rendered simultaneously. Colormaps must be thoughtfully chosen to ensure features can be seen. Choosing colormaps for co-rendering becomes even more complicated as features can become obscured; choosing proper colors takes time and practice. To help overcome these limitations, we choose to implement ML/DR techniques to increase the number of attributes evaluated, the speed of analysis, and enhance the delineation of deepwater channel features.

Geologic Location:

The deepwater channel of study is located off the NW coast of New Zealand, Taranaki Basin. The channel actively migrated leaving behind complex architectural elements (AE) whose larger features can be easily mapped, however, there are more subtle and convoluted features that are better visualized through more advanced workflows. For more detailed geologic information, see: Bull et al. 2020; Di et al. 2020; Grahame 2015; Kamp et al. 2004; King and Thrasher 1992, King 1988; Kroeger et al. 2019; 108 Naewboonnien 2021; Quan et al. 2017; Reilly et al. 2015; Strogen et al. 2017; Strogen et al. 2019.

Workflow:

There are several steps to our workflow, as seen in Figure 2. Moreno-Ward et al. 2025 illustrate how DR improves the interpretation resolution for this dataset (Step 2 of the workflow). This improved resolution should in turn improve the unsupervised clustering results when clustering metrics are reviewed. In Step 3, the number of clusters is validated, and results are reviewed both pre– and post– DR. To ensure a balanced evaluation, results are reviewed using both quantitative and qualitative metrics.

Where Do We Start:

With so many variables and techniques to consider, a natural question for someone new to the implementation of ML and DR may wonder where to begin. We begin by reviewing our seismic data and any available wells. Next, the area of interest should be chosen and well-ties completed. Attribute selection and ML techniques should also be done with the geologic features and goals in mind; not all attributes and ML/DR methods will provide optimal results for all datasets.

When choosing to cluster data, we must determine the best number of clusters and carefully consider which method of analysis is best to validate the workflow. Do we rely solely on quantitative results or do we instead rely just on geologic experience. We should aim to find a balanced approach that extends to data visualization. The ultimate goal is to properly weigh available information to optimize data resolution, enhance interpretability, and increase confidence in our interpretations.

Importance of Attribute Selection:

Attribute selection is a key first step and should be determined by the feature of interest. For our deepwater channel we selected Cosine Instantaneous Phase (CSP), GLCM Dissimilarity, sweetness, envelope, and K1– and K2- curvature. These were selected for their ability to provide information regarding the channel boundaries and indications of the lithologic characteristics. Figure 4 illustrates the difference in DR resolution for PCA, Kernal-PCA (K-PCA), and uniform manifold approximation and projection (UMAP) with all attributes used on the left, and without CSP on the right. We see a distinct difference in the ability of the algorithms to capture the internal channel structure based on one attribute choice alone. See Moreno-Ward and Bedle 2025 for more in depth information regarding CSP for channel feature identification.

Understanding our Algorithms

Having a fundamental understanding of how each algorithm works and how they weigh the input data is an important part in ensuring confidence in our interpretations. PCA focuses on preserving the variance with the data with higher variance indicating more information is being preserved. The loadings of the data also can be reviewed to see how much influence each attribute has.

K-PCA looks at the percent contribution and the component magnitude where the higher the % indicates the attribute has a higher importance to the output and the magnitude provides relative indication of the strength of the input attribute. UMAP looks at the feature importance (FI) where a high FI indicates a strong influence on the project, however, we need to also evaluate the permutation importance (PI) which indicates how much the projection is affected by small shuffling of the input attribute.

Large PI values indicate high importance for preserving relationships between data points. See Figure 5 for a summary. For more information on K-PCA and UMAP see: Wang et al. 2021, Wang 2012, Mika et al. 1998, McInnes et al. 2018, Bengio et al. 2005, and Bellman 1956, 1964, 1969.

Qualitative Evaluation (QLE):

QLE is a crucial part of an interpreter's role and is more than just a quick visual inspection. For channel systems there should be a strong understanding of channel movement, deposition, and the ability to recognize architectural elements (AE). These skills come overtime through studies of modern examples of various channel systems and repeated practical application of interpretations. The evaluation of the original DR outputs lean heavily on the prior experience of the interpreter with the overall evaluation pointing to K-PCA showing the highest internal channel feature resolution followed by UMAP for lateral feature delineation.

Moving to the quantitative evaluation, we see that the data support those findings.

Quantitative Evaluation (QE):

Using silhouette score and DBI, we see that of our 55 models, K-PCA placed in 9 of the top 10 rankings (followed by UMAP models). The overall top ranked number of clusters for all models, was K-PCA 2 components with 8 clusters showing a silhouette score of 0.904 and a DBI score of 0.145. For comparison, our top-ranking k-means prior to DR was k-means with 4 clusters showing a silhouette score of 0.352 and a DBI score or 0.965. For ranking, the higher the silhouette score the better the performance and the lower the DBI score the better the performance. Comparing improvements in choosing 4 vs. 5 clusters we see that out of all 11 groups, 4 clusters showed higher performance only 3 times. The highest improvement increasing from 4 to 5 clusters was 8.4%, UMAP 2 components.

Clustering– pre and post-DR

At this stage of the workflow, we have reviewed the statistical metrics and model rankings finding that the initial results support the findings that K-PCA outperforms the other DR algorithms for improving interpretability of channel features.

Next, we should review the cluster results and visually inspect these to see what the distribution looks like in terms of intra– and inter-cluster spacing. Figure 6 shows a 3D perspective for UMAP and K-PCA for 4 and 5 clusters. Using visual inspection alone, we may perceive that the clusters for UMAP are easier to distinguish as compared to K-PCA and determine that UMAP must be the better algorithm. However, as noted previously, using quantitative metrics, that is not the case.

Conclusion

After evaluating all models' post-DR k-means clustering results, we find that K-PCA outperforms UMAP and PCA taking the top 9 spots using both the silhouette score and DBI

method. Combined, K-PCA and UMAP dominate the top 33 of 55 model rankings with the first PCA model ranking at number 34 and our top pre-DR k-means ranking at number 49.

These quantitative results support our findings that K-PCA, followed by UMAP, improves the interpretation resolution for deepwater channel features. Our future work will include reviewing these post-DR cluster models rendered for geologic interpretation.

References:

Bellman, r., & rand corp santa monica ca. (1956). Dynamic programming.

Bellman, R. (1964). Control theory. Scientific American, 211(3), 186-201.

Bellman, R. (1969). A new type of approximation leading to reduction of dimensionality in control processes. Journal of Mathematical Analysis and Applications, 27(2), 454-459.

Bengio, Y., Delalleau, O., & Le Roux, N. (2005). The curse of dimensionality for local kernel machines. Techn. Rep, 1258(12), 1.

Bull, Suzanne AU - Browne, Greg AU - Arnot, Malcolm AU - Strachan, Lorna PY - 2020/01/17 SP - SP500 EP - 2019 T1 - Influence of Mass Transport Deposit (MTD) surface topography on deep-water deposition: an example from a predominantly fine-grained continental margin, New Zealand VL - 500 DO - 10.1144/SP500-2019-192 JO - Geological Society, London, Special Publications ER –

Di, H., Truelove, L., Li, C., & Abubakar, A. (2020). Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand. The Leading Edge, 39(10), 727-733.

Grahame, J.L. (2015). Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity.

Kamp, P. J., Vonk, A. J., Bland, K. J., Hansen, R. J., Hendy, A. J., McIntyre, A. P., ... & Nelson, C. S. (2004). Neogene stratigraphic architecture and tectonic evolution of Wanganui, King Country, and eastern Taranaki Basins, New Zealand. New Zealand journal of geology and geophysics, 47(4), 625-644.

King, P. R., & Thrasher, G. P. (1992). Post-Eocene development of the Taranaki Basin, New Zealand: Convergent overprint of a passive margin: Chapter 7: Southwest Pacific and Eastern Indian Ocean Margins.

King, P. R., and P. H. Robinson, 1988, An Overview of Taranaki Region Geology, New Zealand: Energy Exploration & Exploitation, 6, no. 3, 213-232.

Kroeger, K. F., Thrasher, G. P., & Sarma, M. (2019). The evolution of a Middle Miocene deep-water sedimentary system in northwestern New Zealand (Taranaki basin):

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Mika, S., Schölkopf, B., Smola, A., Müller, K. R., Scholz, M., & Rätsch, G. (1998). Kernel PCA and de-noising in feature spaces. Advances in neural information processing systems, 11.

Moreno-Ward, A. and Bedle, H., Fences and Fence Posts: Choosing the right attributes for the application of machine learning in seismic interpretation. AAPG Explorer Geophysical Corner March 2025.

Moreno-Ward, A., Vera-Arroyo, A., & Bedle, H. (2025). Improving interpretability: The role of dimension reduction in multiattribute seismic facies identification. *Interpretation*, *13*(3), T475-T496.

Naewboonnien, K. (2021). The Use of Seismic Attributes for Submarine Channel Interpretation in Deepwater Taranaki Basin, New Zealand. Bulletin of Earth Sciences of Thailand, 11(2), 134–143. Retrieved from https://ph01.tci-thaijo.org/index.php/bestjournal/article/view/246780

Quan Li, Shui Yu, Wei Wu, Liqing Tong & Hongquan Kang (2017) Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: a case study from the Taranaki Basin, New Zealand, New Zealand Journal of Geology and Geophysics, 60:3, 199-208, DOI: 10.1080/00288306.2017.1307230

Reilly, C., Nicol, A., Walsh, J. J., & Seebeck, H. (2015). Evolution of faulting and plate boundary deformation in the Southern Taranaki Basin, New Zealand. Tectonophysics, 651, 1-18.

Strogen, D. P., Seebeck, H., Nicol, A., & King, P. R. (2017). Two-phase Cretaceous–Paleocene rifting in the Taranaki Basin region, New Zealand; implications for Gondwana break-up. Journal of the Geological Society, 174(5), 929-946. References Cont.:

Strogen, D. P., Higgs, K. E., Griffin, A. G., & Morgans, H. E. (2019). Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression. Geological Magazine, 156(10), 1751-1770.

Wang, Q. (2012). Kernel principal component analysis and its applications in face recognition and active shape models. arXiv preprint arXiv:1207.3538.

Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. (2021). Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. The Journal of Machine Learning Research, 22(1), 9129-9201.

Appendix 3: The Algorithm Doesn't Know Geology — A Practitioner's Guide for ML-Assisted Seismic Facies Interpretation*

April Moreno-ward* and Heather Bedle, The School of Geosciences, The University of Oklahoma

*This paper is in review for the *International Meeting for Applied Geoscience & Energy Expanded Abstracts* 2026

Summary:

Machine learning (ML) techniques promise to revolutionize seismic interpretation, yet most workflows stop at the statistical output and call it geology. That gap between a defensible algorithm output and a meaningful geologic interpretation is where most ML workflows quietly fail. Bridging this gap requires more than novel technological approaches; it requires a deliberate, geologically informed framework that treats ML as what it is. A tool, not a geologic oracle.

We present exactly that. Using unsupervised k-means++ clustering paired with dimension reduction (DR) we present a multi-step seismic facies interpretation framework applied to a deepwater channel system in the Taranaki Basin, New Zealand. This work represents the fourth step in an ongoing multi-phase study. Prior phases established the seismic attribute selection, well ties, and initial facies definitions (step1): evaluated dimension reduction for qualitative interpretation (step 2); and benchmarked clustering performance across 55 models (step 3) finding that post-DR clustering outperformed pre-DR clustering in every case. This phase asks the harder question: does the best statistical model mean the best geologic model?

The answer is nuanced, and that nuance is the point. We demonstrate how unsupervised cluster outputs can be translated into context-based facies interpretation by integrating seismic reflection character, stratigraphic position, and depositional context. This strategic approach ultimately helps to identify facies tied to larger transgressive-regressive cycles and places

channel complex development and mass transport deposits into the geologic history of the environment.

Introduction:

Incorporating ML algorithms such as dimension reduction (DR) and unsupervised clustering techniques (k-means++), into 3D seismic interpretation workflows has demonstrated improved visual delineation of deepwater channel systems (Moreno-Ward et al. 2024 and 2025). While clustering alone can partition seismic attribute data into discrete facies groups, the quality and geologic meaningfulness of the resulting clusters is strongly dependent on the input data structure. Kernal principal component analysis (K-PCA) addresses this by transforming complex high-dimensional seismic attributes components that better capture non-linear variability prior to clustering, improving the performance of k-means ++ .

The geologic interpretability of the resulting clustering models is central to the overall workflow. The cluster outputs are not the final product, rather they are used by the interpreter in the context of seismic reflection character, stratigraphic position, and depositional setting to help identify geologically significant features within the deepwater channel system. This multi-attribute approach provides a model that is more defensible based on the critical analysis that goes into each step. This interpretive final stage transforms statistical results into a broader geologic framework, providing more robust insight into channel architecture, facies distribution, and reservoir heterogeneity that may be difficult to resolve through conventional attribute analysis.

Input/Output Evaluation for Geologic Implementation:

Before diving into implementation, whether as a researcher or practitioner of novel geologic workflows, one must consider that ML is a statistical tool, not a geologic answer. With this in mind, we can reframe our approach. First, we must acknowledge that as a tool, it must be used in conjunction with geologic knowledge. There, the primary question that we are asking is, “where does it fit in the geoscience workflow”. The statistical algorithms are there to quantify

how data relates to data, or in the geophysical realm, how data can be extracted from amplitude values through seismic attributes and be organized into meaningful patterns. Next, how are those statistically meaningful patterns related to geologically significant features, or even more simply, how well preserved is the data after applying ML.

We begin the implementation of ML by considering what the inputs are or what they need to be to optimize the strengths of the algorithms. Since we are working with seismic data, how can seismic attributes be used as inputs, and what do the outputs represent. For deepwater channel facies identification, we began with kmeans ++ unsupervised clustering (for more in-depth explanation on kmeans ++, see Arthur and Vassilvitskii 2007) with our attributes as the initial inputs, Figure 1, option 1. However, after evaluating the clusters using quantitative and qualitative metrics, we note subpar compaction and cohesion results. We determined that using the seismic attributes as a direct input for clustering is not optimal. Considering the complexity of the seismic data, we determined that dimension reduction should be the first output, Figure 1, option 2. The seismic attributes were then used as inputs for dimension reduction. For more information on how we selected DR methods and evaluated statistical vs. qualitative metrics, see Moreno-Ward et al. 2024 and 2025. The DR outputs were then

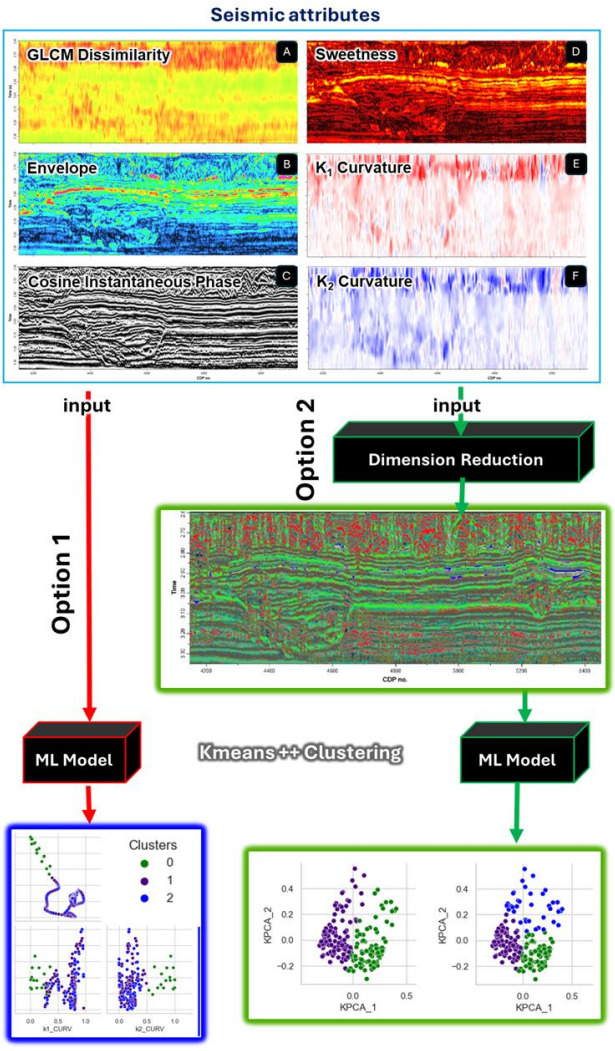


Figure 1: Consideration for which input/output set is optimized for preserving and delineating the most geologically significant data. Option1 uses the seismic attributes as the input for the unsupervised clustering algorithm. Results indicate that clusters do not perform well in cluster cohesion and separation. Option 2 takes the seismic attributes and uses those as the input for dimension reduction first. The output from this is evaluated for improved qualitative interpretation. The dimensionally reduced data is then used as the input for unsupervised clustering. The post dimension reduction models show improved performance for clustering cohesion and compaction.

evaluated qualitatively for improved interpretability of deepwater channel architectural elements. DR was shown to improve interpretability and was then used as the input for kmeans ++, with the resulting statistical outputs evaluated for cluster performance. Of all 55 models considered, the post DR-clustering outperformed pre-DR clustering in all instances for this dataset.

Now that we have optimized the input for unsupervised clustering, we will consider the output data. However, before we begin with interpretation, we need to ensure we understand the depositional system, beginning with geologic location information.

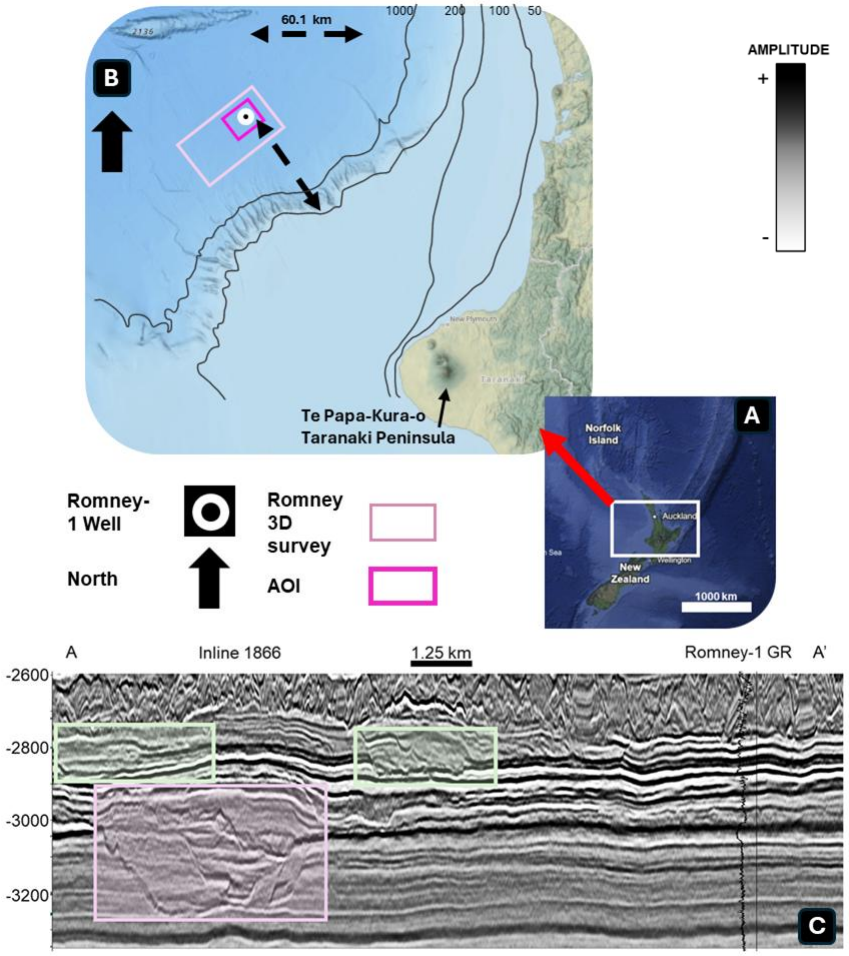


Figure 2: Deepwater channel of interest is located in the A) Taranaki Basin, NW New Zealand in approximately 1550 m of water and B) 60 km from the shelf edge. C) There are several channel systems in the seismic volume, however we are focusing on the channel complex located in the pink rectangle.

Geologic Location:

The deepwater channel of study is located approximately 160km northwest of the Taranaki peninsula and 60 km from the present-day shelf edge in roughly 1,450 m of water depth. Repeated lateral migration and vertical movement of the channel produced a suite of complex architectural elements (AEs) that reflect multiple phases of channel migration. While the broader structural elements are resolvable through conventional seismic interpretation, depositional complexities common to channel systems require more sophisticated workflows to better delineate and map the reservoir. While several iterations of channelized flow are seen in the seismic volume, we will focus our attention on channel system A.

Workflow:

This brings us to step 4 in an evolving, multi-part seismic interpretation workflow developed for deepwater channel characterization. Step 1 utilizes existing seismic workflows to properly select attributes, tie available wells, and then create a baseline for identifying facies. Step 2 examined how DR improved qualitative interpretation and data input for unsupervised clustering (Moreno-Ward et al. 2025). Step 3 analyzed unsupervised clustering performance (k-means ++) applied to pre– and post-DR inputs. Step 4 shifts our focus from statistical model to geologic significance, recognizing that statical metrics alone are insufficient to validate a clustering model's utility for geologic interpretation. We interpret the clustering model in its depositional and stratigraphic context, assessing the geologic meaningfulness of the identified facies and their contribution to understanding channel system A's architecture. This interpretive step is essential to closing the loop between ML outputs and actionable geologic importance.

Defining Facies:

Our first priority in step 4 is considering how we define facies, Figure 3. In previous steps, we (1) focused on our single well and ground truth it provides. Facies A-C are identified using well log curves and mud log descriptions. The understanding of lithofacies also helps us to consider the depositional environment and use that information when developing our interpretation model. Facies C sits on top of a limestone marker and begins a fining up sequence through Facies B, before being capped by another limestone maker, Facies A. From previous geologic study, we can verify what the well information indicates, major transgressive-regressive-transgressive cycles.

After the well is tied, we defined the seismic facies (2) determining what seismic response patterns look like for facies A-C. (3) Next, we pull in the model to identify what seismic attribute responses are preserved in the model and how they relate to these three facies. These three steps do not account for the full facies identification and interpretation workflow, however.

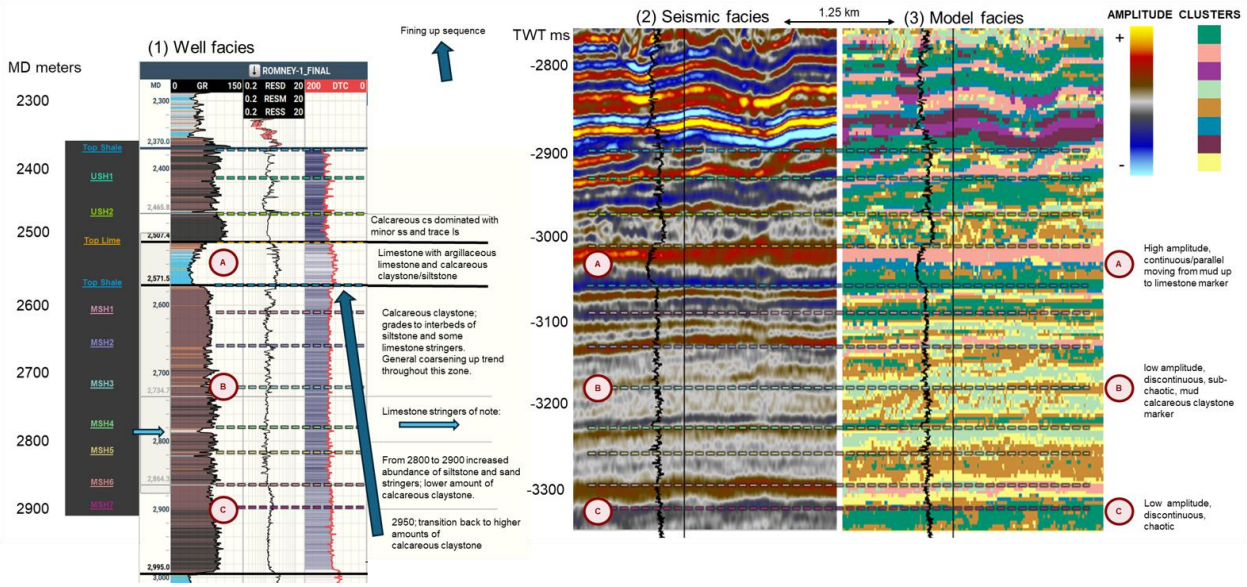


Figure 3: Initial 3-step process for defining facies. (1) Consider what lithologic/well facies can be defined. We identified 3 main lithofacies (A-C) within the depth interval for our channel. (2) After well-tie, we considered what seismic response(s) could be identified for each for A-C. And lastly, (3) can we identify facies within the model. As the well is outside of the channel complex, interpolation will be required.

Figure 4 focuses on the top performing statistical model, post-K-PCA clustering (2 components, 8 clusters). We begin the interpretation by reviewing deepwater channel architectural elements in a timeslice at –3204 ms using amplitude (4A). With a background knowledge in channel morphology and deposition, we can begin to infer what grain size we may expect within channel axis, margins, and elsewhere in the channel complex system. Higher amplitude is associated with coarser grains while lower amplitude is associated with finer grains. We can improve interpretability by applying bump mapping (4B) and then compare this interpretability to the model (4C). Black arrows indicate oldest portions of channel flow in this area indicated by the straight geometry and high amplitude. Orange arrows indicate and older portion of channel deposition where flow velocity has decreased as indicated by more sinuous

morphology and lower amplitude. Pink and yellow arrows indicate locations where seemingly anomalous shaped features are preserved in the model.

Next, we view the seismic data/models in cross-sectional view to continue facies identification (Figure 4(D) amplitude, (E) model, and (F) model corendered with cosφ). We identify 6 additional facies and begin marking where we see facies A-C within the channel. Facies 1*-6 are important for understanding the larger depositional environment and represent transgressive-regressive-transgressive cycles capped with mass transport deposits (MTDs) at the top. For more detailed information on the geologic history of the Taranaki Basin, see Kamaruzaman et al. 2023; Grahame 2015; King and Robinson (1988); Thrasher (1992); Kamp et al. (2004); Cathal et al. (2015); Li et al. (2017); Quan et al. (2017); Strogen et al. (2017 and 2019); Di et al. (2020); Naewboonnien (2021); Chopra et al. (2022) .

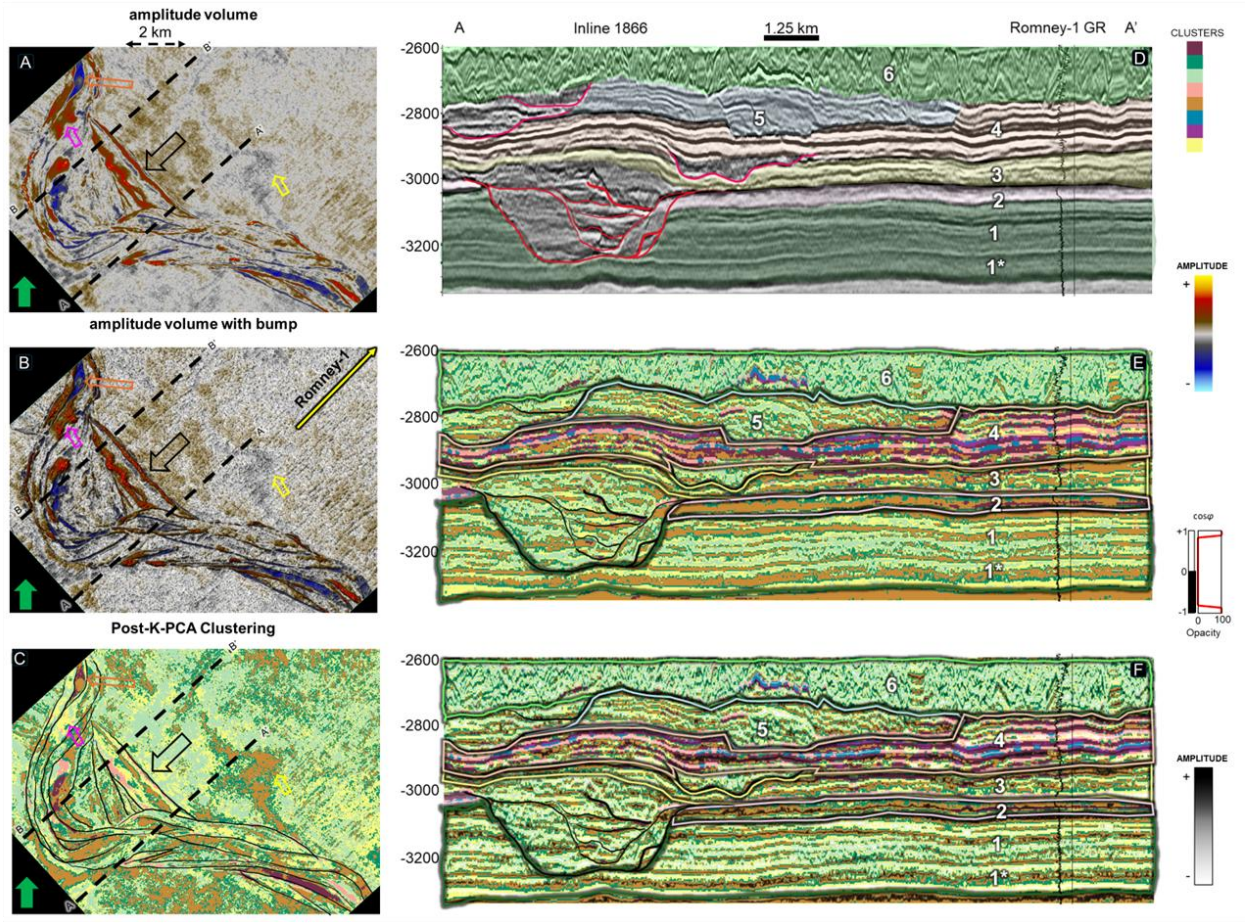


Figure 4: Top performing statistical model is evaluated for geologic interpretation. A) Amplitude timeslice at –3204 is used to identify channel features. B) We can improve the delineation of laterally identifiable channel features by including bump mapping. C) We evaluated the KPCA cluster model (2 components, 8 data clusters). Black arrows indicate oldest portion of channel were flow velocity was higher and indicated by straight geometry. Orange arrows indicate an older portion of the channel once flow velocity decreased and channel shifted to more sinuous morphology. Pink arrows indicate a distinct pattern that can be identified visually in A-C, indicating that the clustering model is retaining geologic feature information. View the data in cross-sectional view (D) amplitude, (E) cluster model, and (F) cluster model corendered with cosφ. We have identified 6 facies packages: (1 and 1) low energy parallel deposition, (2) transgression leading to limestone layer, (3 and 4) beginning of regression with more siltstone and sandstone coming from the shelf. (5) Slumping off the high and (6) MTD.*

Conclusion

When incorporating ML into traditional geologic interpretation, the practitioner must be deliberate about what question they are asking the algorithm to solve. In this workflow, that question was specific: identify similarity in seismic attributes within a deepwater channel system. The algorithm did a phenomenal job, but it could not tell us what those similarities mean. That meaning requires the interpreter. And that is precisely where the risk lives. Interpreter

bias is not eliminated by using an unsupervised algorithm. It may be unknowingly introduced in the data preparation space or simply relocated to the interpretation stage.

This is not a flaw in the workflow, but it is a vulnerability that must be acknowledged and managed. A clustering model will reflect whatever pattern exists in the input data. If the input was poorly selected, if the interpreter forces a geologic narrative onto statistically ambiguous clusters, or if the interpreter misunderstands the question they actually asked, the result may not represent true geologic importance.

In this workflow, we managed that risk by grounding every interpretive decision in seismic reflection character, stratigraphic position, and depositional context. The result was concrete. Within channel system A, the post-K-PCA clustering model successfully delineated channel axis from margin facies, resolved distinctions between high-energy straight-geometry flows and lower-energy sinuous phases, and preserved subtle morphologic features not resolvable through conventional amplitude analysis alone. Six broader facies packages were identified representing transgressive-regressive-transgressive cycles capped by mass transport deposits — stratigraphic context essential for understanding the time and activity of the channel complex.

ML does not replace geologic knowledge. It amplifies it, for better or worse. Used thoughtfully, these methods accelerate architectural element mapping and serve as a powerful visual communication tool. Used carelessly, they may produce confident-looking answers to the wrong questions or result in the interpreter getting lost in the results. The value lies not in the algorithm. It lies in the geologist behind it.

References:

Arthur, D., & Vassilvitskii, S. (2007, January). k-means++: The advantages of careful seeding. In Soda (Vol. 7, pp. 1027-1035).

Cathal Reilly, Andrew Nicol, John J. Walsh, Hannu Seebeck, Evolution of faulting and plate boundary deformation in the Southern Taranaki Basin, New Zealand, Tectonophysics, Volumes 651–652, 2015, Pages 1-18, ISSN 0040-1951, https://doi.org/10.1016/j.tecto.2015.02.009.

Chopra, S., Sharma, R. K., Marfurt, K. J., Bedle, H., & Verma, S. (2022). Attempts at seismic characterization of a deepwater turbidite channel in Taranaki Basin, New Zealand. First Break, 40(10), 27-39.

Di, H., Truelove, L., Li, C., & Abubakar, A. (2020). Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand. The Leading Edge, 39(10), 727-733.

Grahame, J.L. (2015). Deepwater Taranaki Basin, New Zealand—New Interpretation and Modelling Results for Large Scale Neogene Channel and Fan Systems: Implications for Hydrocarbon Prospectivity

Kamp, P. J., Vonk, A. J., Bland, K. J., Hansen, R. J., Hendy, A. J., McIntyre, A. P., ... & Nelson, C. S. (2004). Neogene stratigraphic architecture and tectonic evolution of Wanganui, King Country, and eastern Taranaki Basins, New Zealand. New Zealand journal of geology and geophysics, 47(4), 625-644.

Kamaruzaman, E. H., La Croix, A. D., & Kamp, P. J. (2023). Quantitative seismic geomorphology of sediment conduits on an evolving Miocene slope in Taranaki Basin (New Zealand): The influence of increasing slope gradient through time. Marine and Petroleum Geology, 152, 106233.

King, P. R., & Robinson, P. H. (1988). An Overview of Taranaki Region Geology, New Zealand. Energy Exploration & Exploitation. https://doi.org/10.1177/014459878800600304

Li, Q., Yu, S., Wu, W., Tong, L., & Kang, H. (2017). Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: A case study from the Taranaki Basin, New Zealand. New Zealand Journal of Geology and Geophysics, 60(3), 199-208.

Moreno-Ward, A., Bedle, H., & Vera-Arroyo, A. (2024, August). Enhancing seismic delineation of obscured geologic architectural elements in a deepwater channel complex through multi-attribute analysis: A study from the Taranaki Basin. In SEG International Exposition and Annual Meeting (pp. SEG-2024). SEG.

Moreno-Ward, A., Vera-Arroyo, A., & Bedle, H. (2025). Improving interpretability: The role of dimension reduction in multiattribute seismic facies identification. Interpretation, 13(3), T475-T496.

Naewboonnien, K. (2021). The Use of Seismic Attributes for Submarine Channel Interpretation in Deepwater Taranaki Basin, New Zealand. Bulletin of Earth Sciences of Thailand, 11(2), 134–143. Retrieved from https://ph01.tci-thaijo.org/index.php/bestjournal/article/view/246780

Quan Li, Shui Yu, Wei Wu, Liqing Tong & Hongquan Kang (2017) Detection of a deep-water channel in 3D seismic data using the sweetness attribute and seismic geomorphology: a case study from the Taranaki Basin, New Zealand, New Zealand Journal of Geology and Geophysics, 60:3, 199-208, DOI: 10.1080/00288306.2017.1307230

Strogen, D. P., Seebeck, H., Nicol, A., & King, P. R. (2017). Two-phase Cretaceous–Paleocene rifting in the Taranaki Basin region, New Zealand; implications for Gondwana break-up. Journal of the Geological Society, 174(5), 929-946.

Strogen, D. P., Higgs, K. E., Griffin, A. G., & Morgans, H. E. (2019). Late Eocene–Early Miocene facies and stratigraphic development, Taranaki Basin, New Zealand: the transition to plate boundary tectonics during regional transgression. Geological Magazine, 156(10), 1751-1770.

Thrasher, Glenn P. (1992). Late Cretaceous Geology of Taranaki Basin, New Zealand. Thesis, Victoria University of Wellington. http://hdl.handle.net/10063/525

Appendix 4: Fences and Fenceposts – Choosing the right attributes for the application of machine learning in seismic interpretation*

April Moreno-ward* and Heather Bedle, The School of Geosciences, The University of Oklahoma

*This paper is published in AAPG Explorer Geophysical Corner, March 2025.

Introduction:

Effective seismic interpretation requires understanding geologic boundary-defining attributes, which we can think of as fence posts, and lithology or reservoir-characterizing attributes, or the fence boards. With this analogy in mind, we explore how the use or non-use of these two categories of attributes can affect the interpretability of deepwater channel features in dimension reduction techniques. By evaluating traditional and machine learning interpretation approaches, we demonstrate how the preservation of both the distinct boundaries (fence posts) and lithological variations (fences) can be impacted in multi-attribute ML by the interpreter's choice of seismic attributes.

Seismic attribute selection is a critical first step for the effective geologic interpretation of seismic data. Furthermore, the ability to align appropriate attributes with specific geologic objectives is fundamental in the identification of meaningful geologic features. As each attribute represents a calculation derived from the original seismic data, it is critical to understand both its mathematical foundation and the targeted feature it was designed to enhance. Recognition of the inherent uncertainty of each attribute underscores the interpreter's role in verifying the use of the attribute for interpretations. To implement this understanding methodically, one should first consider the hierarchical categories of seismic attributes (figure 1a) and select those whose application most closely meets the interpretive goal.

To demonstrate this thought process, we will show an example for channel facies discrimination in the deepwater channel system within the 3-D Romney seismic volume, in the

Taranaki Basin, west of New Zealand. Based on our objectives for channel identification, we identify geometrical, instantaneous and multi-trace seismic attributes as optimal for our study goals. Once the categories are selected, we chose seven initial attributes: GLCM dissimilarity (GLCM-D), sweetness, envelope, cosine of instantaneous phase (CSP), structural curvedness (K-curvedness), most positive- and negative-curvature (K1 and K2) (figure 1b). The definitions and applications of these attributes can be found in the November 2007, December 2007, April 2014, November 2014 and December 2015 Geophysical Corner installments. After a preliminary analysis of these attributes on our data, we further narrowed them down to six attributes, opting to retain most-positive and most-negative curvature attributes in lieu of structural curvedness. Before moving forward with any ML workflows, it is important to examine each attribute and conduct an initial interpretation of internal channel architecture to verify the efficacy of the chosen attributes for our objective (figure 1c).

On the Usefulness of CSP Here

As each attribute should contribute positively to the end interpretation goal, which is channel facies identification, some discussion was needed among the research team regarding the usefulness of CSP in this project. So, let's discuss this further. CSP is an instantaneous attribute that focuses on the phase of the seismic trace at each individual sample and is independent of the amplitude. This attribute is often used to provide a detailed view of the continuity and/or discontinuity of geological features and is often used for seismic stratigraphic analysis, as it is useful for identifying pinch-outs and terminations, and other stratigraphic geometries.

Using the analogy of fences and fence posts can help illustrate the utility of CSP in seismic data interpretation. If one imagines the sedimentary layers as shown in the seismic data as a long fence stretching across space, the CSP attribute represents the individual fence posts. Similarly, as each fence post signifies a specific point along the fence, marking boundaries and supporting the structure, the CSP pinpoints where significant phase changes in the seismic signal

a

CATEGORY	MEASURES
Fourier:	frequency content and spectrum of the seismic signal, indicates cyclic sedimentary patterns or frequency anomalies.
Geometrical:	geometric relationships and configurations of subsurface structures, useful for identification of geologic features.
Instantaneous:	instantaneous characteristics of the seismic wave, such as amplitude, frequency, or phase.
Multi-trace:	coherence, continuity, and the lateral extension of subsurface features; assists in delineation and interpretation.
Physical:	physical properties, aids in identification of lithological changes.
Post-stack:	composite seismic responses, often used for structural interpretation.
Pre-stack	variations in seismic responses due to changes in lithology, fluid content, or gas presence.
Wavelet:	properties of the seismic wavelet to interpret the reflector series and changes in the subsurface.
Window:	attributes within a confined subsurface segment; can be critical for stratigraphic or local feature analysis.

b

CATEGORY	ATTRIBUTE	USE
Geometrical	Most Positive Curvature	Most positive geologic relief (K1-Curvature)
	Most Negative Curvature	Most negative geologic relief (K2-Curvature)
	Structural Curvedness	quantifies seismic refection curvature, with respect to 0 (K-Curvature)
Instantaneous	Cosine instantaneous phase	illustrates continuity of the seismic traces
	Envelope	measures the strength or intensity of the seismic amplitude
	Sweetness	enhance bright spots; lithology indicator
Multi-Trace/ Textural	GLCM dissimilarity	measures changes in seismic amplitude often associated with structural changes and/or lithologic boundaries

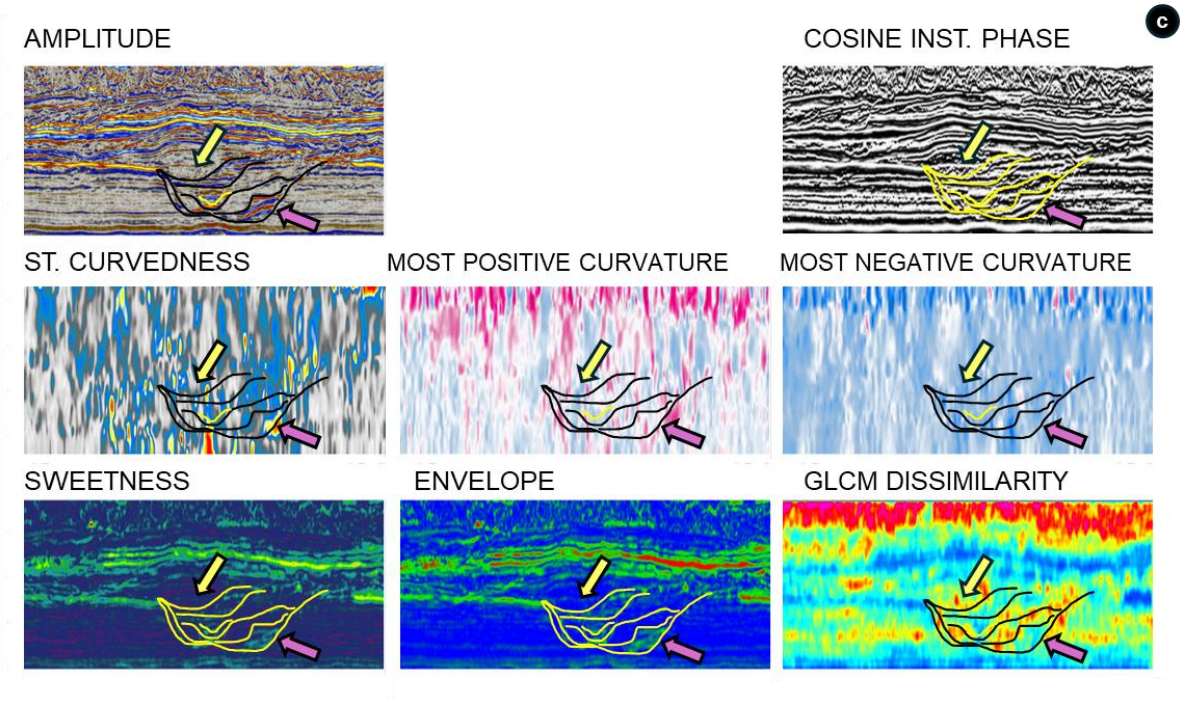


Figure 1: (a) Hierarchical seismic attribute classes, (b) selected attributes for identification of deepwater channel seismic facies in the Romney dataset, and (c) interpreted section of internal channel architecture for selected attributes (amplitude included as a baseline for channel features). (Data courtesy of NZP&M)

occur. These points of change can indicate either continuous geological features, akin to the uninterrupted stretch of a fence, or discontinuities, like breaks or gaps in the fence.

CSP versus Variance

Let us quickly consider another example of a fence-post attribute: variance. When compared to CSP for the identification of channel features, its performance is sub-par (figure 2a and b). The difference in highlighting those sharp channel edges is an important reminder to ensure visual inspection of potential seismic attributes before continuing with your workflow. In figure 2a, CSP illustrates distinct changes in the lateral reflectors and clearly delineates the unconformable surface of the channel complex edge. Within the channel of interest, we can further begin to delineate lateral channel migration and younger portions as well. The sweetness and envelope attribute displayed in figures 2d and e, as fence attributes, give us more insight into how to interpret the seismic geologically. Rather than only focusing on abrupt changes, we can use the attributes to infer something about the lithologic characteristics. In figures 2c and g we change the viewing perspective for our fence-post attribute (CSP) and one fence attribute. In this new perspective, we can follow those indications of channel features onto a time slice and get a sense of what the lateral channel morphology looks like. We have also adapted our color map for the display in figure 2g by adding bump-mapping.

Bump-mapping alters the lights and shadows, essentially creating a textured look, or relief to the seismic, which helps to delineate the channel's architecture. In figure 2f, we note that while GLCM-D might have seemed underwhelming in the cross-sectional view for helping to really tease out the intricacies of the complex channel system, we gain a better appreciation of its ability to inform the interpreter of the changes happening in the channel system by showing it both in the bench view and by adding in the texture. For example, in figure 2g note the location indicated by the black arrow. In the cross-sectional view, we note a small area with low GLCM-D between areas with much higher values. When we map that feature onto the time-slice we see that it corresponds to a smaller-scale channel feature which is filled with fine-grained material and is clear to see.

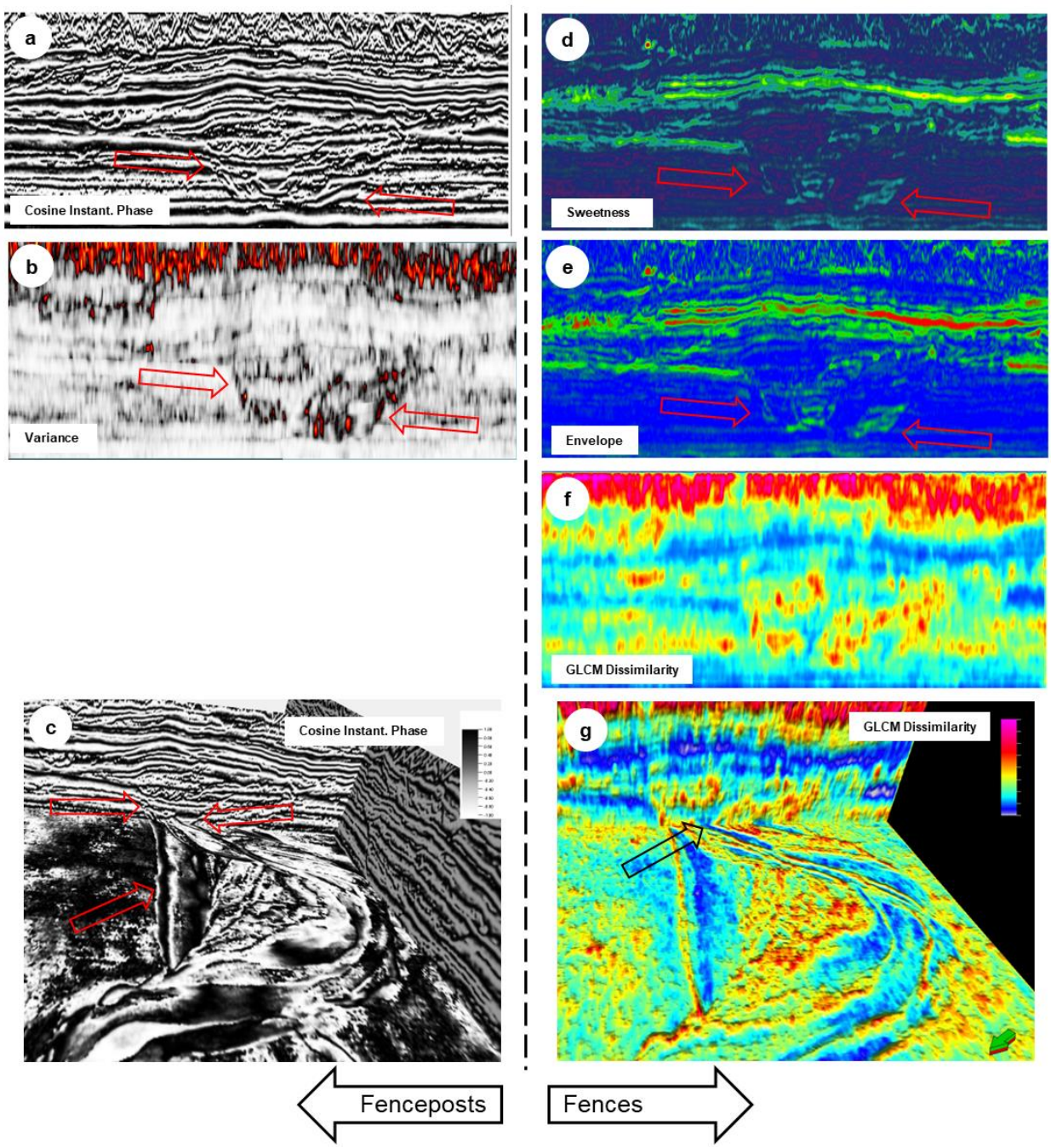


Figure 2: Examples of fencepost attribute displays on the left (a to c), and fence attributes on the right (d to g). Using a cross-sectional view of our channel of interest, we note the clear delineation of the outer bounding edge (red arrow) for the channel complex in (a) depicting cosine of instantaneous phase. Viewing the same feature on the equivalent variance display in (b), the bounding edge (red arrow) is somewhat less sharp but no less evident. The sweetness, envelope, and GLCM dissimilarity attribute displays in (d to f) respectively, we can find the bounding edge with a little more effort. It should be noted that a change in viewing dimension can be very effective when delineating those features of interest. Perspective views are shown for the cosine of instantaneous phase attribute as fencepost in (c) and GLCM-D attribute as fence attribute in (g). Notice, the fencepost clearly denotes sharp changes laterally, and the fence attribute illustrates the changes in channel migration. (Data courtesy of NZP&M)

Algorithm Selection

Once we are confident in our selection of seismic attributes for our geologic objectives, we can begin evaluating ML algorithms to select those that align with our goals. For this, we use a common ML technique to serve as a baseline to validate newer techniques. For this purpose, we selected principal component analysis. PCA is a dimensionality reduction (DR) technique that is commonly applied to 3-D seismic data when analyzing more than three seismic attributes simultaneously. While this technique has been utilized in seismic workflow for some time, it has two major limitations. First, PCA assumes that the data are linear. Second, it assumes input attributes follow a Gaussian distribution, which is not the case for many seismic attributes (sweetness and envelope excepted). Additionally, its focus on reducing dimensional complexity by focusing on variance within the data can lead to overlooking more subtle patterns that may be geologically important.

To overcome these inherent limitations of traditional PCA, we evaluate additional DR techniques specifically designed to handle more complex non-linear data, which is more similar to what we find in 3-D seismic datasets. We chose kernel principal component analysis (K-PCA) and uniform manifold approximation and projection (UMAP) for this area. Figure 3 provides a comparison between these three methods.

Method	PCA	KPCA	UMAP
Type	Dimensionality Reduction (DR)		
DR, how?	While preserving variance	Using kernel trick	While preserving local structure
Output	Principal components (linear combinations of features)	Principal components in a transformed space	Lower-dimensional embedding of data
Scalability	Good	Moderate	Good
Interpretability	High	Moderate	Moderate to low
Non-linear handling	Poor	Good	Excellent
Strengths	Easy to interpret, reduces overfitting	Captures complex patterns, non-linear relationships	Preserves local data structure, effective on complex data
Weaknesses	Assumes linear relationships, can miss non-linear structures	Computationally intensive, sensitive to kernel choice	Sensitive to hyperparameters, can be computationally intensive

Figure 3: Comparison of selected machine learning dimension reduction algorithms. During the selection process, it is important to note how each algorithm works, its scalability, interpretability, handling of nonlinearity, and their strengths and weaknesses.

Application:

With our ML DR algorithms selected, we use them to simplify the dimensionality of the original six seismic attributes and qualitatively evaluate each model's ability to preserve and/or enhance the interpreter's ability to resolve both large-scale channel features and more subtle features.

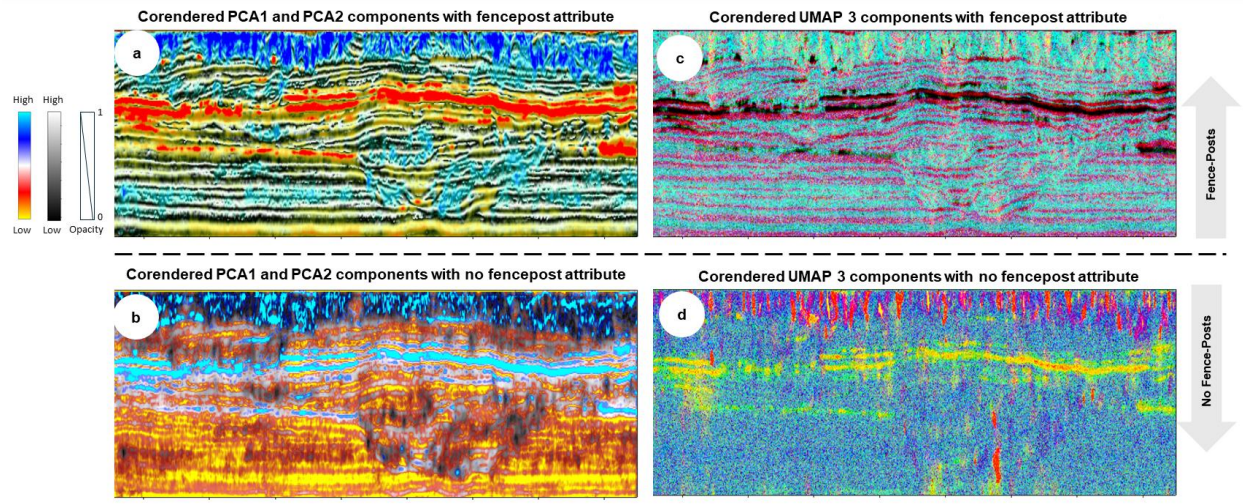


Figure 4: Comparing results with and without cosine of instantaneous phase (CSP). (a) Utilizing PCA to reduce the dimensionality of the input data, the 2D PCA display, PCA 1 (in color) corendered with PCA 2 (in grayscale) when it (a) includes, and (b) does not include the fencepost attribute. Similarly, utilizing UMAP to reduce the dimensionality of the input data, the UMAP displays are shown when the fence post attribute is (c) included, and (d) not included in the analysis. The RGB corendering was effective in delineating seismic features. (Data courtesy of NZP&M)

Building on this initial evaluation, we specifically assess the impact of including the fence-post attribute (CSP). In figure 4, we compare the performance of PCA and UMAP both with and without the fence-post attribute (CSP). From our original six-dimensional dataset (each attribute representing one dimension), each DR model is reduced to two components and three components (each component representing one dimension), that preserve what the model identifies as the most significant patterns in our data while simultaneously reducing the overall dimensionality. By comparing the results of these reductions both with and without the CSP, we can evaluate how significant the fence-post attribute is and also how many dimensions best preserve our geologic feature of interest. The success of these comparisons ultimately depends on the interpreter's ability to identify both larger-scale channel boundaries and nested boundaries within the complex. CSP aids in identifying stark boundary edges and excels at defining both internal channel edges and larger-scale bounding features.

We found through this comparison that including the fence-post attribute (CSP) in DR methods plays a critical role in preserving the geologic interpretability of deepwater channel features. Including CSP in both PCA and UMAP enabled clear preservation of channel

boundaries while maintaining the visibility of internal architectural elements. The notable contrast between results with and without CSP clearly shows that just as a sturdy fence needs both posts and boards, effective seismic interpretation benefits from both boundary-defining and lithology-characterizing attributes working together. While phase-based attributes are sometimes overlooked in ML workflows, our results suggest that using both types of attributes can lead to improved geological interpretations. This approach is good to keep in mind when moving from traditional interpretation methods to emerging ML techniques.

Acknowledgements

We would like to acknowledge Kurt Marfurt for spurring the conversation regarding fences and fence-post attributes. Taking the time to explore and visualize the differences has been rewarding and insightful.