UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MACHINE ASSISTED QUANTITATIVE SEISMIC INTERPRETATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

TAO ZHAO
Norman, Oklahoma
2017

MACHINE ASSISTED QUANTITATIVE SEISMIC INTERPRETATION

A DISSERTATION APPROVED FOR THE
CONOCOPHILLIPS SCHOOL OF GEOLOGY AND GEOPHYSICS

BY

_____

Dr. Kurt J. Marfurt, Chair

_____

Dr. Deepak Devegowda

_____

Dr. Shankar Mitra

_____

Dr. Xiaowei Chen

_____

Dr. Vikram Jayaram

To my family and everyone that cares about me

ACKNOWLEDGMENTS

My life has changed dramatically over the past four years. I married. I received the first job offer of my life. I graduated with my PhD.

People sometimes attribute marriage and finding a job to luck or destiny, but we never say that for getting a PhD. Because we have to have perseverance, inspiration, and more importantly, the support from everyone around us to earn that “Dr.”. That is exactly why I am eager to sincerely acknowledge those who have helped me through the past four years, without whom none of my achievements would be possible.

Dr. Kurt Marfurt, my PhD advisor, as well as the one who gave me the opportunity to spend the memorable years at the University of Oklahoma, deserves the first place on my list of recognition. He shows great respect to his students, and always treats our needs as the top priority. His critical thinking, everlasting energy, passion to research, and charming personality have established him as an idol of my future career. For everything that I might have expected from a PhD advisor, he has done that beyond my expectation. Maybe I will never become a prestigious professor and geophysicist as he is, but he is a lighthouse that guides me towards the right direction in becoming one.

When I first started my PhD at the University of Oklahoma, Dr. Vikram Jayaram guided me over my first research topic, which has a tremendous impact on developing my own research. He is always encouraging, and thinks from my perspective. Being originally from an electrical engineering background and having experience from multiple industries, he also builds an excellent example of adapting knowledge from one domain and applying to another. Beyond the academic guidance, he is also one of my most reliable resources that I would like to turn to when I seek for advices for my life and career. He is truly my teacher for life, and friend for life.

I would also like to show my greatest gratitude to Dr. Shankar Mitra, Dr. Xiaowei Chen, and Dr. Deepak Devegowda, who serve on my PhD committee. Dr. Mitra is a world famous structural geologist who explains the complex concepts exceptionally well to someone with little to no structural geology background. Dr. Chen and Dr. Devegowda are experts in their own domain, and contribute to the integrity of my PhD research. Both of them are great examples of young professionals from whom I have learned a lot. It is my honor to have them all on my committee.

I want to thank all team members and industrial sponsors of the Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their continuously support and constructive critique on my research. Especially, I want to appreciate Fangyu Li and Dr. Bradley Wallet, for constantly inspiring me with new research ideas, and Dr. Atish Roy, who had graduated before I joined AASPI so that I really did not have a chance to work with, for his creative work in AASPI that I built on.

Finally, I want to thank my wife, Yun Liu, for giving all these work a purpose. She is always supportive, encouraging, and is my best teacher for the truth of life. I also want to thank my family, who are on the other side of the Earth, for always being my strongest support in the life path I choose.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

During the past decades, the size of 3D seismic data volumes and the number of seismic attributes have increased to the extent that it is difficult, if not impossible, for interpreters to examine every seismic line and time slice. Reducing the labor associated with seismic interpretation while increasing the reliability of the interpreted result has been an on going challenge that becomes increasingly more difficult with the amount of data available to interpreters. To address this issue, geoscientists often adopt concepts and algorithms from fields such as image processing, signal processing, and statistics, with much of the focus on auto-picking and automatic seismic facies analysis. I focus my research on adapting and improving machine learning and pattern recognition methods for automatic seismic facies analysis. Being an emerging and rapid developing topic, there is an endless list of machine learning and pattern recognition techniques available to scientific researchers. More often, the obstacle that prevents geoscientists from using such techniques is the "black box" nature of such techniques. Interpreters may not know the assumptions and limitations of a given technique, resulting in subsequent choices that may be suboptimum. In this dissertation, I provide a review of the more commonly used seismic facies analysis algorithms. My goal is to assist seismic interpreters in choosing the best method for a specific problem. Moreover, because all these methods are just generic mathematic tools that solve highly abstract, analytical problems, we have to tailor them to fit seismic interpretation problems. Self-organizing map (SOM) is a popular unsupervised learning technique that interpreters use to explore seismic facies using multiple seismic attributes as input. It projects the high dimensional seismic attribute data onto a lower dimensional (usually 2D) space in which interpreters are able to identify clusters of seismic facies. In this dissertation, using SOM as an example, I provide three

improvements on the traditional algorithm, in order to present the information residing in the seismic attributes more adequately, and therefore reducing the uncertainly in the generated seismic facies map.

CHAPTER 1

INTRODUCTION[1]

In the 21st century, pattern recognition has become part of everyday life. Amazon or Alibaba analyzes the clothes you buy, Google analyzes your driving routine, and your local grocery store knows the kind of cereal you eat in the morning. Big companies and big government, attempting to identify patterns in our spending habits and the people with whom we associate, are analyzing "big data" with "deep learning algorithms" all the time.

Successful seismic interpreters are experts at pattern recognition, identifying features such as channels, mass transport complexes, and collapse features where our engineering colleagues only see wiggles. Our challenge as interpreters is that the data volumes we need to analyze keep growing in size and dimensionality, while the number of experienced interpreters has remained relatively constant. One solution to this dilemma is for these experienced interpreters to teach their skills to the next generation of geologists and geophysicists, either through traditional education or on-the-job training. An alternative and complimentary solution is for these experienced interpreters to teach theirs skills to a machine. Turing (1950), whose scientific contributions and life has recently been popularized in a movie, asked whether "Machines can think?" Whether machines will ever be able to think is a question for scientists and philosophers to answer (e.g. Eagleman, 2012), but machines can be taught to perform repetitive tasks, and even

[1] *This chapter contains contents from a published article - Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58.*

to unravel the relationships that underlay repetitive patterns, in an area called machine learning.

25 years ago, skilled interpreters delineated seismic facies on a suite of 2D lines by visually examining seismic waveforms, frequency, amplitude, phase, and geometric configurations. Facies would then be posted on a map and hand contoured to generate a seismic facies map. With the introduction of 3D seismic data and volumetric attributes, such analysis has become both more quantitative and more automated. In this dissertation, I focus on pattern recognition on large 3D seismic data whereby like patterns in the seismic response (seismic facies) are assigned similar values. Based on the availability of training data (ground truth of the pattern to be discovered), we define supervised learning as situations with training data, and unsupervised learning as situations without. In either case, the ultimate goal is always discovering the hidden relation. In supervised learning, we want to build a model to discover the relation between a given input and output data pair (the aforementioned training data); in unsupervised learning, we want to build a model to characterize the relation among the input data (no known output data are given in this case). Seismic interpreters and petrophysicists use similar technology (the simpler being crossplots) to define specific rock properties, such as brittleness, TOC, or porosity. Pattern recognition is common to many industries, from using cameras to identify knotholes in plywood production to tracking cell phone communications to identify potential narcotics traffickers.

I modify a workflow from the classical Duda et al. (2001) textbook in Figure 1.1 to summarize the process. In this figure, “sensing” consists of seismic, well log, completion, and production measurements. For interpreters “segmentation” will usually

mean focusing on a given stratigraphic formation or suite of formations. Seismic data lose both temporal and lateral resolution with depth, such that a given seismic facies changes its appearance, or is nonstationary, as we go deeper in the section. The number of potential facies also increases as we analyze larger vertical windows incorporating different depositional environments, making classification more difficult. For computer assisted facies classification, "feature extraction" means attributes, be they simple measurements of amplitude and frequency, geometric attributes that measure reflector configurations, or more quantitative measurements of lithology, fractures, or geomechanical properties provided by prestack inversion and azimuthal anisotropy analysis. "Classification" assigns each voxel to one of a finite number of classes (also called clusters), each of which represents a seismic facies that may or may not correspond to a geological facies. Finally, using validation data, the interpreter makes a "decision" that determines whether a given cluster represents a unique seismic facies, if it should be lumped in other clusters having a somewhat similar attribute expression, or whether it should be further subdivided, perhaps through the introduction of additional attributes.

Pattern recognition of seismic features is fundamental to human based interpretation, where our job may be as "simple" as identifying and picking horizons and faults, or more advanced such as the delineation of channels, mass transport complexes, carbonate buildups, or potential gas accumulations. The use of computer-assisted tools began soon after the development of seismic attributes in the 1970s (Balch, 1971; Taner et al., 1979), with the work by Sonneland (1983) and Justice et al. (1985) being two of the first. After decades of development, pattern recognition tools available to seismic interpreters have grown tremendously, so that it is infeasible for interpreters to try all

available methods before coming to a solution. Therefore, a comprehensive review on the popular contemporary pattern recognition methods is necessary and beneficial to the interpretation community. In this dissertation, using the same dataset, I provide in Chapter 2 a comparison of unsupervised and supervised learning techniques that have been implemented for seismic data, aiming to help interpreters make an educated decision when selecting a pattern recognition method. I will discuss K-means, self-organizing map (SOM), and generative topographic mapping (GTM) as candidates for unsupervised learning, as well as artificial neural network and support vector machine (SVM) as candidates for supervised learning.

Because all these methods are just generic mathematic tools that solve highly abstracted, analytical problems, we have to tailor them to fit seismic interpretation problems. SOM is probably the most commonly used unsupervised learning method in seismic interpretation that provides good performance with high efficiency. SOM projects multiattribute data from a high dimensional space (we define the number of dimension by the number of attributes) to a low dimensional (usual 2D) space, and defines clusters in the 2D space. In this dissertation, I introduce three improvements over the traditional SOM so that it handles seismic data more appropriately.

Firstly, traditional SOM only preserves topology during the projection, which means after projection, it provides the order of similarity among clusters, but loses the proportion of similarity. In other words, the distance between two clusters in a 2D SOM space cannot properly represent the distance between two multiattribute data vectors that belong to those two clusters in the original multiattribute space. Such lack of distance information may confuse interpreters in that the facies on a SOM facies map may be too

separated or not separated enough. To address this issue, I adopt a distance-preserving step in SOM, which helps to define facies more appropriately. I demonstrate the value of such modifications by using an example on mapping a turbidite channel system. Compared with traditional SOM, I am able to better differentiate specific architectural elements.

Secondly, traditional SOM defines facies only using input attribute response, and is spatially/temporally unaware. The pitfall in being spatially and temporally unaware is the potential of mixing geology across different formations during the SOM process. Adding information of stratigraphy (sedimentary cycle), which provides temporal (or spatial, if seismic data are in depth domain) constraint on the vertical axis, may help define layers that are otherwise not well defined by seismic attributes. In this dissertation, I adopt the workflow described in Li et al. (2016) to derive a sedimentary cycle model using a mode decomposition method, and use this model as a constraint on SOM facies analysis. I test the proposed stratigraphy constrained SOM to a Barnett Shale survey, with the objective of recovering more subtle lithologic variations than using the unconstrained SOM.

Thirdly, all such multiattribute SOM analyses and subsequent improvements share one implicit assumption: input attributes are all independent and have the same contribution/importance to the SOM facies. Often times, interpreters qualitatively choose input attributes for multiattribute facies analysis based on their experience, but once an input attribute group is determined, every attribute is treated equally. Treating all input attributes equally is the somewhat "normal" practice in pattern recognition; unfortunately, seismic attributes are not created equally, and the importance of an attribute varies greatly

with exploration perspective. In this dissertation, I introduce a novel attribute selection approach for unsupervised seismic facies analysis. I assume that candidate attributes are chosen by an experienced interpreter. Once chosen, rather than assuming the selected attributes contribute equally to the facies map, I weight them based on their response from the unsupervised learning algorithm used to generate the facies map. as well as interpreter's preference. Applying this strategy to the Barnett Shale, I find that the weighted attribute selection method better differentiates facies, retaining features seen on the previously insufficiently weighted attributes.

I structure the dissertation as follows. In Chapter 2, I will provide an in-depth review and comparison of some of the most popular pattern recognition techniques that have been applied to seismic data. Then I will introduce the three improvements that I made on the traditional SOM algorithm to better fit problems in seismic exploration. In Chapter 3, I will introduce the distance-preserving SOM using a case study on data from offshore New Zealand. I will then discuss stratigraphy constrained SOM using an example from the Barnett Shale, United States in Chapter 4. In Chapter 5, I will demonstrate the previously mentioned attribute selection scheme for SOM, which incorporates both SOM response and the interpreter's knowledge. Finally, I sum up the materials from Chapter 2 to 5 with conclusions in Chapter 6.

CHAPTER 1 FIGURES

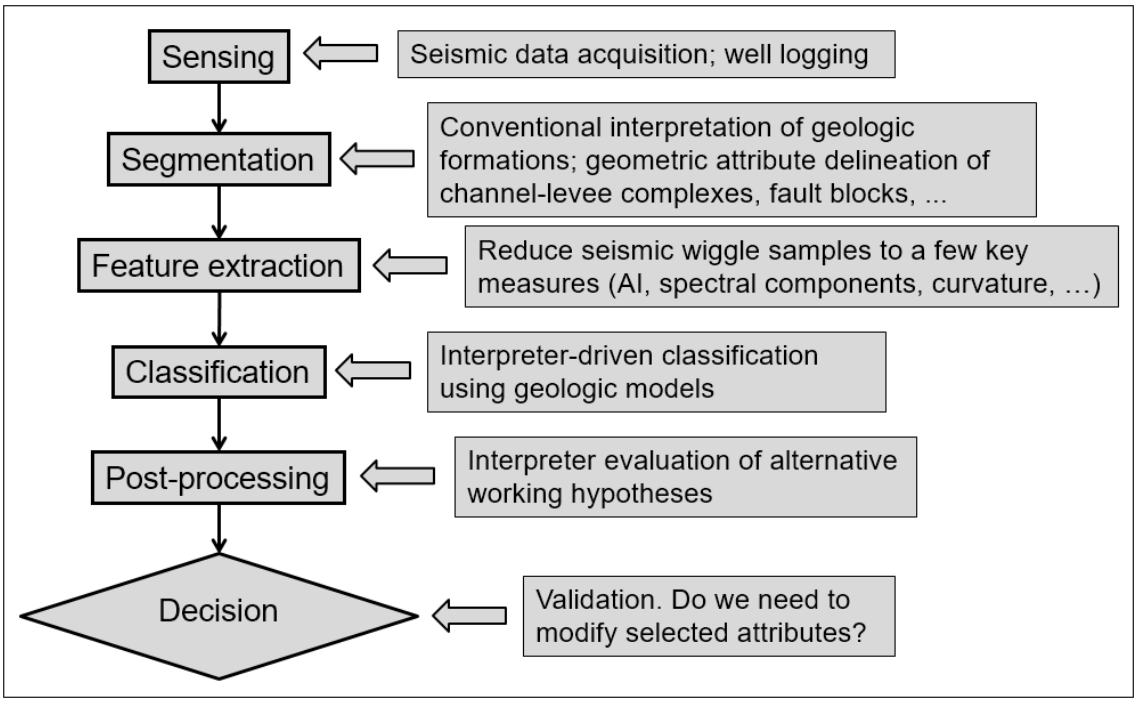


Figure 1.1. Pattern recognition as applied to the interpretation of seismic facies (Modified from Duda et al., 2000).

REFERENCES

Balch, A. H., 1971, Color sonagrams: A new dimension in seismic data interpretation: Geophysics, **36,** 1074-1098.

Duda, R.O., P. E. Hart, and D. G. Stork, 2001, Pattern Classification, 2nd Edition, John Wiley & Sons, New York, United States.

Eagleman, D., 2012, Incognito: The secret lives of the brain: Pantheon Books, New York, USA.

Justice, J. H., D. J. Hawkins, and D. J. Wong, 1985, Multidimensional attribute analysis and pattern recognition for seismic interpretation: Pattern Recognition, **18**, 391-407.

Li, F., T. Zhao, X. Qi, K. Marfurt, and B. Zhang, 2016, Lateral consistency preserved variational mode decomposition (VMD): 86th Annual International Meeting, SEG, Expanded Abstracts, 1717–1721,

Sonneland, L., 1983, Computer aided interpretation of seismic data: 53rd Annual International Meeting, SEG, Expanded Abstracts, 546-549.

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

Turing, A.M., 1950, Computing machinery and intelligence: Mind, **59**, 433-460.

CHAPTER 2

A REVIEW OF PATTERN RECOGNITION TECHNIQUES FOR SEISMIC FACIES ANALYSIS[2]

INTRODUCTION

As briefly introduced in Chapter 1, the amount of pattern recognition techniques that available at interpreters' disposal is so huge that only the specialized experts are able to manipulate these techniques on seismic data confidently. Therefore, a comparison of the popular techniques deems necessary and important for average interpreters to embrace these powerful tools. In this chapter, I introduce three unsupervised learning techniques (K-means, SOM, and GTM) and two supervised learning techniques (neural network and SVM) with their mathematical backgrounds, and apply these techniques to a same seismic dataset from offshore New Zealand. By directly comparing the results using the same dataset, I am able to deliver the merits and demerits for each algorithm.

K-means (Forgy, 1965; Jancey, 1966) was one of the earliest clustering algorithms developed, and was quickly applied by service companies and today is common to almost all interpretation software packages. K-means is an unsupervised learning algorithm in that the interpreter provides no prior information other than the selection of attributes and the number of desired clusters.

Barnes and Laughlin (2002) reviewed several unsupervised learning techniques, including K-means, fuzzy clustering, and SOM. Their primary finding was that the clustering algorithm used was less important than the choice of attributes used. Among

---

*[2] This study is published as - Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58.*

the clustering algorithms, they favored SOM since there is topologically ordered mapping of the clusters with similar clusters lying adjacent to each other on a manifold and in the associated latent space. In the examples to be discussed, a "manifold" is a deformed 2D surface that best fits the distribution of N attributes lying in an N-dimensional attribute space. The clusters are then mapped to a simpler 2D rectangular "latent" (Latin for "hidden") space upon which the interpreter can either interactively define clusters or simply map the projections using a 2D color map. A properly chosen latent space can help identify data properties that are otherwise difficult to observe in the original input space. Coleou et al.'s (2003) seismic "waveform classification" algorithm is implemented using SOM, where the "attributes" are seismic amplitudes that lie on a suite of 16 phantom horizon slices. Each (x,y) location in the analysis window provides a 16-dimensional vector of amplitudes. When plotted one element after the other, the mean of each cluster in 16-dimensional space looks like a waveform. These waveforms lie along a 1D deformed string (the manifold) that lies in 16D. This 1D string is then mapped to a 1D line (the latent space) which in turn is mapped against a 1D continuous color bar. The proximity of like waveforms to each other on the manifold and latent spaces results in similar seismic facies appearing as similar colors. Coleou et al. (2003) also generalized their algorithm to attributes other than seismic amplitude, constructing vectors of dip magnitude, coherence, and reflector parallelism. Strecker and Uden (2002) were perhaps the first to use 2D manifolds and 2D latent spaces with geophysical data, using multidimensional attribute volumes to form N-dimensional vectors at each seismic sample point. Typical attributes included envelope, bandwidth, impedance, AVO slope and intercept, dip magnitude, and coherence. These attributes were projected onto a 2D

latent space and their results plotted against a 2D color table. Gao (2007) applied a 1D SOM to GLCM texture attributes to map seismic facies offshore Angola. Overdefining the clusters with 256 prototype vectors, he then used 3D visualization and his knowledge of the depositional environment to map the "natural" clusters. These natural clusters were then calibrated using well control, giving rise to what is called a posteriori supervision. Roy et al. (2013) built on these concepts and developed an SOM classification workflow of multiple seismic attributes computed over a deep-water depositional system. They calibrated the clusters a posteriori using classical principles of seismic stratigraphy on a subset of vertical slices through the seismic amplitude. A simple but very important innovation was to project the clusters onto a 2D nonlinear Sammon space (Sammon, 1969). This projection was then colored using a gradational 2D color-scale like that of Matos et al. (2009) thus facilitating the interpretation. Roy et al. (2013) introduced a Euclidean distance measure to correlate predefined unsupervised clusters to average data vectors about interpreter defined well log facies.

Generative topographic mapping (GTM) is a more recent unsupervised classification innovation, providing a probabilistic representation of the data-vectors in the latent space (Bishop et al., 1998). There has been very little work on the application of GTM technique to seismic data and exploration problems. Wallet et al. (2009) are probably the first to apply the GTM technique to seismic data, using a suite of phantom horizon slices through a seismic amplitude volume generating a "waveform classification". While generating excellent images, Roy et al. (2013, 2014) found the introduction of well control to SOM classification to be somewhat limited, and instead applied generative topographic mapping (GTM) to Mississippian tripolitic chert reservoir

in the Midcontinent USA and a carbonate wash play in the Sierra Madre Oriental of Mexico. They found that GTM provided not only the most likely cluster associated with a given voxel, but also the probability that that voxel belongs each of clusters, providing a measure of confidence or risk in the prediction.

K-means, SOM, and GTM are all unsupervised learning techniques, where the clustering is driven only by the choice of input attributes and the number of desired clusters. If we wish to teach the computer to mimic the facies identification previously chosen by a skilled interpreter, or link seismic facies to electro-facies interpreted using wireline logs, we need to introduce "supervision" or external control to the clustering algorithm. The most popular means of supervised learning classification are based on artificial neural networks (ANN). Meldahl et al. (1999) used seismic energy and coherence attributes coupled with interpreter control (picked seed points) to train a neural network to identify hydrocarbon chimneys. West et al. (2002) used a similar workflow where the objective was seismic facies analysis of a channel system and the input attributes were textures. Corradi et al. (2009) used GLCM (gray level co-occurrence matrix) textures and ANN, with controls based on wells and skilled interpretation of some key 2D vertical slices to map sand, evaporate, and sealing vs. non-sealing shale facies offshore west Africa.

Support vector machine (SVM, where the word "machine" is due to Turing's (1950) mechanical decryption machine) is a more recent introduction to (e.g. Li and Castagna, 2004; Kuzma and Rector, 2004, 2005; Zhao et al., 2005; Al-Anazi and Gates, 2010). Originating from maximum margin classifiers, SVMs have gained great popularity for solving pattern classification and regression problems since the concept of a "soft

margin" was first introduced by Cortes and Vapnik (1995). SVMs map the N-dimensional input data into a higher dimensional latent (often called feature) space, where clusters can be linearly separated by hyperplanes. Detailed description on SVMs can be found in Cortes and Vapnik (1995), Cristianini and Shawe-Taylor (2000), and Schölkopf and Smola (2002). Li and Castagna (2004) used SVM to discriminate alternative AVO responses while Zhao et al. (2014) and Zhang et al. (2015) used a variation of SVM using mineralogy logs and seismic attributes to predict lithology and brittleness in a shale resource play.

I begin the remaining of this chapter by providing a summary of the more common clustering techniques used in seismic facies classification, emphasizing their similarities and differences. I start from the unsupervised learning K-means algorithm, progress through projections onto principal component hyperplanes, and end with projections onto SOM and GTM manifolds, which are topological spaces that resemble Euclidean space near each point. Next, I provide a summary of supervised learning techniques including artificial neural networks and support vector machines. Given these definitions, I apply each of these methods to identify seismic facies in the same data volume acquired in the Canterbury Basin, New Zealand. I conclude this chapter with a discussion on the advantages and limitations of each method and areas for future algorithm development and workflow refinement. At the very end, I also provide an appendix containing some of the mathematical details to better quantify how each algorithm works.

REVIEW OF UNSUPERVISED LEARNING TECHNIQUES

Crossplotting

Crossplotting one or more attributes against each other is an interactive and perhaps the most common clustering technique. In its simplest implementation, one computes and then displays a 2D histogram of two attributes. In most software packages, the interpreter then identifies a cluster of interest and draws a polygon around it. While several software packages allow crossplotting of up to three attributes, crossplotting more than three attributes quickly becomes intractable. One workflow to address this visualization limitation is to first project a high number of attributes onto the first two or three eigenvectors, and then crossplot the principal components. Principal components will be discussed later in the section on projection methods.

K-means clustering

K-means (MacQueen, 1967) is perhaps the simplest clustering algorithm and is widely available in commercial interpretation software packages. The method is summarized in the cartoons shown in Figure 2.1. One drawback of the method is that the interpreter needs to define how many clusters reside in the data. Once the number of clusters is defined, the cluster means or centers are defined either on a grid or randomly to begin the iteration loop. Since attributes have different units of measurement (e.g. Hz for peak frequency, 1/km for curvature, and mV for RMS amplitude) the distance of each data point to the current means are computed by scaling the data by the inverse of the covariance matrix, giving us the "Mahalanobis" distance (see Appendix). Each data point is then assigned to the cluster to whose mean it is closest. Once assigned, new cluster

means are computed from the newly assigned data clusters and the process repeated. If there are Q clusters, the process will converge in about Q iterations.

K-means is fast and easy to implement. Unfortunately, the clustering has no structure such that there is no relationship between the cluster numbering (and therefore coloring) and the proximity of one cluster to another. This lack of organization can result in similar facies appearing in totally different colors, confusing the interpretation. Tuning the number of clusters to force similar facies into the same cluster is a somewhat tedious procedure that also decreases the resolution of the facies map.

Projection Techniques

Although not defined this way in the pattern recognition literature, since this is a review, I will lump the following methods, principal component analysis (PCA), self-organizing maps, and generative topographic maps together and call them "projection techniques". Projection techniques project data residing in a higher dimensional space (say a 5D space defined by five attributes) onto a lower dimensional space (say a 2D plane or deformed 2D surface). Once projected, the data can be clustered in that space by the algorithm (such as SOM) or interactively clustered by the interpreter by drawing polygons (routine for PCA, and the preferred analysis technique for both SOM and GTM).

Principal Component Analysis

Principal component analysis is widely used to reduce the redundancy and excess dimensionality of the input attribute data. Such reduction is based on the assumption that most of the signals are preserved in the first few principle components (eigenvectors), while the last principal components contain uncorrelated noise. In this study, I use PCA

as the first iteration of the SOM and GTM algorithms. Many workers use PCA to reduce redundant attributes into "meta attributes" to simplify the computation. The first eigenvector is a vector in N-dimensional attribute space that best represents the attribute patterns in the data. Cross-correlating (projecting) the N-dimensional data against the first eigenvector at each voxel gives us the first principal component volume. If we scale the first eigenvector by the first principal component and subtract it from the original data vector, we obtain a residual data vector. The second eigenvector is that vector that best represents the attribute patterns in this residual. Cross-correlating (projecting) the second eigenvector against either the original data or residual data vector at each voxel gives us the second principal component volume. This process continues for all N-dimensions resulting in N eigenvectors and N principal components. In this study, I limit myself to the first two eigenvectors which thus define the plane that least-squares fits the N-dimensional attribute data. Figure 2.2c shows a numerical example of the first two principle components defining a plane in a 3-dimensional data space.

Self-organizing maps

While many workers (e.g. Coleou et al., 2003) describe SOM as a type of neural network, for the purposes of this review, I prefer to describe SOM as a manifold projection technique. Kohonen (1982) SOM, originally developed for gene pattern recognition, is one of the most popular classification techniques, and it has been implemented in at least four commercial software packages for seismic facies classification. The major advantage of SOM over K-means is that the clusters residing on the deformed manifold in N-dimensional data space are directly mapped to a rectilinear

or otherwise regularly gridded latent space. I provide a brief summary of the mathematical formulations of the SOM implementation used in this study in the Appendix.

Although SOM is one of the most popular classification technique there are several limitations to the SOM algorithm. First, the choice of neighborhood function at each iteration is subjective, with different choices resulting in different solutions. Second, the absence of a quantitative error measure does not let us know whether the solution has converged to an acceptable level, thus providing confidence in the resulting analysis. Third, while we find the most likely cluster for a given data vector, we have no quantitative measure of confidence in the facies classification, and no indication if the vector could be nearly as well represented by other facies.

Generative topographic mapping

GTM is a nonlinear dimensionality reduction technique that provides a probabilistic representation of the data vectors on a lower L-dimensional deformed manifold that is in turn mapped to an L-dimensional latent space. While SOM seeks the node or prototype vector that is closest to the randomly chosen vector from the training or input dataset, in GTM each of the nodes lying on the lower dimensional manifold provides some mathematical support to the data and is considered to be to some degree "responsible" for the data vector (Figure 2.3). The level of support or "responsibility" is modeled with a constrained mixture of Gaussians. The model parameter estimations are determined by maximum likelihood using the Expectation Maximization (EM) algorithm (Bishop et al., 1998).

Because GTM theory is deeply rooted in probability, it can also be used in modern risk analysis. We can extend the GTM application in seismic exploration by projecting

the mean posterior probabilities of a particular window of multiattribute data (say, about a producing well) onto the 2D latent space. By project the data vector at any given voxel onto the latent space, we obtain a probability estimates of whether it falls into the same category (Roy et al., 2014). We thus have a probabilistic estimate of how similar any data vector is to attribute behavior (and hence facies) about a producing or non-producing well of interest.

Other Unsupervised Learning Methods

There are many other unsupervised learning techniques, several of which were evaluated by Barnes and Laughlin (2002). I do not currently have access to software to apply independent component analysis and Gaussian mixture models to seismic facies classification problem, but mention them as possible candidates.

Independent component analysis

Like PCA, independent component analysis (ICA) is a statistical technique used to project a set of N-dimensional vectors onto a smaller L-dimensional space. Unlike PCA which is based on Gaussian statistics, whereby the first eigenvector best represents the variance in the multidimensional data, ICA attempts to project data onto subspaces that result in non-Gaussian distributions which are then easier to separate and visualize. Honorio et al. (2014) successfully apply ICA to multiple spectral components to delineate architectural elements of an offshore Brazil carbonate terrain. Both PCA and ICA are commonly used to reduce a redundant set of attributes to form a smaller set of independent meta-attributes (e.g. Gao, 2007).

Gaussian mixture models

Gaussian mixture model (GMM), are parametric models of probability distributions which can provide greater flexibility and precision in modeling than traditional unsupervised clustering algorithms. Lubo et al. (2014) apply this technique to a suite of well logs acquired over Horseshoe Atoll, west Texas, to generate different lithologies. These GMM lithologies are then used to calibrate 3D seismic prestack inversion results to generate a 3D rock property model. At present, I do not know of any GMM algorithms applied to seismic facies classification using seismic attributes as input data.

REVIEW OF SUPERVISED LEARNING TECHNIQUES

Artificial Neural Networks

Artificial neural networks can be used in both unsupervised and supervised mulitattribute analysis (van der Baan and Jutten, 2000). The multilayer perceptron (MLP) and the radial basis function (RBF) are two popular types of neural networks used in supervised learning. Probabilistic neural network, PNN, which also uses radial basis functions, forms the basis of additional neural network geophysical applications. In terms of network architecture, the supervised algorithms are feed-forward networks. In contrast, the unsupervised SOM algorithm described earlier is a recurrent (or feed-backward) network. An advantage of feed-forward networks over SOMs is the ability to predict both continuous values (such as porosity) as well as discrete values (such as facies class number). Applications of neural networks can be found in seismic inversion (Röth and Tarantola, 1994), well log prediction from other logs (Huang et al., 1996; Lim, 2005), waveform recognition (Murat and Rudman, 1992), seismic facies analysis (West et al.,

2002), and reservoir property prediction using seismic attributes (Yu et al., 2008; Zhao and Ramachandran, 2013). For the last application listed above, however, due to the resolution difference between seismic and well logs, structural and lithologic variation of inter-well points, and the highly nonlinear relation between these two domains, achieving a convincing prediction result can be challenging. In this case, geostatistical methods such as Bayesian analysis can be used jointly to provide a probability index, giving interpreters an estimate of how much confidence they should have in the prediction.

Artificial neural networks are routinely used in the exploration and production industry. ANN provides a means to correlate well measurements such as gamma ray logs to seismic attributes (e.g. Verma, 2012) where the underlying relationship is a function of rock properties, depositional environment, and diagenetic alteration. Although it has produced reliable classification in many applications during its service, defects such as converging to local minima and difficult in parameterization are not negligible. In both industrial and scientific applications, we prefer a constant and robust classifier once the training vectors and model parameters have been determined. This leads to the more recent supervised learning technique developed in the late 20^{th} century, the support vector machines.

Support Vector Machines

The basic idea of SVMs is straightforward. First, we transform the training data vectors into a still higher dimensional "feature" space using nonlinear mapping. Then we find a hyperplane in this feature space that separates the data into two classes with an optimal "margin". The concept of a margin is defined to be the smallest distance between the separation hyperplane (commonly called a decision boundary) and the training vectors

(Bishop, 2006) (Figure 2.4). An optimal margin balances two criteria: maximizing the margin, thereby giving the classifier the best generalization, and minimizing the number of misclassified training vectors if the training data are not linearly separable. The margin can also be described as the distance between the decision boundary and two hyperplanes defined by the data vectors which have the smallest distance to the decision boundary. These two hyperplanes are called the "plus-plane" and the "minus-plane". The vectors which lie exactly on these two hyperplanes mathematically define or "support" them and are called support vectors. Tong and Koller (2002) show that the decision boundary is dependent solely on the support vectors, resulting in the name "support vector machines".

SVMs can be used in either a supervised or in a semi-supervised learning mode. In contrast to supervised learning, semi-supervised learning defines a learning process that utilizes both labeled and unlabeled vectors. When there are a limited number of interpreter classified data vectors, the classifier may not act well due to insufficient training. In semi-supervised training, some of the nearby unclassified data vectors are automatically selected and classified based on a distance measurement during the training step, as in an unsupervised learning process. These vectors are then used as additional training vectors (Figure 2.5), resulting in a classifier that will perform better for the specific problem. The generalization power is sacrificed by using unlabeled data. In this study I focus on SVM; however, the future of semi-supervised SVM in geophysical applications is quite promising.

Proximal Support Vector Machines

Proximal support vector machine (PSVM) (Fung and Mangasarian, 2001, 2005) is a recent variant of SVM, which, instead of looking for a separating plane directly,

builds two parallel planes that approximate two data classes; the decision-boundary then falls between these two planes (Figure 2.6). Other researchers have found that PSVM provides comparable classification correctness to standard SVM but at considerable computational savings (Fung and Mangasarian, 2001, 2005; Mangasarian and Wild, 2006). In this study, I use PSVM as the implementation of SVM. Details on the PSVM algorithm are provided in the Appendix.

We may face problems in seismic interpretation that are linearly inseparable in the original input multidimensional attribute space. In SVM, we map the data vectors into a higher dimensional space where they become linearly separable (Figure 2.7), where the increase in dimensionality may result in significantly increased computational cost. Instead of using an explicit mapping function to map input data into a higher dimensional space, PSVM achieves the same goal by manipulating a kernel function in the input attribute space. In this implementation, I use a Gaussian kernel function, but in principal many other functions can be used (Shawe-Taylor and Cristianini, 2004).

SVM can be used either as a classifier or as a regression operator. Used as a regression operator, SVM is capable of predicting petrophysical properties such as porosity (Wong et al., 2005), V_p, V_s and density (Kuzma and Rector, 2004), and permeability (Al-Anazi and Gates, 2010; Nazari et al., 2011). In all such applications, SVM shows comparable or superior performance to neural networks with respect to prediction error and training cost. When used as a classifier, SVM is suitable in predicting lithofacies (Al-Anazi and Gates, 2010; Torres and Reveron, 2013; Wang et al., 2014; Zhao et al., 2014) or pseudo rock properties (Zhang et al., 2015), either from well log data, core data, or seismic attributes.

GEOLOGIC SETTING

In this study I use the Waka-3D seismic survey acquired over the Canterbury Basin, offshore New Zealand, generously made public by New Zealand Petroleum and Minerals. Readers can request this data set through their website for research purposes. Figure 2.8 shows the location of this survey, where the red rectangle corresponds to time slices shown in subsequent figures. The study area lies on the transition zone of continental slope and rise, with abundance of paleocanyons and turbidite deposits of Cretaceous and Tertiary ages. These sediments are deposited in a single, tectonically driven transgressive – regressive cycle (Uruski, 2010). Being a very recent and underexplored prospect, publically available comprehensive studies of the Canterbury Basin are somewhat limited. The modern seafloor canyons shown in Figure 2.8 are good analogs of the deeper paleocanyons illuminated by the 3D seismic amplitude and attribute data.

ATTRIBUTE SELECTION

In their comparison of alternative unsupervised learning techniques, Barnes and Laughlin (2002) concluded that the appropriate choice of attributes was the most critical component of computer assisted seismic facies identification. Although interpreters are skilled at identifying facies, such recognition is often subconscious and hard to define (see Eagleman's 2012 discussion on differentiating male from female chicks and identifying military aircraft from silhouettes). In supervised learning, the software does some of the work during the training process, though we must always be wary of false correlations if we provide too many attributes (Kalkomey, 1999). For the prediction of continuous data such as porosity, Russell (1997) and others suggest that one begin with

exploratory data analysis, where one simply cross-correlates a candidate attribute with the desired property at the well. Such cross−correlation does not work well when trying to identify seismic facies, which are simply "labeled" with an integer number or alphanumeric name.

Table 2.1 summarizes how interpreters perceive each of the seismic facies of interest. Once we have enumerated the seismic expression, the quantification using attribute expression is relatively straightforward. In general, amplitude and frequency attributes are lithology indicators and may provide direct hydrocarbon detection in conventional reservoirs, geometric attributes delineate reflector morphology such as dip, curvature, rotation, and convergence, while statistical and texture attributes provides information about data distribution that quantifies subtle patterns that are hard to define (Chopra and Marfurt, 2007). Attributes such as coherence provide images of the edges of seismic facies rather than a measure of the facies themselves, although slumps often appear as a suite of closely spaced faults separating rotated fault blocks. Finally, what we see as interpreters and what our clustering algorithms see can be quite different. While we may see a slump feature as exhibiting a high number of faults per km, the clustering algorithms are applied voxel by voxel and see only the local behavior. Extending the clustering to see such large scale textures requires the development of new texture attributes.

The number of attributes should be as small as possible to discriminate the facies of interest, and each attribute should be mathematical independent from the others. While it may be fairly easy to represent three attributes with a deformed 2D manifold, increasing the dimensionality results in increased deformation, such that the manifold may fold on

itself or may not accurately represent the increased data variability. Because the Waka-3D survey is just recently released to the public, interpreters have tested numerous attributes that they think may highlight different facies in the turbidite system. Among these attributes, I find the shape index to be good for visual classification but dominates the unsupervised classifications with valley and ridge features across the survey. After such analysis I chose four attributes that are mathematically independent but should be coupled through the underlying geology: peak spectral frequency, peak spectral magnitude, GLCM homogeneity, and curvedness, as the input to the classifiers. The peak spectral frequency and peak spectral magnitude form an attribute pair that crudely represents the spectral response. Peak frequency of spectrally whitened data is sensitive to tuning thickness while peak magnitude is a function of both tuning thickness and impedance contrast. GLCM homogeneity is a texture attribute that has a high value for adjacent traces with similar (high or low) amplitudes and measures the continuity of a seismic facies. Curvedness defines the magnitude of reflector structural or stratigraphic deformation, with dome-, ridge-, saddle-, valley-, and bowl-shaped features exhibiting high curvedness and planar features exhibiting zero curvedness.

Figure 2.9 shows a time slice at t=1.88 s through the seismic amplitude volume on which I identify channels (white arrows), high amplitude deposits (yellow arrows), and slope fans (red arrows). Figure 2.10 shows an equivalent time slice through peak spectral frequency co-rendered with peak spectral magnitude that emphasizes the relative thickness and reflectivity of the turbidite system and surrounding slope fan sediments into which it was incised. The edges of the channels are delineated by Sobel filter similarity. I show equivalent time slices through (Figure 2.11) GLCM homogeneity, and (Figure

2.12) co-rendered shape index and curvedness. In Figure 2.13 I show a representative vertical slice at line AA' in Figure 2.14 cutting through the channels through (Figure 2.13a) seismic amplitude, (Figure 2.13b) seismic amplitude co-rendered with peak spectral magnitude/peak spectral frequency, (Figure 2.13c) seismic amplitude co-rendered with GLCM homogeneity, and (Figure 2.13d) seismic amplitude co-rendered shape index and curvedness. White arrows indicate incised valleys, yellow arrows high amplitude deposits, and red arrows a slope fan. We note several of the incised values are visible at time slice t=1.88 s.

In a conventional interpretation workflow, the geoscientist would examine each of these attribute images and integrate them within a depositional framework. Such interpretation takes time and may be impractical for extremely large data volumes. In contrast, in seismic facies classification the computer either attempts to classify what it sees as distinct seismic facies (in unsupervised learning) or attempts to emulate the interpreter's classification made on a finite number of vertical sections, time, and/or horizon slices and apply the same classification to the full 3D volume (in supervised learning). In both cases, the interpreter needs to validate the final classification to determine if they represent seismic facies of interest. In this example I will use Sobel filter similarity to separate the facies and then evaluate how they fit within my understanding of a turbidite system.

APPLICATION

Given these four attributes, we now construct four-dimensional attribute vectors as input to the previously described classification algorithms. To better illustrate the performance of each algorithm, I summarize the data size, number of computational

processors, and runtime in Table 2.2. All the algorithms are developed by the authors except ANN, which is implemented using MATLAB® toolbox.

I begin with K-means. As previously discussed, a limitation of K-means is the lack of any structure to the cluster number selection process. I illustrate this limitation by computing K-means with 16 (Figure 2.14) and 256 (Figure 2.15) clusters. On Figure 2.14, we can identify high amplitude overbank deposits (yellow arrows), channels (white arrows), and slope fan deposits (red arrows). A main limitation of K-means is that there is no structure linking the clusters, which leads to a somewhat random choice of color assignment to clusters. This problem becomes more serious when more clusters are selected: the result with 256 clusters (Figure 2.15) is so chaotic that we can rarely separate the overbank high amplitude deposits (yellow arrows) and slope fan deposits (red arrows) that were easily separable in Figure 2.14. For this reason, modern K-means applications focus on estimating the correct number of clusters in the data.

In contrast to K-means, SOM restricts the cluster centers to lie on a deformed 2D manifold. While clusters may move closer or further apart, they still form (in this implementation) a deformed quadrilateral mesh which maps to a rectangular mesh on the 2D latent space. Mapping the latent space to a continuous 1D (Coleou et al., 2003) or 2D color bar (Strecker and Uden, 2002), reduces the sensitivity to the number of clusters chosen. I follow Gao (2007) and avoid guessing at the number of clusters necessary to represent the data by overdefining the number of prototype vectors to be 256 (the limit of color levels in our commercial display software). These 256 prototype vectors (potential clusters) reduce to only three or four distinct “natural” clusters through the SOM neighborhood training criteria. The 2D SOM manifold is initialized using the first two

principle components, defining a plane through the N-dimensional attribute space (Figure 2.16). The algorithm then deforms the manifold to better fit the data. Overdefining the number of prototype vectors results in clumping into a smaller number natural clusters. These clumped prototype vectors project onto adjacent locations in the latent space are therefore appear as subtle shades of the same color as indicated by the limited palette of 256 colors shown in Figure 2.17. On the classification result shown on Figure 2.17, we can clearly identify the green colored spill-over deposits (yellow arrows). The difference between channel fill (white arrows) and slope fans (red arrows) is insignificant. However, by co-rendering with similarity, the channels are delineated nicely, allowing us to visually distinguish channel fills and the surrounded slope fans. We can also identify some purple color clusters (orange arrows) which I interpret to be crevasse splays at this moment.

Next, I apply GTM to the same four attributes. I compute two "orthogonal" projections of data onto the manifold and thence onto the two dimensions of the latent space. Rather than define explicit clusters, we project the mean a posteriori probability distribution onto the 2D latent space and then export the projection onto the two latent space axes. I crossplot the projections along axes 1 and 2 and map them against a 2D color bar (Figure 2.18). In this slice, we see channels delineated by purple colors (white arrows), point bar and crevasse splays in pinkish colors (yellow arrows), and slope fans in lime green colors (red arrows). We can also identify some thin, braided channels at the south end of the survey (blue arrow). Similarly to the SOM result, similarity separates the incised valleys from the slope fans. However, the geological meaning of the orange colored facies is somehow vague. This is the nature of unsupervised learning techniques in that the clusters represent topological differences in the input data vectors, which are

not necessarily the facies differences we wish to delineate. We can ameliorate this shortcoming by adding a posteriori supervision to the GTM manifold. The simplest way to add supervision is to compute the average attribute vectors about a given seismic facies and map it to the GTM crossplot. Then, the interpreter can manually define clusters on the 2D histogram by constructing one or more polygons (Figure 2.19), where I cluster the data into four facies: multistoried channels (blue), high-energy point bar and crevasse splay deposits (yellow), slope fans (green), and "everything else" (red). A more quantitative methodology is to mathematically project these average clusters onto the manifold, and then cross multiply the probability distribution of the control vectors against the probability distribution function of each data vector, thereby forming the Bhattacharya distance (Roy et al., 2013, 2014). Such measures then provide a probability ranging between 0 and 100% as to whether the data vector at any seismic sample point is like the data vectors about well control (Roy et al., 2013, 2014) or like the average data vector within a facies picked by the interpreter.

The a posteriori supervision added to GTM is the critical prior supervision necessary for supervised classification such as ANN and SVM. In this study I used the same four attributes as input for both unsupervised and supervised learning techniques. The supervision consists of picked seed points for the three main facies previously delineated using the unsupervised classification results, which are multistoried channel, point bar and crevasse splay deposits, and slope fans, plus an additional channel flank facies. The seed points are shown in Figure 2.20. Seed points should be picked with great caution to correctly represent the corresponding facies, any false picking (a seed point that does not belong to the intended facies) will greatly compromise the classification

result. I then compute averages of the four input attributes within a 7 trace X 7 trace X 24 *ms* window about each seed point to generate a training table which consists of 4-dimensional input attribute vectors and one dimensional targets (the labeled facies).

For the ANN application, I used the neural networks toolbox in MATLAB®, and generated a probabilistic neural network (PNN) composed of 20 neurons. Because of the relatively small size of the training data, the training process only took a second or so; however, since a PNN may converge to local minima, I am not confident that the first trained network has the best performance. The workflow is then to rerun the training process 50 times and choose the network exhibiting the lowest training and cross-validation errors. Figure 2.21 and 2.22 show the PNN performance during training, while Figure 2.23 shows the PNN classification result. We notice that all the training, testing, and cross-validation performance are acceptable, with training and cross-validation correctness being around 90%, and testing correctness being over 86%. We identify blue channel stories within the relatively larger scale incised valleys (white arrows), and yellow point bars and crevasse splays (yellow arrows). However, many of the slope fan deposits are now classified as channel flanks or multistoried channels (blue arrows), which need to be further calibrated with well log data. Nevertheless, as a supervised learning technique, ANN provides classification with explicit geological meaning, which is its primary advantage over unsupervised learning techniques.

Finally, I cluster the four-dimensional input data using SVM, using the same training data (interpreter picks) as for ANN. The workflow is similar to ANN in that I ran 20 passes of training, varying the Gaussian kernel standard deviation, σ, and misclassification tolerance, ε, parameters for each pass. These parameter choices are

easier than selecting the number of neurons for ANN, since the SVM algorithm solves a convex optimization problem that converges to a global minima. The training and cross-validation performance is comparable to ANN, with roughly 92% training correctness and 85% cross-validation correctness. Figure 2.24 shows the SVM classification result at time t = 1.88 s. The SVM map follows the same pattern as we have seen on the ANN map, but is generally cleaner, with some differences in details. Compared to ANN, SVM successfully mapped more of the slope fans (white arrows), but missed some crevasse splays that were correctly picked by ANN (yellow arrow). We also see a great amount of facies variation within the incised valleys, which is reasonable because of the multiple course changes of a paleochannel during its deposition that results in multiple channel stories. Finally, we note some red lines following NW-SE direction (red arrows) which correspond to acquisition footprint.

CONCLUSION AND DISCUSSION

In this study I have compared and contrasted some of the more important multiattribute facies classification tools, including four unsupervised (PCA, K-means, SOM, GTM) and two supervised (ANN, SVM) learning techniques. In addition to highlighting the differences in assumptions and implementation, I have applied each method to the same Canterbury Basin survey, with the goal of delineating seismic facies in a turbidite system to demonstrate the effectiveness and weaknesses of each method. K-means and SOM move the user-defined number of cluster centers towards the input data vectors. PCA is the simplest manifold method, where the data variability in previous examples is approximated by a 2D plane defined by the first two eigenvectors. GTM is more accurately described as a mapping technique, like PCA, where the clusters are

formed either in the human brain as part of visualization or through crossplotting and the construction of polygons. SOM and GTM manifolds deform to fit the N-dimensional data. In SOM, the cluster centers (prototype vectors) move along the manifold towards the data vectors, forming true clusters. In all four methods, any labeling of a given cluster to a given facies happens after the process is completed. In contrast, ANN and SVM build a specific relation between the input data vectors and a subset of user-labeled input training data vectors thereby explicitly labeling the output clusters to the desired facies. Supervised learning is constructed from a limited group of training samples (usually at certain well locations or manually picked seed points) which generally are insufficient to represent all the lithologic and stratigraphic variations within a relatively large seismic data volume. A pitfall of supervised learning is that unforeseen clusters will be misclassified as clusters that have been chosen.

For this reason, unsupervised classification products can be used to construct not only an initial estimate of the number of classes, but also a validation tool to determine if separate clusters have been incorrectly lumped together. I advise computing unsupervised SOM or GTM prior to picking seed points for subsequent supervised learning, to clarify the topological differences mapped by the choice of attributes. Such mapping will greatly improve the picking confidence, because the seed points are now confirmed by both human experience and mathematical statistics.

The choice of the correct suite of attributes is critical. Specifically, images that are ideal for multiattribute visualization may be suboptimal for clustering. I made several poor choices in previous iterations of writing this paper. The image of inline (SW-NE) structural dip illustrates this problem directly. While a skilled interpreter sees a great deal

of detail in Figure 2.25, there is no clear facies difference between positive and negative dips, such that this component of vector dip cannot be used to differentiate them. A better choice would be dip magnitude, except that a long wavelength overprint (such as descending into the basin) would again bias our clustering in a manner that is unrelated to facies. Therefore, I tried to use relative changes in dip – curvedness and shape indices measure lateral changes in dip, and reflector convergence which differentiates conformal from nonconformal reflectors.

Certain attributes should never be used in clustering. Phase, azimuth, and strike have circular distributions, where a phase value of -180 indicates the same value as +180. No trend can be found. While the shape index, s, is not circular, ranging between -1 and +1, the histogram has a peaks about the ridge (s=+0.5) and about the valley (s=-0.5). I speculate that shape components may be more amenable to classification. Reflector convergence follows the same pattern as curvedness. For this reason I only used curvedness as a representative of these three attributes. The addition of this choice improved the clustering.

Edge attributes like the Sobel filter similarity and coherence are not useful for the example show here; instead, I have visually added them as an edge “cluster” and co-rendered with the images shown in Figure 2.14-2.20, 2.23, and 2.24. In contrast, when analyzing more chaotic features such as salt domes and karst collapse, coherence is a good input to clustering algorithms. I do wish to provide an estimate of continuity and randomness to the clustering. To do so, I follow Corradi et al. (2009) and West et al. (2002) and use GLCM homogeneity as an input attribute.

Theoretically, no one technique is superior to all the others in every aspect, and each technique has its inherent advantages and defects. K-means with a relatively small numbers of clusters is the easiest algorithm to implement, provides rapid interpretation, but lacks the relation among clusters. SOM provides a generally more "interpreter friendly" clustering result with topological connections among clusters, but is computationally more demanding than K-means. GTM relies on probability theory and enables the interpreter to add posteriori supervision by manipulating the data's posterior probability distribution; however, it is not widely accessible to the exploration geophysicist community. Rather than displaying the conventional cluster numbers (or labels), I suggest displaying the cluster coordinates projected onto the 2D SOM and GTM latent space axes. Doing so not only provides greater flexibility in constructing a 2D color bar but also provides data that can be further manipulated using 2D crossplot tools.

For the two supervised learning techniques, ANN suffers from the convergence problem and requires expertise to achieve the optimal performance, while the computation cost is relatively low. SVM is mathematically more robust and easier to train, but is more computationally demanding.

Practically, if no software limitations are set, we can make suggestions on how an interpreter can incorporate these techniques to facilitate seismic facies interpretation at different exploration and development stages. To identify the main features in a recently acquired 3D seismic survey on which limited to no traditional structural interpretation is done, K-means is a good candidate for exploratory classification starting with a small K (typically K = 4) and gradually increase the number of class. As more data are acquired (e.g. well log data and production data) and detailed structural interpretation has been

performed, SOM or GTM focusing in the target formations will provide more refined classification, which needs to be calibrated with wells. In the development stage when most of the data have been acquired, with proper training process, ANN and SVM provide targeted products, characterizing the reservoir by mimicking interpreters' behavior. Generally, SVM provides superior classification than ANN but at a considerably higher computational cost, so choosing between these two requires balancing performance and runtime cost. As a practical manner, no given interpretation software platform provides all five of these clustering techniques, such that many of the choices are based on software availability.

Because I wish this study to serve as an inspiration of interpreters, I do want to reveal one drawback of this work: all the classifications are performed volumetrically but not along a certain formation. Such classification may be biased by the bonding formations above and below the target formation (if we do have a target formation), therefore contaminates the facies map. However, I want to make the point that such classification can happen at a very early stage of interpretation, when both structural interpretation and well logs are very limited. And even in such situation, I can still use classification techniques to generate facies volumes to assist subsequent interpretation.

In the 1970s and 1980s much of geophysical innovation in seismic processing and interpretation was facilitated by the rapid evolution of computer technology – from mainframes to minicomputers to workstations to distributed processing. I believe similar advances in facies analysis will be facilitated by the rapid innovation in "big data" analysis, driven by needs in marketing and security. While we may not answer Turing's

(1950) question “Can machines think?”, we will certainly be able to teach them how to emulate a skilled human interpreter.

CHAPTER 2 TABLES

Facies	Appearance to Interpreter	Attribute Expression
Levee	Structurally high	Stronger dome or ridge shape structural components
	locally continuous	Higher GLCM homogeneity; lower GLCM entropy
	Higher amplitude	Dome or ridge shape component
	Possibly thicker	Lower peak spectral frequency
Channel thalwegs	Shale-filled with negative compaction	Stronger bowl or valley shape structural components; higher peak spectral frequency
	Sand-filled with positive compaction	Stronger dome or ridge shape structural components; lower peak spectral frequency
Channel flanks	Onlap onto incisement, canyon edges	Higher reflector convergence magnitude
Gas-charged sands	High amplitude, continuous reflections	Higher GLCM homogeneity; lower GLCM entropy; high high peak magnitude
Incised floodplain	Erosional truncation	Higher reflector convergence magnitude, Higher curvedness
Floodplain	Lower amplitude	Lower spectral magnitude
	Higher frequency	Higher peak spectral frequency
	Continuous	Higher GLCM homogeneity; lower GLCM entropy
	Near planar events	Lower amplitude structural shape components; lower reflector convergence magnitude
Slumps	Chaotic reflectivity	Higher reflector convergence magnitude; higher spectral frequency; lower GLCM homogeneity; higher GLCM entropy

Table 2.1. Attribute expressions of seismic facies.

Algorithm	Number of classes	MPI processors*	Dataset size (samples)		Runtime (sec.)		
			Training	Total	Training	Applying to the entire dataset	Total
K-means	16	50	809,600	101,200,000	65	20	85
K-means	256	50	809,600	101,200,000	1060	70	1130
SOM	256	1	809,600	101,200,000	4125	6035	10160
GTM	-	50	809,600	101,200,000	9582	1025	10607
ANN^{+}	4	1	437	101,200,000	2	304	306
SVM	4	50	437	101,200,000	24	12092	12116

*SOM is not run under MPI in our implementation. ANN is run using MATLAB® and is not under MPI. All other three are run under MPI when applying the model to the entire dataset.
$^{+}$ANN is implemented using MATLAB® toolbox.

Table 2.2. Algorithm settings and runtimes.

Variable Name	Definition
n, N	attribute index and number of attributes
j, J	voxel (attribute vector) index and number of voxels
k, K	manifold index and number of grid points
$\mathbf{a}_j$	the j^{th} attribute data vector
$\mathbf{p}$	matrix of principle components
C	attribute covariance matrix
μ_n	mean of the n^{th} attribute
λ_m, $\mathbf{v}_m$	the m^{th} eigenvalue and eigenvector pair
$\mathbf{m}_k$	the k^{th} grid point lying on the manifold (prototype vector for SOM, or Gaussian center for GTM)
$\mathbf{u}_k$	the k^{th} grid point lying on the latent space
r_{jk}	the Mahalanobis distance between the j^{th} data vector and the k^{th} cluster center or manifold grid point
$\mathbf{I}$	Identity matrix of dimension defined in the text

Table 2.3. List of shared mathematical symbols.

CHAPTER 2 FIGURES

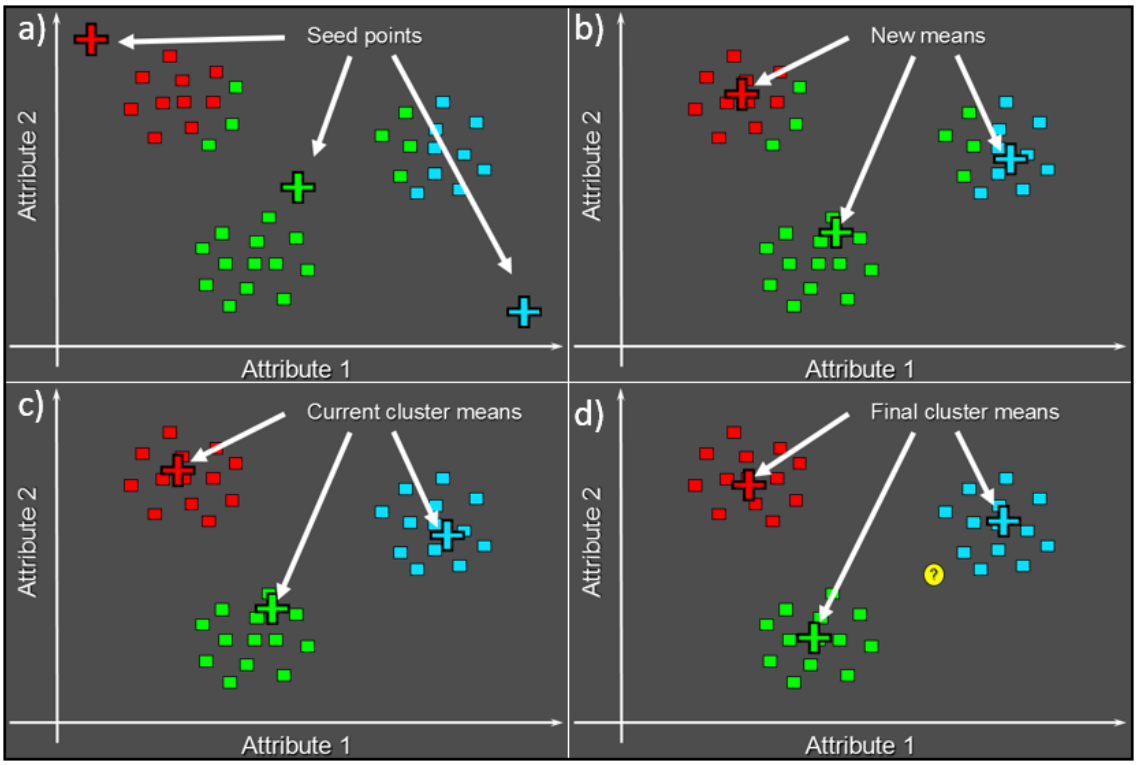


Figure 2.1. Cartoon illustration of a K-means clustering of 3 clusters. (a) Select 3 random or equally spaced, but distinct seed points, which serve as the initial estimate of the vector means of each cluster. Next, compute the Mahalanobis distance between each data vector and each cluster mean. Then color code or otherwise label each data vector to belong to the cluster that has the smallest Mahalanobis distance. (b) Recompute the means of each cluster from the previously defined data vectors. (c) Recalculate the Mahalanobis distance from each vector to the new cluster means. Assign each vector to the cluster that has the smallest distance. (d) The process continues until the changes in means converge to their final locations. If we now add a new (yellow) point, we will use a Bayesian classifier to determine into which cluster it falls (Figure courtesy of Scott Pickford).

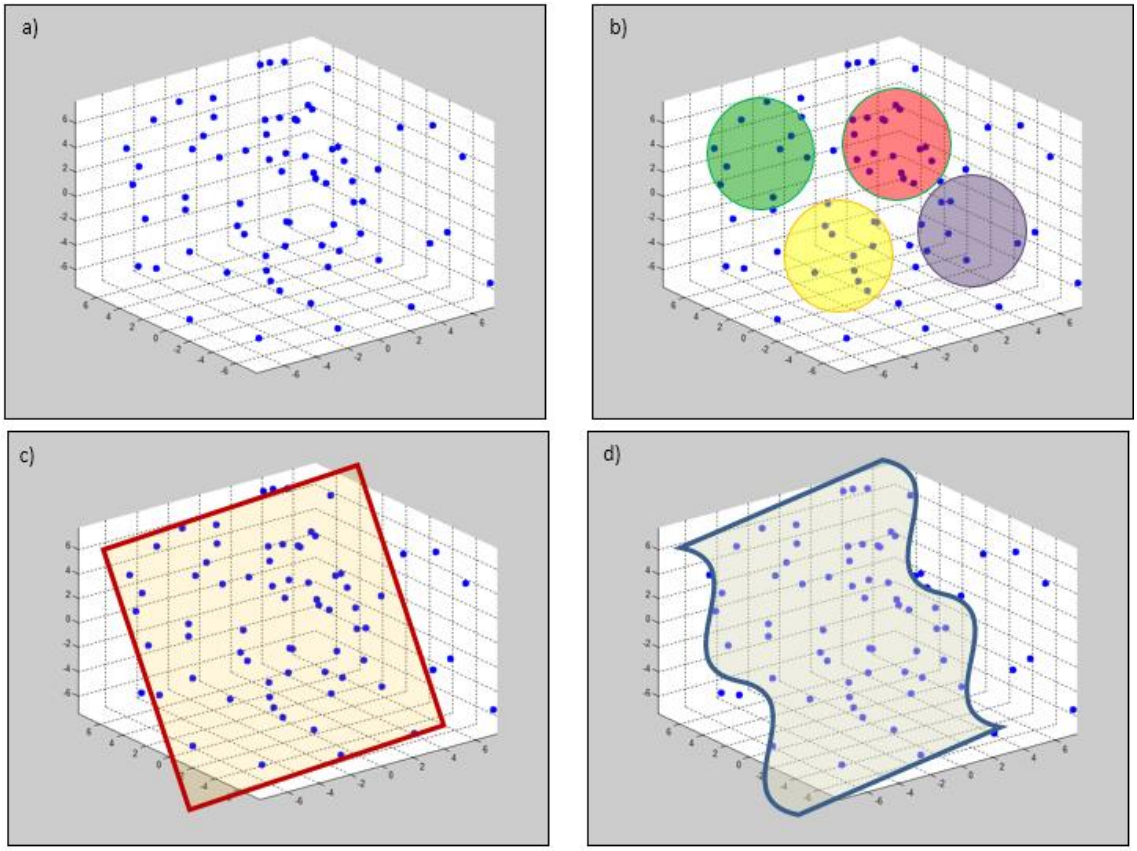


Figure 2.2. (a) A distribution of data points in 3-dimensional attribute space. The statistics of this distribution can be defined by the covariance matrix. (b) K-means will cluster data into a user-defined number of distributions (4 in this example) based on Mahalanobis distance measure. (c) The plane that best fits these data is defined by the first two eigenvectors of the covariance matrix. The projection of the 3D data onto this plane provides the first two principle components of the data as well as the initial model for both the SOM and GTM algorithms. (d) SOM and GTM deform the initial 2D plane into a 2D "manifold" that better fits the data. Each point on the deformed 2D manifold is in turn mapped to a 2D rectangular "latent" space. Clusters are color-coded or interactively defined on this latent space.

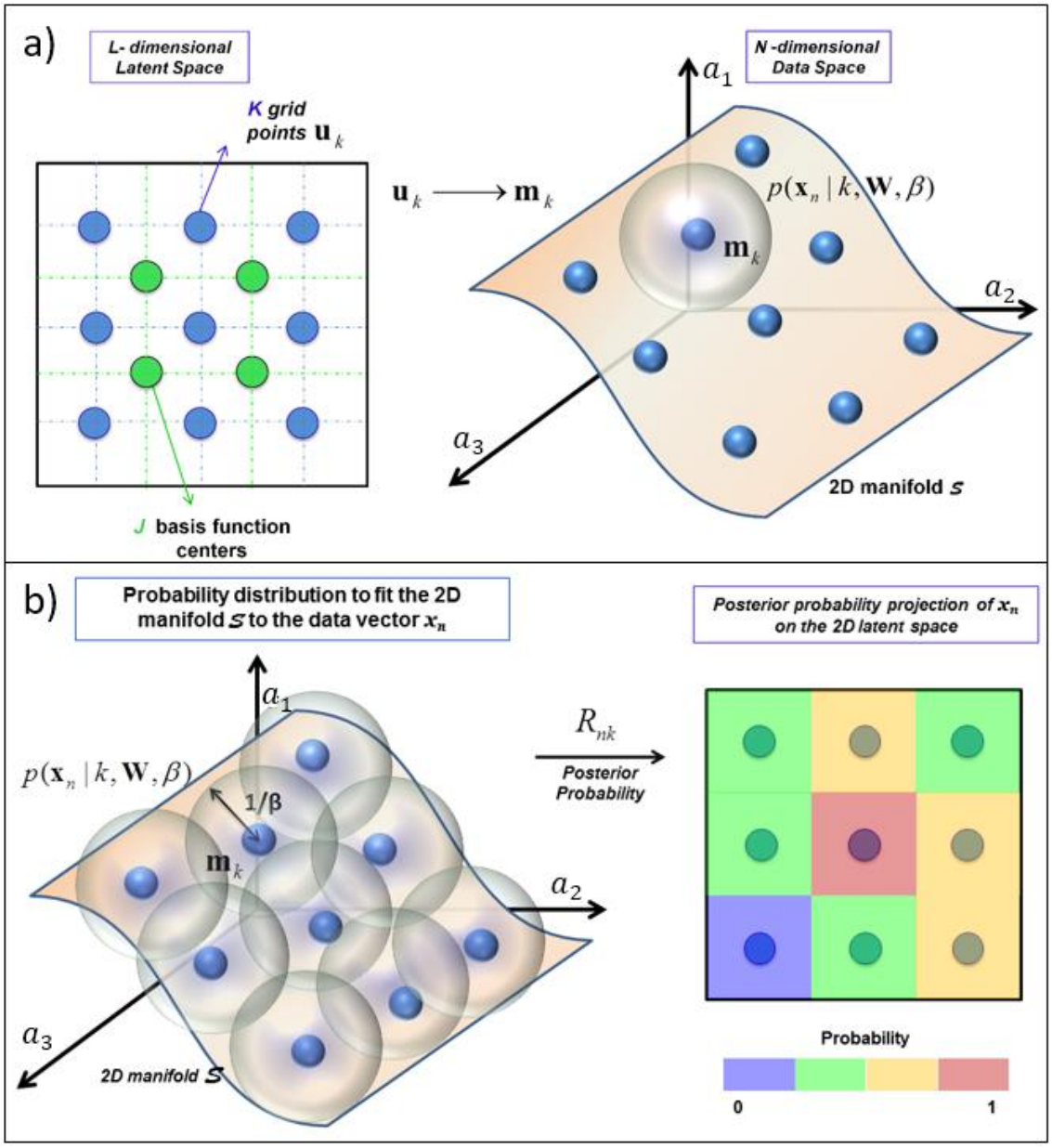


Figure 2.3. (a) K grid points $\mathbf{u}_k$ defined on a L-dimensional latent space grid are mapped to K grid points $\mathbf{m}_k$ lying on a non-Euclidean manifold in N-dimensional data space. In this paper, L=2 and will be mapped against a 2-dimensional color bar. The Gaussian mapping functions are initialized to be equally spaced on the plane defined by the first two eigenvectors. (b) Schematic showing the training of the latent space grid points to a data vector $\mathbf{a}_j$ lying near the GTM manifold using an expectation maximization algorithm. The posterior probability of each data vector is calculated for all Gaussian centroids points $\mathbf{m}_k$ and are assigned to the respective latent space grid points $\mathbf{u}_k$. Grid points with high probabilities are displayed as bright colors. All variables are discussed in Appendix.

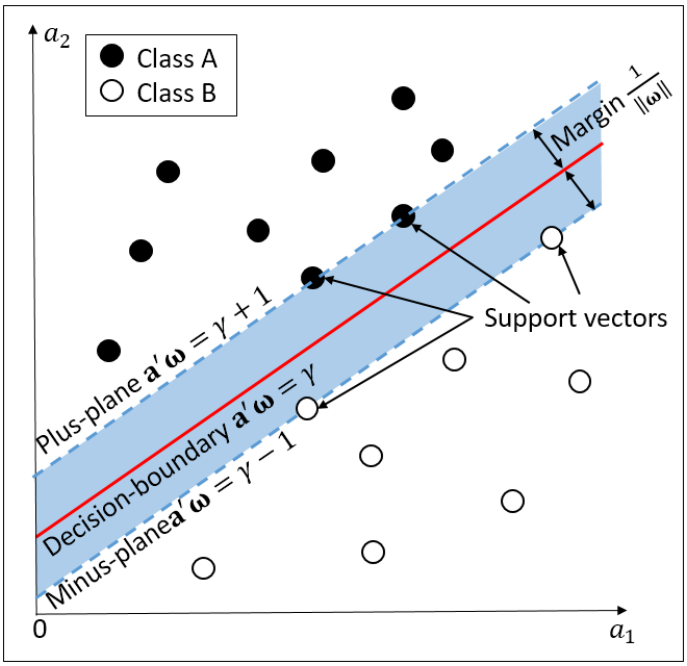


Figure 2.4. Cartoon of a linear SVM classifier separating black from white data vectors. The two dashed lines are the margins defined by support vector data points. The red decision boundary falls midway between the margins, separating the two clusters. If the data clusters overlap, no margins can be drawn. In this situation the data vectors will be mapped to a higher dimensional space where they can be separated.

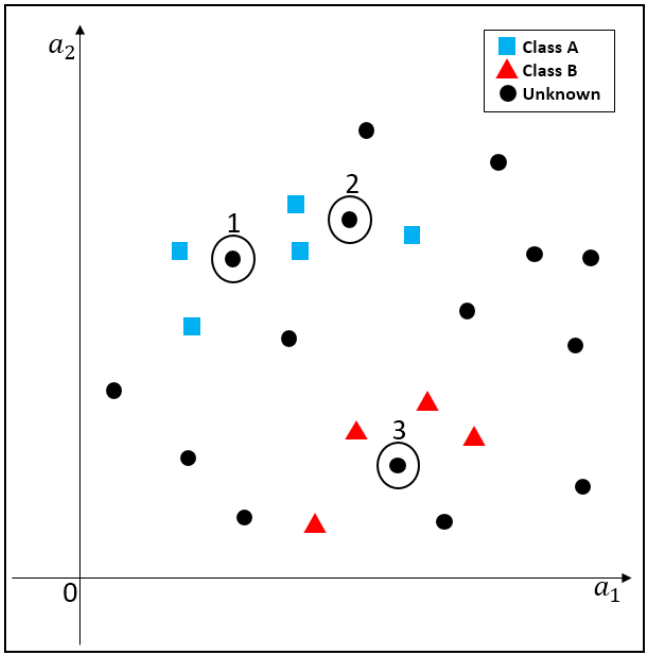


Figure 2.5. Cartoon describing semi-supervised learning. Blue squares and red triangles indicate two different interpreter defined classes. Black dots indicate unclassified points. In semi-supervised learning, unclassified data vectors 1 and 2 are classified to be class "A" while data vector 3 is classified to be class "B" during the training process.

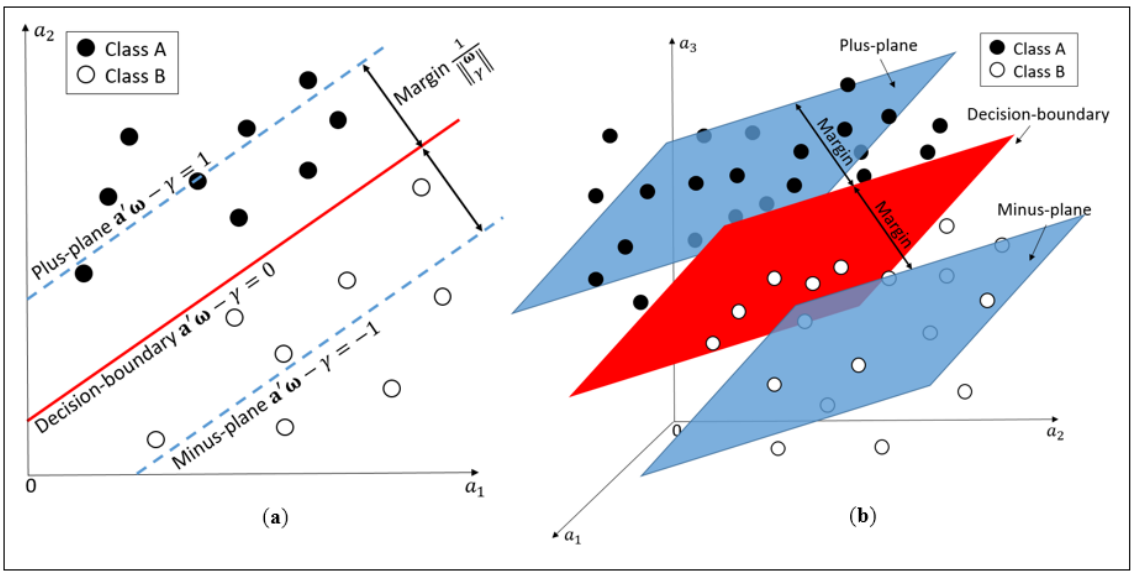


Figure 2.6. (a) Cartoon showing a two-class PSVM in 2D space. Classes "A" and "B" are approximated by two parallel lines that have been pushed as far apart as possible forming the cluster "margins". The red decision-boundary lies midway between the two margins. Maximizing the margin is equivalent to minimizing $(\boldsymbol{\omega}^T\boldsymbol{\omega} + \gamma^2)^{1/2}$. (b) A two-class PSVM in 3D space. In this case the decision-boundary and margins are 2D planes.

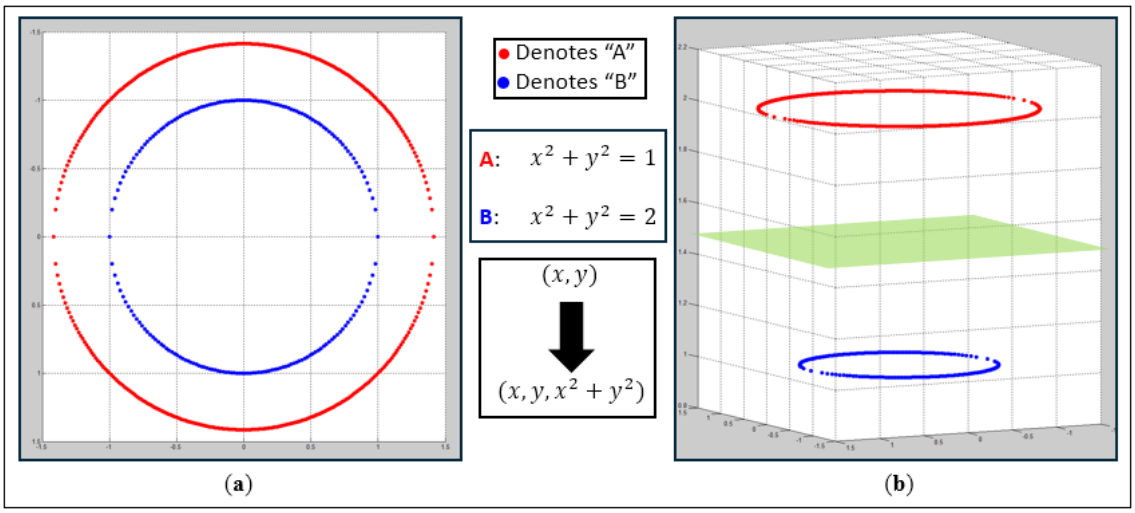


Figure 2.7. Cartoon show how one SVM can map two linearly inseparable problem into a higher dimensional space in which they can be separated. (a) Circular classes "A" and "B" in a 2D space cannot be separated by a linear decision-boundary (line). (b) Mapping the same data into a higher 3-dimensional "feature" space using the given projection. This transformation allows the two classes to be separated by the green plane.

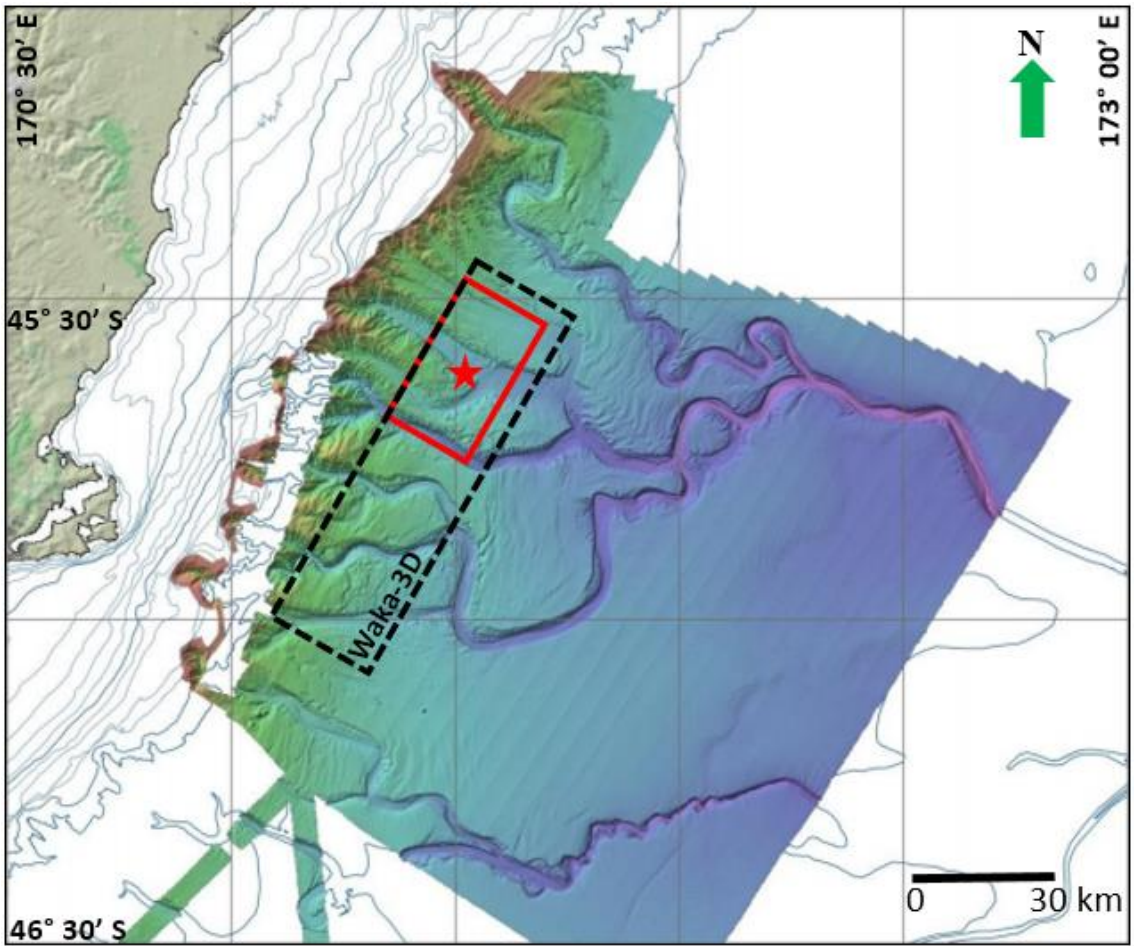


Figure 2.8. A map showing the location of the 3D seismic survey acquired over the Canterbury Basin, offshore New Zealand. The black rectangle denotes the limits of the Waka-3D survey, while the smaller red rectangle denotes the part of the survey shown in subsequent figures. Colors represent the relative depth of the current seafloor, warm being shallower and cold being deeper. Current seafloor canyons are delineated in this map, which are good analogs for the paleocanyons in Cretaceous and Tertiary ages (Modified from Mitchell and Neil, 2012).

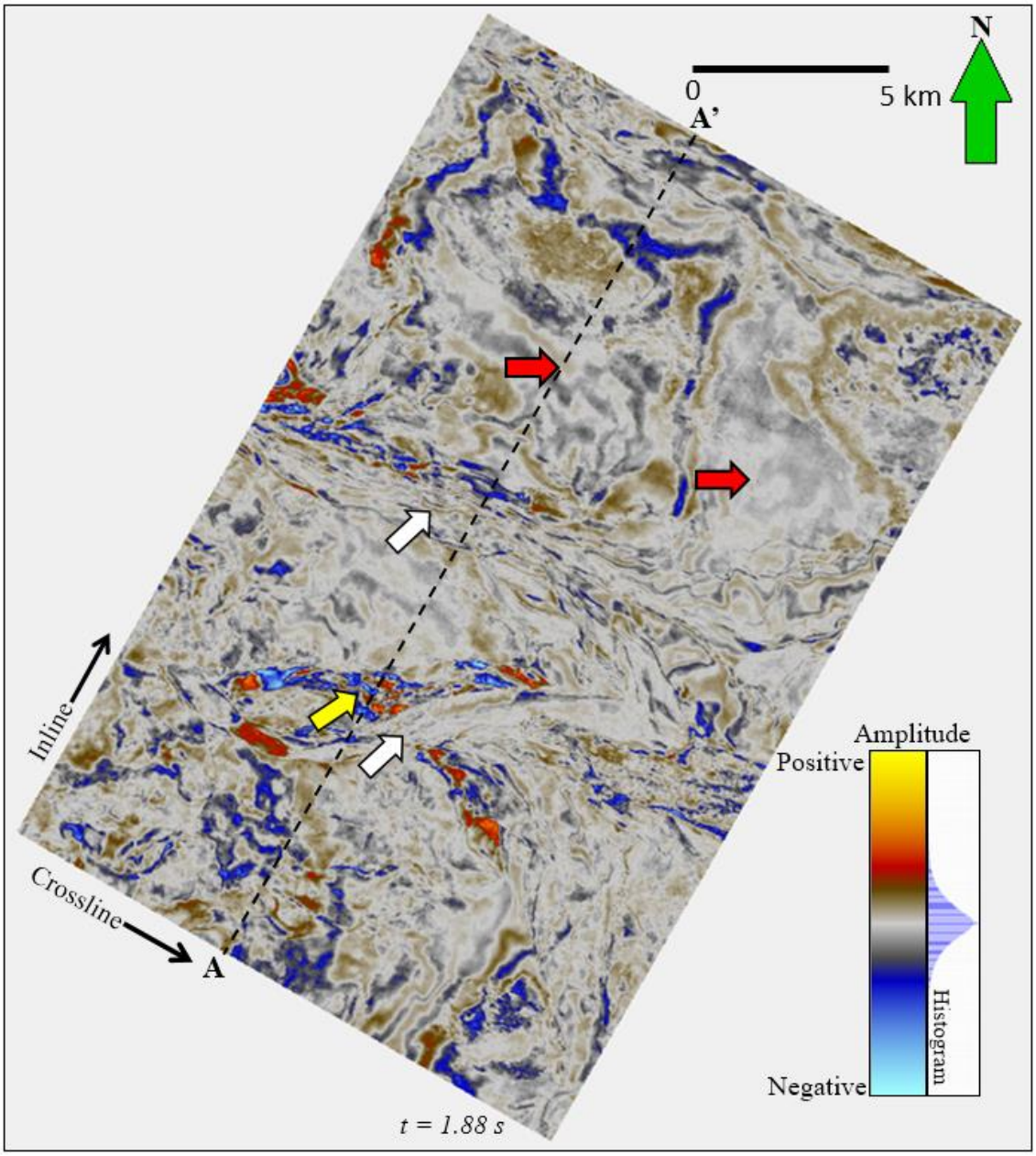


Figure 2.9. Time slice at t=1.88 s through the seismic amplitude volume. White arrows indicate potential channel/ canyon features. The yellow arrow indicates a high amplitude feature. Red arrows indicate relatively low energy, gently dipping area. AA' denotes a cross section shown in Figure 2.13.

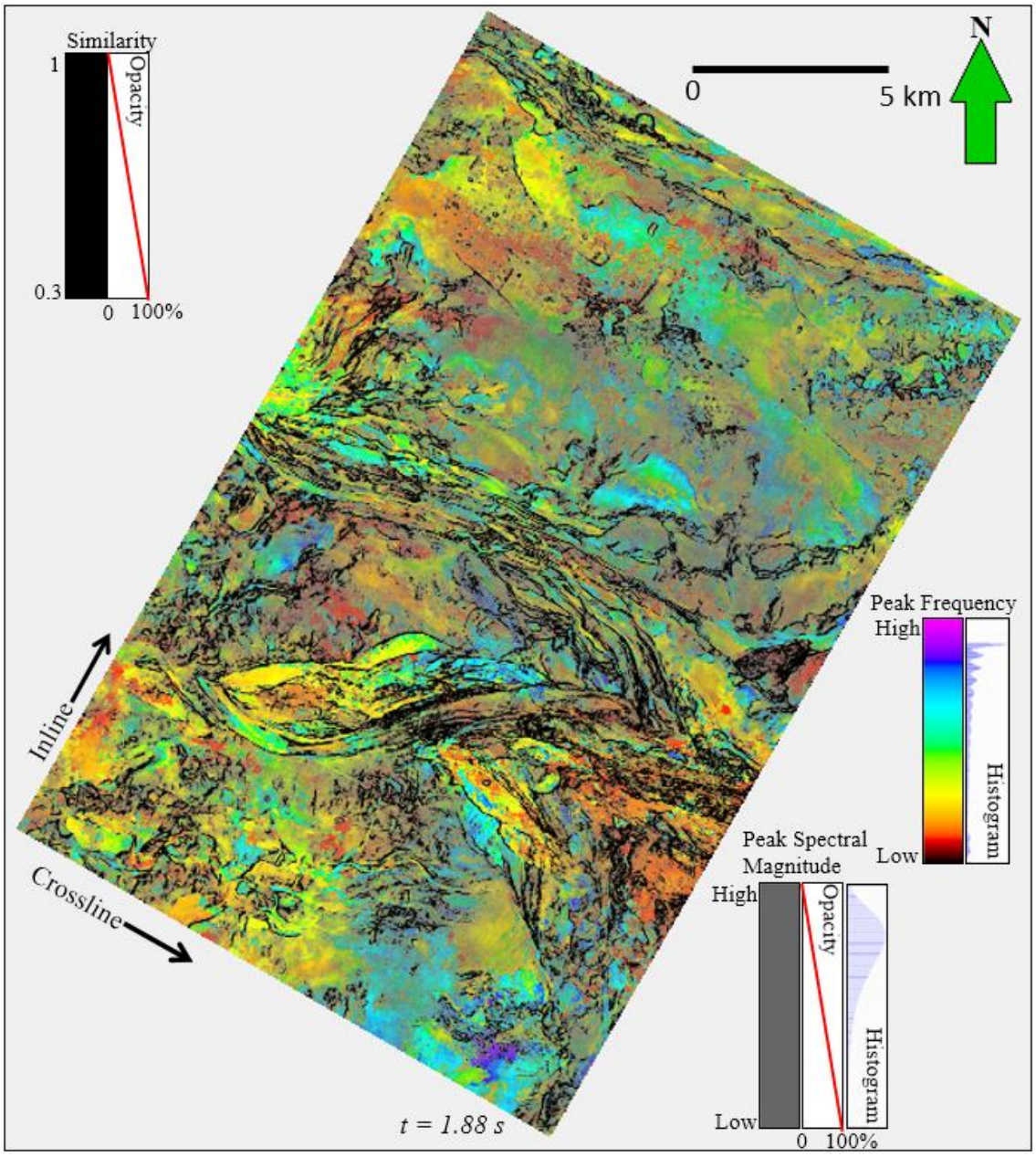


Figure 2.10. Time slice at t=1.88 s through peak spectral frequency co-rendered with peak spectral magnitude that emphasizes the relative thickness and reflectivity of the turbidite system and surrounding slope fan sediments into which it was incised. The two attributes are computed using a continuous wavelet transform algorithm. The edges of the channels are delineated by Sobel filter similarity.

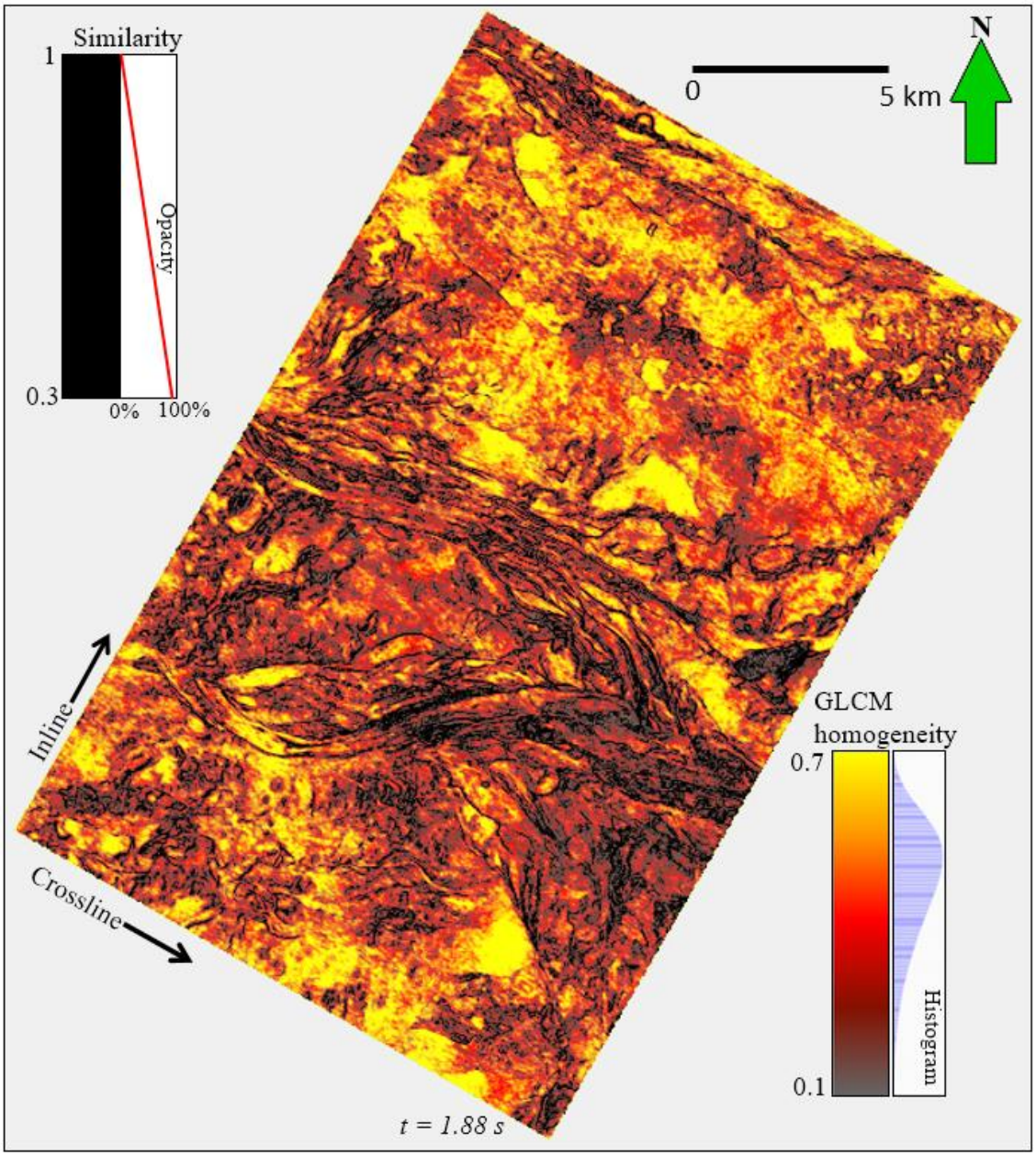


Figure 2.11. Time slice at t=1.88 s through the GLCM homogeneity attribute co-rendered with Sobel filter similarity. Bright colors highlights areas with potential fan sand deposits.

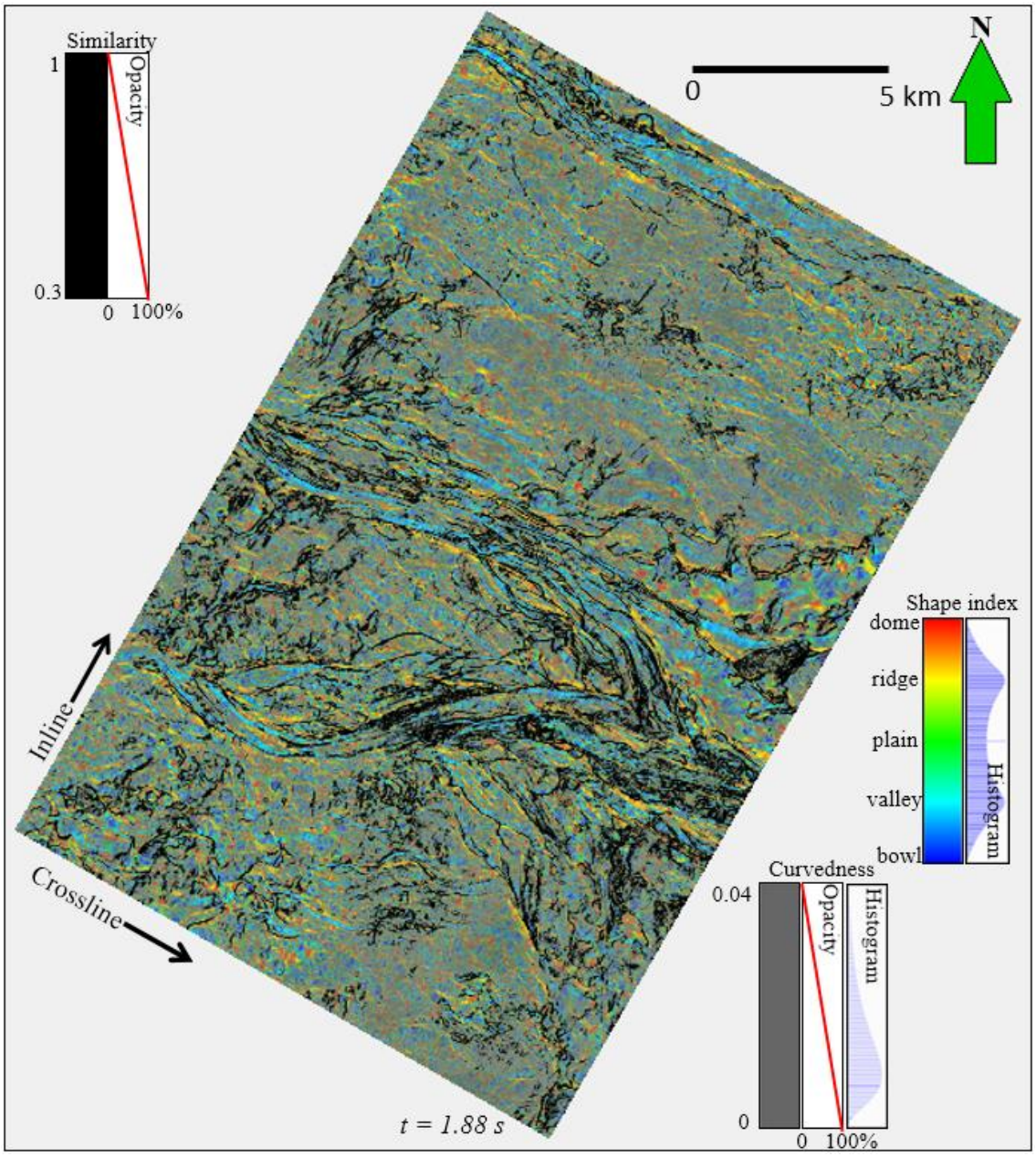


Figure 2.12. Time slice at t=1.88 s through the co-rendered shape index, curvedness, and Sobel filter similarity. The shape index highlights incisement, channel flanks, and levees providing an excellent image for interactive interpreter-driven classification. However, the shape index dominates the unsupervised classifications, highlighting valley and ridge features and minimizing more planar features of interest in the survey.

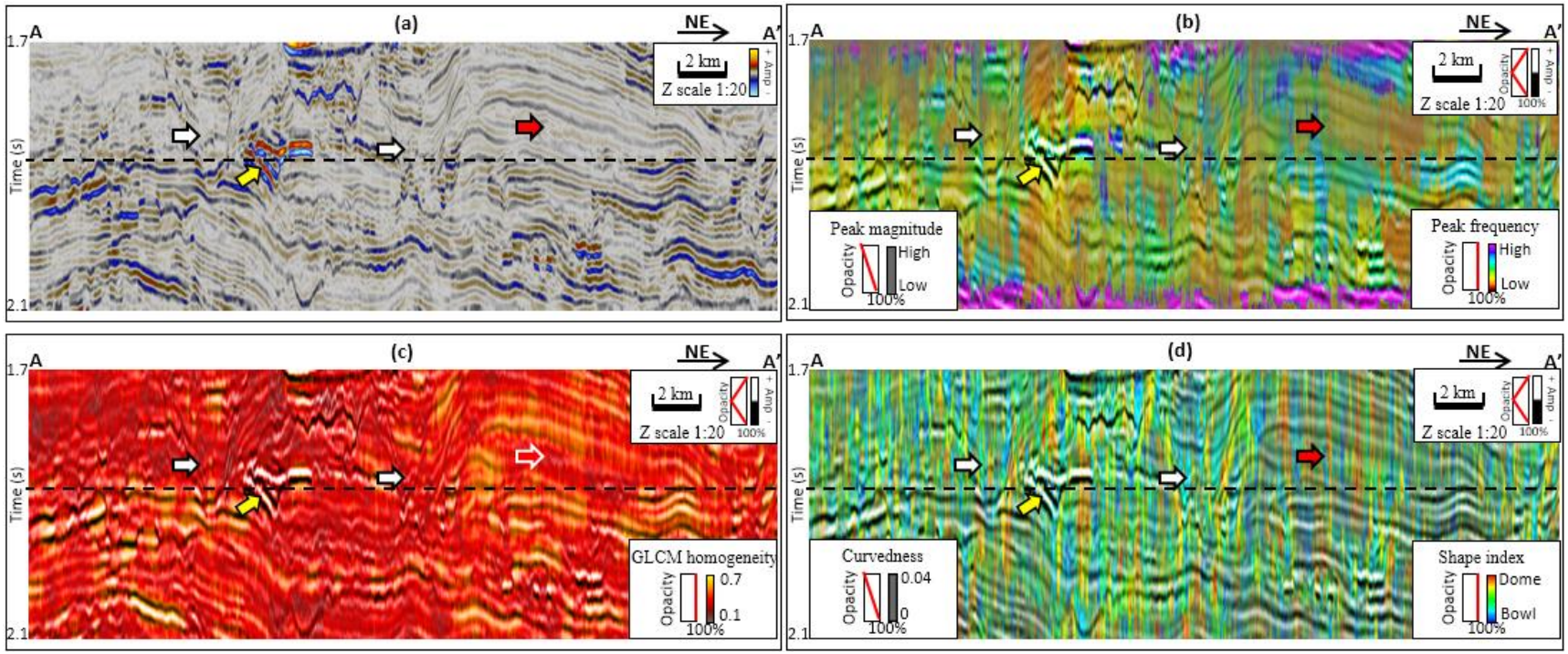


Figure 2.13. Vertical slices along line AA' (location shown in Figure 10) through (a) seismic amplitude, (b) seismic amplitude co-rendered with peak spectral magnitude and peak spectral frequency, (c) seismic amplitude co-rendered with GLCM homogeneity, and (d) seismic amplitude co-rendered with shape index and curvedness. White arrows indicate incised channel and canyon features. The yellow arrow indicates at a high amplitude reflector. Red arrows indicate relatively low amplitude, gently dipping areas.

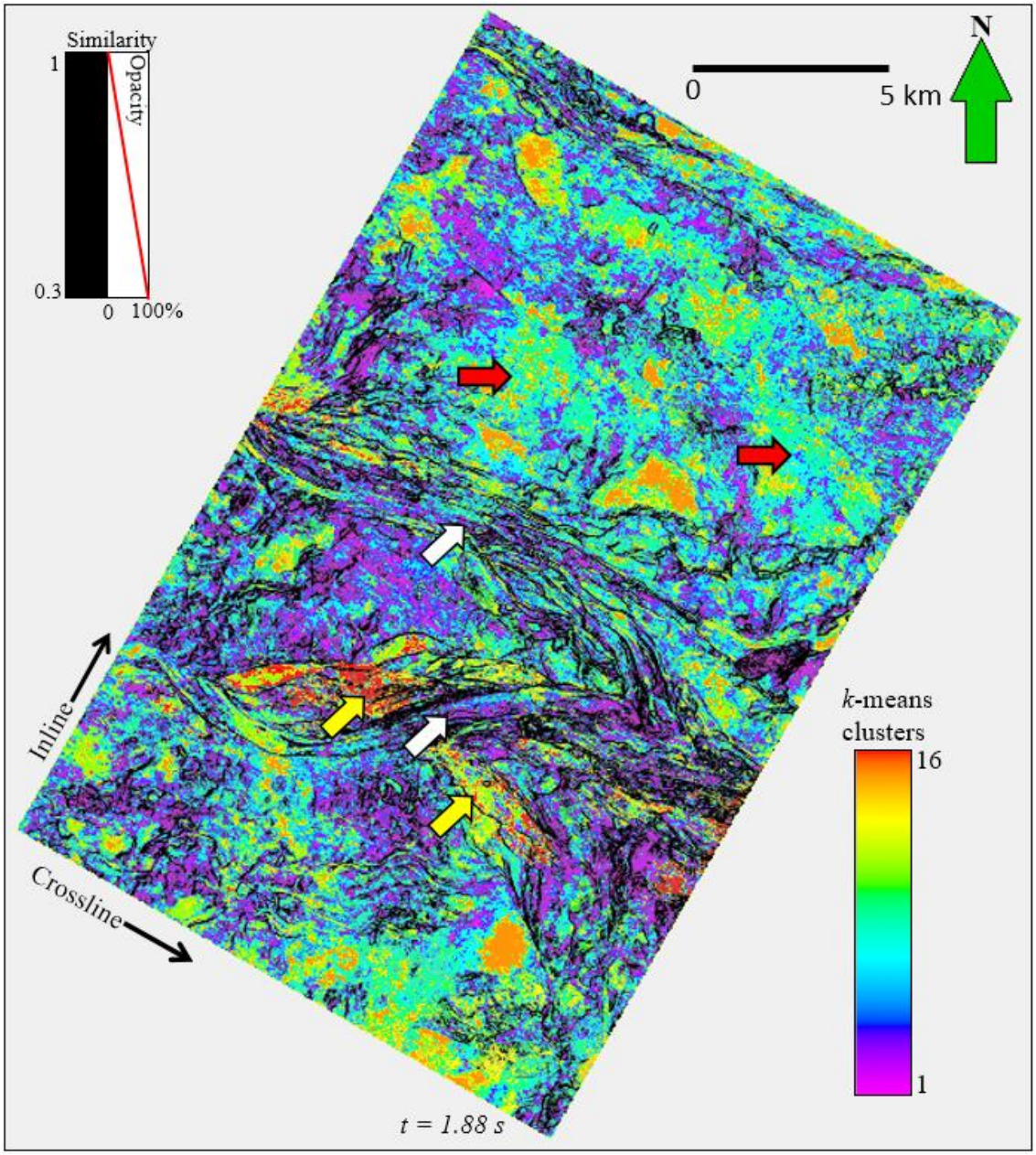


Figure 2.14. Time slice at t=1.88 s through K-means clustering volume with K=16. White arrows indicate channel-like features. Yellow arrows indicate high amplitude overbank deposits. Red arrows indicate possible slope fans. The edges of the channels are delineated by Sobel filter similarity.

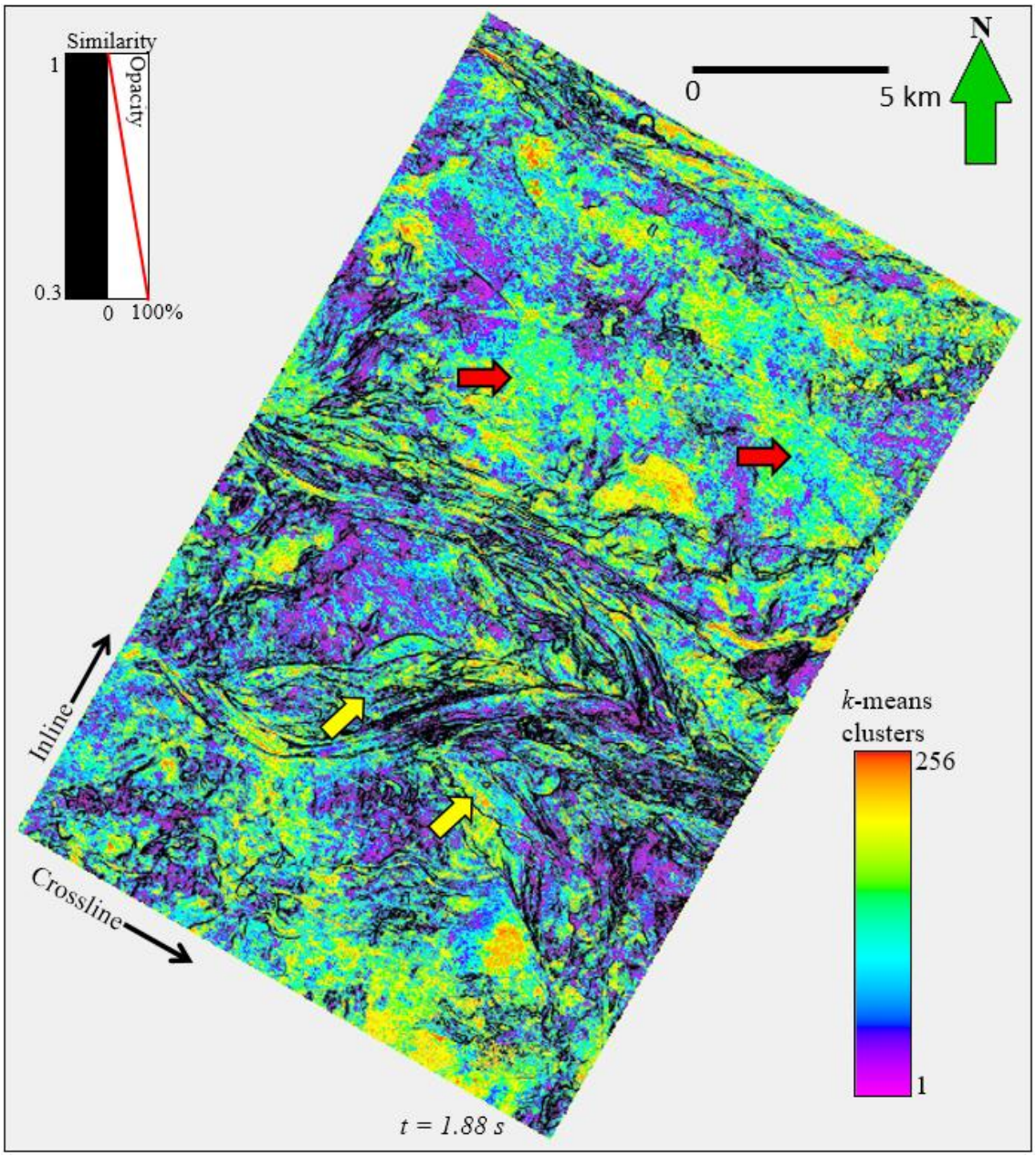


Figure 2.15. Time slice at t=1.88 s through K-means clustering volume with K=256. The classification result follows the same pattern as K=16 but is more chaotic since the classes are computed independently and are not constrained to fall on a lower dimensional manifold. Note the similarity between clusters of high amplitude overbank (yellow arrows) and slope fan deposits (red arrows) which were separable in Figure 2.14.

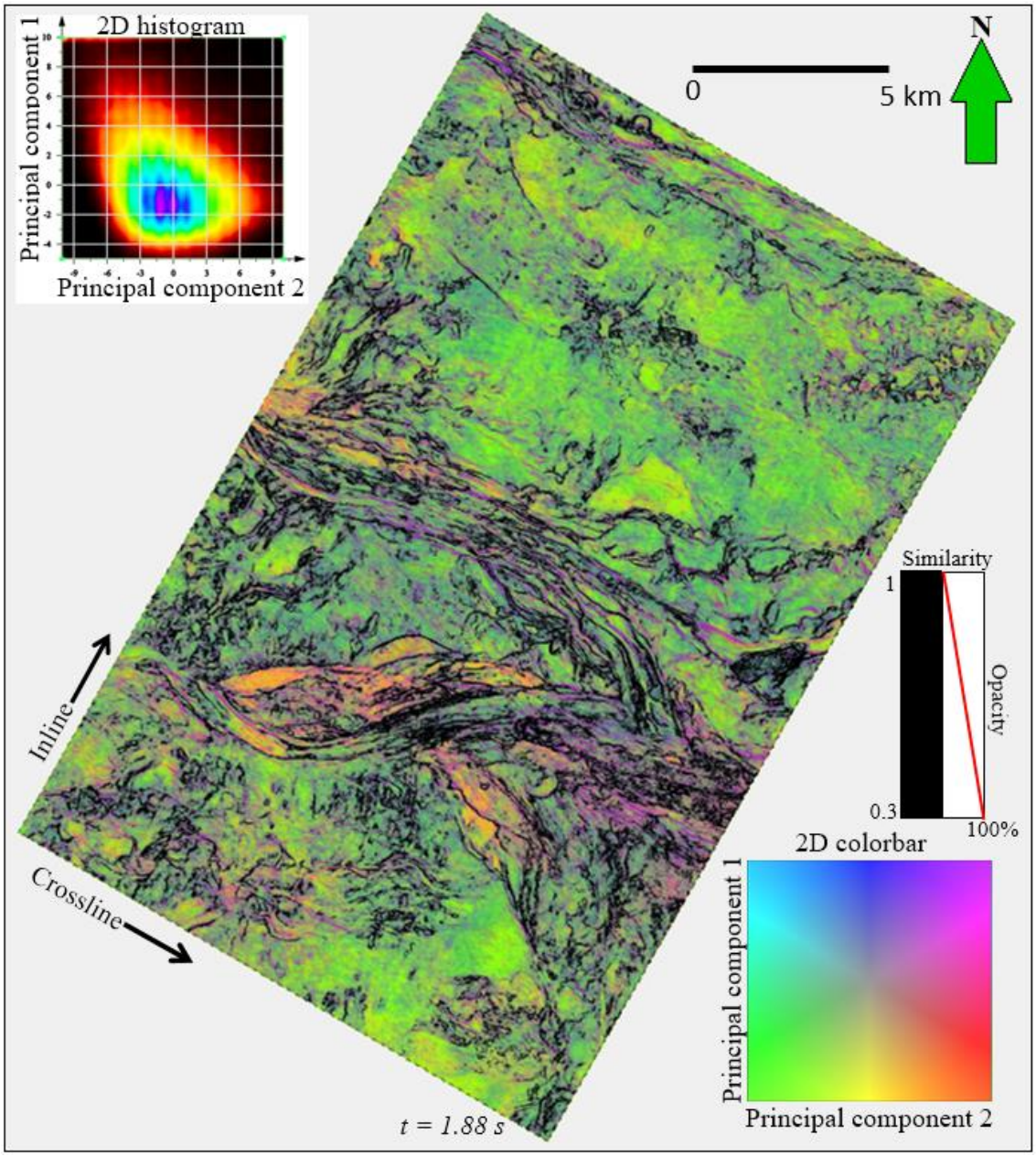


Figure 2.16. Time slice at t=1.88 s of the first two principle components plotted against a 2D colorbar. These two principal components serve as the initial model for both the SOM and GTM images that follow. With each iteration, the SOM and GTM manifolds will deform to better fit the natural clusters in the input data.

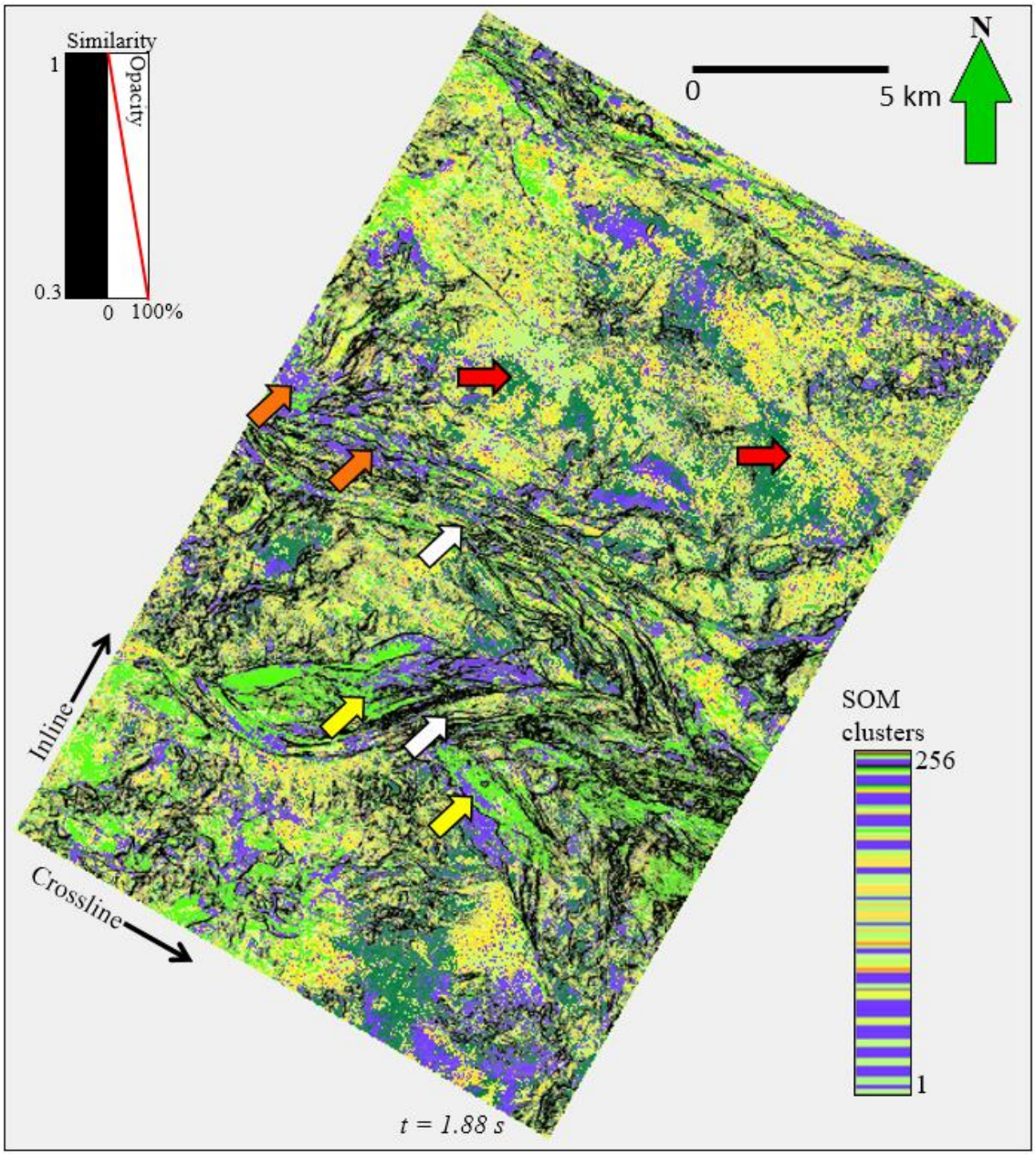


Figure 2.17. Time slice at t=1.88 s through an SOM classification volume using 256 clusters. White arrows indicate channel-like features. Combined with vertical sections through seismic amplitude, I interpret overbank deposits (yellow arrows), crevasse splays (orange arrows), and slope fan deposits (red arrows). The data are mapped to a 2D manifold initialized by first two principle components and are somewhat more organized than the K-means image shown in the previous figures.

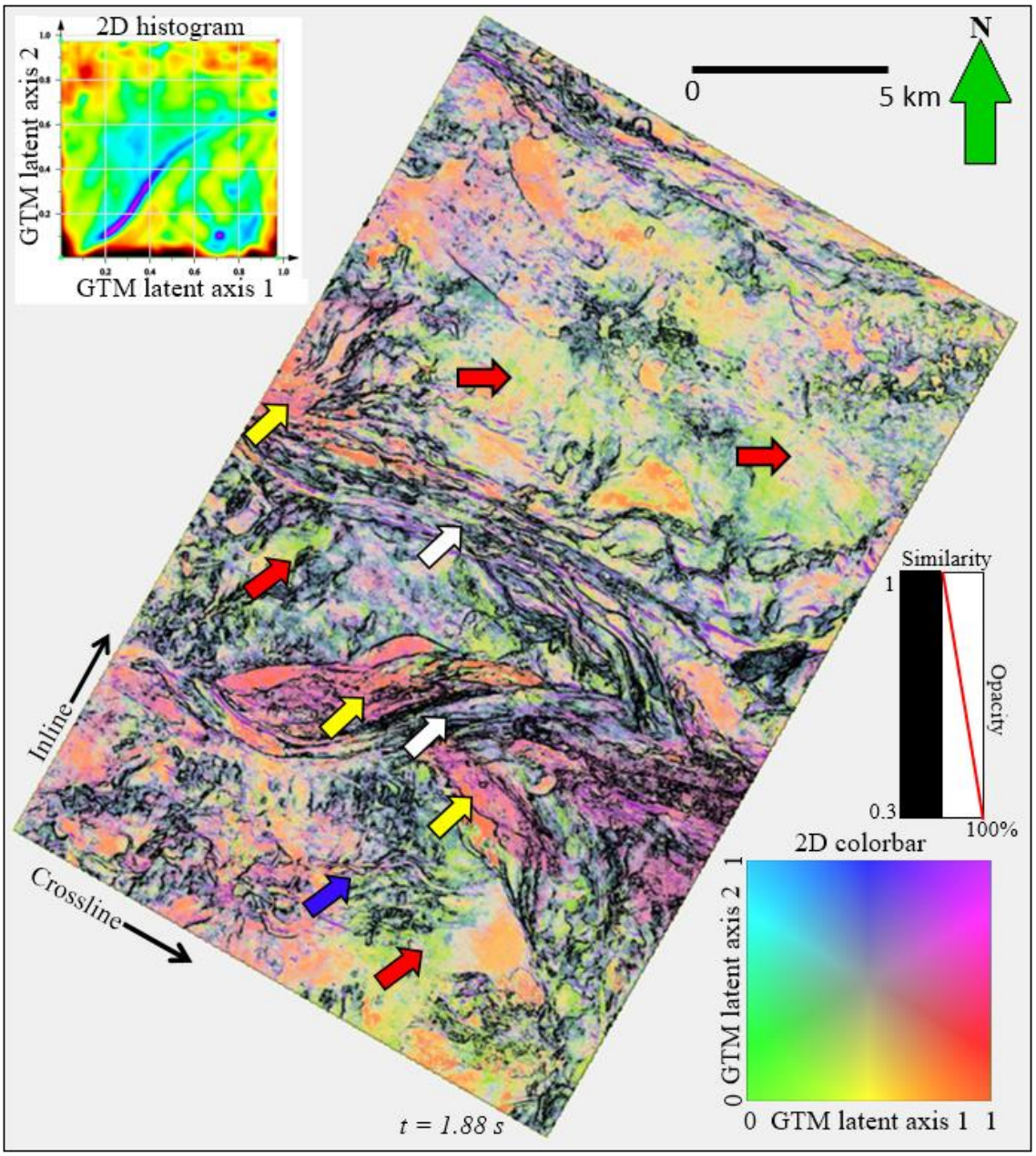


Figure 2.18. Time slice at t=1.88 s through crossplotting GTM projection 1 and 2 using a 2D colorbar. White arrows indicate channel-like features, yellow arrows overbank deposits, and red arrows slope fan deposits. The blue arrow indicates a braided channel system that can be seen on PCA but cannot be identified from K-means or SOM classification maps. The color indicates the location of the mean probability of each data vector mapped into the 2D latent space.

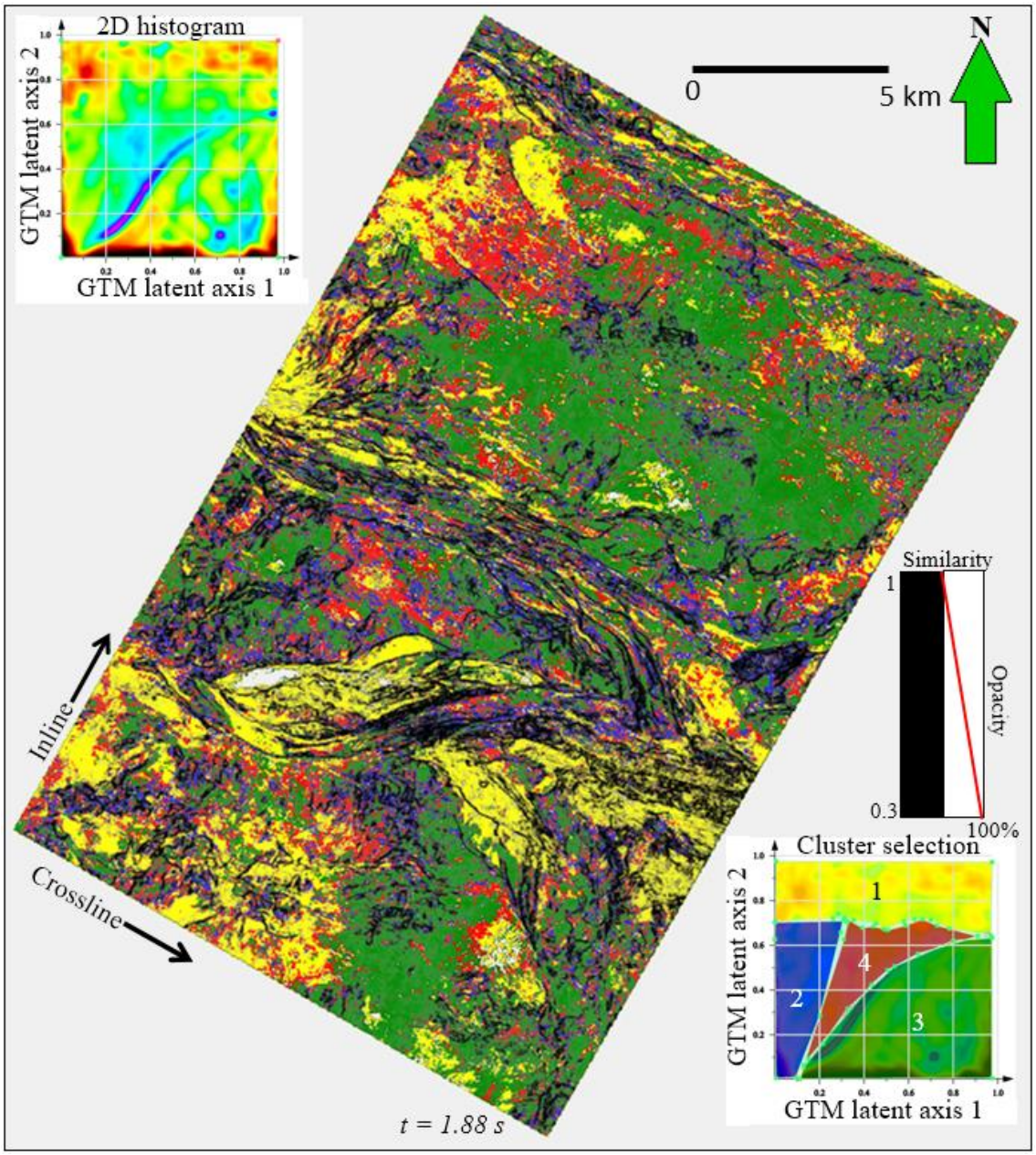


Figure 2.19. The same time slice through the GTM projections shown in the previous image but now displayed as four seismic facies. To do so, I first create two GTM "components" aligned with the original first two principal components. I then pick four colored polygons representing four seismic facies on the histogram generated using a commercial crossplot tool, This histogram is a map of the GTM posterior probability distribution in the latent space. The yellow polygon represents overbank deposits, the blue polygon channels /canyons, the green polygon slope fan deposits, and the red polygon "everything else".

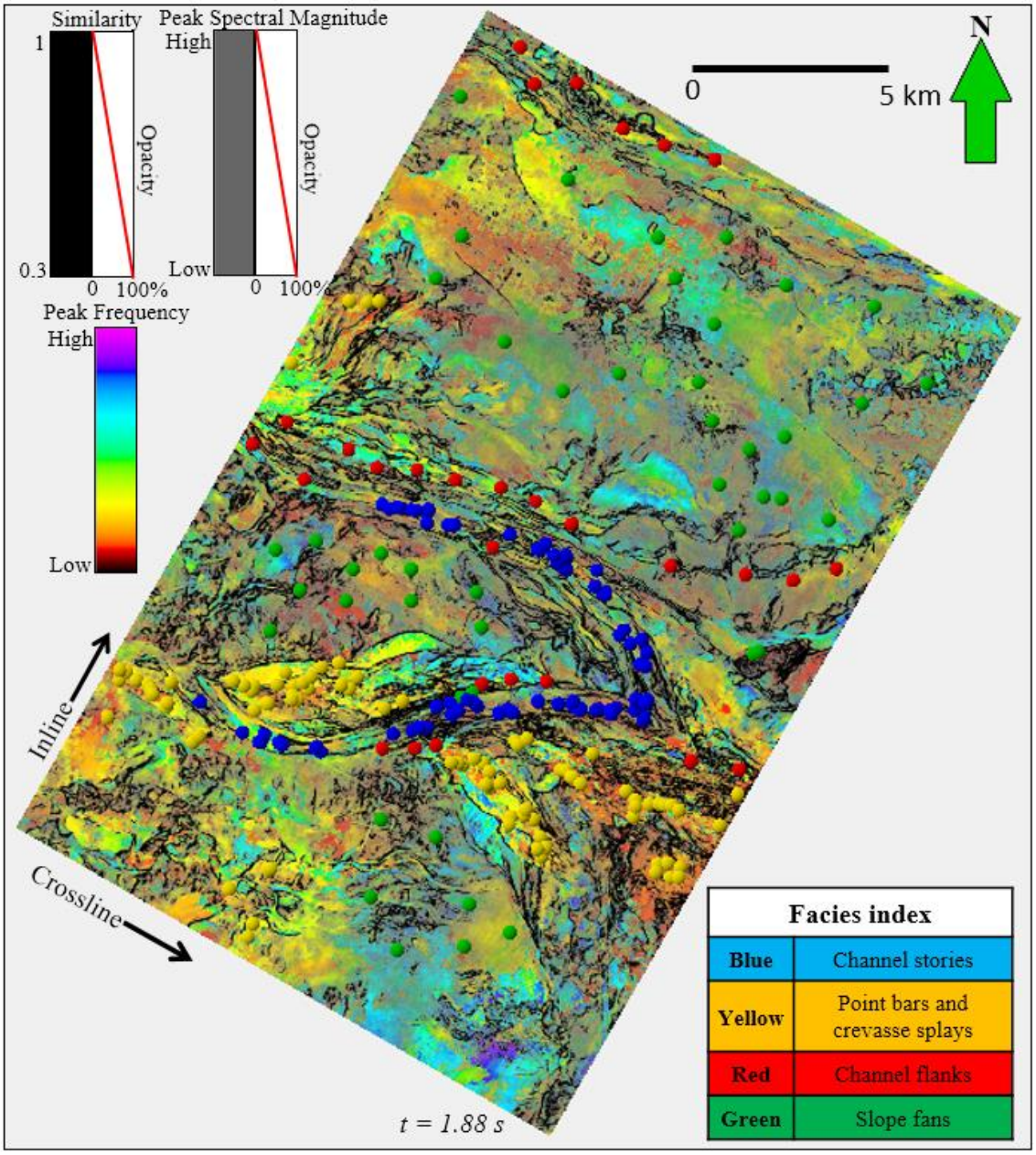


Figure 2.20. Time slice at t=1.88 s through co-rendered peak spectral frequency, peak spectral magnitude, and Sobel filter similarity volumes. Seed points (training data) are shown with colors for the picked four facies, blue indicating multistoried channels, yellow point bars and crevasse splays, red channel flanks, and green slope fans. Attribute vectors at these seed points are used as training data in supervised classification.

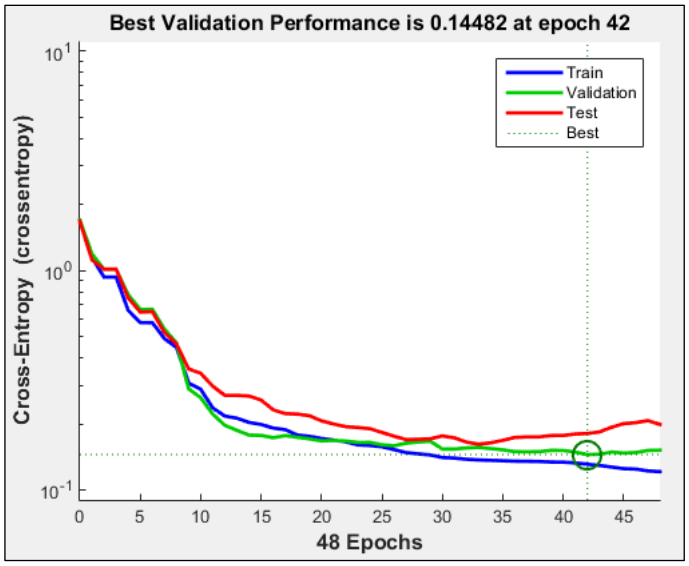


Figure 2.21. PNN errors through the training epochs. The neural network reaches its best performance at epoch 42.

Figure 2.22. Confusion tables for the same PNN shown in Figure 2.21. From these tables I find the training correctness to be 90%, the testing and cross-validation correctness to be 86% and 91%, warranting a reliable prediction.

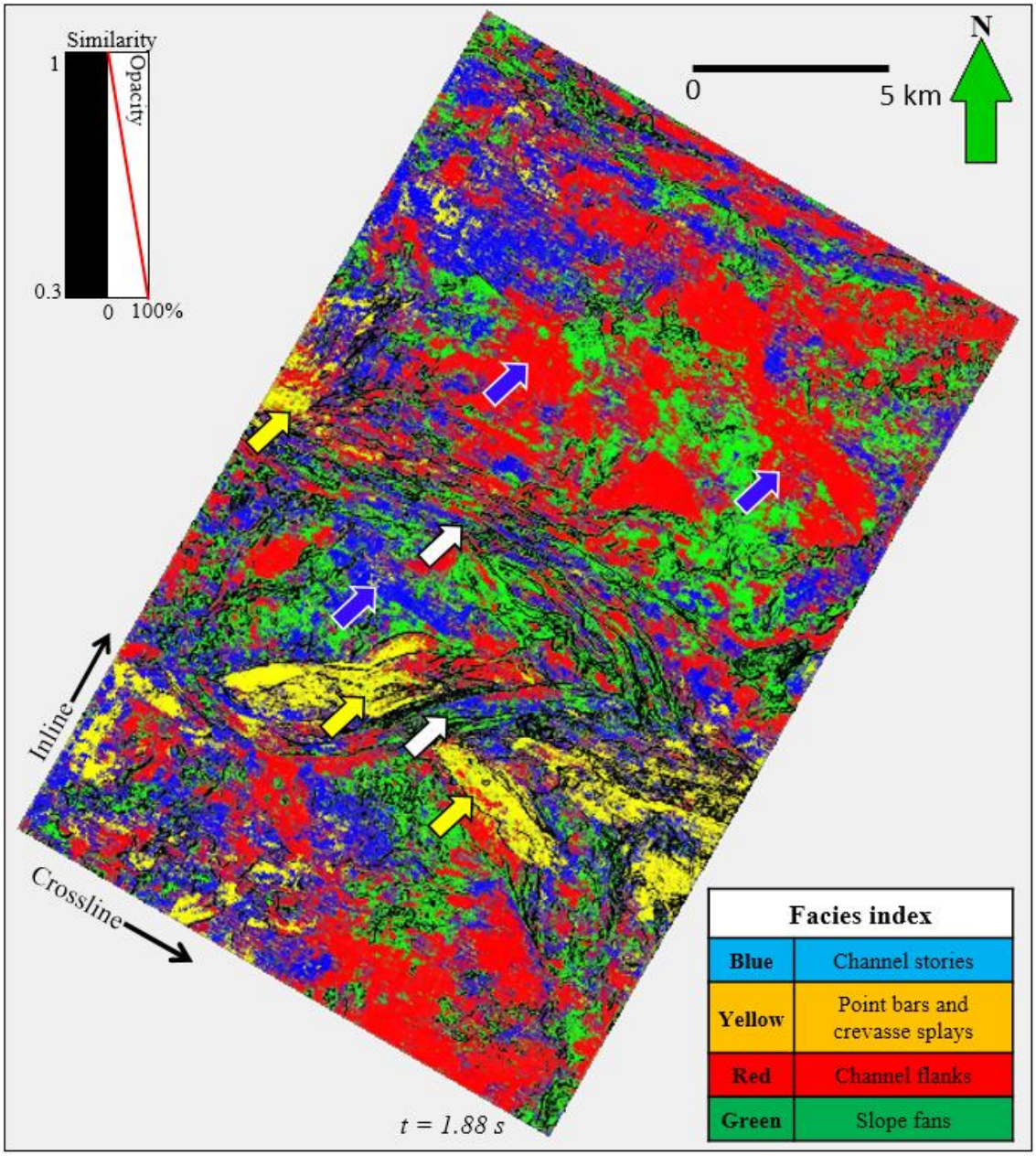


Figure 2.23. Time slice at t=1.88 s through the ANN classification result. White arrows indicate channels/canyons. Yellow arrows indicate point bars and crevasse splays.

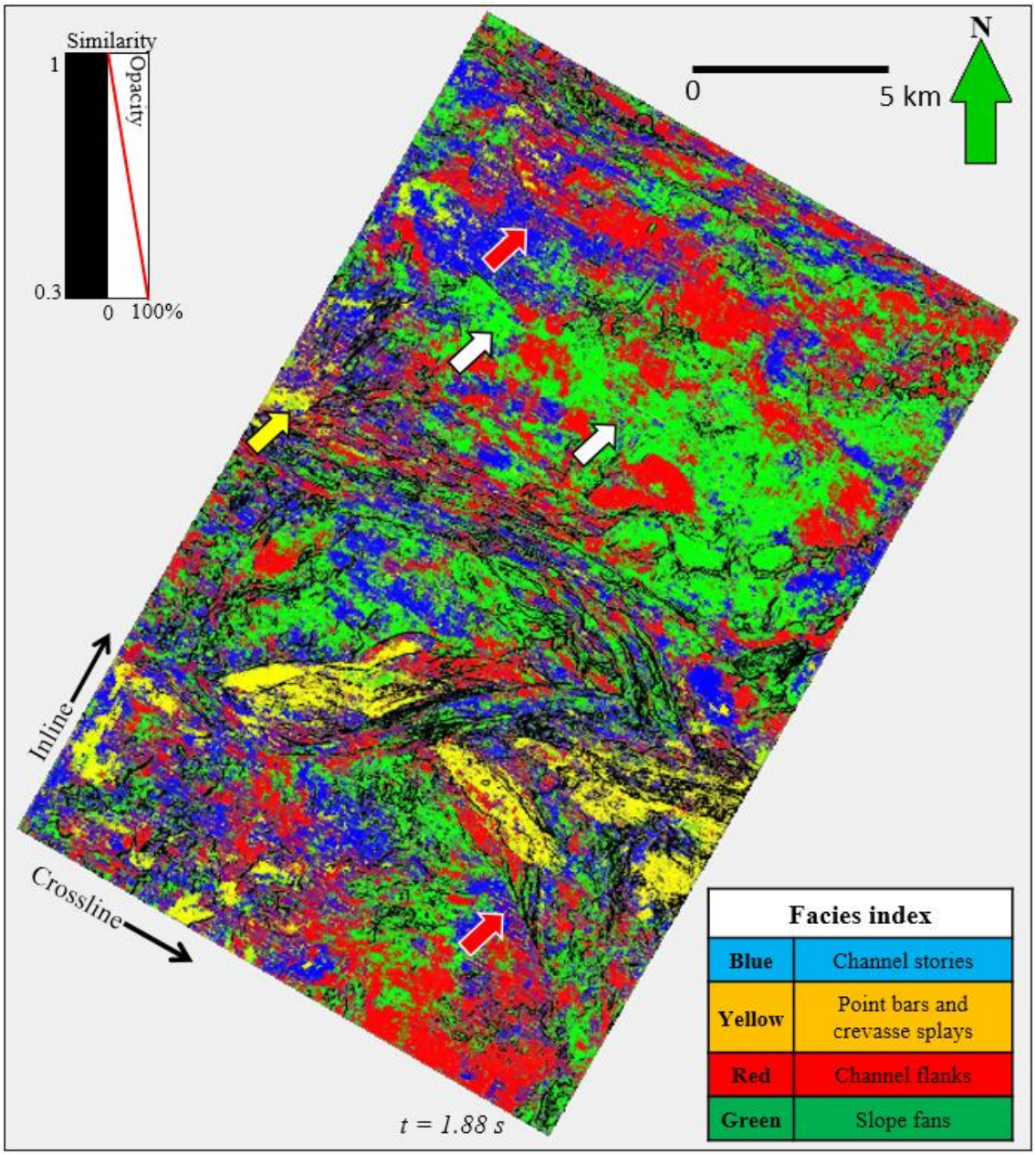


Figure 2.24. Time slice at t=1.88 s through SVM classification result. White arrows indicate more correctly classified slope fans. Yellow arrow indicates crevasse splays. Red arrows show the misclassifications due to possible acquisition footprint.

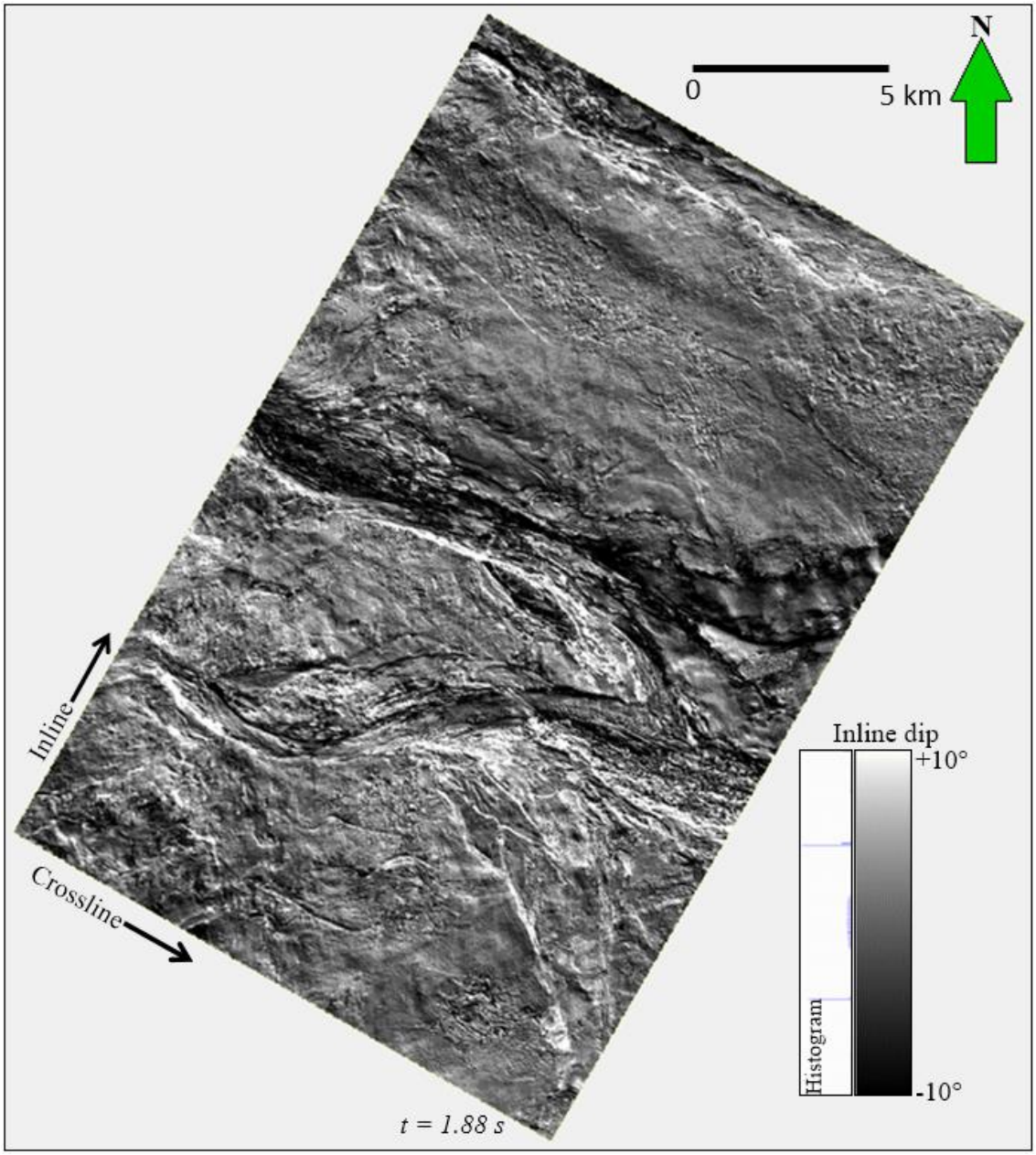


Figure 2.25. Time slice at t=1.88 s through inline dip component of reflector dip. Inline dip magnitude provides a photo-like image of the paleocanyons.

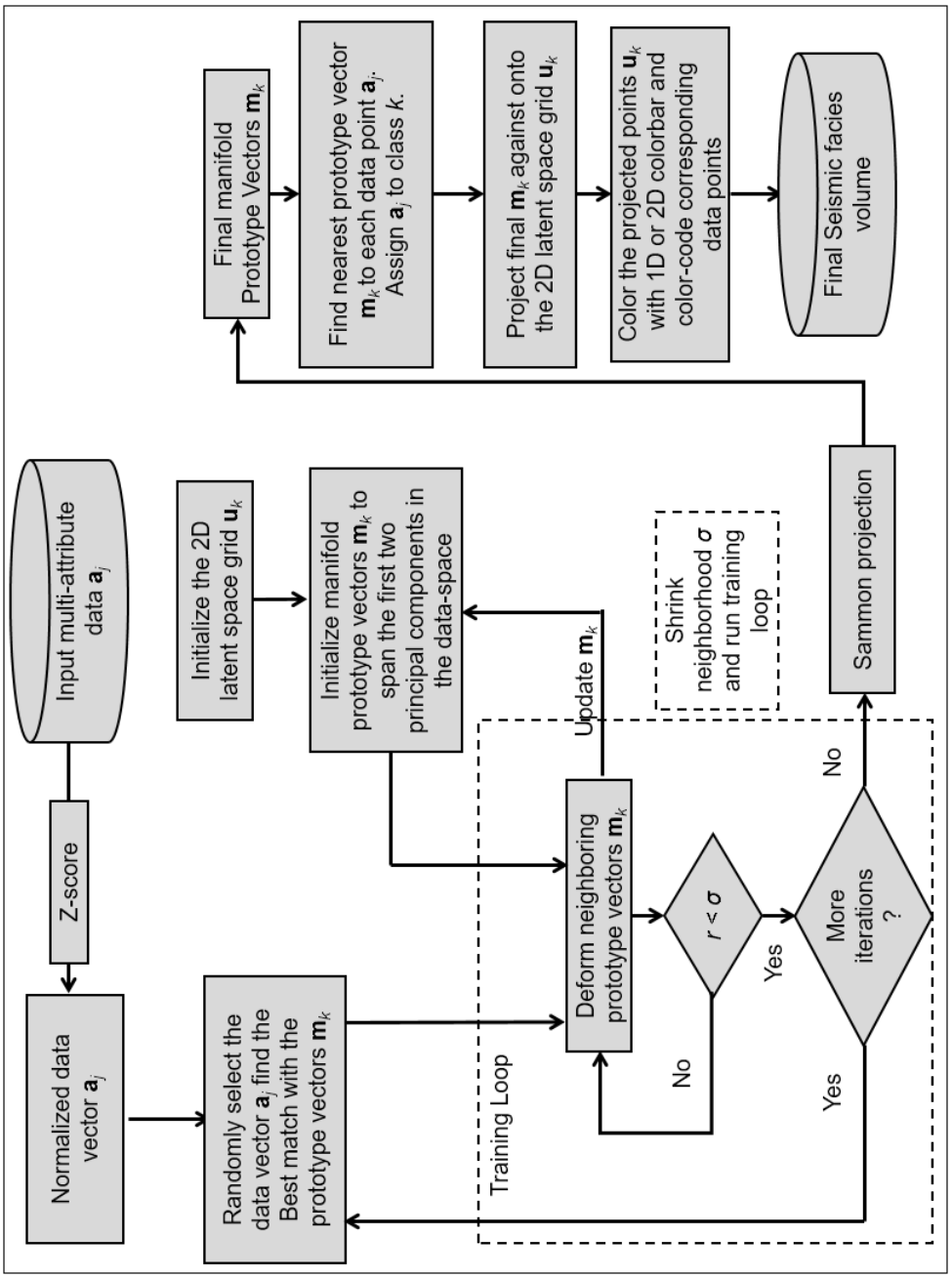


Figure 2.26. Self-organizing maps (SOM) workflow.

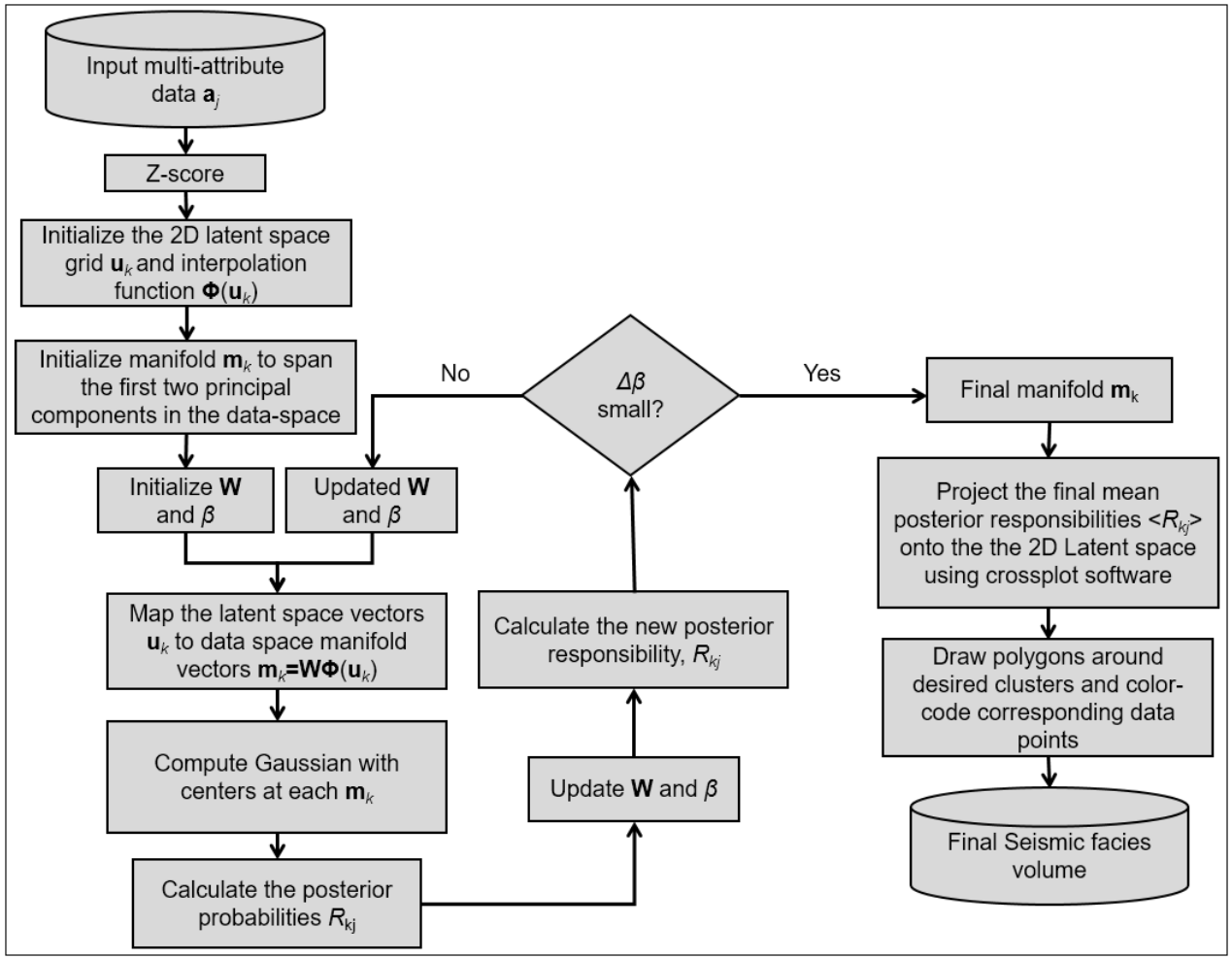


Figure 2.27. Generative topographic mapping (GTM) workflow.

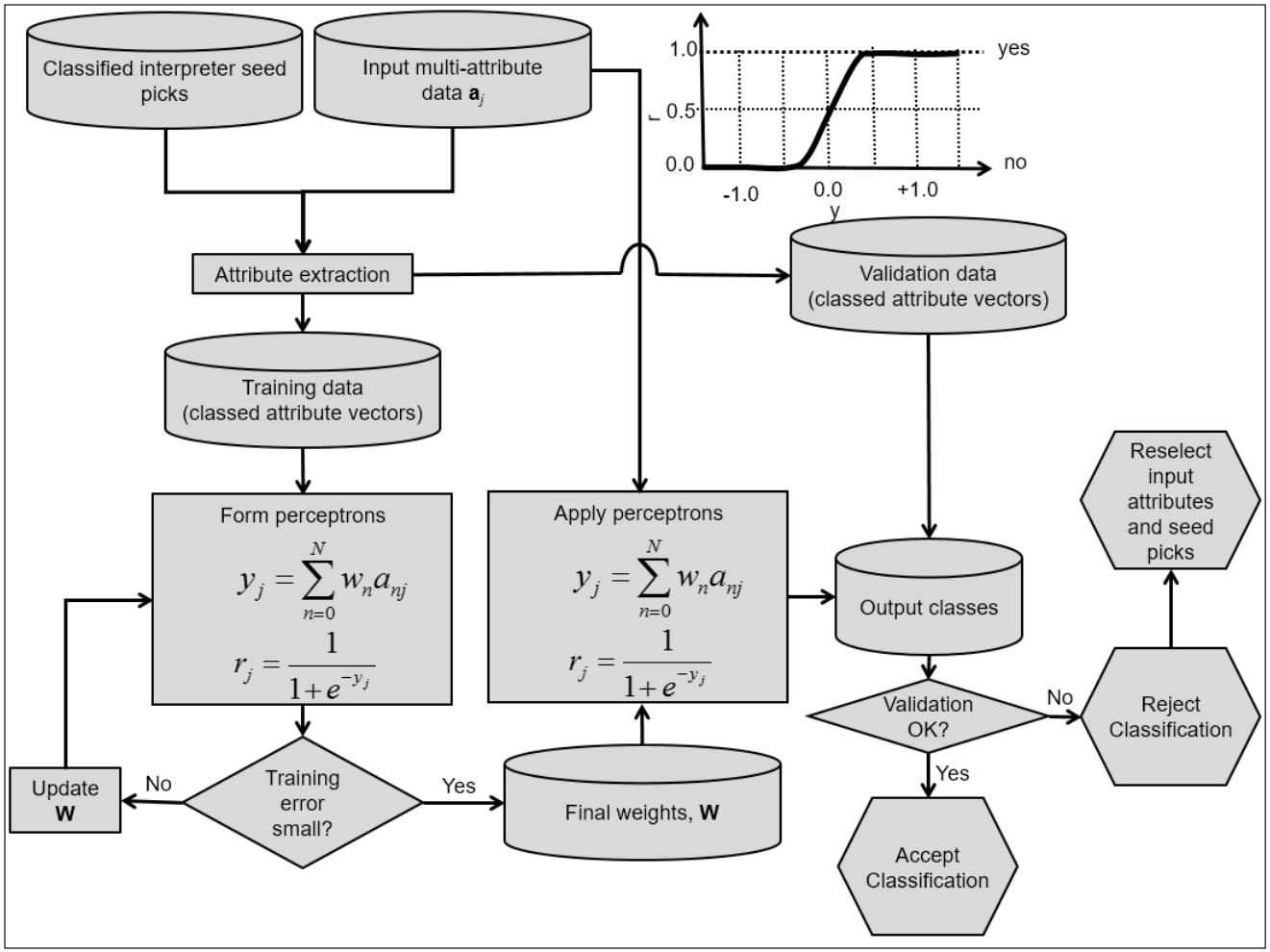
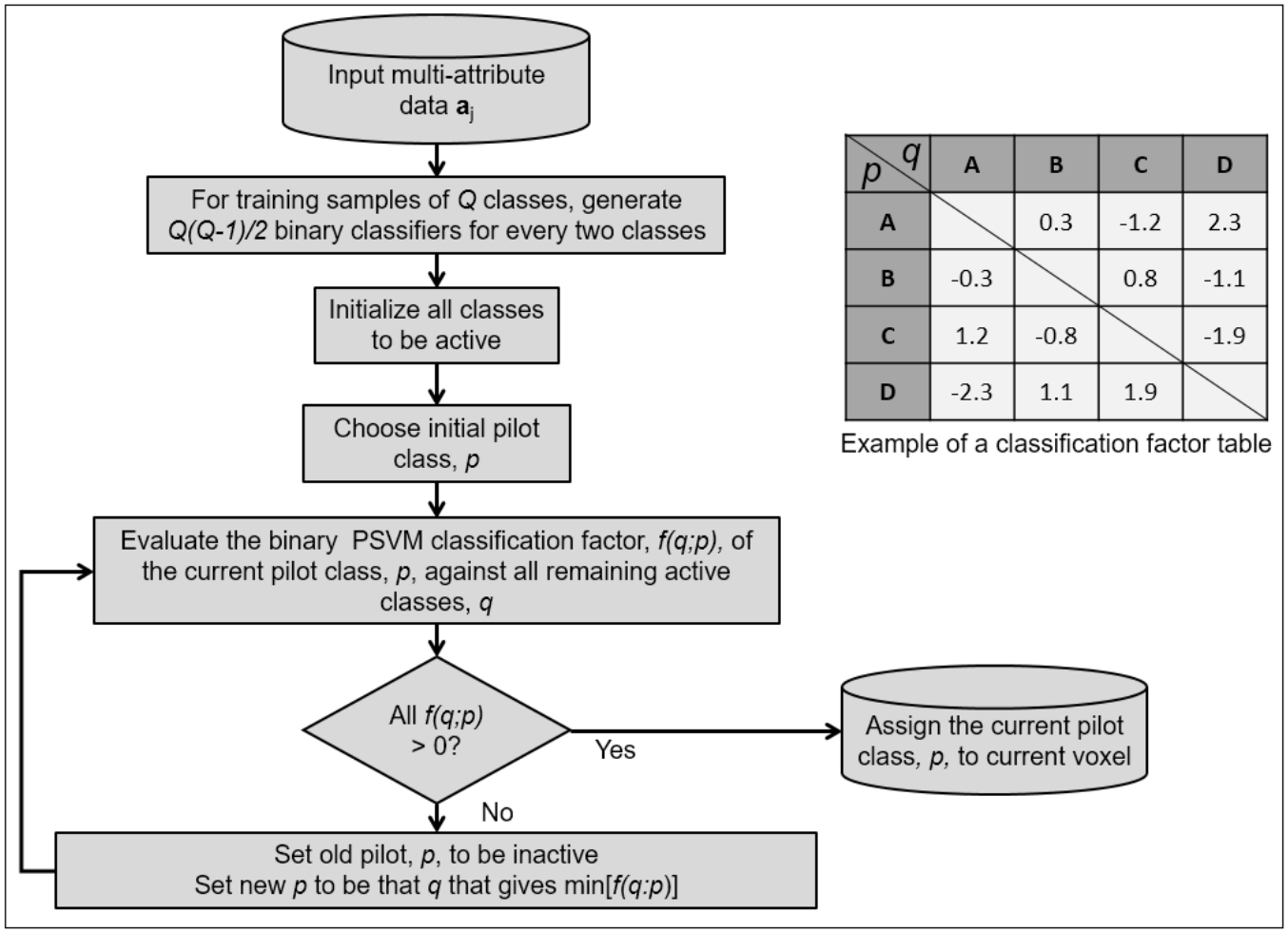


Figure 2.28. Artificial neural network (ANN) workflow.



p \ q	A	B	C	D
A		0.3	-1.2	2.3
B	-0.3		0.8	-1.1
C	1.2	-0.8		-1.9
D	-2.3	1.1	1.9	

Example of a classification factor table

Figure 2.29. Proximal support vector machine (PSVM) workflow.

REFERENCES

Al-Anazi, A. and I. D. Gates, 2010, A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs: Engineering Geology, **114**, 267-277.

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221-2224.

Bennett, K. P. and A. Demiriz, 1999, Semi-supervised support vector machines: Advances in Neural Information Processing Systems 11: Proceedings of the 1998 Conference, 368-374.

Bishop, C. M., 2006, Pattern recognition and machine learning: Springer, New York, United States.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, **10**, 215-234.

Chao, J., M. Hoshino, T. Kitamura, and T. Masuda, 2001, A multilayer RBF network and its supervised learning: International Joint Conference on Neural Networks, INNS/IEEE, Expanded Abstracts, 1995–2000.

Chapelle, O., B. Schölkopf, and A. Zien, 2006, Semi-supervised learning: MIT press, Cambridge, United States.

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

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.

Corradi, A., P. Ruffo, A. Corrao, and C. Visentin, 2009, 3D hydrocarbon migration by percolation technique in an alternative sand-shale environment described by a seismic facies classification volume: Marine and Petroleum Geology, **26**, 495-503.

Cortes, C. and V. Vapnik, 1995, Support-vector networks: Machine Learning, **20**, 273-297.

Cristianini, N. and J. Shawe-Taylor, 2000, An introduction to support vector machines and other kernel-based learning methods: Cambridge University Press, New York, United States.

Dempster, A.P., N. M. Laird, and D. B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm: Journal of Royal Statistical Society, Series B, **39**, 1-38.

Eagleman, D., 2012, Incognito: The secret lives of the brain: Pantheon Books, New York, USA.

Forgy, E. W., 1965, Cluster analysis of multivariate data: efficiency vs interpretability of classifications: Biometrics, **21**, 768-769.

Fung, G. and O. L. Mangasarian, 2001, Proximal support vector machine classifiers: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM 2001, 77-86.

Fung, G. M. and O. L. Mangasarian, 2005, Multicategory proximal support vector machine classifiers: Machine Learning, **59**, 77-97.

Gao, D., 2007, Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: An example from offshore Angola, West Africa: AAPG Bulletin, **91**, 1665-1683.

Heggland, R., Meldahl, P., Bril, B. and de Groot, P., 1999, The chimney cube, an example of semi-automated detection of seismic objects by directive attributes and neural networks: Part II; interpretation, 69th Annual International Meeting, SEG, Expanded Abstracts, 935-937.

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

Hsu, C. and C. Lin, 2002, A comparison of methods for multiclass support vector machines: IEEE Transactions on Neural Networks, **13**, 415-425.

Huang, Z., J. Shimeld, M. Williamson, and J. Katsube, 1996, Permeability prediction with artificial neural network modeling in the Ventura gas field, offshore eastern Canada: Geophysics, **61**, 422–436.

Jancey, R. C., 1966, Multidimensional group analysis: Australian Journal of Botany, **14**, 127-130.

Jayaram, V. and B. Usevitch, 2008, Dynamic mixing kernels in Gaussian mixture classifier for hyperspectral classification: SPIE Optics + Photonics 2008, 70750L-70750L.

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

Kreßel, U., 1999, Pairwise classification and support vector machines: Advances in Kernel Methods - Support Vector Learning, 255-268.

Kuzma, H. A. and J. W. Rector, 2004, Nonlinear AVO inversion using Support Vector Machines: 74th Annual International Meeting, SEG, Expanded Abstracts, 203-206.

Kuzma, H. A. and J. W. Rector, 2005, The Zoeppritz equations, information theory, and support vector machines: 75th Annual International Meeting, SEG, Expanded Abstracts, 1701-1704.

Kuzma, H. A. and J. W. Rector, 2007, Support Vector Machines implemented on a Graphics Processing Unit: 77th Annual International Meeting, SEG, Expanded Abstracts, 2089-2092.

Li, J. and J. Castagna, 2004, Support Vector Machine (SVM) pattern recognition to AVO classification: Geophysical Research Letters, **31**, L02609.

Lim, J., 2005, Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea: Journal of Petroleum Science and Engineering, **49**, 182-192.

Lubo, D., K. Marfurt, and V. Jayaram, 2014, Statistical characterization and geological correlation of wells using automatic learning Gaussian mixture models: Unconventional Resources Technology Conference, Extended Abstract, 2014, 774-783.

Mangasarian, O. L. and E. W. Wild, 2006, Multisurface proximal support vector machine classification via generalized eigenvalues: IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**, 69-74.

MacQueen, J., 1967, Some methods for classification and analysis of multivariate observations: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281--297,

Matos, M. C., K. J. Marfurt., and P. R. S. Johann, 2009, Seismic color self-organizing maps: 11th International Congress of the Brazilian Geophysical Society, Expanded Abstracts.

Meldahl, P., R. Heggland, B. Bril, and P. de Groot, 1999, The chimney cube, an example of semi-automated detection of seismic objects by directive attributes and neural networks: Part I; Methodology: 69th Annual International Meeting, SEG, Expanded Abstracts, 931-934.

Mitchell, J. and H. L. Neil, 2012, OS20/20 Canterbury – Great South Basin TAN1209 voyage report: National Institute of Water and Atmospheric Research Ltd (NIWA).

Murat, M. E. and A. J. Rudman, 1992, Automated first arrival picking: A neural network approach: Geophysical Prospecting, **40**, 587–604.

Nazari, S., H. A. Kuzma, and J. W. Rector, 2011, Predicting permeability from well log data and core measurements using Support Vector Machines: 81st Annual International Meeting, SEG, Expanded Abstracts, 2004-2007.

Platt, J. C., 1998, Sequential minimal optimization: A fast algorithm for training support vector machines: Microsoft Research Technical Report, MSR-TR-98-14.

Platt, J. C., N. Cristianini, and J. Shawe-Taylor, 2000, Large margin DAGs for multiclass classification: NIPS, **12**, 547-553.

Röth, G. and A. Tarantola, 1994, Neural networks and inversion of seismic data: Journal of Geophysics Research, **99**, 6753–6768.

Roy, A., 2013, Latent space classification of seismic facies: PhD Dissertation, The University of Oklahoma.

Roy, A., B.L. Dowdell, and K.J. Marfurt, 2013, Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma: Interpretation, **1**, SB109-SB124.

Roy, A., Araceli, S. R, J. T. Kwiatkowski, and K. J. Marfurt, 2014, Generative Topographic Mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, Southern Mexico: Interpretaion, **2**, SA31-SA47.

Sammon, W. J., 1969, A nonlinear mapping for data structure analysis, IEEE Transaction on Computers, **C-18**, 401-409.

Schölkopf, B. and A. J. Smola, 2002, Learning with kernels: Support vector machines, regularization, optimization, and beyond: MIT Press, Cambridge, United States.

Shawe-Taylor, J. and N. Cristianini, 2004, Kernel methods for pattern analysis: Cambridge University Press, New York, United States.

Slatt, M. R., and Y. Abousleiman, 2011, Merging sequence stratigraphy and geomechanics for unconventional gas shales: The Leading Edge, **30**, 274-282.

Strecker, U., and R. Uden, 2002, Data mining of 3D post- stack attribute volumes using Kohonen self-organizing maps: The Leading Edge, **21**, 1032-1037.

Tong, S. and D. Koller, 2002, Support vector machine active learning with applications to text classification: The Journal of Machine Learning Research, **2**, 45-66.

Torres, A. and J. Reveron, 2013, Lithofacies discrimination using support vector machines, rock physics and simultaneous seismic inversion in clastic reservoirs in the Orinoco Oil Belt, Venezuela: 83rd Annual International Meeting, SEG, Expanded Abstracts, 2578-2582.

Turing, A.M., 1950, Computing machinery and intelligence. Mind, **59**, 433-460.

Uruski, C. I., 2010, New Zealand's deepwater frontier: Marine and Petroleum Geology, **27**, 2005-2026.

van der Baan, M. and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032-1047.

Vapnik, V., 1998, Statistical learning theory: Wiley & Sons, Hoboken, United States.

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

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

Wang, G., T. R. Carr, Y. Ju, and C. Li, 2014, Identifying organic-rich Marcellus Shale lithofacies by support vector machine classifier in the Appalachian basin: Computers and Geosciences, **64**, 52-60.

West, P. B., R. S. May, E. J. Eastwood, and C. Rossen, 2002, Interactive seismic facies classification using textural attributes and neural networks: The Leading Edge, **21**, 1042-1049.

Wong, K. W., Y. S. Ong, T. D. Gedeon, and C. C. Fung, 2005, Reservoir characterization using support vector machines: Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on, **2**, 354-359.

Yu, S., K. Zhu, and F. Diao, 2008, A dynamic all parameters adaptive BP neural networks model and its application on oil reservoir prediction: Applied Mathematics and Computation, **195**, 66-75.

Zhang, B., T. Zhao, X. Jin, and K. J. Marfurt, 2015, Brittleness evaluation of resource plays by integrating petrophysical and seismic data analysis (accepted by Interpretation).

Zhao, T., V. Jayaram, K. J. Marfurt, and H. Zhou, 2014, Lithofacies classification in Barnett Shale using proximal support vector machines: 84th Annual International Meeting, SEG, Expanded Abstracts, 1491-1495.

Zhao, T. and K. Ramachandran, 2013, Performance evaluation of complex neural networks in reservoir characterization: applied to Boonsville 3-D seismic data: 83rd Annual International Meeting, SEG, Expanded Abstracts, 2621-2624.

Zhao, B., H. Zhou, and F. Hilterman, 2005, Fizz and gas separation with SVM classification: 75th Annual International Meeting, SEG, Expanded Abstracts, 297-300.

APPENDIX: Mathematical details

In this appendix I summarize many of the mathematical details defining the various algorithm implementations. Although insufficient to allow a straightforward implementation of each algorithm, I hope to more quantitatively illustrate the algorithmic assumptions as well as algorithmic similarities and differences. Because K-means and artificial neural networks have been widely studied, in this appendix I only give some principle statistical background, and brief reviews of SOM, GTM, and PSVM algorithms involved in this tutorial. I begin this appendix by giving statistical formulations of the covariance matrix, principal components and the Mahalanobis distance when applied to seismic attributes. I further illustrate the formulations and some necessary theory for SOM, GTM, ANN, and PSVM. Because of the extensive use of mathematical symbols and notations, a table of shared mathematical notations is given in Table 2.3. All other symbols are defined in the text.

Covariance matrix, principal components, and the Mahalanobis distance

Given a suite of *N* attributes, the covariance matrix is defined as

$$C_{mn} = \frac{1}{J}\sum_{j=1}^{J}\left(a_{jm}(t_j, x_j, y_j) - \mu_m\right)\left(a_{jn}(t_j, x_j, y_j) - \mu_n\right), \qquad (2.1)$$

where a_{jm} and a_{jn} are the *m*th and *n*th attributes, *J* is the total number of data vectors, and where

$$\mu_n = \frac{1}{J}\sum_{j=1}^{J} a_{jn}(t_j, x_j, y_j), \qquad (2.2)$$

is the mean of the *n*th attribute. If we compute the eigenvalues, λ_i, and eigenvectors, $\mathbf{v}_i$, of the real, symmetric covariance matrix, **C**, the *i*th principal component at data vector *j* is defined as

$$p_{ji} = \sum_{n=1}^{N} a_{jn}(t_j, x_j, y_j)\, v_{ni}, \tag{2.3}$$

where v_{ni} indicates the *n*th attribute component of the *i*th eigenvector. In this chapter, the first two eigenvectors and eigenvalues are also used to construct an initial model in both the self-organizing map (SOM) and generative topographic mapping (GTM) algorithms.

The Mahalanobis distance, r_{jq}, of the *j*th sample from the *q*th cluster center, θ_q, is defined as

$$r_{jq}^2 = \sum_{n=1}^{N} \sum_{m=1}^{N} (a_{jn} - \theta_{nq}) C_{nm}^{-1} (a_{jm} - \theta_{mq}), \tag{2.4}$$

where the inversion of the covariance matrix, **C**, takes place prior to extracting the *mn*th element.

Self-organizing map

Rather than computing the Mahalanobis distance, both SOM and GTM first normalize the data using a z-score. If the data exhibit an approximately Gaussian distribution, the z-score of the *n*th attribute is obtained by subtracting the mean and dividing by the standard deviation (the square root of the diagonal of the covariance matrix, $\mathbf{C}_{nn}$). To z-score non-Gaussian distributed data, such as coherence, one needs to first break the data using histograms that approximate a Gaussian. The objective of the SOM algorithm is to map the input seismic attributes onto a geometric manifold called the self-organized map. The SOM manifold is defined by a suite of prototype vectors $\mathbf{m}_k$ lying on a lower-dimensional (in this case, 2D) surface which fit the N-dimensional

attribute data. The prototype vectors $\mathbf{m}_k$ are typically arranged in 2D hexagonal or rectangular structure maps that preserve their original neighborhood relationship, such that neighboring prototype vectors represent similar data vectors. The number of prototype vectors in the 2D map determines the effectiveness and generalization of the algorithm. One strategy is to estimate the number of initial clusters, and then to either divide or join clusters based on distance criteria. In this case, I follow Gao (2007) and overdefine the number of clusters to be the maximum number of colors supported by our visualization software. Interpreters then either use their color perception or construct polygons on 2D histograms to define a smaller number of clusters.

The implementation of the SOM algorithm is summarized in Figure 2.26. After computing z-scores of the input data, the initial manifold is defined to be a plane defined by the two first principal components. Prototype vectors $\mathbf{m}_k$ are defined on a rectangular grid to the first two eigenvalues to range between $\pm 2(\lambda_1)^{1/2}$ and $\pm 2(\lambda_2)^{1/2}$. The seismic attribute data are then compared to each of the prototype vectors, finding the nearest one. This prototype vector and its nearest neighbors (those that fall within a range σ, defining a Gaussian perturbation) are moved towards the data point. After all the training vectors have been examined, the neighborhood radius, σ, is reduced. Iterations continue until σ approaches the distance between the original prototype vectors. Given this background, Kohonen (2001) defines the SOM training algorithm using the following five steps:

Step 1: Randomly chose a previously z-scored input attribute vector, $\mathbf{a}_j$, from the set of input vectors.

Step 2: Compute the Euclidean distance between this vector $\mathbf{a}_j$ and all prototype vectors $\mathbf{m}_k$, *k=1,2,…,K*. The prototype vector which has the minimum distance to the input vector $\mathbf{a}_j$, is defined to be the "winner" or the Best Matching Unit, $\mathbf{m}_b$:

$$\left|\left|\mathbf{a}_j - \mathbf{m}_b\right|\right| = \min_k \left\{\left|\left|\mathbf{a}_j - \mathbf{m}_k\right|\right|\right\}. \tag{2.5}$$

Step 3: Update the "winner" prototype vector and its neighbors. The updating rule for the weight of the *k*th prototype vector inside and outside the neighborhood radius $\sigma(t)$ is given by

$$\mathbf{m}_k(t+1) = \begin{cases} \mathbf{m}_k(t) + \alpha(t)h_{bk}(t)\left[\mathbf{a}_j - \mathbf{m}_k(t)\right], & if\ \left|\left|\mathbf{r}_k - \mathbf{r}_b\right|\right| \le \sigma(t) \\ \mathbf{m}_k(t), & if\ \left|\left|\mathbf{r}_k - \mathbf{r}_b\right|\right| > \sigma(t) \end{cases} \tag{2.6}$$

where the neighborhood radius defined as $\sigma(t)$ is predefined for a problem and decreases with each iteration *t*. $\mathbf{r}_b$ and $\mathbf{r}_k$ are the position vectors of the winner prototype vector $\mathbf{m}_b$ and the *k*th prototype vector $\mathbf{m}_k$ respectively. We can also define the neighborhood function, $h_{bk}(t)$, the exponential learning function, $\alpha(t)$, and the length of training, *T*. $h_{bk}(t)$ and $\alpha(t)$ decrease with each iteration in the learning process and are defined as

$$h_{bk}(t) = e^{-(||\mathbf{r}_b - \mathbf{r}_k||^2/2\sigma^2(t)}, \tag{2.7}$$

and

$$\alpha(t) = \alpha_0\left(\frac{0.005}{\alpha_0}\right)^{t/T}. \tag{2.8}$$

Step 4: Iterate through each learning step (steps 1-3) until the convergence criterion (which depends on the predefined lowest neighborhood radius and the minimum distance between the prototype vectors in the latent space) is reached.

Step 5: Project the prototype vectors onto the first two principal components and color code using a 2D color bar (Matos et al. 2009).

Generative topographic mapping

In GTM, the grid points of the 2D deformed manifold in N-dimensional attribute space define the centers, $\mathbf{m}_k$, of Gaussian distributions of variance $\sigma^2=\beta^{-1}$. These centers, $\mathbf{m}_k$, are in turn projected onto a 2D latent space, defined by a grid of nodes $\mathbf{u}_k$ and nonlinear basis functions, $\mathbf{\Phi}$:

$$\mathbf{m}_k = \sum_{m=1}^{M} W_{km}\Phi_m(\mathbf{u}_k)\ , \tag{2.9}$$

where $\mathbf{W}$ is a K X M matrix of unknown weights, $\Phi_m(\mathbf{u}_k)$ is a set of M nonlinear basis functions, $\mathbf{m}_k$ are vectors defining the deformed manifold in the N-dimensional data space, and $k=1,2,\ldots,K$ is the number of grid points arranged on a lower L-dimensional latent space (in this case, $L=2$). A noise model (the probability of the existence of a particular data vector $\mathbf{a}_j$ given weights $\mathbf{W}$ and inverse variance β) is introduced for each measured data vector. The probability density function, p, is represented by a suite of K radially symmetric N-dimensional Gaussian functions centered about $\mathbf{m}_k$ with variance of $1/\beta$:

$$p\left(\mathbf{a}_j\middle|\mathbf{W},\beta\right) = \sum_{k=1}^{K}\frac{1}{K}\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\beta}{2}\left|\left|\mathbf{m}_k-\mathbf{a}_j\right|\right|^2}. \tag{2.10}$$

The prior probabilities of each of these components are assumed to be equal with a value of $1/K$, for all data vectors $\mathbf{a}_j$. Figure 2.3 illustrates the GTM mapping from an $L=2D$ latent space to the 3D data space.

The probability density model (GTM model) is fit to the data $\mathbf{a}_j$ to find the parameters **W** and β using a maximum likelihood estimation. One of the popular techniques used in parameter estimations is the Expectation Maximization (EM) algorithm. Using Bayes' theorem, and the current values of the GTM model parameters **W** and β, we can calculate the *J X K* posterior probability or responsibility, R_{jk}, for each of the *K* components in latent space for each data-vector:

$$R_{jk} = \frac{e^{-\frac{\beta}{2}\left|\left|\mathbf{m}_k - \mathbf{a}_j\right|\right|^2}}{\sum_i e^{-\frac{\beta}{2}\left|\left|\mathbf{m}_i - \mathbf{a}_j\right|\right|^2}}. \tag{2.11}$$

Equation 2.11 forms the "E-step" or Expectation step in the EM algorithm. The E-step is followed by the Maximization or "M-step", which uses these responsibilities to update the model for a new weight matrix **W** by solving a set of linear equations (Dempster et al., 1977):

$$\left(\mathbf{\Phi}^T\mathbf{G}\mathbf{\Phi} + \frac{\alpha}{\beta}\mathbf{I}\right)\mathbf{W}^T{}_{new} = \mathbf{\Phi}^T\mathbf{R}\mathbf{X}, \tag{2.12}$$

where

$G_{kk} = \sum_{j=1}^{J} R_{jk}$ are the non-zero elements of the *K X K* diagonal matrix **G**,

Φ is a *K X M* Matrix with elements $\mathbf{\Phi} = \Phi_m(\mathbf{u}_k)$,

α is a regularization constant to avoid division by zero, and

I is the *M X M* identity matrix.

The updated value of β is given by

$$\frac{1}{\beta_{new}} = \frac{1}{JN}\sum_{j=1}^{J}\sum_{k=1}^{K} R_{jk}\ \left|\left|\mathbf{W}_{km_{new}}\Phi_m(\mathbf{u}_k) - \mathbf{a}_j\right|\right|^2. \tag{2.13}$$

The initialization of **W** is done so that the initial GTM model approximates the principal components (largest eigenvectors) of the input data, $\mathbf{a}_j$. The value of β^{-1} is initialized to be the larger of the $(L+1)^{th}$ eigenvalue from PCA where L is the dimension of the latent space. In Figure 2.3, $L=2$, such that we can initialize β^{-1} to be the inverse of the third eigenvalue. Figure 2.27 summarizes this workflow.

Artificial Neural Networks

The artificial neural networks are a class of pattern recognition algorithms that were derived separately in different fields such as statistics and artificial intelligence. Artificial neural networks are easily accessible for most of the geophysical interpreters, so I only provide a general workflow of applying an ANN to seismic facies classification for completeness of this tutorial. The workflow is shown in Figure 2.28.

Proximal support vector machines

Because SVMs are originally developed to solve binary classification problems, the arithmetic I begin with a summary of the arithmetic describing a binary PSVM classifier.

Similarly to SVM, a PSVM decision condition is defined as (Figure 2.6):

$$\mathbf{x}^T\boldsymbol{\omega} - \gamma \begin{cases} > 0, & \mathbf{x} \in X+; \\ = 0, \mathbf{x} \in X+ \text{ or } X-; \\ < 0, & \mathbf{x} \in X-, \end{cases} \tag{2.14}$$

where **x** is an N-dimensional attribute vector to be classified, **ω** is an $N\ X\ 1$ vector implicitly defines the normal of the decision-boundary in the higher dimensional space, γ defines the location of the decision-boundary, and "$X+$" and "$X-$" indicate the two classes of the binary classification. PSVM solves an optimization problem and takes the form of (Fung and Mangasarian, 2001):

$$\min_{\boldsymbol{\omega},\gamma,\mathbf{y}} \varepsilon \frac{1}{2}\|\mathbf{y}\|^2 + \frac{1}{2}(\boldsymbol{\omega}^T\boldsymbol{\omega} + \gamma^2), \tag{2.15}$$

subject to

$$\mathbf{D}(\mathbf{a}\boldsymbol{\omega} - \mathbf{e}\gamma) + \mathbf{y} = \mathbf{e}. \tag{2.16}$$

In this optimization problem, $\mathbf{y}$ is a *J X 1* error variable; $\mathbf{a}$ is a *J X N* sample matrix composed of *J* attribute vectors, which can be divided into two classes, "$X+$" and "$X-$". $\mathbf{D}$ is a J X J diagonal matrix of labels with a diagonal composed of "$+1$" for "$X+$" and "-1" for "$X-$". ε is a non-negative parameter. Finally, $\mathbf{e}$ is a *J X 1* column vector of ones. This optimization problem can be solved by using a *J X 1* Lagrangian multiplier $\mathbf{t}$:

$$L(\boldsymbol{\omega},\gamma,\mathbf{y},\mathbf{t}) = \varepsilon\frac{1}{2}\|\mathbf{y}\|^2 + \frac{1}{2}(\boldsymbol{\omega}^T\boldsymbol{\omega} + \gamma^2) - \mathbf{t}^T(\mathbf{D}(\mathbf{a}\boldsymbol{\omega} - \mathbf{e}\gamma) + \mathbf{y} - \mathbf{e}). \tag{2.17}$$

By setting the gradients of L to zero, I obtain expressions for $\boldsymbol{\omega}$, γ and $\mathbf{y}$ explicitly in the knowns and $\mathbf{t}$, where $\mathbf{t}$ can further be represented by $\mathbf{a}$, $\mathbf{D}$ and ε. Then by substituting $\boldsymbol{\omega}$ in Equation 2.15 and 2.16 using its dual equivalent $\boldsymbol{\omega} = \mathbf{a}^T\mathbf{D}\mathbf{t}$, we can arrive at (Fung and Mangasarian, 2001):

$$\min_{\boldsymbol{\omega},\gamma,\mathbf{y}} \varepsilon \frac{1}{2}\|\mathbf{y}\|^2 + \frac{1}{2}(\mathbf{t}^T\mathbf{t} + \gamma^2), \tag{2.18}$$

subject to

$$\mathbf{D}(\mathbf{a}\mathbf{a}^T\mathbf{D}\mathbf{t} - \mathbf{e}\gamma) + \mathbf{y} = \mathbf{e}. \tag{2.19}$$

Equation 2.18 and 2.19 provide a more desirable version of the optimization problem since one can now insert kernel methods to solve nonlinear classification problems made possible by the term $\mathbf{a}\mathbf{a}^T$ in Equation 2.19. Utilizing the Lagrangian multiplier again (this time we denote the multiplier as $\boldsymbol{\tau}$), we can minimize the new optimization problem against $\mathbf{t}$, γ, $\mathbf{y}$ and $\boldsymbol{\tau}$. By setting the gradients of these four variables to zero, we can express $\mathbf{t}$, γ and $\mathbf{y}$ explicitly by $\boldsymbol{\tau}$ and other knowns, where $\boldsymbol{\tau}$ is solely a

dependent on the data matrices. Then for N-dimensional attribute vector **x** we write the decision conditions as

$$\mathbf{x}^T\mathbf{a}^T\mathbf{D}\mathbf{t} - \gamma \begin{cases} > 0, & \mathbf{x} \in X+; \\ = 0, \mathbf{x} \in X+ or\ X-; \\ < 0, & \mathbf{x} \in X-, \end{cases} \tag{2.20}$$

with

$$\mathbf{t} = \mathbf{D}\mathbf{K}^T\mathbf{D}\left(\frac{\mathbf{I}}{\varepsilon} + \mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{e}, \tag{2.21}$$

$$\gamma = \mathbf{e}^T\mathbf{D}\left(\frac{\mathbf{I}}{\varepsilon} + \mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{e}, \tag{2.22}$$

and

$$\mathbf{G} = \mathbf{D}[\mathbf{K} \quad -\mathbf{e}]. \tag{2.23}$$

Instead of **a**, we have **K** in Equation 2.21 and 2.23, which is a Gaussian kernel function of **a** and $\mathbf{a}^T$ that has the form of:

$$\mathbf{K}(\mathbf{a}, \mathbf{a}^T)_{ij} = exp\left(-\sigma\left\|\mathbf{a}^T{}_{i\cdot} - \mathbf{a}^T{}_{j\cdot}\right\|^2\right), i, j \in [1, J], \tag{2.24}$$

where σ is a scalar parameter. Finally, by replacing $\mathbf{x}^T\mathbf{a}^T$ by its corresponding kernel expression, the decision condition can be written as:

$$\mathbf{K}(\mathbf{x}^T, \mathbf{a}^T)\mathbf{D}\mathbf{t} - \gamma \begin{cases} > 0, & \mathbf{x} \in X+; \\ = 0, \mathbf{x} \in X+ or\ X-; \\ < 0, & \mathbf{x} \in X-. \end{cases} \tag{2.25}$$

and

$$\mathbf{K}(\mathbf{x}^T, \mathbf{a}^T)_{ij} = exp(-\sigma\|\mathbf{x} - \mathbf{a}^T{}_{i\cdot}\|^2), i \in [1, J]. \tag{2.26}$$

The formulations above represent a nonlinear PSVM classifier.

To extend this binary classifier to handle multiclass classification problems, some strategies have been developed by researchers, which generally fall into three categories: “one-versus-all”, “one-versus-one” and “all together”. For Q classes, the former two

strategies build a suite of binary classifiers individually: $(Q(Q-1)/2$ for the "one-versus-one" and Q for the "one-versus-all" algorithm, and then use these classifiers to construct the final classification decision. The "all together" attempts to solve multiclass problem in one step. Hsu and Lin (2002) found "one-versus-one" method to be superior for large problems. There are two particular algorithms for "one-versus-one" strategies, namely the "Max Wins" (Kreßel, 1999) and directed acyclic graph (DAG) (Platt et al., 2000) algorithms. Both algorithms provide comparable results while surpassing the "one-versus-all" method in accuracy and computational efficiency.

The approach in this study uses a classification factor table to assign classes to unknown samples (Figure 2.29). A classification factor of an unknown sample point for a certain pilot class "X" is the normalized distance to the binary decision boundary between "X" and the other class used when generating this binary decision boundary. An example of a classification factor table is shown in Figure 2.29, and based on this table, the unknown sample point belongs to class "D".

CHAPTER 3

DISTANCE PRESERVING SELF-ORGANIZING MAP AND ITS APPLICATION ON CHARACTERIZING A TURBIDITE SYSTEM IN CANTERBURY BASIN, OFFSHORE NEW ZEALAND[3]

INTRODUCTION

During the last two decades, seismic attributes have improved substantially, not only in innovative algorithm development, but also in their use in integrated interpretation. In addition to highlighting features of geological interest, multiple attributes can be combined for seismic facies classification. 3D co-rendering of multiple attributes is an effective way to visualize subtle features that may be overlooked on conventional amplitude data. Unfortunately, 32-bit color model (either RGB, HLS, or CMY color gamut modulated by alpha blending) limits us to combining at most four attributes. Furthermore, while some attribute combinations form natural pairs or triplets, such as dip magnitude, dip azimuth, and coherence (e.g. Marfurt, 2015), others do not, resulting in co-rendered images that are difficult to interpret. Seismic facies classification algorithms attempt to mimic human pattern recognition and can be applied to a more arbitrary collection and greater number of attributes. Feeding multiple attributes into a classification algorithm enables interpreters to analyze different aspects of seismic response (energy, frequency, phase, geometry, texture, etc.) simultaneously, generating a

[3]*This study is published as - Zhao, T., J. Zhang, F. Li, and K. J. Marfurt, 2016, Characterizing a turbidite system in Canterbury basin, New Zealand using seismic attributes and distance-preserving self-organizing maps: Interpretation, **4**, SB79-SB89.*

map of facies or correlating seismic responses to engineering/ production data (Roy et al., 2013; Zhang et al., 2015).

The rapid invention of fresh seismic attributes has greatly expanded the interpreters' arsenal for seismic interpretation. On the other hand, such great number of attributes put a serious challenge on inexperienced seismic interpreters: which attributes are most suitable to use? Generally, we prefer mathematically well designed, independent attributes with intuitive geologic and geophysical meaning (Barnes, 2007). Also, attribute selection is highly target oriented, which means we may use formation attributes to map lithology change, structure attributes for deformation, and edge detection attributes for fractures. In contrast, it may be inappropriate to use spectral magnitude to map a fault. Modern seismic interpretation often involves facies classification in which attributes serve as the input. The results of such applications are highly sensitive to the attribute selection process. During the analysis, I chose several attributes that as a human interpreter I thought represented important architectural elements. The classification scheme saw things differently such that these attributes needed to be rejected. In this chapter I therefore not only present attributes I used in seismic facies analysis to characterize the turbidite systems, but discuss attributes that were eliminated during the classification.

Seismic attributes are routinely used by interpreters. In contrast, seismic facies classification, deeply rooted in statistics, and requiring significant computing power, is less commonly used. During the last twenty years, several alternative algorithms have been proposed and successfully applied to computer-assisted seismic facies classification (Zhao et al., 2015). Kohonen self-organizing map (SOM) (Kohonen, 1982) is one of the

most widely accessible technique in commercial software packages that produces reasonably high quality seismic facies maps. SOM preserves the topological connections among clusters, which is a preferred attribute for visualization when using similar colors for similar facies (Strecker and Uden, 2002; Roy et al., 2013). However, traditional Kohonen SOM does not preserve the distance in input space, which may result in an over/under shrinking of the clusters in the SOM latent space. A latent space is a lower dimensional space, in which the original input data are projected into. Analyze data in a particular latent space may discover data properties that are easily overlooked in the original space. A frequently used example would be projecting multiattribute data using principle component analysis (PCA) and crossplotting the first two principle components. Such crossplot serves as a 2D latent space of the original multiattribute data. Intuitively, data vectors with similar seismic attribute responses should belong to similar facies. However, after mapping into a latent space in which the SOM nodes live in, the projection of such data vectors may no longer maintain the same degree of "similarity" between each other. For this reason, I need to introduce some constraints to maintain the consistency of "similarity" among clusters in both n-dimensional input and (in this example) two-dimensional latent space.

During the development of SOM, several variants have been introduced to address this issue, including grouping neuron SOM (GNSOM) (Rubio and Gimenez, 2003), adaptive coordinate SOM (ACSOM) (Merkl and Rauber, 1997), and double SOM (DSOM) (Wang et al., 2002), to name a few. All these algorithms are based on Himberg's contraction model (Himberg, 2000), which means the adjustment of distance among SOM nodes in the SOM latent space is only contraction, but not expansion. In this study I

implement Shao and Yang's (2012) distance-preserving SOM (DPSOM) (in which the distance adjustment among SOM nodes can be both contraction and expansion.

I begin this chapter by discussing the geologic setting in the Canterbury basin, followed by the attribute expression of a turbidite system. Then I will introduce DPSOM, apply it to a simple synthetic, and compare it with traditional Kohonen SOM. I then apply DPSOM to a Canterbury basin, New Zealand, turbidite, with the goal of mapping key architectural elements. I conclude with a final comparison to traditional SOM, summarizing the value and limitations of SOM in general.

GEOLOGIC SETTING

I examine a turbidite system imaged in the Waka-3D seismic survey acquired over the Canterbury basin, offshore New Zealand, which has been generously made public by the New Zealand Petroleum and Minerals. Figure 3.1 shows the location of this survey, where the red rectangle corresponds to the area shown in subsequent figures. The study area lies on the transition zone of continental slope and rise, with an abundance of paleocanyons and turbidite deposits of Cretaceous and Tertiary ages. These sediments were deposited in a single, tectonically driven transgressive – regressive cycle (Uruski, 2010). Potential source rocks in the Canterbury basin are Late Jurassic coaly sediments, Late Cretaceous coaly sediments, and Paleocene marine shales. Reservoir rocks are fluvial, estuarine and marine sandstones of Cretaceous and Tertiary (Sutherland and Browne, 2003). Being a very recent and underexplored prospect, publically available comprehensive studies of the Canterbury basin are somewhat limited. The modern seafloor canyons shown in Figure 3.1 are good analogs of the deeper paleocanyons illuminated by the 3D seismic amplitude and attribute data.

In this study my goal is to analyze a turbidite system of Miocene age characterized by abundant submarine channels. This Miocene turbidite system is characterized by abundant of stacked submarine channel fills such that there are only piecewise continuous horizons within the zone of interest. Such complexity is more amenable to 3D volumetric interpretation using 3D visualization, geobodies, and in this paper, volumetric facies mapping. Animation through time slices provides a quick look at the geomorphology (Figure 3.2), but because of subsidence, no single time slice displays the complete depositional system. In this study, I interpreted a continuous reflector below the zone of interest, then used this horizon to generate a phantom horizon (horizon A) within the Miocene turbidite system. I then assume that deposition along horizon A represents a consistent geologic time interval. Figure 3.3 shows seismic amplitude along horizon A, on which I identify stacked channels (white arrows), high amplitude deposits (blue arrow), and slope fans (red arrow) which are also shown on Figure 3.2. On this phantom horizon slice another weaker, sinuous channel (black arrow), can be identified which cannot be seen on the time slice in Figure 3.2.

ATTRIBUTE EXPRESSION

Rather than replacing the human interpreters, a good facies analysis workflow builds on the interpreters' experience and objectives. Given sufficient time, interpreters should be able to identify all the facies in a high quality data volume. With less time, important but subtle facies and spatial relationships might be overlooked. In this work, seismic attributes quantify key features in the amplitude data for subsequent computer analysis. Amplitude and frequency attributes are often lithology indicators that may be hydrocarbon indicators in conventional reservoirs; geometric attributes quantify

reflector morphology such as dip, curvature, rotation, and convergence; statistical attributes provide information about data distribution that quantifies subtle patterns that may be easy to identify but hard to define (Chopra and Marfurt, 2007). Attributes such as coherence provide images of the edges of seismic facies rather than a measure of the facies themselves, although slumps often appear as a suite of closely spaced faults separating rotated fault blocks. Since the objective is to map a deep water channel system consisting of incised and multistoried channels, levees, point bars, channel flanks, slope fans, and slumps, I generate attributes that highlight these features. Peak spectral frequency and peak spectral magnitude can be used to differentiate thick from thin channels as well as overbank deposits. Reflector convergence differentiates conformal reflectors from pinchouts and angular unconformities. Gray-level co-occurrence matrix (GLCM) attributes are texture attributes representing the degree of disorder, quantifying subtle variation in the depositional environment and useful to characterize geomorphology changes within channel beds, channel flanks, and slope fans. Similarity (coherence) attributes help to detect edges compartmentalizing different deposition features.

Proper display of generated attributes is almost as important as selecting the suitable attributes, and co-rendering certain attributes correctly will greatly enhance the interpretability of an image. By employing a hue-lightness-saturation (HLS) color model, I follow Marfurt (2015) to display the generated attributes in the subsequent combinations, and the rule of thumb is to display one background attribute against hue, one modulating attribute against saturation, and one calibration attribute against lightness. Figure 3.4 shows phantom horizon A through peak spectral frequency co-rendered with

peak spectral magnitude that emphasizes the relative thickness and reflectivity of the turbidite system and the surrounding slope fan. The edges of the channels are delineated by Sobel filter similarity. I show the same phantom horizon A through (Figure 3.5) co-rendered shape index and curvedness, and through (Figure 3.6) GLCM homogeneity co-rendered with coherent energy. In Figure 3.7 I show vertical slices at line AA' in Figure 3.3 through (Figure 3.7a) seismic amplitude, (Figure 3.7b) seismic amplitude co-rendered with peak spectral magnitude/peak spectral frequency, (Figure 3.7c) seismic amplitude co-rendered with shape index and curvedness, and (Figure 3.7d) seismic amplitude co-rendered with GLCM homogeneity and coherent energy. Block arrows indicate several of the key facies, white being channels, blue being high amplitude deposits, and red being a slope fan.

From the generated attributes we note that two main channels (white arrows in Figure 3.4, 3.5, 3.6, and 3.7) exhibit medium-low frequency, low peak spectral magnitude, a valley shape, and low coherent energy, suggesting mud fill. Depending on which part of the channel we examine, GLCM homogeneity ranges from low to medium, indicating significant complexity within the meander valley because of the frequent change in channel route. Also note several sinuous channel complexes developed at this level (black arrows in Figure 3.4, 3.5, and 3.6), with attribute responses generally similar to those of the two main channels. Some high amplitude reflectors are indicated by blue arrows in Figure 3.4, 3.5, and 3.6, which suggest sand deposits, but it is unclear whether they are point bar or levee deposits. In most of the survey area, there are highly homogeneous, medium energy, and relatively flat reflectors which are probably slope fans (red arrows in Figure 3.4, 3.5, 3.6, and 3.7).

Using co-rendering, we are able to display and analyze two or three attributes simultaneously. However, we cannot use visualization to meaningfully combine the information content of all the attributes into a single image. Seismic facies classification techniques use multiple attributes as input to generate a facies map. Barnes and Laughlin (2002) concluded that the appropriate choice of attributes was the most critical component of computers assisted seismic facies identification. The number of attributes should be as small as possible to discriminate the facies of interest. Each attribute should be mathematically independent of the others. In contrast, attributes should be correlated through the underlying geology, forming a "pattern" or waveform in the attribute space. We also need to be aware that because human eyes and computers see the same image differently (a human sees features as combination of different textures, while a computer only see "voxels"), attributes that are suitable for visual interpretation may not be appropriate for use in classification. In this study, I find shape index dominates the unsupervised classifications with valley and ridge features across the survey, which "contaminate" the facies map. In this particular case, GLCM homogeneity anomalies follow a similar trend as the low coherent energy, such that I decide not to use GLCM homogeneity as an input attribute to reduce the redundancy. The final group of input attributes for classification consists of peak spectral frequency, peak spectral magnitude, coherent energy, and curvedness. Figure 3.8 shows crossplots for every attribute pair on which we see very limited correlation. Such independence is key to successful seismic classification techniques.

SOM AND DPSOM

Seismic attributes serve as input to modern seismic classification techniques. Most geophysicists are familiar with SOM as a waveform classification technique, where the input "attributes" are amplitudes along adjacent phantom horizon slices (Poupon et al., 1999). Coléou et al. (2003) and Strecker and Uden (2002) showed how a suite of attributes can form a vector at each voxel. With this insight we can indeed conclude that waveform classification and multiattribute analysis are one and the same. The waveforms or attribute vectors are represented by a finite number of prototype vectors, each of which is assigned a unique color.

In traditional Kohonen SOM, the position of an SOM node in the SOM latent space is only based on the distance between the corresponding prototype vector (the projection of an SOM node in the input data space) and the nearest data vector in the input space. DPSOM adds a step of adjusting the position of all SOM nodes according to their distances from the current winning node (best matching unit) in both input data space and SOM latent space. The adjustment rule is (Shao and Yang, 2012):

$$\mathbf{p}_k(t+1) = \mathbf{p}_k(t) + \alpha(t) \cdot \left(1 - \frac{\delta_{vk}}{d_{vk}}\right) \cdot \left(\mathbf{p}_v(t) - \mathbf{p}_k(t)\right), \forall\, k \neq v. \qquad (1)$$

In Equation 1, $\mathbf{p}_k(t)$ is the positon of an SOM node before adjustment; $\mathbf{p}_k(t+1)$ is the position of an SOM node after adjustment; $\mathbf{p}_v(t)$ is the position of the current winning node; δ_{vk} and d_{vk} are the distance between an SOM node and the current winning node in input data space and SOM latent space, respectively. $\alpha(t)$ is the learning rate which exponentially decays over iterations.

In this implementation, the SOM latent space is a 2D space initialized using the first two principal components from principal component analysis (PCA). Defined by

Equation 1, an SOM node may either move towards or away from the current winning node in the SOM latent space to preserve their distance found in the input data space.

I generate a synthetic consisting of three attributes to test the proposed DPSOM algorithm and compare it with the traditional SOM (Figure 3.9). All attribute values are randomly generated within a certain range for each facies. Table 1 summarizes the ranges for each facies on each attribute. This model consists of three facies, a mud-filled channel, a gas-charged levee, and a flood plain.

I feed these three synthetic attributes into traditional Kohonen SOM as well as DPSOM with (over defined) 256 prototype vectors, and show the results in Figure 3.10. Note the traditional SOM generates a facies map with two facies (Figure 3.10a), splitting the flood plain into levee and channel facies. In contrast, DPSOM generates three facies, even though there are some local variations in the flood plain and levee facies (Figure 3.10b). The corresponding SOM latent spaces are shown in Figure 3.10c and 3.10d, where the color for each cluster represents the histogram count. We clearly see two clusters as end members in the traditional SOM, where we have three clusters in the DPSOM, two of which are further subdivided into two adjacent sub-clusters. Moreover, in Figure 3.10d we clearly see the distance between the green cluster (channel) and the cadet blue cluster (flood plain) is less than the distance between the cadet blue cluster and the deep pink cluster (levee). This is consistent with the model setting in which the attribute value ranges of flood plain are closer to that of channel than to that of levee. Such result confirms the advantage of the distance preserving character in DPSOM over the traditional SOM.

APPLICATION

Given the selected four attributes, I form four-dimensional attribute vectors at each voxel and feed them into the DPSOM classification algorithm. As an interpreter, I need to validate the final classification to determine if they delineate the seismic facies of interest. In this example I use Sobel filter similarity to separate the facies and then evaluate how they fit within interpreters' understanding of a turbidite system.

To ensure the DPSOM classifier analyzing data in a relatively constant depositional environment, I take a 100 *ms* analysis window centered about phantom horizon A. In this window, each voxel (inline, crossline, time) is represented by a four-dimensional data vector (peak spectral frequency, peak spectral magnitude, coherent energy, curvedness). After normalizing the input data along each dimension using the Mahalanobis distance (Mahalanobis, 1936), a 2D SOM latent space is initialized using the first two principle components. I then take 1 of every 125 data vectors (every 5^{th} inline, crossline, and time sample) as training data to construct the SOM model. For each SOM training iteration, the position of all SOM nodes are adjusted based on Equation 1, resulting in SOM model that honors the distance among data vectors in the 4D input space. Once the algorithm reaches the desired number of iteration (in this case, 5 iterations), the SOM model is applied to the data within the analysis window. Using only a single processor, the total calculation time is around 10 minutes, for a 25 sample by 600 inline by 1800 crossline data slab of 4-attribute data.

Figure 3.11 shows the DPSOM classification result along phantom horizon A. In the upper right corner I show a 2D colorbar and a 2D histogram representing the classes in the SOM latent space, with the two axes being the first two principle components of

the 4D input data. I then color code the classes using the 2D colorbar, and paint each sample point in the 3D seismic volume using the same color as the class to which it belongs. To better illustrate the distance-preserving characteristic of DPSOM, I compare with the result from classic Kohonon-SOM using the same parameters (Figure 3.12). It is observed that different geologic features in Figure 3.11 are represented by more distinct colors than in Figure 3.12. For example, the sinuous channel complex in the far north is more visually segmented from the surrounding slope fan in Figure 3.11 than in Figure 3.12. This is attributed to the distance between these two facies being better preserved in DPSOM, therefore the contrast between colors representing such facies are greater in Figure 12. Back to Figure 3.11, we then identify two main slope channels (white arrows) which are classified as cyan that converge downstream. Vertical slices show these both to be multistory channels. As the channels move downslope, the slope becomes gentler, such that sediments lose momentum, spread out, and form a lobate feature. Black arrows indicate several sinuous channel complexes. On the facies map we see the northernmost large channel complex bifurcating into two minor channels. Most of the channel fill appear as cyan, similar to the two main channels, which suggests they are probably mud filled. The coherent slope fans (indicated by red arrows) are characterized by brownish colors. The purplish color facies (orange arrows) are less coherent and may indicate massive turbidite current or slump deposits. The yellow arrow indicates an older, probably sand-filled channel within the multistoried system. Blue arrows indicate the same color facies that I interpreted to be an older, high energy, sand filled channel developed earlier than the mud filled channel cutting through it. This sand-filled channel

spreads out and contributes to the lobe further downslope where it is covered by mud deposits transported by later stage channels.

In Figure 3.11a-f, I also show several vertical slices through co-rendered seismic amplitude with SOM facies volumes. In each image, phantom horizon A is vertically centered within the SOM analysis window. Figure 3.11a shows the smaller scaled channels in the northern channel complex. I use red lines to delineate the longitudinal profile of the channels. I find the channels are mapped in cyan to purple colors, with the surrounding overbank complex deposits mapped in yellow to brown colors. This contrast in color has greatly simplified the characterization of such channel complexes, which otherwise are extremely tedious to interpret on traditional seismic amplitude data. Figure 3.11b displays the lateral migration of the northern main channel complex. In this vertical section we identify at least four channel stories migrating from northeast to southwest (yellow lines highlight different stories). Notice the oldest story is mapped as lime green, while the younger channels are cyan (with the youngest one out of the SOM analysis window). This suggests a change of lithology during deposition, with the oldest story sand-filled, and the younger stories mud-filled. Figure 3.11c shows the vertical profile of the more distal northern main channel. Comparing to section Figure 3.11d, which is more proximal, the distal part of the channel is more spread out, forming a lobe mixed with the southern main channel. We also see an oxbow-like feature (probably another sandy channel fill) in the older deposits (outlined in green). Figure 3.11d-f show three vertical profiles through the proximal to the distal portions of the southern main channel (outlined in red), in which we clearly see it cutting through an older sand-filled channel in lime green (outlined in green). Similar to the northern main channel, we see the width of the

southern channel (as well as the old sand-filled channel) has expanded dramatically from proximal to distal, merging with the northern lobe. The yellow line in Figure 3.11f highlights a younger mud-filled incised channel in the southern main channel.

In Figure 3.13 and 3.14, I show four prototype vectors representing four key facies. The location of the prototype vector in the 3D inline-crossline-time space and 2D SOM latent space is shown in Figure 3.13, while the normalized attribute responses of these four prototype vectors are in Figure 3.14. To ensure that the prototype vectors correctly represent the facies they belong to, all the attribute responses in Figure 3.14 are averaged from a 5 sample X 5 trace X 5 trace window centered at the target prototype vectors. As expected, the prototype vectors in similar facies (colors) have similar attribute responses. Prototype vector 1 is selected to represent sand-filled channel deposits, which has a similar response to that of prototype vector 3, which is selected to represent sandy overbank complexes. The inter-channel overbank complex (prototype vector 2) and the mud-filled sinuous channel complex (prototype vector 4) are in similar facies (blue to purple colors) and have different attribute responses. I also show the difference in reflection characteristics of the multistoried channel (yellow circle) and older sand filled channel (blue circle) that are in similar SOM facies. All such interpretations are solely based on seismic amplitude data, which need to be further verified by well log or core analysis.

CONCLUSIONS

In this study, I am able to characterize a Miocene deep water turbidite system in Canterbury basin, New Zealand, by incorporating seismic attributes, seismic facies classification, and a limited amount of structural interpretation (picking one horizon).

Phantom horizon slices through seismic attribute volumes effectively delineate depositional features where there are no continuous surfaces to pick. Co-rendering seismic attributes against an HLS color model facilitates the visual integration of two or three attributes at the same time, which greatly increases the value of seismic attributes. DPSOM is an innovated classification technique that produces an improved classification result by preserving the distance in input data space. The seismic facies map generated from DPSOM using appropriate seismic attributes provides statistics based machine assistance to identify depositional facies in a turbidite system. Multiple sinuous channel complexes and multistory channel stacking pattern are delineated precisely along with their depositional pattern, which can be further utilized to locate potential reservoirs.

CHAPTER 3 TABLES

Attribute / Facies	Attribute 1	Attribute 2	Attribute 3
Shale-filled channel	25 ~ 30	-0.7 ~ -0.4	0 ~ 2
Gas-charged levee	10 ~ 15	0.4 ~ 0.7	10 ~ 15
Flood plain	40 ~ 45	-0.1 ~ 0.1	4 ~ 6

Table 3.1. Value ranges for each facies on each attribute. All values are randomly generated within the given range.

CHAPTER 3 FIGURES

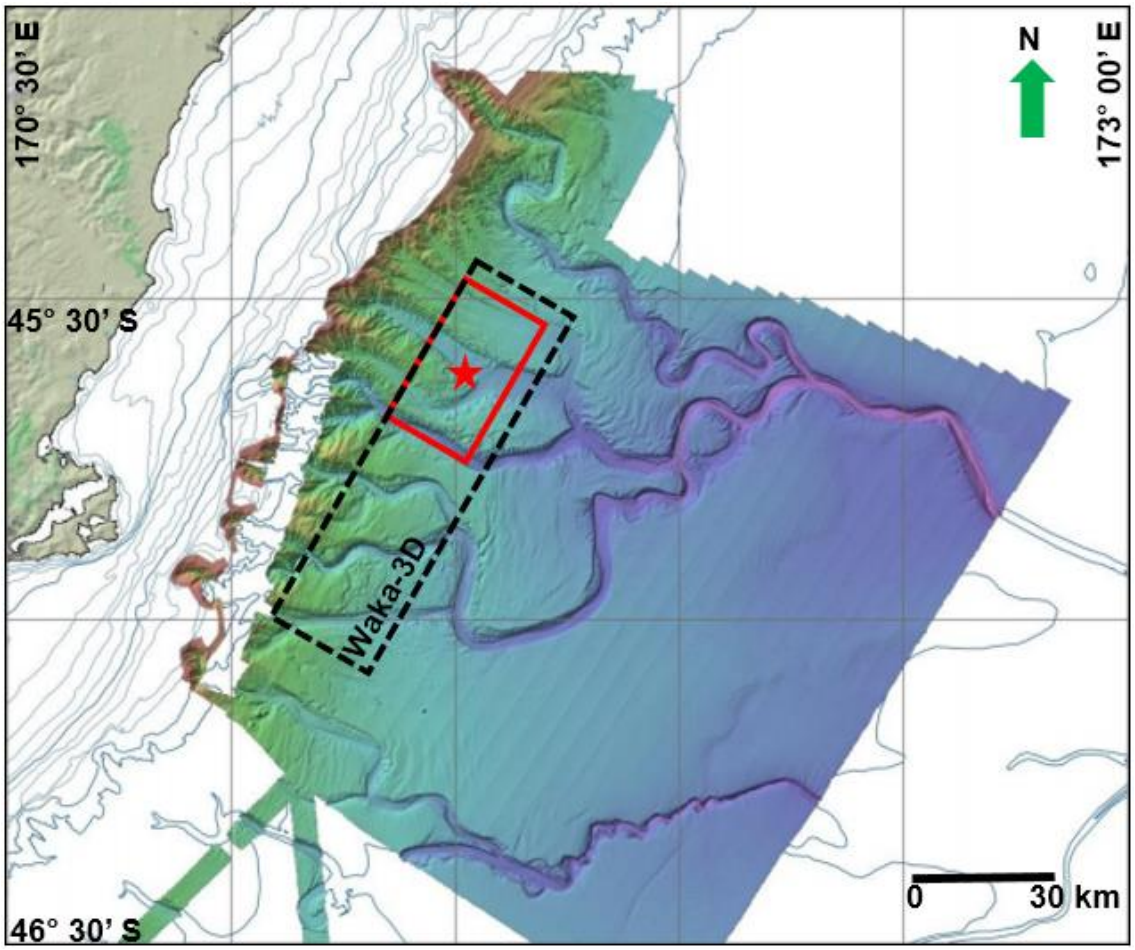


Figure 3.1. A map showing the location of the 3D seismic survey acquired over the Canterbury basin, offshore New Zealand. The black rectangle denotes the limits of the Waka-3D survey, while the smaller red rectangle denotes the part of the survey shown in subsequent figures. Colors represent the relative depth of the current seafloor, warm being shallower and cold being deeper. Current seafloor canyons are delineated in this map, which are good analogs for the paleocanyons in Cretaceous and Tertiary turbidites. (Modified from Mitchell and Neil, 2012)

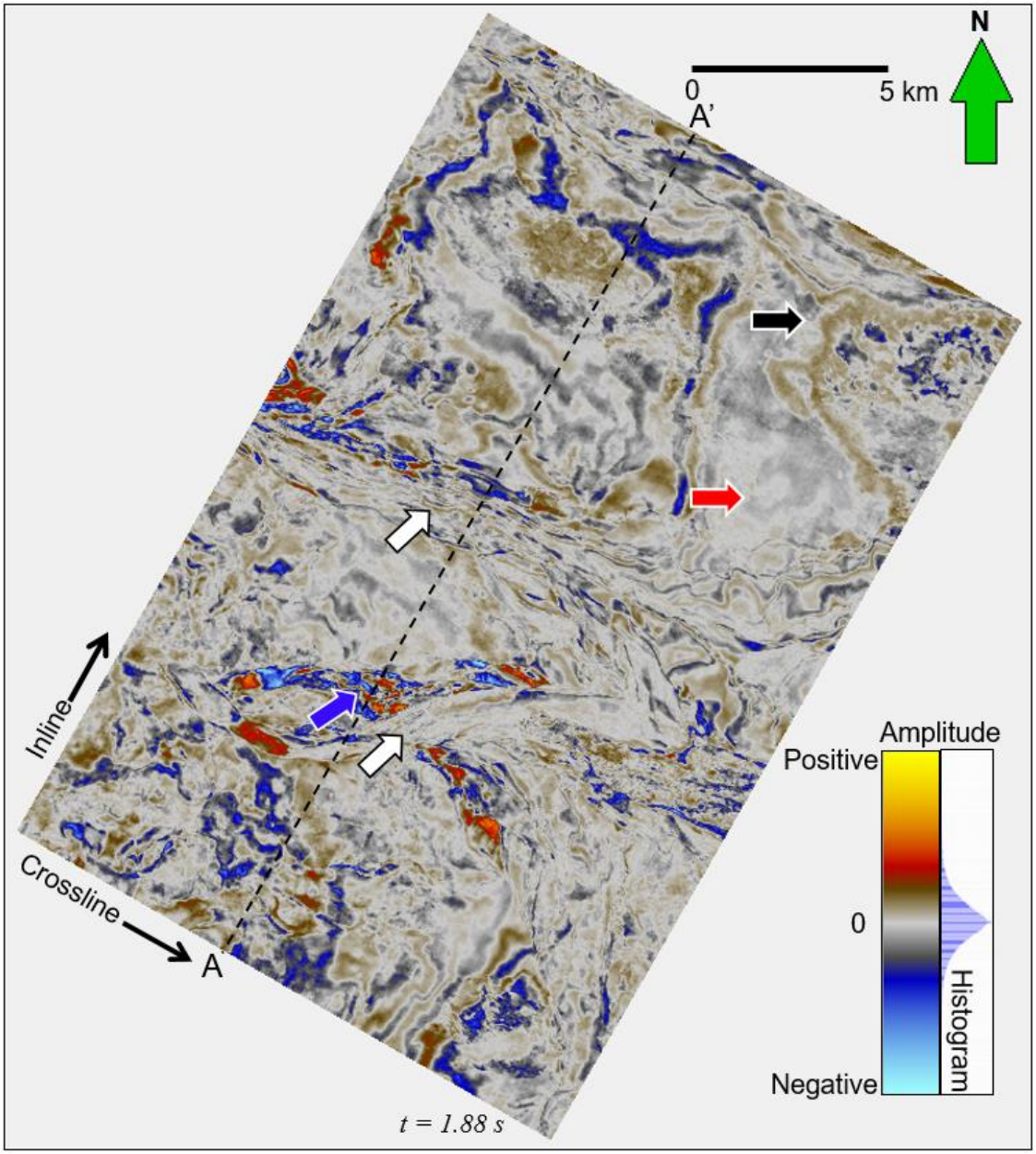


Figure 3.2. Time slice at t=1.88 s through the seismic amplitude volume. White arrows indicate potential channel/ canyon features. The blue arrow indicates a high amplitude feature. The red arrow indicates a relatively low energy, gently dipping area. AA' denotes a vertical cross section shown in Figure 3.7.

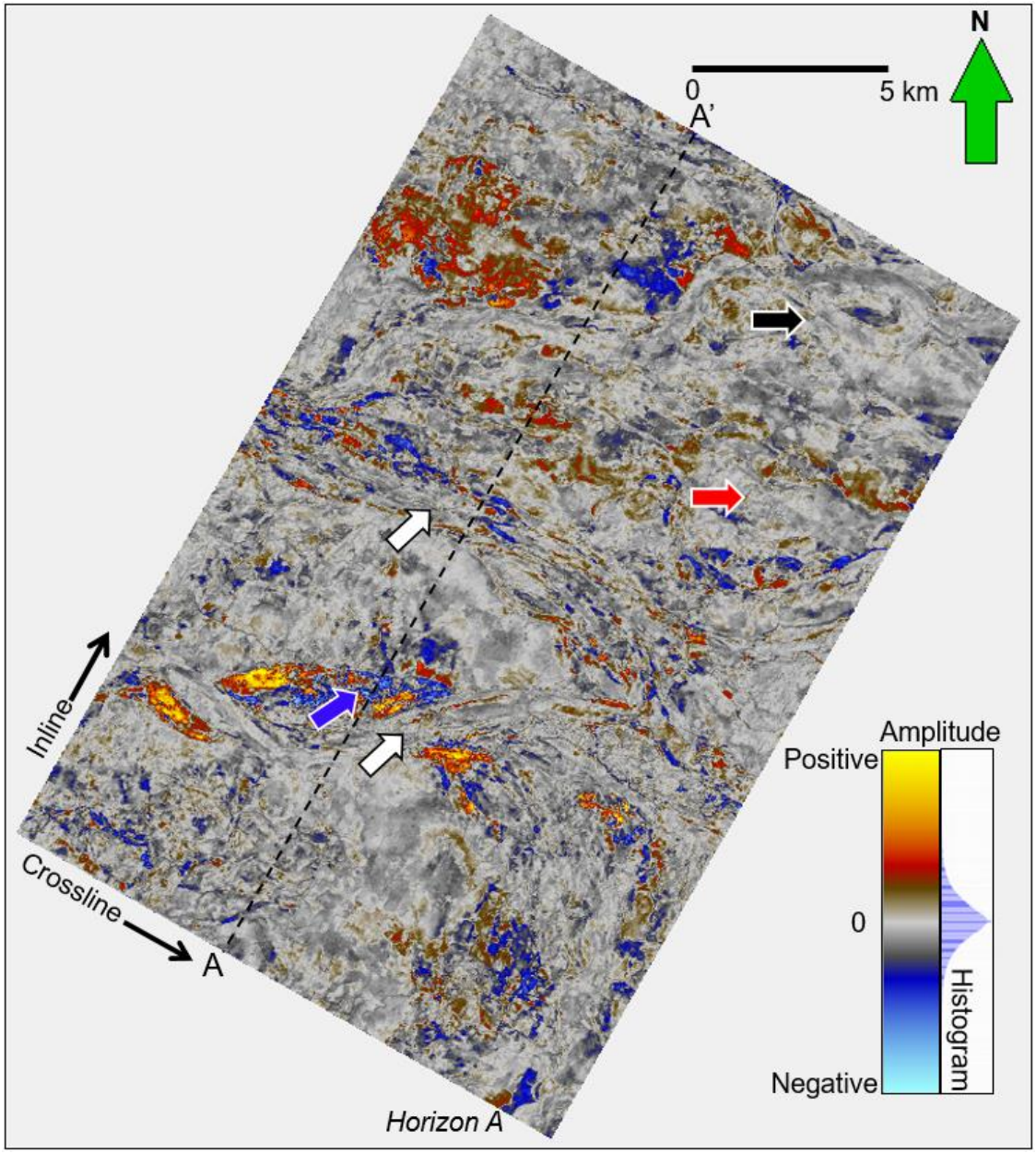


Figure 3.3. Horizon slice along horizon A through seismic amplitude volume. White arrows indicate potential channel/ canyon features. The blue arrow indicates a high amplitude feature. The red arrow indicates a relatively low energy, gently dipping area. The black arrow indicates a sinuous channel complex which is not seen in the time slice shown in Figure 3.2. AA' denotes a cross section shown in Figure 3.7.

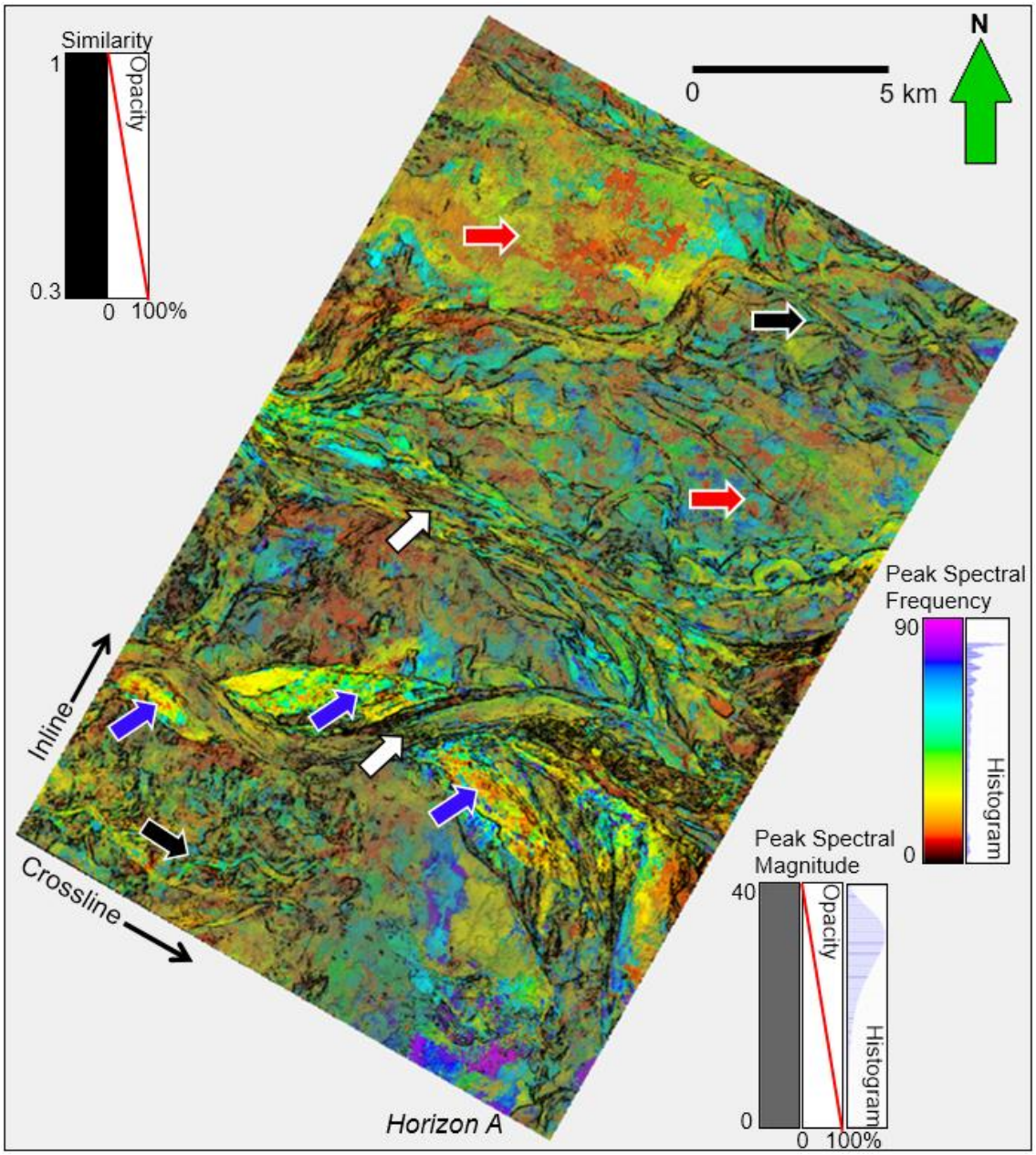


Figure 3.4. Horizon slice along horizon A through the peak spectral frequency co-rendered with peak spectral magnitude volumes. These spectral attributes computed using a continuous wavelet transform emphasize the relative thickness and reflectivity of the turbidite system and surrounding slope fan sediments into which it was incised. The edges of the channels are delineated by Sobel filter similarity. White arrows indicate two straight to meandering main channels; black arrows indicate sinuous channel complexes; blue arrows indicate high amplitude reflectors; and red arrows indicate gently dipping slope fans.

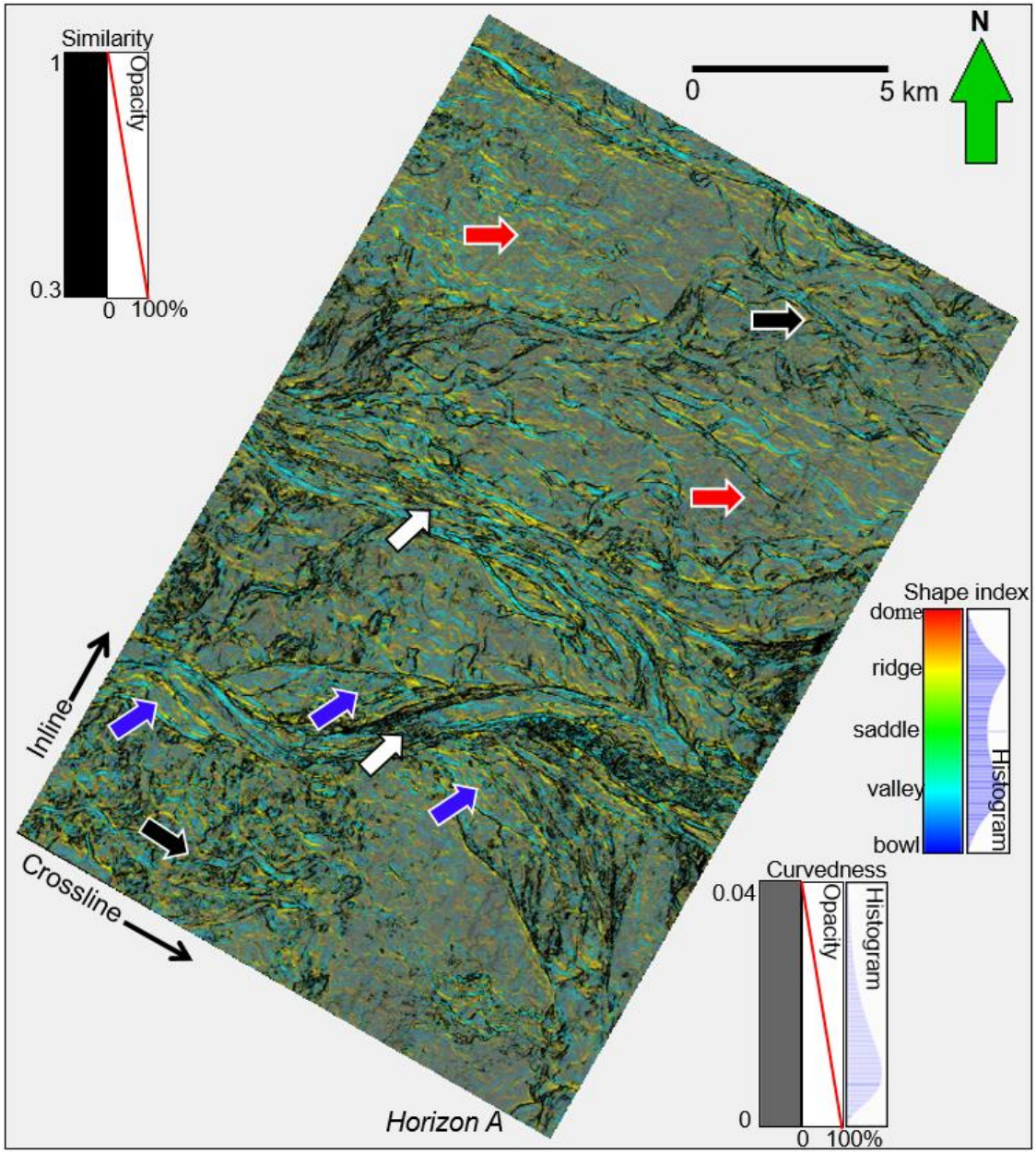


Figure 3.5. Horizon slice along horizon A through the co-rendered shape index, curvedness, and Sobel filter similarity volumes. The shape index co-rendered with curvedness is good for visual classification but dominates the unsupervised classifications with valley and ridge features across the survey. White arrows indicate two straight to meandering main channels; black arrows indicate sinuous channel complexes; blue arrows indicate high amplitude reflectors; and red arrows indicate gently dipping slope fans.

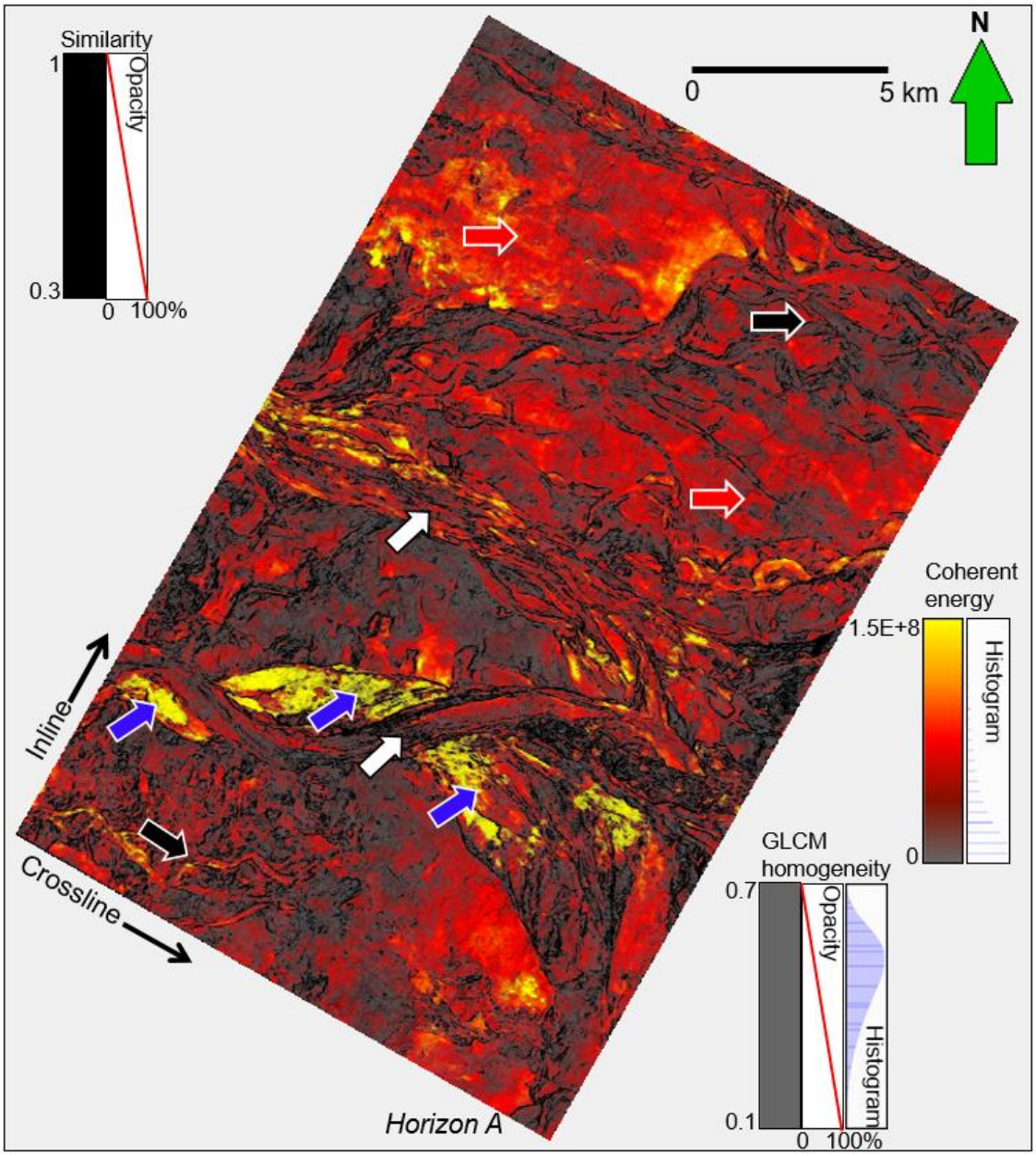


Figure 3.6. Horizon slice along horizon A through the co-rendered GLCM homogeneity, coherent energy, and Sobel filter similarity volumes. Bright colors highlight coherent, high energy areas that indicate potential sand deposits. White arrows indicate two straight to meandering main channels; black arrows indicate sinuous channel complexes; blue arrows indicate high amplitude reflectors; and red arrows indicate gently dipping slope fans.

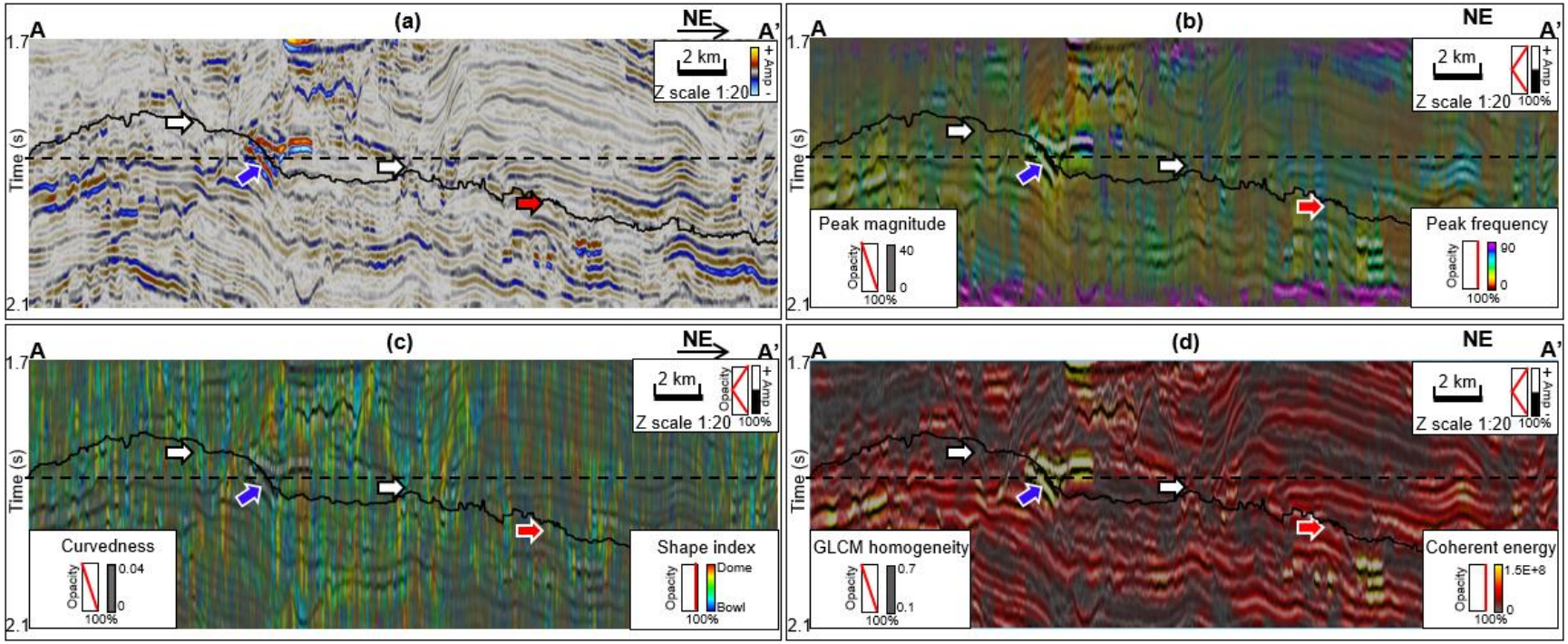


Figure 3.7. Vertical sections along line AA' (location shown in Figure 3.2 and 3.3) through (a) seismic amplitude, (b) seismic amplitude co-rendered with peak spectral magnitude and peak spectral frequency, (c) seismic amplitude co-rendered with shape index and curvedness, and (d) seismic amplitude co-rendered with GLCM homogeneity and coherent energy. White arrows indicate channel and canyon features. Blue arrows indicates a high amplitude reflector. Red arrows indicate relatively low amplitude, gently dipping areas. Solid black line indicates horizon A, and the dashed black line indicates *t = 1.88 s.*

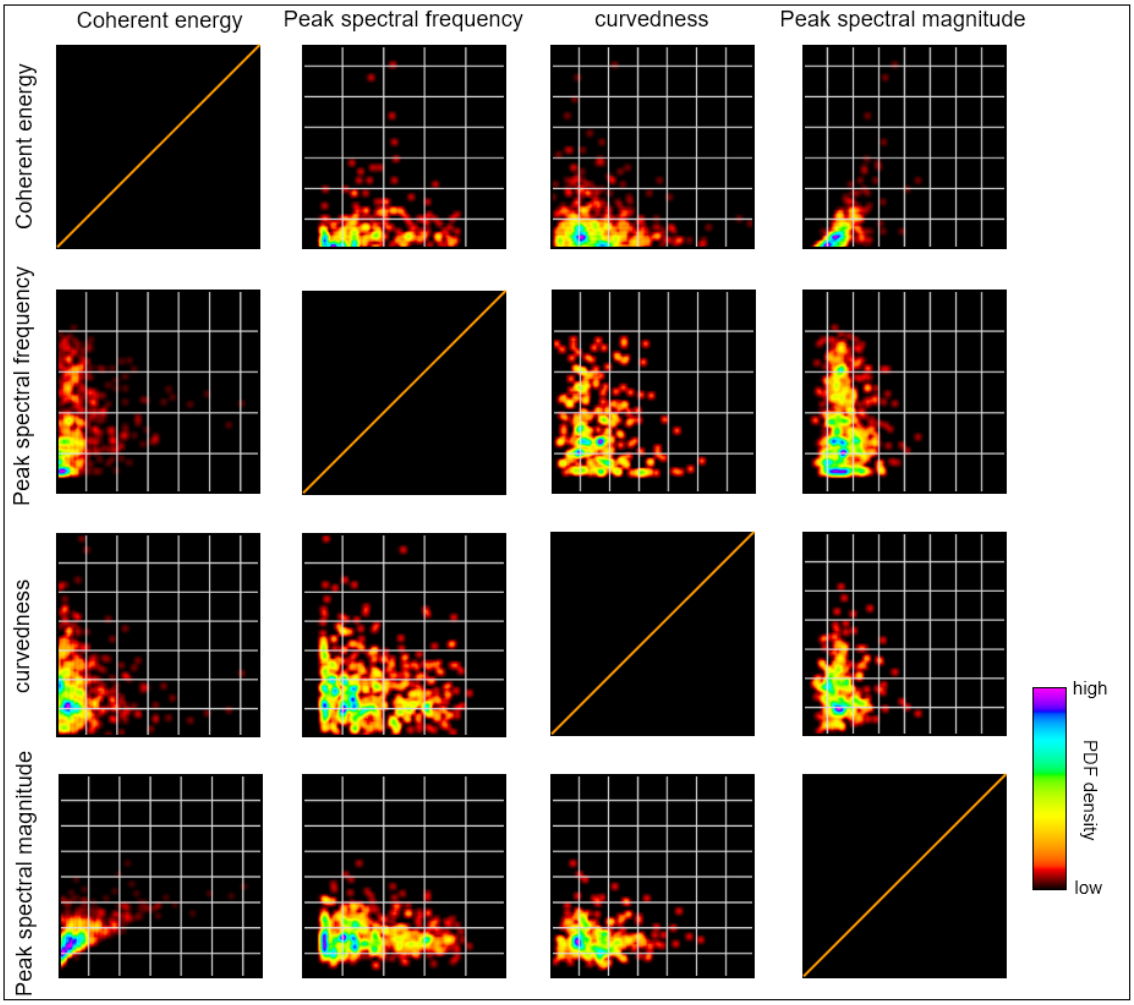


Figure 3.8. Crossplot among all input attribute pairs. Note all crossplots are in a cloud shape instead of following a straight diagonal line. Such behavior demonstrates that these four attributes are relatively independent.

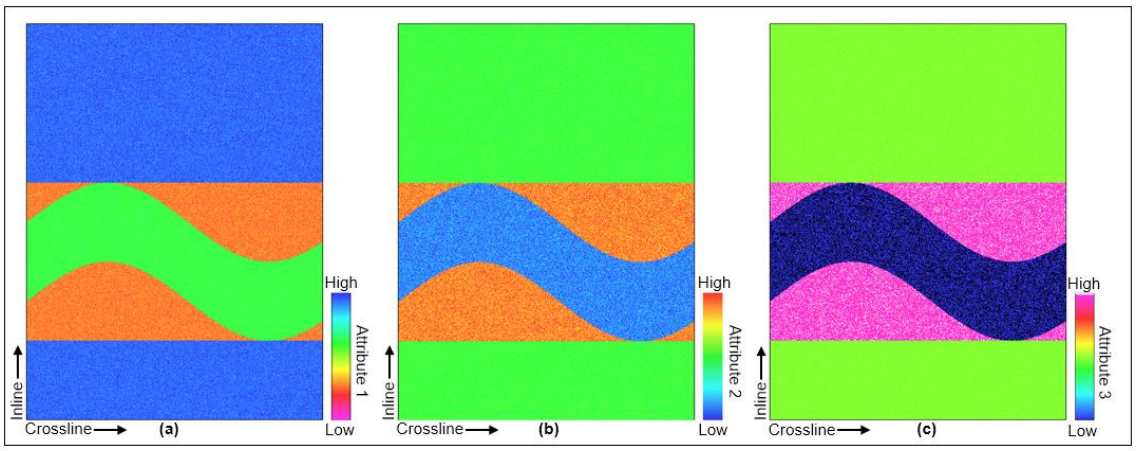


Figure 3.9. Time slices through three synthetic seismic attribute volumes, delineating three seismic facies.

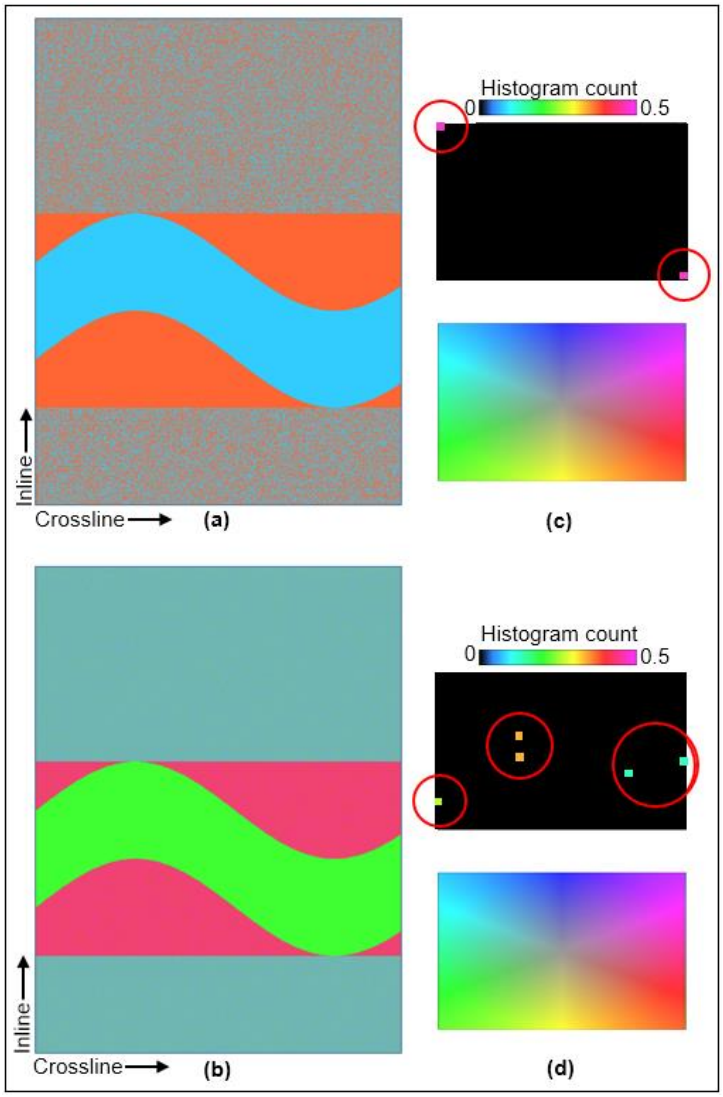


Figure 3.10. Classification result using (a) traditional SOM and (b) DPSOM for the synthetic data shown in Figure 3.9. (c) Corresponding SOM latent spaces and 3D colorbars for (a) traditional SOM and (d) DPSOM. Clusters in the upper figures in (c) and (d) are painted in colors according to their corresponding position in the 2D colorbar. For example, a cluster appears in the upper left corner of the SOM latent space is painted in orange color (the color in the upper left corner of the 2D colorbar) in the seismic faces map.

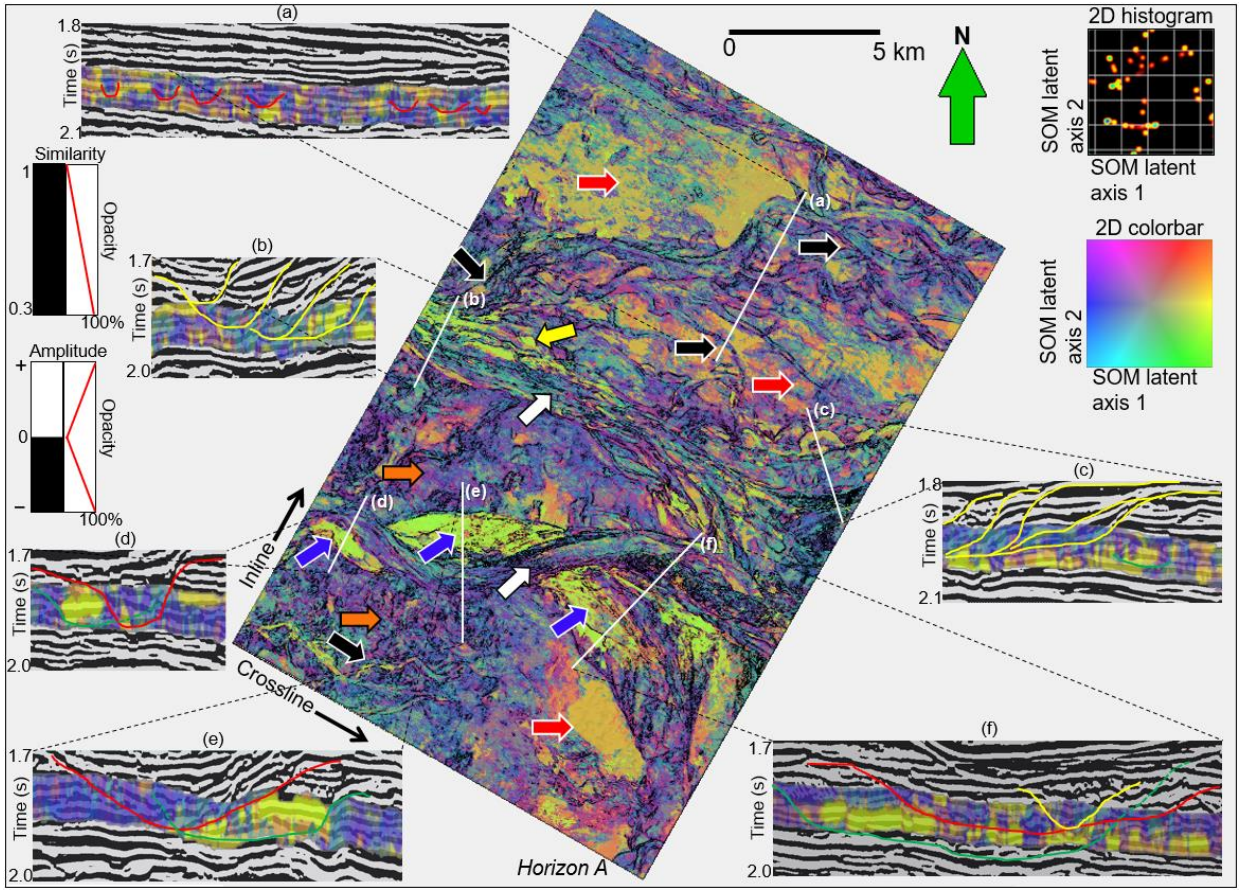


Figure 3.11. Slice along horizon A through the seismic facies map generated from DPSOM using a 2D colorbar. White arrows indicate two straight to meandering mud filled channels. Blue arrows indicate a possible sand filled channel that is cut through by one of the mud filled channels. Red arrows indicate possible slope fan deposits. The black arrows indicates sinuous channel complexes. Yellow arrow indicates possible sand filled lateral accretion packages. Sections (a) to (f) are vertical sections of co-rendered seismic amplitude and SOM facies demonstrating the morphology of the channel complexes in vertical profiles. The displayed SOM facies volume is centered at horizon A.

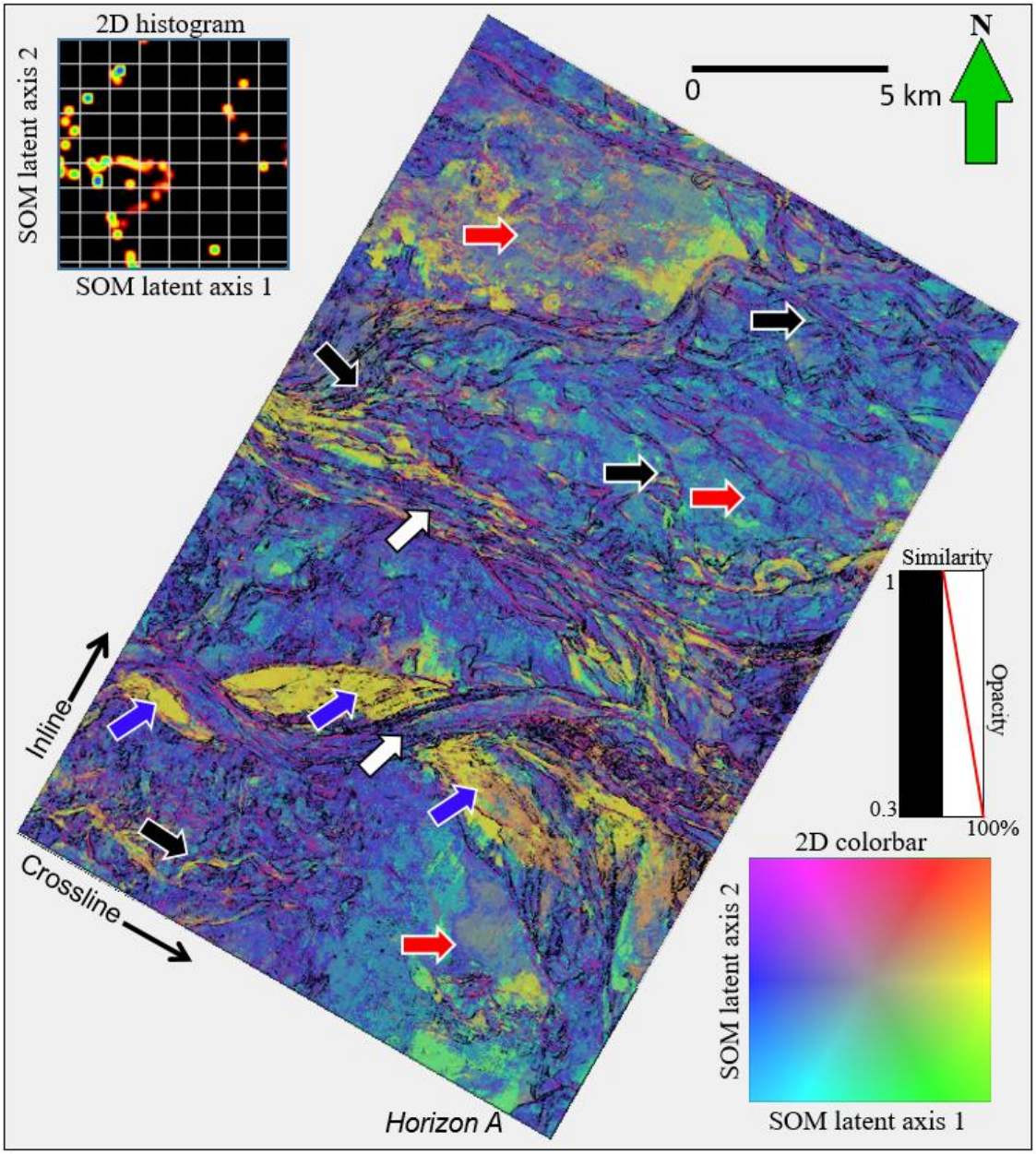


Figure 3.12. Slice along horizon A through the seismic facies map generated from classic Kohonen-SOM using a 2D colorbar. White arrows indicate two straight to meandering mud filled channels. Blue arrows indicate a possible sand filled channel that is cut through by one of the mud filled channels. Red arrows indicate possible slope fan and lobe deposits. The black arrows indicate sinuous channel complexes. Note the less separated facies represented by more similar colors comparing to Figure 3.11.

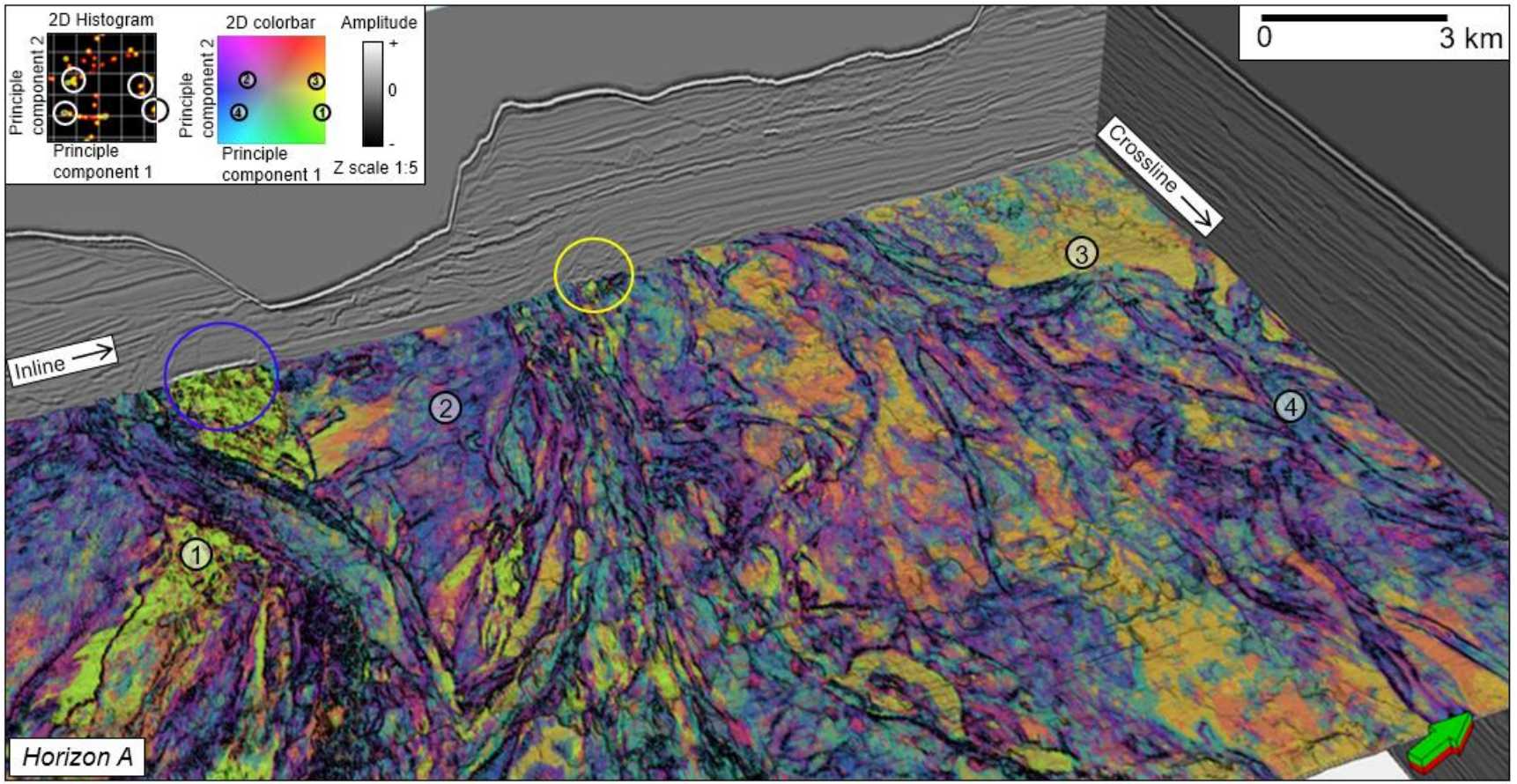


Figure 3.13. 3D chair display of the SOM facies map along horizon A with orthogonal vertical slices through seismic amplitude. The location of four prototype vectors are shown on the facies map as well as on the 2D histogram and colorbar. Also note the similarity in classification of what I interpret to be an older sand-filled channel (blue circle) and a discrete sand-filled channel that forms part of a multistoried channel complex (yellow circle).

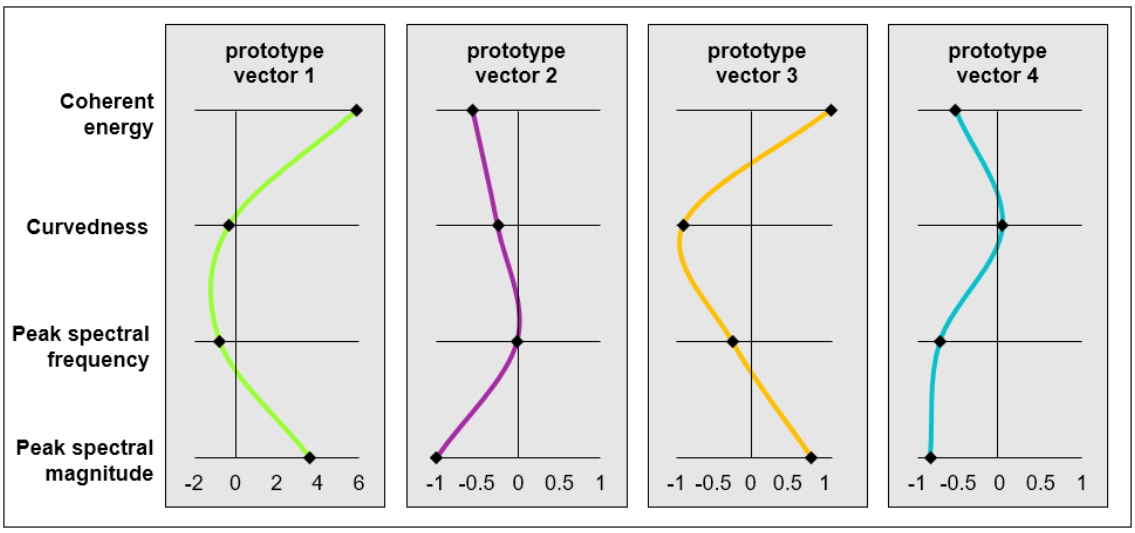


Figure 3.14. Seismic attribute responses of the four prototype vectors in Figure 3.13. Input attributes are coherent energy, curvedness, peak spectral frequency, and peak spectral magnitude. All attribute values are normalized using z-score.

REFERENCES

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221–2224.

Barnes, A. E., 2007, Redundant and useless seismic attributes: Geophysics, **72**, P33–P38.

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

Coléou, 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.

Himberg, J., 2000, A SOM based cluster visualization and its application for false coloring: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, **3**, 587–592.

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

Mahalanobis, P. C., 1936, On the generalized distance in statistics: Proceedings of the National Institute of Sciences of India, **2**, 49–55.

Marfurt, K. J., 2015, Techniques and beset practices in multiattribute display: Interpretation, **3**, B1–B23.

Merkl, D. and A. Rauber, 1997, Alternative ways for cluster visualization in self-organizing maps: Proceedings of the Workshop on Self-Organizing Maps (WSOM97), 4–6.

Mitchell, J. and H. L. Neil, 2012, OS20/20 Canterbury – Great South Basin TAN1209 voyage report: National Institute of Water and Atmospheric Research Ltd (NIWA).

Poupon, M., K. Azbel, and G. Palmer, 1999, A new methodology based on seismic facies analysis and litho-seismic modeling: The Elkhorn Slough field pilot project, Solano County, California: 69th Annual International Meeting, SEG, Expanded Abstracts, 927–930.

Roy, A., B.L. Dowdell, and K.J. Marfurt, 2013, Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma: Interpretation, **1**, SB109–SB124.

Rubio, M. and V. Gimnez, 2003, New methods for self-organising map visual analysis: Neural Computing & Applications, **12**, 142–152.

Shao, C. and Y. Yang, 2012, Distance-preserving SOM: a new data visualization algorithm: Journal of Software, **7**, 196–203.

Strecker, U., and R. Uden, 2002, Data mining of 3D post- stack attribute volumes using Kohonen self-organizing maps: The Leading Edge, **21**, 1032–1037.

Sutherland, R. and G. Browne, 2003, Canterbury basin offers potential on South Island, New Zealand: Oil & Gas Journal, **101**, 45–49.

Uruski, C. I., 2010, New Zealand's deepwater frontier: Marine and Petroleum Geology, **27**, 2005–2026.

Wang, D., H. Ressom, M. Musavi, and C. Domnisoru, 2002, Double self-organizing maps to cluster gene expression data: Proceedings of 2002 European Symposium on Artificial Neural Networks, 45–50.

Zhang, B., T. Zhao, X. Jin, and K. J. Marfurt, 2015, Brittleness evaluation of resource plays by integrating petrophysical and seismic data analysis: Interpretation, **3**, T81–T92.

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 4

CONSTRAINING SELF-ORGANIZING MAP FACIES ANALYSIS WITH STRATIGRAPHY[4]

INTRODUCTION

Skilled seismic interpreters identify seismic facies by examining spatial variations in seismic reflection amplitude, phase, frequency, continuity, and orientation. Modern seismic attributes and impedance inversion algorithms quantify the local variation of these reflectors voxel by voxel, allowing them to be statistically analyzed by a computer. Zhao et al. (2015) reviewed some of the more commonly used supervised and unsupervised seismic facies classification techniques, including principal component analysis, K-means clustering, self-organizing maps, generative topological maps, artificial neural networks, and support vector machines. In general, unsupervised classification techniques are solely driven by the input data, while supervised techniques incorporate external control provided either by wells or as data labels defined by the interpreter. In either case, validation of the results is critical to gaining confidence in the prediction. When there is a great deal of well control or interpreter-generated labels (geomorphological/petrophysical facies), statistical validation may be sufficient. However, when the well control is limited or the interpreter suspect, correlation to well-accepted geologic and petrophysical models provides added confidence. In a conventional reservoir, such a correlation may be that of a gas-sand facies exhibiting high

[4]*This study is published as - Zhao, T., F. Li, and K. J. Marfurt, 2017, Constraining self-organizing map facies analysis with stratigraphy: an approach to increase the credibility in automatic seismic facies classification: Interpretation,* ***5****, T163-T171.*

negative reflectivity, low Poisson's ratio, high local continuity, and occurring near the top of an anticlinal fold. For an unconventional shale reservoir such as the Barnett Shale, such a correlation may be a pattern of laterally continuous brittle and ductile layers that correspond to the cyclicity of the Fort Worth Basin evolution, interrupted by faults and collapse features associated with the deeper Ellenburger dolomite.

Self-organizing map (SOM) (Kohonen, 1982) is one of the most popular seismic facies analysis algorithm that projects N-dimensional multiattribute data vectors, one for each voxel, onto a deformed lower dimensional surface that attempts to best fit the data distribution. Poupon et al. (1999) described one of the earliest SOM-based seismic facies analysis. In their application, the "attributes" consisted of seismic amplitude from a suite of 16 phantom horizon slices. Data vectors in this 16-dimensional space were then projected onto a 1D shoestring "manifold" that best represented the data. The mean of each cluster along this shoestring therefore could be interpreted as a 16-sample seismic waveform. They then subsequently plotted the location of the cluster center along the 1D manifold against a 1D rainbow colorbar to delineate multiple turbidite deposits. Later, Strecker and Uden (2002) clustered volumetric attributes such as envelope, frequency, and coherence rather than amplitudes extracted about a surface and projected them onto a 2D deformed surface (manifold) rather than a 1D shoestring, plotting the cluster centers against a continuous 2D color table. Coleou et al. (2003) furthered this work by using more sophisticated attributes. More recently, Matos et al. (2009) used concepts of inter-cluster distances and a 2D HLS color table to improve the visualization of SOM facies. Roden et al. (2015) showed how to incorporate principal component analysis (PCA) to select the mathematically most meaningful input attributes for SOM. Because the

traditional Kohonen SOM only preserves topology but not distance, the distance information in the input attribute space is lost once projected into the 2D SOM latent space. Here the "distance" is not the spatial/temporal distance defined in (x, y, z) space in which the interpreter lives but rather the mathematical L2 norm between two N-dimensional data vectors or between an N-dimensional data vector and an N-dimensional cluster center. Based on this definition, two data vectors that have similar seismic attribute response should be adjacent to each other, while their projections onto the deformed 2D manifold should also be close. To preserve the consistency of distance from input attribute space to 2D SOM space, Zhao et al. (2016) adopted a distance-preserving step, constraining the SOM facies to better reflect the degree of diversity as found in the input attribute space. However, unless the interpreter provides attributes that somehow measure spatial proximity between two data vectors, the analysis remains spatially and temporally unaware.

The pitfall in being spatially and temporally unaware is the potential of mixing geology across different formations during the SOM process. In practice, an experienced interpreter can mitigate the possibility of mixing data samples from different formations by manipulating an operation window just localized at the target formation, yet there would always be vertical variations defined by sedimentary cycles in different scale (ranges from hundreds of millions of years to one tenth of million years), e.g. pinchouts and thin layers, within the operation window. Adding information of stratigraphy (sedimentary cycle), which provides temporal (or spatial, if seismic data are in depth domain) constraint on the vertical axis, may help define layers that are otherwise not well defined on seismic attributes. Relative geologic time (RGT) volume generated from

unwrapping instantaneous phase (Stark, 2003, 2004; Wu and Zhong, 2012) is an appropriate candidate for constraining SOM facies analysis with geologic time. However, in such RGT volumes, the vertical axis is monotonically increasing, on which sedimentary cycles controlled by changing sea level are not easily identified. In fact, there is very limited published research, if any, on calibrating RGT volumes with wells to identify sedimentary cycles which are routinely interpreted on well logs (e.g. gamma ray logs). An oscillation curve that directly links to periodic change in grain size is more preferred, and due to the limited resolution, presence of noise, and layer-interface relation in seismic data, such periodic change in grain size in most cases cannot be matched to seismic events. Empirical mode decomposition (EMD) (Huang et al., 1998) is an effective method to decompose seismic signal into several band limited modes, and researchers have discovered that some of the decomposed modes (which are named intrinsic mode functions or IMFs) corresponded with sedimentary cycles derived from well logs (Liu et al., 2010; Liu et al., 2015). However, EMD is a recursive model decomposition method, and is sensitive to noise and sampling and therefore not so robust. To overcome this issue, Dragomiretskiy and Zosso (2014) proposed a novel mode decomposition method, variational mode decomposition (VMD), which decomposes a signal concurrently and is robust to noise and sampling. Lateral consistency-preserved VMD has been successfully applied to seismic amplitude data to derive a sedimentary cycle model (Li et al., 2016). In this study, I adopt the workflow described in Li et al. (2016) to derive a sedimentary cycle model, and use this model as a constraint on SOM facies analysis. I test the proposed stratigraphy constrained SOM to a Barnett Shale survey in the United States, with the

objective of recovering more subtle lithologic variations than using the unconstrained SOM.

I begin this chapter by introducing the workflow of stratigraphy constrained SOM, followed by the geologic background of the study area, which is in the Fort Worth Basin, United States. I then apply the stratigraphy constrained SOM to analyze the lithofacies distribution in the Barnett Shale reservoir intervals. To illustrate the effectiveness of the proposed method, I also compare with the unconstrained SOM using the same input attributes. Finally, I summarize the values and limitations of the proposed method in the conclusion.

STRATIGRAPHY CONSTRAINED SOM

The proposed method starts from extracting VMD modes (IMFs) from seismic amplitudes to build a sedimentary cycle model. In this implementation, I choose to decompose the seismic signal into four modes, aiming to represent the sedimentary cycles with varies orders. Because the IMFs are decomposed from seismic amplitude signal which is the response of the interfaces, I further calculate the gradient of IMFs, assuming that the oscillation in the IMF gradients is a more suitable candidate to match with the sedimentary cycle caused by cyclic sea level change, comparing to the oscillation of IMFs. To verify this assumption and choose the appropriate gradient, I use gamma ray logs as the representation of sedimentary cycle pattern, and compare with the derived IMF gradients to choose one that matches the gamma ray pattern the best. Such IMF gradient with the highest correlation to gamma ray log (referred in the later text as best matching IMF gradient) is the sedimentary cycle model to be used to constrain the SOM.

In traditional Kohonen SOM, the distance used to find the best matching unit (BMU) for a given multiattribute data sample vector is calculated using only attribute values. As discussed in the Introduction, the lack of a temporal/spatial constraint may sometimes lead to unreasonable classification results. As an improvement, I add a term defined by the best matching IMF gradient into the distance metric, which now becomes:

$$d = (1-\lambda)\sum_{i=1}^{N}\|\mathbf{a}_i - \hat{\mathbf{a}}_i\| + \lambda\|g - \hat{g}\|, \tag{4.1}$$

where d is the weighted distance between a multiattribute data sample and a prototype vector, N is the number of attributes, $\mathbf{a}$ and $\hat{\mathbf{a}}$ are N-dimensional vectors at each voxel and of prototype vector, respectively, g and $\hat{g}$ are the best matching IMF gradients for a data sample and a prototype vector, respectively, and λ is a weight between zero and one.

Unfortunately, the weight λ is handpicked by interpreters. A larger weight means the generated SOM facies will follow the trend of IMF gradient more, and vice versa. A weight too large will generate a SOM facies map that follows the IMF gradient strictly and therefore suppress the response found in input attributes. A weight too small will not add the amount of constraint interpreters need. So an optimal weight is really up to an interpreter's choice and depends on project objective. In practice, I find values of λ between 0.6 and 0.7 to provide good results. When $\lambda = \frac{1}{N+1}$, the IMF gradient is weighted equally with other attributes. Such a modification, although in an extremely simple form, introduces a degree of stratigraphic constraint, which I will later illustrate through the application. The complete workflow of the modified SOM facies analysis is shown in Figure 4.1.

GEOLOGIC SETTING

The seismic data used to evaluate the stratigraphy constrained SOM were acquired to image the Barnett Shale unconventional reservoir of the Fort Worth Basin, United States. The Fort Worth Basin is a shallow north-south-elongated foreland basin that encompasses roughly 15,000 mi2 in north-central Texas and formed during the late Paleozoic Ouachita orogeny (Walper, 1982). Being one of the most classic unconventional shale reservoir, the Barnett Shale is an organic-rich petroliferous shale formation deposited in Mississippian age, when an alternating series of black, organic-rich shales and shallow marine limestones were deposited (Montgomery et al., 2005). The Barnett formation is deposited directly over the Viola Limestone formation, and topped by the Marble Falls Limestone formation. In the study area, the Barnett Shale is further divided into an upper and lower interval by a limestone interval, known as the Forestburg Limestone (Montgomery et al., 2005). All the three limestone intervals are water bearing and behave highly ductile, which make them hydraulic fracturing barriers. The general stratigraphy of the Barnett Shale is shown in Figure 4.2. The upper and lower Barnett Shale formations are not homogeneous, which can be subdivided into siliceous shale, argillaceous shale, calcareous shale, and limestone layers, with the presence of dolomite (Singh, 2008). Lithofacies stacking patterns (usually systematic) are also identified in the Barnett Shale, which are controlled by the eustatic sea level change over multiple scales of geological time (Slatt and Abousleiman, 2011). Slatt and Abousleiman (2011) further defined brittle-ductile couplets at different scales from such systematic lithofacies patterns identified on both gamma ray logs and core cuttings, ranging from over 100 m thick to only a few cms thick. The lithofacies patterns and brittle-ductile couplets are

directly linked to sedimentary cycles that are a result of eustatic sea level change. Therefore, if we are able to replicate the sedimentary cycles on seismic data and use such stratigraphic pattern to constrain the SOM analysis process, it well definitely benefit the facies classification, even though the sedimentary cycles are of high order due to the limited frequency band in seismic data.

APPLICATION

I apply the proposed stratigraphy constrained SOM facies analysis to the aforementioned Barnett Shale seismic survey, and compare the facies analysis result with one from unconstrained SOM, using the same input attributes and model parameters. In unconventional shale reservoirs, where interpreters are usually interested in lithology changes more than structural deformations, we prefer seismic inversion attributes as input attributes for automatic facies analysis algorithms. Zhang et al. (2014) and Verma et al. (2016) both used such inversion attributes to estimated brittleness and total organic carbon (TOC) in a supervised fashion. Spectral decomposition attributes are also routinely used as lithology indicators, and are especially helpful in mapping horizontal extension of different facies and geobodies. However, due to the window-based nature when generating such attributes, spectral decomposition attribute, such as peak spectral frequency and magnitude, have suboptimal vertical resolution, therefore are not suitable to represent vertical stacking patterns resulted from cyclic sea level change. In this study, attributes from prestack simultaneous inversion were used as inputs, which are P-impedance, ZP, S-impedance, ZS, the ratio of incompressibility and shear modulus, λ/μ, and Poisson's ratio, ν. I selecte these attributes with the understanding that such attributes directly correlate with mineral contents, grain size, and elastic properties of the rocks.

The prestack seismic data were carefully processed and preconditioned with the workflow described by Zhang et al. (2013) and Zhang et al. (2015).

With the input attributes at my disposal, the next step is to generate the four IMFs using VMD. Figure 4.3 gives an example trace along well A (location shown in later figures) together with its decomposed four IMFs, and the gradient of IMF 3, which, after further analysis, was able to match the pattern found in gamma ray logs. Figure 4.4 gives vertical sections along seismic amplitude and the gradient of IMF 3 plotted with the gamma ray log at well A, and Figure 4.5 shows how the composite trace of the gradient of IMF 3 matches the gamma ray log along well A and well B(locations of both wells are shown in later figures). In Figure 4.4, formation tops are marked as solid colored curves, gamma ray logs as solid blue curve, and the well trajectory as dashed red line. From top to bottom, the marked formation tops are: Marble Falls Limestone, Upper Barnett Limestone, Upper Barnett Shale, Forestburg Limestone, and Lower Barnett Shale. The formations tops are displayed in the same color scheme for all remaining figures. We identify a similar cyclic pattern in the IMF 3 gradient as in the gamma ray log, which is commonly used by sequence stratigraphers to interpret sedimentary cycles controlled by eustatic sea level change, subsidence rate of the basin, and sediment supply. Therefore, the gradient of IMF 3 provides a volumetric approximation of sedimentary cycles, providing information that is not easily acquired from tradition spectral decomposition.

I provide the input attributes with and without the constraint to a SOM classifier defined in the workflow discussed in Figure 4.1. Figure 4.6 shows time slices from unconstrained SOM (Figure 4.6a) and constrained SOM (Figure 4.6b), on which the location of well A and B, as well as vertical sections AA', BB', CC' (discussed later) are

displayed. The facies maps are generated by crossplotting the data projected onto two SOM axes, and such crossplot enables the use of a 2D color map, providing better visualization. By looking at these two time slices alone, it is nearly impossible to draw any conclusions comparing the quality of facies, as the contribution of the stratigraphy constraint is in the vertical direction. Moving to vertical section AA', Figure 4.7 gives the unconstrained SOM facies, and Figure 4.8 gives the constrained SOM facies, both of which are overlapped by the gamma ray log at well B. The formation tops are marked as colored curves. We identify the gamma ray value increases from the top to the bottom in the Marble Falls Limestone (black arrows in Figure 4.7 and 4.8), which translates into a color change from purplish to magenta in the unconstrained SOM facies (Figure 4.7), and a color change from orange to lime green in the constrained SOM facies (Figure 4.8). Adding the stratigraphy constraint makes such lithology variation more obvious in the form of colors with higher contrast, and matches the gamma ray trend better. The white arrows in Figure 4.7 and 4.8 show a local variation within the Upper Barnett formation, which is more obvious on the constrained SOM facies. Figure 4.9 displays vertical section AA' along VP/VS ratio. Although VP/VS is derived from ZP and ZS, it is not directly used as an input attribute for SOM. We clearly identify a high VP/VS ratio layer corresponding to the high gamma ray at the bottom of Marble Falls Limestone, and a very low VP/VS zone within the Upper Barnett at the white arrow.

Figure 4.10 and 4.11 are the vertical sections of unconstrained and constrained SOM facies maps along line BB', respectively. Here, besides the higher color contrast at the Marble Falls Limestone and Upper Barnett Limestone, the constrained SOM also shows more details in the Upper Barnett Shale formation, comparing to the unconstrained

SOM. As previously discussed, the Barnett Shale is deposited over multiple cycles of sea level change, and thin layers at different scales are developed. In the unconstrained SOM facies map (Figure 4.10), the middle section of Upper Barnett Shale is a thick layer of grayish colors, with a hint of yellow. However, in the same region of the constrained SOM facies map (Figure 4.11), the two black arrows point at two facies with different colors in a stacking pattern, which are nearly identical in Figure 4.10. Figure 4.12 gives a vertical section of VP/VS ratio along the same BB' line, which does not contain the sedimentary cycle information, can still provide an indication of thin layers of different VP/VS ratio, and the green facies in Figure 4.11 clearly correlates with the relatively higher VP/VS regions in the middle part of Upper Barnett Shale. Figure 4.13 provides a zoom-in around traces X1 and X1', which are extracted from the unconstrained and constrained SOM facies volumes, respectively, at the same location, and overlaid with curve display of these two traces. The values on the traces are "facies numbers", which are over-defined with 4096 SOM prototype vectors to ensure a smooth visualization. Such 4096 "facies" are then arranged over a 64 × 64 2D space, and color-coded using the 2D color map shown in the lower right corner. This translates to the fact that facies N and facies N±64 have similar colors, so the curve display alone may sometime be misleading and has to be verified with color display. In practice, I am still able to see different layers (black arrows) on trace X1', but nearly impossible to see on trace X1. Such details provide interpreters more insights of the small scale stratigraphy distribution in the Upper Barnett Shale formation.

Figure 4.14 -4.16 show the unconstrained SOM facies, constrained SOM facies, and VP/VS ratio along line CC', respectively. Similar to line AA' and BB', I am still able

to identify more facies with higher color contrast in the constrained SOM facies; while in the unconstrained SOM facies map, the facies are more smeared. The black arrows point to some high VP/VS regions in the Upper Barnett formation which are better delineated in the constrained SOM map as bright green spots (Figure 4.15). The constrained SOM map also shows the high VP/VS layer in the Lower Barnett (white arrows) clearer than in the unconstrained SOM map. Better delineation of such local elastic property change will greatly facilitate well-planning in the completion stage.

CONCLUSIONS

In this study, I explore the feasibility of constraining the SOM facies analysis using stratigraphy information, in the form of sedimentary cycles. The stratigraphy constrained SOM facies map provides more details, and shows layers that are more likely being overlooked on SOM facies maps without such constraint. The extra features can be calibrated with well log data and the VP/VS ratio attribute, which prove the credibility of the resulted facies. The sedimentary cycle is estimated by decomposing seismic amplitude signal into a finite number of modes using VMD, and I believe the selection of the most appropriate component to represent sedimentary cycle requires calibration with other data, and the most appropriate component may differ from region to region. However, the geological meaning of such modes is not well understood, and these modes need to be carefully calibrated with well logs. The different VMD gradient patterns in adjacent layers are a good indicator of layer geometry, however layers with the same VMD gradient response are not distinguishable. Fortunately, such limitation poses less problem in real application, because adjacent layers rarely have the same VMD gradient in seismic scale.

CHAPTER 4 FIGURES

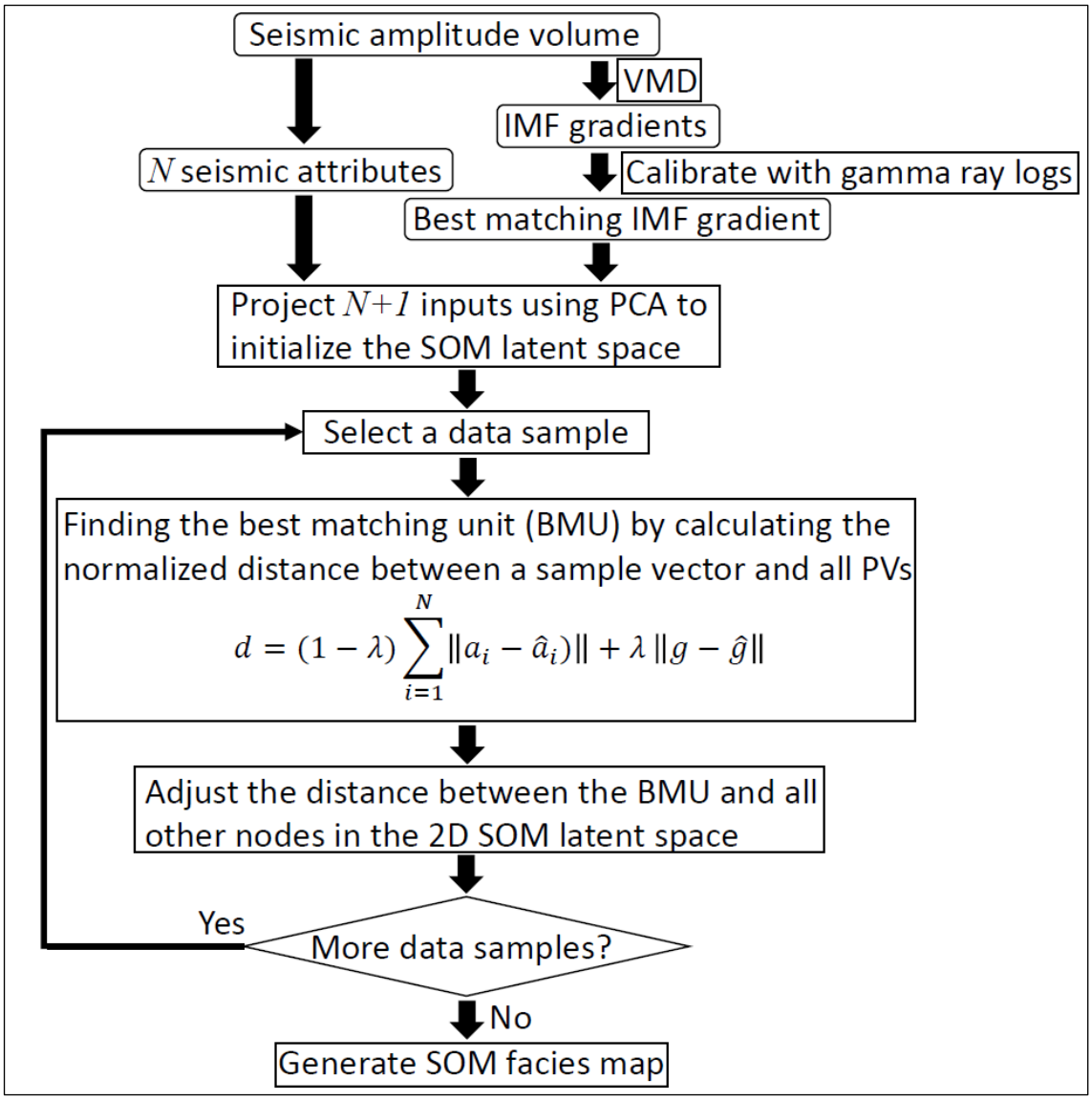


Figure 4.1. Workflow of the stratigraphy constrained SOM facies analysis. All abbreviations are explained in the main text.

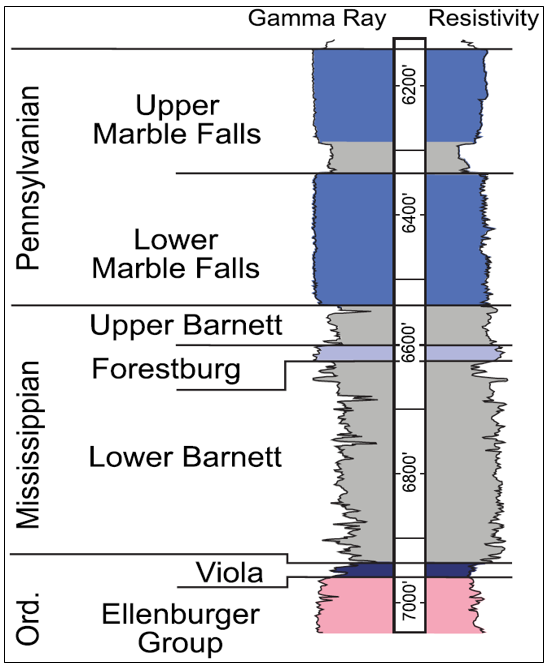


Figure 4.2. General stratigraphy of the Ordovician to Pennsylvanian section in the Fort Worth Basin through a well near the study area (After Loucks and Ruppel, 2007).

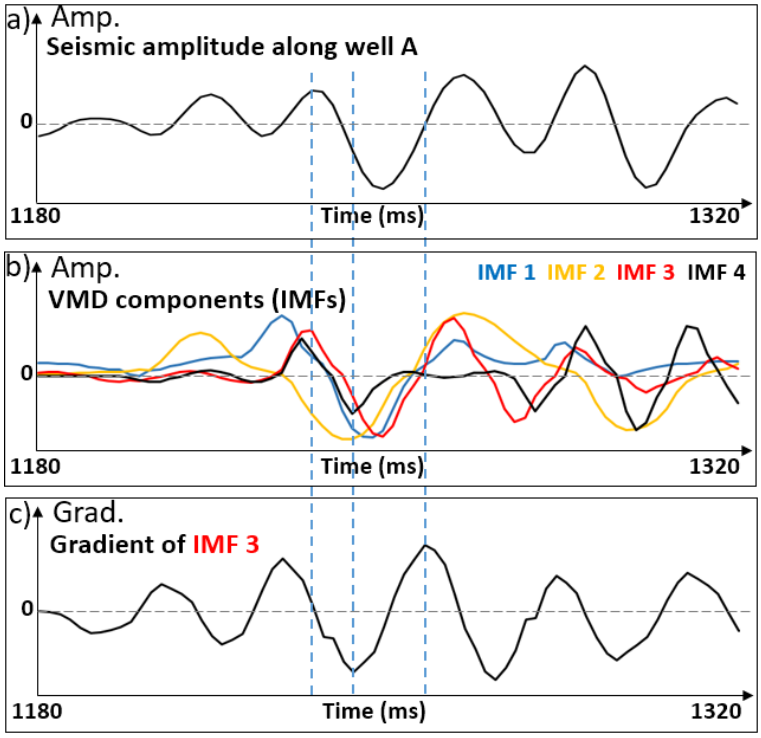


Figure 4.3. (a) Seismic amplitude from a trace along well A (location shown in Figure 4.6). (b) VMD components (IMFs) of the trace above. Four components are used to represent sedimentary cycle at different scales. (c) The gradient of IMF 3. Dashed lines show the correspondence among seismic amplitude, IMF 3, and IMF 3 gradient, when IMF 3 gradient is at zero, local minimum, and local maximum.

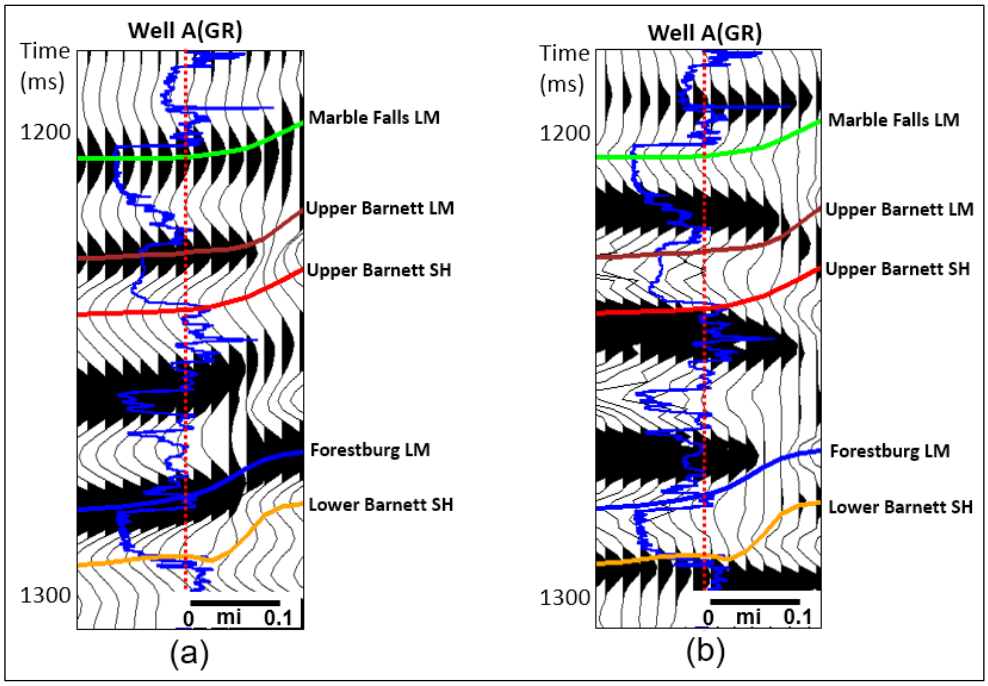


Figure 4.4. Vertical sections along (a) seismic amplitude and (b) IMF 3 gradient plotted with the gamma ray log (blue curve) at well A (location shown in Figure 4.6). The well trajectory is marked by the red dashed line. Formation tops are marked as colored curves. Note the good match in pattern between the gamma ray log and IMF 3 gradient.

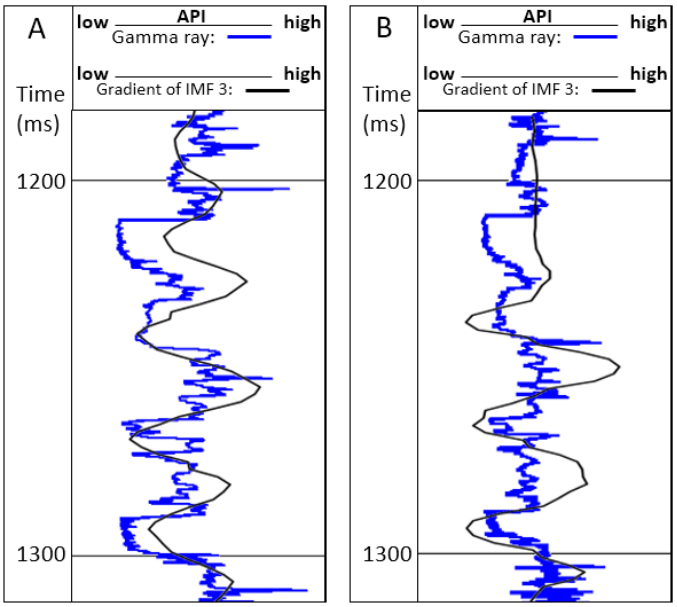


Figure 4.5. Traces of IMF 3 gradient plotted with gamma ray logs (blue curve) at well A (left) and well B (right) (well locations are shown in Figure 4.6). Well A is a vertical well, and the corresponding IMF 3 gradient trace is the most adjacent trace of well A. Well B is a deviated well, and the corresponding IMF 3 gradient trace is a composite trace along the well trajectory. Note the good match in pattern between the gamma ray log and IMF 3 gradient at both wells.

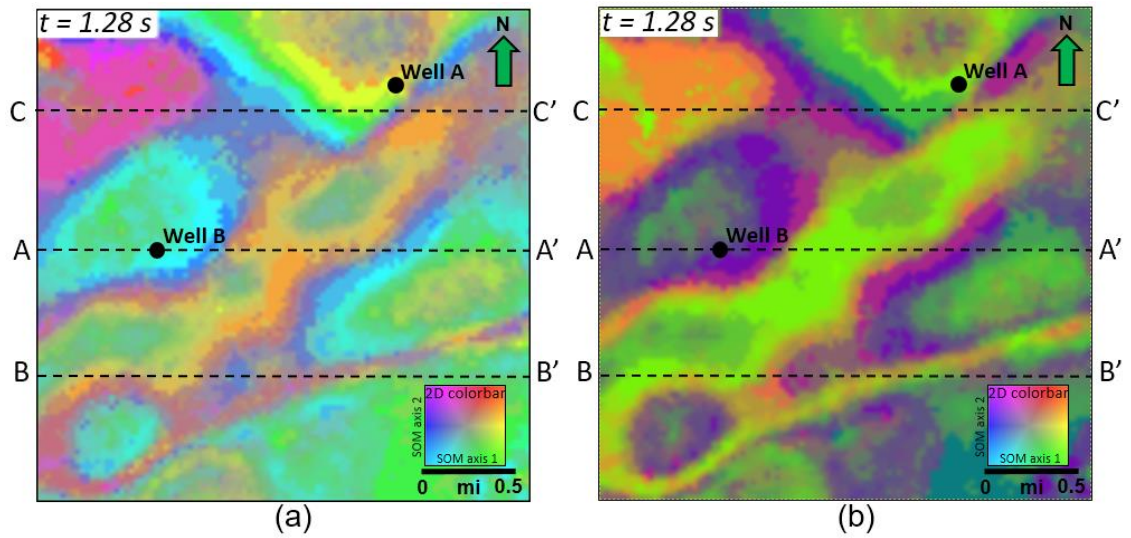


Figure 4.6. Time slices at t=1.28 s along SOM facies maps generated (a) without stratigraphy constraint and (b) with stratigraphy constraint. A 2D colorbar is used for visualization. Note the difference between two time slices is very limited, because the stratigraphy constraint is added on the vertical axis.

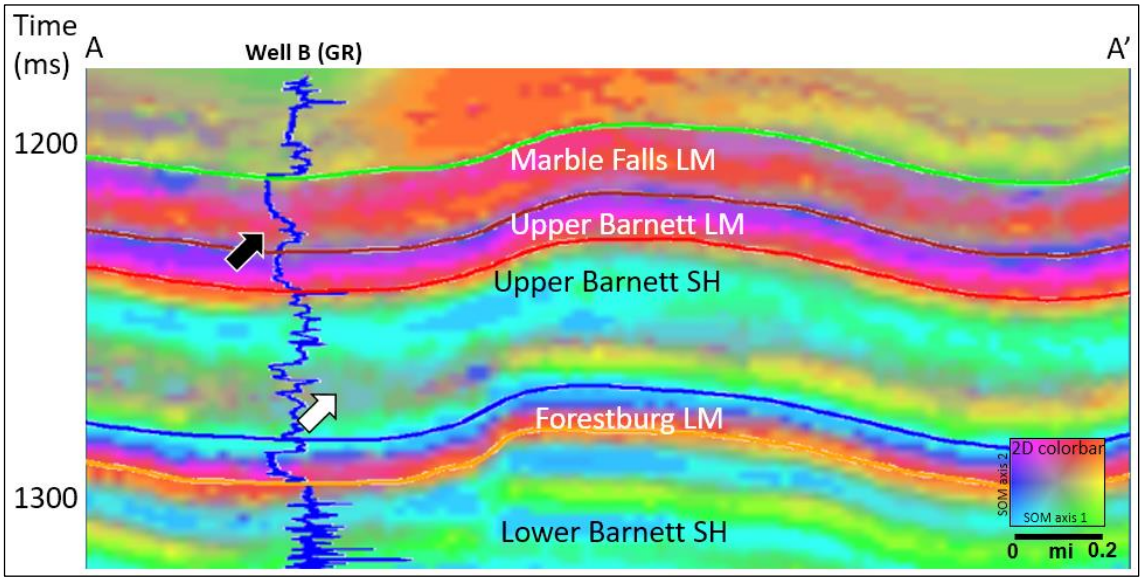


Figure 4.7. Vertical section along line AA' (location shown in Figure 4.6) through unconstrained SOM facies map. Formation tops are marked with colored curves. The black arrow indicates a high gamma ray layer at the bottom of the Marble Falls Limestone formation. The white arrow indicates a local facies change.

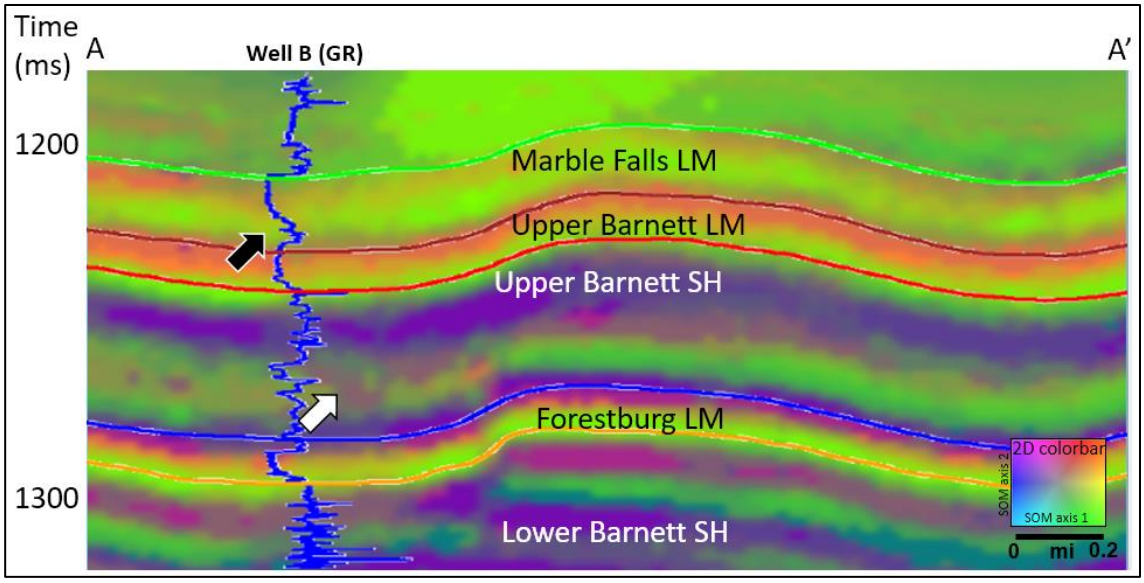


Figure 4.8. Vertical section along line AA' (location shown in Figure 4.6) through constrained SOM facies map. Formation tops are marked with colored curves. The black arrow indicates a high gamma ray layer at the bottom of the Marble Falls Limestone formation. This high gamma ray layer corresponds better to the constrained SOM facies than the unconstrained SOM facies, as the facies show higher contrast in color (orange to lime versus purple to magenta in Figure 4.7). The white arrow indicates a local facies change in the Upper Barnett which corresponds to a low V_P/V_S region.

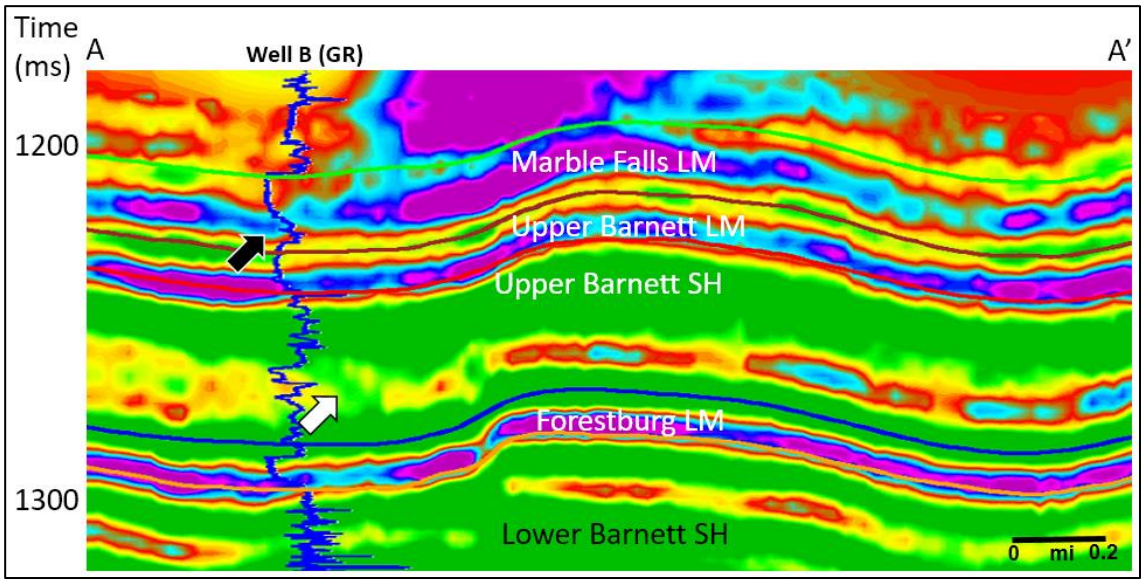


Figure 4.9. Vertical section along line AA' (location shown in Figure 4.6) through V_P/V_S ratio. Formation tops are marked with colored curves. The black arrow indicates a high gamma ray layer at the bottom of the Marble Falls Limestone formation. This high gamma ray layer has a very high V_P/V_S ratio. The white arrow indicates a low V_P/V_S region.

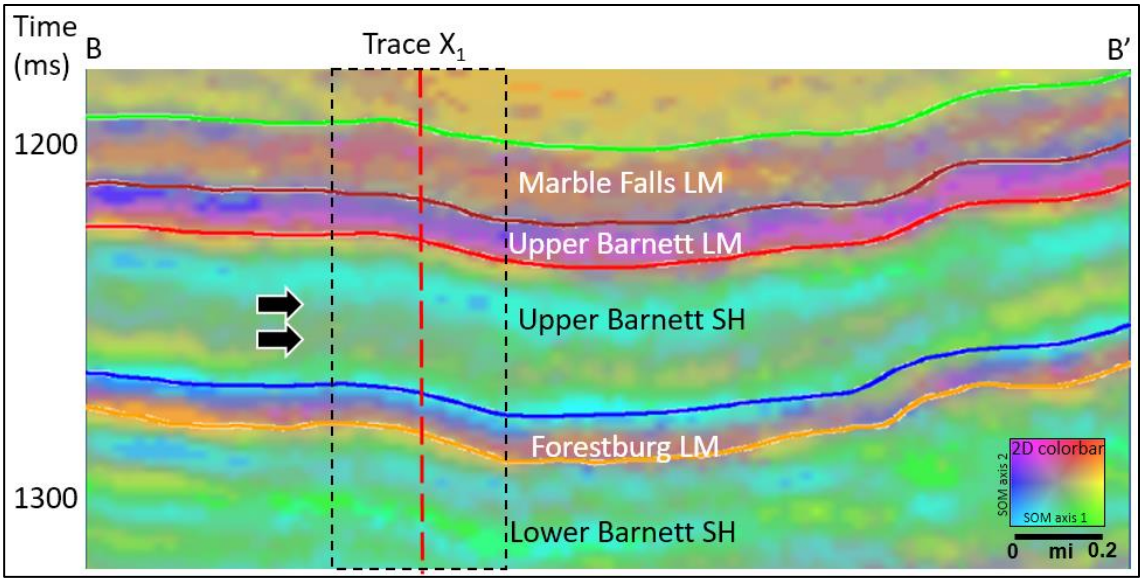


Figure 4.10. Vertical section along line BB' (location shown in Figure 4.6) through unconstrained SOM facies map. Formation tops are marked with colored curves. The black arrows indicate two thin layers in the Upper Barnett Shale formation that are not well defined in the unconstrained SOM facies map. The area in the dashed box is discussed in Figure 4.13. The red dashed line is the location of trace X_1 which is discussed in Figure 4.13.

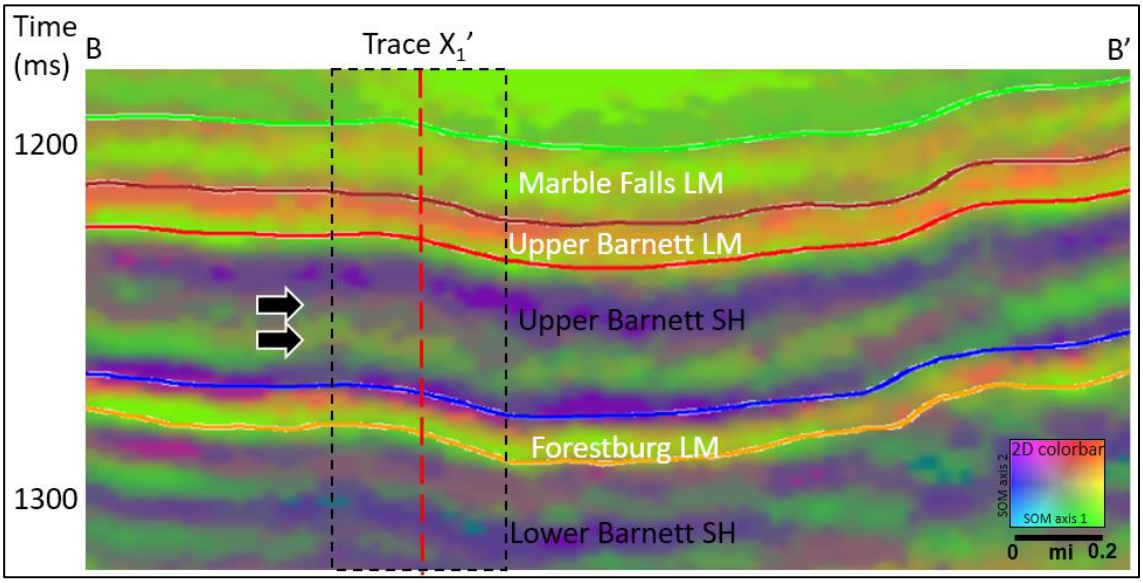


Figure 4.11. Vertical section along line BB' (location shown in Figure 4.6) through constrained SOM facies map. Formation tops are marked with colored curves. The black arrows indicate two thin layers in the Upper Barnett Shale formation that can be identified in the constrained SOM facies map but are not well defined in the unconstrained SOM facies map. The area in the dashed box is discussed in Figure 4.13. The red dashed line is the location of trace X_1' which is discussed in Figure 4.13.

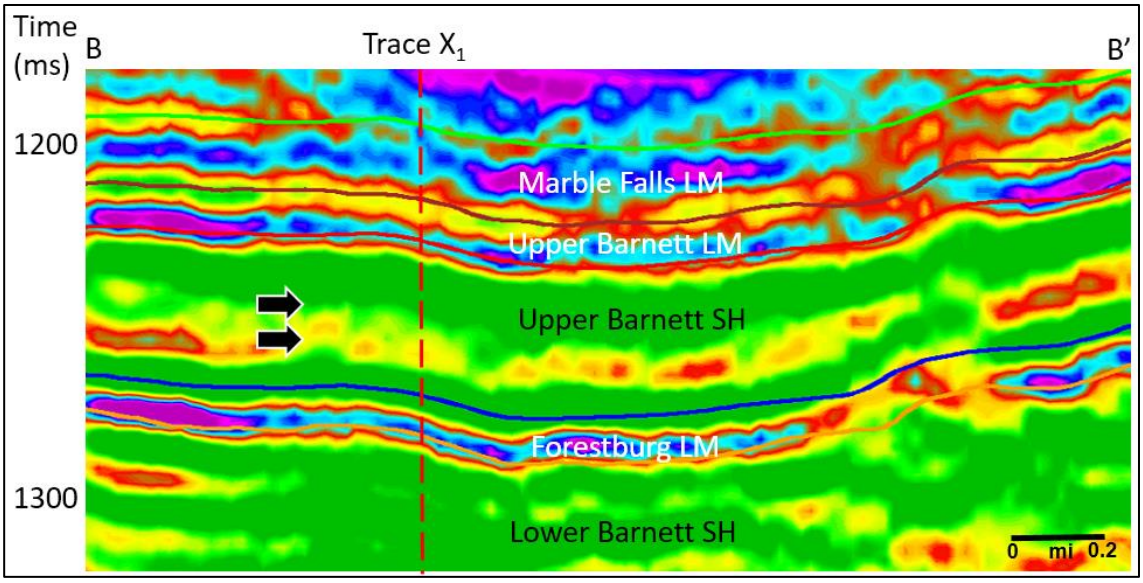


Figure 4.12. Vertical section along line BB' (location shown in Figure 4.6) through V_P/V_S ratio. Formation tops are marked with colored curves. The black arrows indicate two thin layers in the Upper Barnett Shale formation identified in the constrained SOM facies map. Although the stratigraphy constrain is not from V_P/V_S ratio, we do observe a difference in the V_P/V_S ratio between these two layers. The red dashed line is the location of trace X_1/X_1' which are discussed in Figure 4.13.

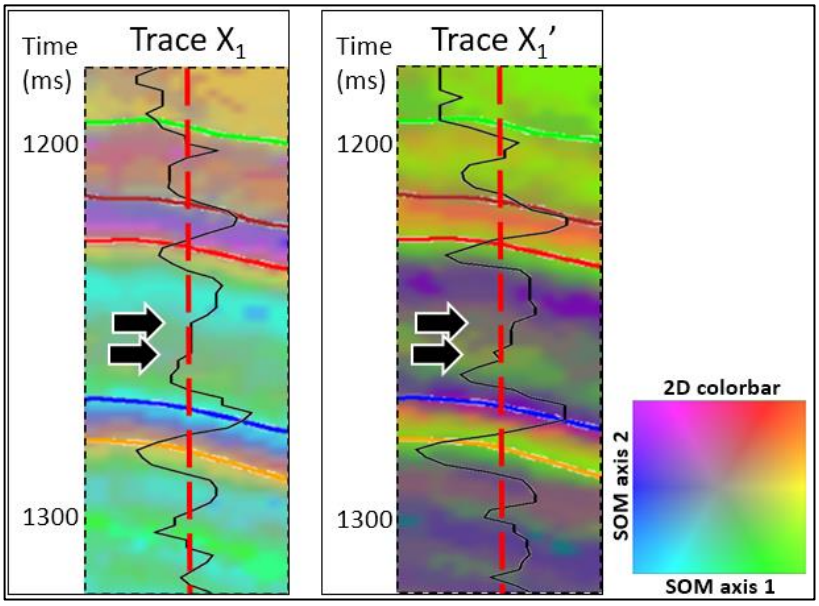


Figure 4.13. Zoom-ins around traces X_1 and X_1' extracted from the unconstrained and constrained SOM facies volume, respectively, at the same location, and overlaid with curve display of these two traces. The values on the traces are "facies numbers", which are over-defined with 4096 SOM prototype vectors to ensure a smooth visualization. Such 4096 "facies" are then arranged over a 64 × 64 2D space, and color-coded using the 2D color map shown in the lower right corner. We identify different layers (black arrows) on trace X_1', but nearly impossible to see on trace X_1.

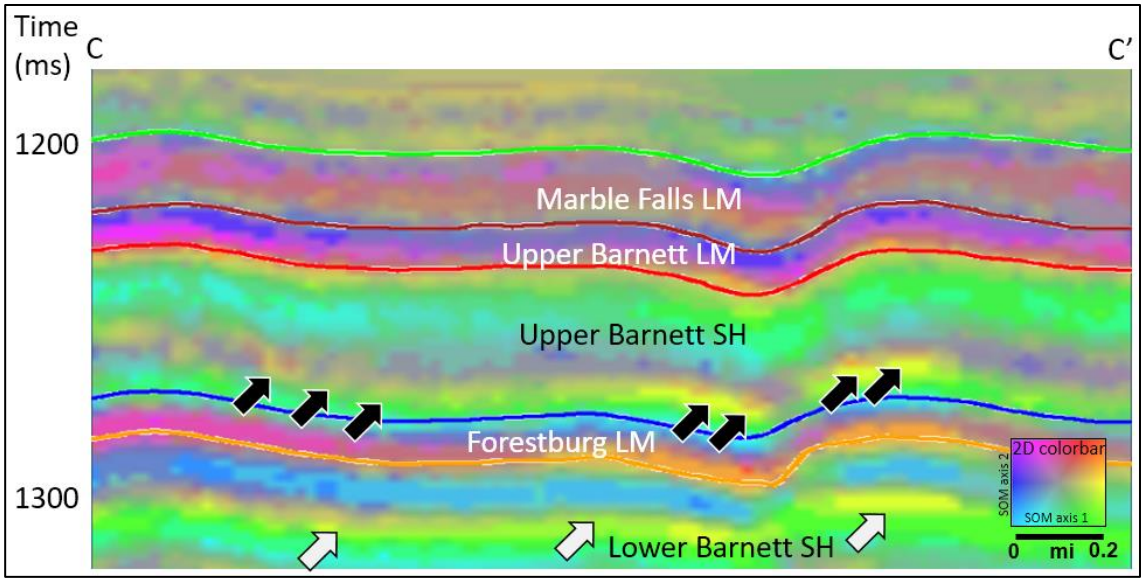


Figure 4.14. Vertical section along line CC' (location shown in Figure 4.6) through unconstrained SOM facies map. Formation tops are marked with colored curves. The black arrows indicate a thin layer with high V_P/V_S ratio in the Upper Barnett Shale formation. The white arrows indicate a thin layer with high V_P/V_S ratio in the Lower Barnett Shale formation.

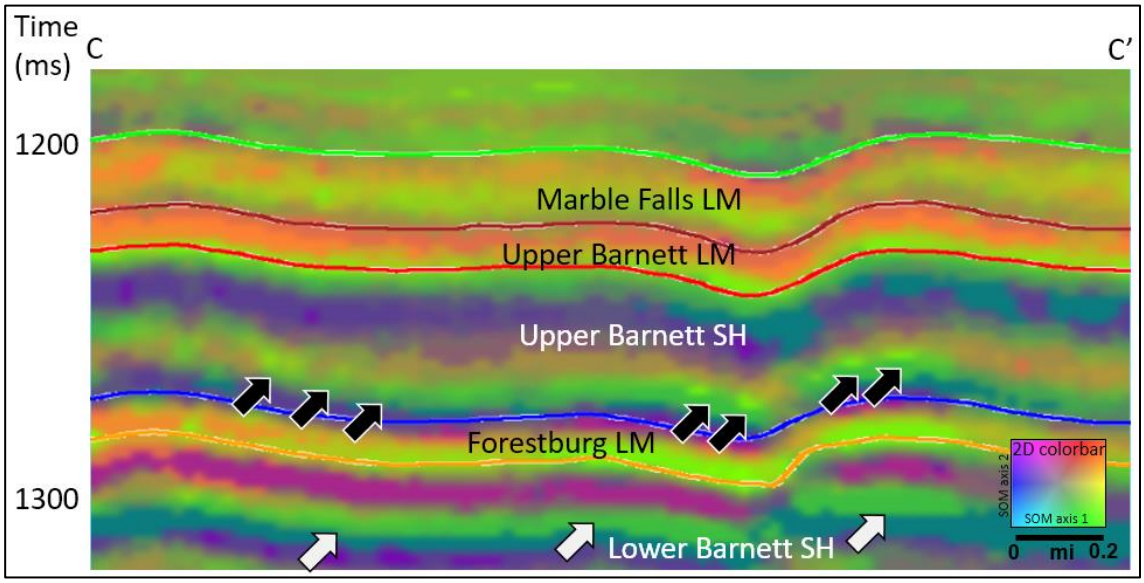


Figure 4.15. Vertical section along line CC' (location shown in Figure 4.6) through constrained SOM facies map. Formation tops are marked with colored curves. The black arrows indicate a thin layer with high V_P/V_S ratio in the Upper Barnett Shale formation. The white arrows indicate a thin layer with high V_P/V_S ratio in the Lower Barnett Shale formation.

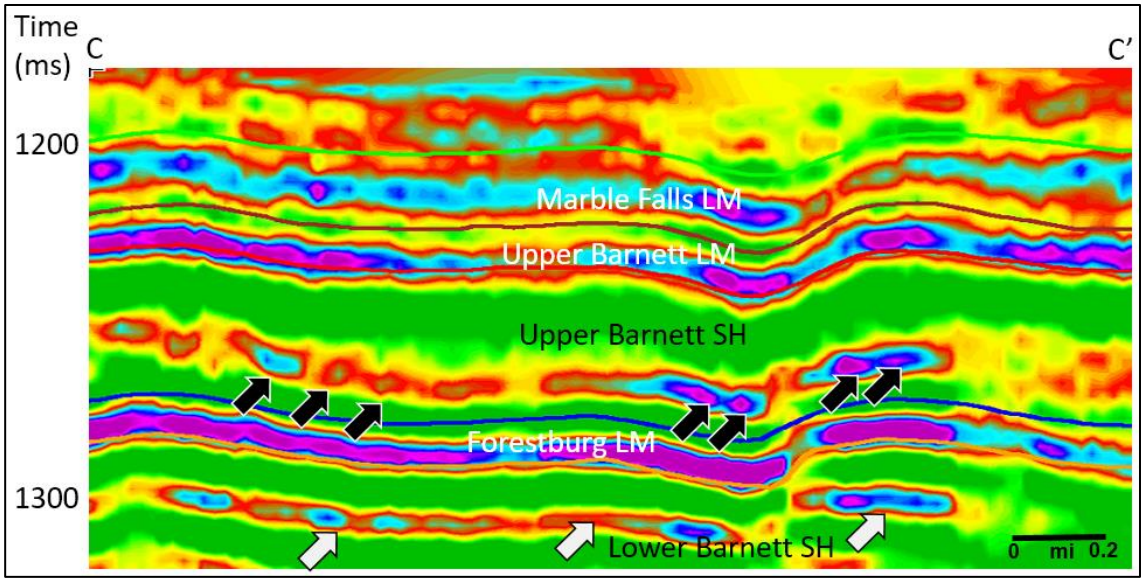


Figure 4.16. Vertical section along line CC' (location shown in Figure 4.6) through V_P/V_S ratio. Formation tops are marked with colored curves. The black arrows indicate a thin layer with high V_P/V_S ratio in the Upper Barnett Shale formation. The white arrows indicate a thin layer with high V_P/V_S ratio in the Lower Barnett Shale formation.

REFERENCES

Coléou, 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.

Dragomiretskiy, K. D. Zosso, 2014, Variational mode decomposition: IEEE Transactions on Signal Processing, **62**, 531-544.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, The Royal Society， **454**, 903-995.

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

Li, F., T. Zhao, X. Qi, and K. Marfurt, 2016, Lateral consistency preserved variational mode decomposition (VMD): 86th Annual International Meeting, SEG, Expanded Abstracts, 1717-1721.

Liu, Y., G. Yang, and W. Cao, 2015, The Division of Sedimentary Cycle based on HHT: 85th Annual International Meeting, SEG, Expanded Abstracts, 1902-1906.

Liu, Q., W. Yang, and L. Tian, 2010, Research and application of seismic facies analysis based on the empirical mode decomposition: 80th Annual International Meeting, SEG, Expanded Abstracts, 2329-2333.

Loucks, G. R, and C. S. Ruppel, 2007, Mississippian Barnett Shale: Lithofacies and depositional setting of a deep-water shale-gas succession in the Fort Worth Basin, Texas: AAPG Bulletin, **91**, 579-601.

Matos, M. C., K. J. Marfurt, and P. R. S. Johann, 2009, Seismic color self-organizing maps: Presented at 11th International Congress of the Brazilian Geophysical Society, Extended Abstracts.

Montgomery, S. L., D. M. Jarvie, K. A. Bowker, and R. M. Pollastro, 2005, Mississippian Barnett Shale, Fort Worth basin, north-central Texas: Gas-shale play with multi–trillion cubic foot potential: AAPG bulletin, **89**, 155-175.

Poupon, M., K. Azbel, and G. Palmer, 1999, A new methodology based on seismic facies analysis and litho-seismic modeling: The Elkhorn Slough field pilot project, Solano County, California: 69th Annual International Meeting, SEG, Expanded Abstracts, 927–930.

Roden, R., T. Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **4**, SAE59-SAE83.

Singh, P., 2008, Lithofacies and sequence stratigraphic framework of the Barnett Shale, Northeast Texas: Ph.D. dissertation, University of Oklahoma.

Slatt, R. M., and Y. Abousleiman, 2011, Merging sequence stratigraphy and geomechanics for unconventional gas shales: The Leading Edge, **30**, 274–282.

Stark, T. J., 2003, Unwrapping instantaneous phase to generate a relative geologic time volume: 73rd Annual International Meeting, SEG, Expanded Abstracts, 1707-1710.

Stark, T. J., 2004, Relative geologic time (age) volumes—Relating every seismic sample to a geologically reasonable horizon: The Leading Edge, **23**, 928-932.

Strecker, U., and R. Uden, 2002, Data mining of 3D post- stack attribute volumes using Kohonen self-organizing maps: The Leading Edge, **21**, 1032–1037.

Verma, S., T. Zhao, K. J. Marfurt, and D. Devegowda, 2016, Estimation of total organic carbon and brittleness volume: Interpretation, **4**, T373-T385.

Walper, J. L., 1982, Plate tectonic evolution of the Fort Worth basin, in Martin, C. A., ed., Petroleum geology of the Fort Worth basin and Bend arch area: Dallas Geological Society, 237–251.

Wu, X. and G. Zhong, 2012, Generating a relative geologic time volume by 3D graph-cut phase unwrapping method with horizon and unconformity constraints: Geophysics, **77**, O21-O34.

Zhang, B., D. Chang, T. Lin, and K. J. Marfurt, 2015, Improving the quality of prestack inversion by prestack data conditioning: Interpretation, **3**, T5-T12.

Zhang, B., K. Zhang, S. Guo, and K. J. Marfurt, 2013, Nonstretching NMO correction of prestack time-migrated gathers using a matching-pursuit algorithm: Geophysics, **78**, U9-U18.

Zhang, B., T. Zhao, X. Jin, and K. J. Marfurt, 2015, Brittleness evaluation of resource plays by integrating petrophysical and seismic data analysis, Interpretation, **3**, T81-T92.

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

Zhao, T., J. Zhang, F. Li, and K. J. Marfurt, 2016, Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps: Interpretation, **4**, SB79-SB89.

CHAPTER 5

USER GUIDED DATA-ADAPTIVE ATTRIBUTE SELECTION FOR UNSUPERVISED SEISMIC FACIES ANALYSIS[5]

INTRODUCTION

Seismic attributes are routinely used in nearly every stage of an oilfield life cycle. However, Barnes (2007) finds that there are many redundant attributes, and many attributes with vague geological meaning rendering them useless. Excluding such redundant and useless attributes from the seismic interpretation workflow is important, but interpreters still face the challenge of selecting which of remaining attributes are appropriate for a given interpretation task. Interpreters have spent a considerable amount of effort on how to select the most suitable attributes for both qualitative and quantitative facies and reservoir property estimation. Chen and Sidney (1997) provide a comprehensive review on attribute selection for reservoir prediction and monitoring, dividing attributes into wave kinematics and reservoir feature categories, and further summarize the applicability for each attribute. Kalkomey (1997) discusses the risk of false correlation between seismic attributes and reservoir properties, in which she suggests that special caution is needed when there are too few wells to correlate with too many attributes. She further suggests to use only those attributes that have a physically justifiable relationship with reservoir property as predictors. Hart and Balch (2000) present a case study on predicting reservoir properties from seismic attributes with limited well control, in which they propose a suite of visual correlation schemes to define the

[5]*A paper on this study is submitted to journal Geophysics as - Zhao, T., F. Li, and K. J. Marfurt, 2017, User guided data-adaptive attribute selection for unsupervised seismic facies analysis.*

attributes of choice. From a more quantitative aspect, Schuelke and Quirein (1998) propose to use cross-validation as a measure of prediction performance, then select attributes that lead to higher cross-validation. Since then, almost all the proposed alternative strategies have shared one fundamental concept, which is to select attributes that lead to the lowest validation error. While sequential correlation (e.g. Hampson et al., 2001) is the most commonly used correlation method to determine the input attributes, Dorrington and Link (2004) use a genetic nonlinear inversion algorithm to automatically determine which combination of input attributes minimizes the error in neural network porosity prediction.

Such prediction error based attribute selection methods require a relatively large number of training samples, either from well control or from interpreter handpicked polygons defining seismic facies, providing supervision to the process. However, interpreters often wish to discover the natural facies distribution in the data in an unsupervised manner, without introducing any bias.

Barnes and Laughlin (2002) find that the selection of input attributes has a higher impact on the facies map than the unsupervised learning algorithm used for classification. Zhao et al. (2015) review several competing unsupervised learning algorithms and through iterative experiments find a subset of attributes to highlight different architectural elements in a turbidite system. Roden et al. (2015) use principal component analysis (PCA) to quantitatively measure the contribution of each seismic attribute to the top principal components which in turn are used in the subsequent facies analysis. Although PCA estimates the contribution of each attribute to represent the data variability as a

whole, it does not provide a means of determining which attributes best differentiate a given facies from other facies and the background behavior.

One of the most commonly used seismic facies analysis algorithms is the Kohonen (1982) self-organizing map (SOM). SOM is an unsupervised learning technique that projects higher dimensional multiattribute data onto a lower dimensional (usually 2D) space called a manifold in which clusters of seismic facies are more easily defined and interpreted. In perhaps the first application of SOM to seismic data, Poupon et al. (1999) use amplitude from a suite of 30 phantom horizons as input projected to a 1D manifold, thereby generating a facies map along a target horizon. Strecker and Uden (2002) and Coleou et al. (2003) extend these concepts to volumetric SOM analysis, in which interpreters take multiple attributes at each voxel as inputs. Since then, researchers have been investigating how to effectively recover and represent the information buried in multiple seismic attributes. Matos et al. (2009) display an SOM facies map using a 2D color map. Hu et al. (2014) and Zhao et al. (2016) introduce improvements to preserve the Euclidean distance from seismic attribute space to 2D SOM latent space. Zhao et al. (2017) constrain SOM analysis using a sedimentary cycle model to introduce spatial/temporal awareness.

At present, all multiattribute SOM analyses share one implicit assumption: input attributes are all independent and have the same contribution/importance to the SOM facies. This assumption has greatly simplified the SOM analysis, as interpreters just need to select which attributes to feed into a SOM, without specifying which of the selected attributes are more important than the others. Treating all input attributes equally is the somewhat “normal” practice in pattern recognition; unfortunately, seismic attributes are

not created equally, and the importance of an attribute varies greatly with the exploration perspective. Geometric attributes (e.g. coherence) attributes are routinely used to map discontinuities (e.g. faults) and stratigraphic edges in seismic data. In the Barnett Shale area, interpreters find coherence and curvature extremely useful when mapping the karst collapse features in the Ellenburger dolomitic formation, which pose drilling hazards to the overlaying Barnett Shale formation and therefore are crucial during completion design (Sullivan et al., 2006). In contrast, Qi et al. (2014) find that spectral decomposition attributes provide estimates of the lateral change in layer thickness and lithology, which can also map karst features in the same region. If an interpreter use both structural and spectral attributes in SOM, with the goal of mapping the facies distribution in the highly karsted Ellenburger formation, he/she would expect these attributes to have varying degrees of impact.

In the structurally less deformed Barnett Shale reservoirs, interpreters are more interested in mapping different lithofacies, which may contain high total organic carbon (TOC), or may be more susceptible to hydraulic fracturing. In this case, seismic inversion attributes (e.g. $\lambda\rho$ and $\mu\rho$) that are closely related to geomechanical behavior, are more favored by interpreters (Altamar and Marfurt, 2015). If the interpreter uses a combination of structural attributes and inversion attributes as inputs for SOM, weaker differences measured by inversion attributes may be suppressed by the stronger differences in structural attributes.

The attribute selection system we use today is in fact simply a weighting system: if we use an attribute, its weight is one; if we reject it, its weight is zero. I therefore ask the question: instead of either selecting or rejecting an attribute, can we define weights

that represent the value of each input attribute in differentiating facies of interest? SOM clusters in data using either a Euclidean distance,

$$r_E = \sqrt{(\mathbf{a}_1 - \mathbf{a}_2)^T \mathbf{I} (\mathbf{a}_1 - \mathbf{a}_2)}, \tag{5.1}$$

or a Mahalanobis (1936) distance,

$$r_M = \sqrt{(\mathbf{a}_1 - \mathbf{a}_2)^T \mathbf{C}^{-1} (\mathbf{a}_1 - \mathbf{a}_2)}, \tag{5.2}$$

where $\mathbf{a}_1$ is an $N \times 1$ multiattribute data vector of N attributes, **I** is a diagonal matrix of ones, $\mathbf{a}_2$ is another multiattribute data vector, **C** is the covariance matrix of multiattribute variables, and r is the Euclidean or Mahalanobis distance between $\mathbf{a}_1$ and $\mathbf{a}_2$. If we assume the input attributes to be independent, the Mahalanobis distance then becomes z-score, where

$$C_{ii} = \sigma_i^2, \tag{5.3}$$

and

$$C_{ij} = 0, \forall\, i \neq j. \tag{5.4}$$

σ_i is the standard deviation of the ith attribute. To emphasize and deemphasize the importance of a given attribute, I define a weight matrix **W** when calculating distance in SOM:

$$r = \sqrt{(\mathbf{a}_1 - \mathbf{a}_2)^T \mathbf{W} \mathbf{C}^{-1} (\mathbf{a}_1 - \mathbf{a}_2)}. \tag{5.5}$$

Now we have a diagonal matrix **W** along with the diagonal normalization matrix **C** used for z-score. A larger value of W_{ii} means the *i*th attribute is more important and is weighted more in SOM, and a smaller value of W_{ii} means the *i*th attribute is less important and therefore has less impact on the SOM facies map. If the diagonal contains only zeros or ones, equation 5 reduces to the traditional (binary attribute weighting) SOM where one

selects (*Wii = 1*) or rejects (*Wii = 0*) an attribute. In Figure 5.1, I show a schematic plot of two attributes to demonstrate how the distance matrix **W** changes clusters in the data.

I organize the remaining of this chapter as follows. I introduce the definition and derivation of the weight matrix in the next section, followed by data description of a field example. I then demonstrate the proposed workflow and compare with the traditional, input attributes equally weighted SOM using a dataset from the Barnett Shale, United States. I further provide a discussion using a second field example, complementing the first experiment with another input attribute scenario. Finally, I sum up the study with conclusions.

WEIGHTING AS A MEANS FOR ATTRIBUTE SELECTION

My goal is to define the weight matrix **W** as a function of interpreter's knowledge and attributes' contribution to SOM. I encourage readers to retrieve details of the SOM implementation on which this study is built from Roy et al. (2013), Zhao et al. (2015), and Zhao et al. (2016).

Inspired by Benabdeslem and Lebbah (2007), given *N* input attributes and *J* prototype vectors (which are the proxies of the 2D SOM neurons in the attribute space), I define ω_i, the ith attribute's contribution to a SOM model, as:

$$\omega_i = \sum_{\mathrm{j}=1}^{J} d_j \frac{|p_{ji}|}{\sum_{k=1}^{N}|p_{jk}|}, \tag{5.6}$$

and

$$d_j = \frac{h_j}{M}, \tag{5.7}$$

where h_j is the number of multiattribute training samples that are nearest to the *j*th prototype vector, *M* is the total number of multiattribute training samples, d_j represents

the density of training samples assigned to the *j*th prototype vectors, and p_{jk} is the value of the *j*th prototype vector along dimension *k* (the dimension of the *k*th attribute). Physically, if a prototype vector has a very large value in the dimension of the target attribute, and a large percentage of training samples are close to this prototype vector, then the target attribute's contribution at this prototype vector is significant. Summing up over all the prototype vectors, we then arrive at the target attribute's contribution to the whole SOM model.

Unfortunately, after some tests, I conclude that this contribution **ω** depends on the combination of attributes, which means the rank of contribution for a given group of attributes changes by adding a new attribute in the group. Such phenomenon is a drawback of using PCA to initialize the SOM neurons. In this SOM implementation, the SOM neurons are initialized uniformly in the 2D space defined by the first two principal components of the input attributes. The first two principal components will change by adding or dropping attributes from the input group; as a result, the projection from input attribute space to the 2D SOM space changes, while the contribution of each attribute changes as well. To ensure an overall optimal contribution measurement for a given group of attributes, I propose to use an exhaustive search over all combinations of three of more attributes,

$$S = \sum_{i=3}^{N} \frac{N!}{i!\,(N-i)!}, \tag{5.8}$$

and then weight by the number of attributes in each combination and take the sum over all used combinations:

$$\omega_i = \sum_{l=1}^{S} N_l \widetilde{\omega}_{il}, \tag{5.9}$$

where ! denotes factorial operation, S is the total number of SOM models to be searched, N_l is the number of attributes in the lth combination, $\widetilde{\omega}_{il}$ is the contribution of the ith attribute to the lth SOM model, and ω_i is the final contribution of the ith attribute to SOM. Although the proposed method involves running SOM multiple times with different input attribute combinations, it is an embarrassingly parallel problem so that the increase in computation time over the traditional SOM is negligible given sufficient amount of threads/processors.

To quantify which attributes are favored by interpreters to be used for SOM facies analysis, we first need to define an ideal facies map. Practically, an interpreter can either use different colors, or delineate boundaries to define different facies. The authors favor color representation, because then the boundary information becomes complimentary once a color change is identified. Attributes representing boundaries or local abrupt deformations such as coherence have much faster local variation than attributes representing bodies, and can dominate in voxel based classification methods. A window based method is proposed by Song et al. (2017), however they use waveforms instead of multiple attributes. Here I loosely define an "edge" attribute to be an attribute representing the variation among neighboring seismic samples, and a "body" attribute to be an attribute representing a property of a seismic sample or samples within a window. Figure 5.2 shows an example of two attributes along a horizon slice from the Ellenburger formation, Fort Worth Basin. The GLCM homogeneity is a texture attribute that provides a measure of the amplitude smoothness along a horizon, which is a "body" attribute. In contrast, the energy ratio similarity highlights edges, but ignores changes outside the edges. Researchers in geostatistics routinely inspect histograms to determine if variables behave

proper distribution for subsequent analysis, and even transform a variable using quantile representation if its distribution is highly skewed (Coburn et al., 2005). Figure 5.2c and 5.2d show the histogram of these two attributes, from which I observe that the body attribute exhibits a flatter and more symmetric histogram, whereas the edge attribute's histogram is tighter and skewed. Based on this observation, I propose to use skewness and kurtosis, which measures the symmetry and sharpness, to quantify the interpreter's preference of body attributes over edge attributes.

Skewness, which is the third moment of the standard score of a variable *x*, is defined as:

$$s(x) = \mathbb{E}\left[\left(\frac{x - \bar{x}}{\sigma_x}\right)^3\right], \tag{5.10}$$

where $\bar{x}$ is the mean of variable *x*, σ_x is the standard deviation, and $\mathbb{E}$ represents expectation. Similarly, kurtosis is the fourth moment of the standard score of a variable *x* and is defined as:

$$k(x) = \mathbb{E}\left[\left(\frac{x - \bar{x}}{\sigma_x}\right)^4\right]. \tag{5.11}$$

In practice, the skewness and kurtosis are precomputed before determining the attribute contribution **ω**. After the computation of **ω**, I further normalize both skewness and kurtosis to range between zero and one. Weighting the previously defined **ω** using skewness and kurtosis, and normalize again using the z-score:

$$w_i = \left(3 - \frac{|s_i| - \min\limits_{i=1,N}|s_i|}{\max\limits_{i=1,N}|s_i| - \min\limits_{i=1,N}|s_i|} - \frac{k_i - \min\limits_{i=1,N} k_i}{\max\limits_{i=1,N} k_i - \min\limits_{i=1,N} k_i}\right)\omega_i, \tag{5.12}$$

$$\widehat{w}_i = \frac{w_i - \overline{w}}{\sigma_w}. \tag{5.13}$$

Here, w_i is the weight of attribute *i* before z-score normalization, $\overline{w}$ is the mean of w_i, σ_w is the standard deviation, and $\widehat{w}_i$ is the weight of attribute *i* after the z-score. In equation 5.12, because the skewness term and kurtosis term are both normalized to range between zero and one, I assume an equal impact of skewness and kurtosis. At the same time, the absolute value of w_i is of less interest, as I further normalize it to be $\widehat{w}_i$ using z-score. Finally, I constrain the weight to range from zero to two using a sigmoid function, and defining the elements of the diagonal weight matrix **W** to be:

$$W_{ii} = \frac{2}{1 + e^{-\widehat{w}_i}}. \tag{5.14}$$

Using this heuristic approach, I have defined a weighting parameter as a function of the interpreter's knowledge and SOM response. I then use the distance definition in equation 5.5 with the updated weight matrix **W** to run SOM a final time, and generate the facies map. I provide a workflow summarizing the steps in Figure 5.3.

DATA DESCRIPTION

In this field example, the Barnett Shale lies directly on top of the dolomitic Ellenburger formation in the western region of the Fort Worth Basin (Figure 5.4). Going through a series of diagenetic episodes (Kupecz and Land, 1991) which include at least five karst events ranging between post Ellenburger to Early Pennsylvanian (Canter et al., 1993), the Ellenburger formation is highly deformed, with extensive development of karst and joints that extend upwards from the water-saturated Ellenburger into the Barnett Shale, posing drilling and completion hazards (Pollastro et al., 2007). I use spectral decomposition, geometric, and texture attributes, which are sensitive to strata thickness, lithology, and structural deformation, to illuminate the architectural elements presented in the shallow part of Ellenburger formation.

To study the geomorphology of the highly karsted Ellenburger formation, I use co-rendered multiple attributes along a phantom horizon (Horizon A) *25 ms* below the top of Ellenburger formation (Figure 5.5 to 5.8). Karst features in the study area appear either as isolated circular to oval shape, or in a cellular network of polygonal karst. The positive (red) on the perimeter and negative (blue) in the center in structural curvature (Figure 5.5) define collapse features. Amplitude curvature provides a complimentary image, enhancing smaller collapse features (e.g., blue arrows in Figure 5.6). Polygonal karst and fractures are delineated in both structural and amplitude curvatures (yellow arrows), allowing the identification of the major faults (red arrows), which compartmentalize the karst into different clusters. Modulating peak spectral frequency by peak spectral magnitude (Figure 5.7), we observe that highly karsted regions exhibit lower frequency compared to the surrounding area, possibly due to the non-specular scattering from the chaotic reflectors. These regions are also low in peak spectral magnitude (dimmer color in Figure 5.7), as a substantial amount of the reflected energy is not properly received by the receivers within the migration aperture. Besides the highly karsted regions, we also notice a shift in frequency in the less deformed surrounding area, indicating a change in layer thickness. Figure 5.8 co-renders GLCM homogeneity with energy ratio similarity. Visually, areas that are less coherent are also less homogeneous, suggesting a more rugose surface. The low GLCM homogeneity outlines the regions of karst, inside which the similarity attribute provides details of the karst geometry at different scales. These visual correlations of mathematically independent attributes for a given geologic feature provide an interactive, interpreter-driven means of selecting the most appropriate attributes for subsequent machine-driven classification.

APPLICATION

I perform SOM analysis within a 50 *ms* time window below the top Ellenburger surface in order to capture the structural deformation and diagenetic alteration of the sallow Ellenburger formation, which has a direct impact on the Barnett Shale reservoir lying unconformably above. While interaction with crossplots of two or three attributes (against x-, y-, and an additional color-axis, if needed) is simple, interacting with eight attributes and an 8D crossplot is intractable. In this implementation, SOM projects this intractable 8D crossplot into a 2D space that can be visualized on a computer screen, while still preserving most of the topological and distance relationship from the original 8D space. After precomputing the attribute weight matrix as described previously, I obtain the weights of the eight input attributes and display with their corresponding histogram in Figure 5.9. Note that highly skewed attributes (curvature and similarity) are penalized; in contrast, "body" attributes are amplified. With such a penalty on curvature and similarity attributes, one may suspect that the anomalies from these attributes would have been overly mitigated. I assume that the contrast in curvature and similarity between features of interest (karst and faults, versus planar features in this example) still allows them to be classified by the SOM facies map. The majority of geology is planar and continuous, such that the mean value of curvature is close to zero and the mean value of similarity close to one. In contrast, structural or curvature anomalies of interest have values several standard deviations from the mean. Therefore, we still expect to observe the anomalies after penalizing on these attributes.

The SOM facies map with equally weighed input attributes is shown in Figure 5.10, and from the proposed workflow using the same group of input attribute in Figure

5.11. In this implementation, I output the SOM facies map as two volumes, one for each axis of the projected 2D space. I then blend these two volumes together to generate a SOM facies map as shown in Figure 5.10 and 5.11. Therefore, every sample in these SOM facies map is no longer a scalar, but a two dimensional vector that represents the coordinates of that sample in the projected SOM 2D space. This vector representation enables us to plot each dimension against two 1D color bars perpendicular to each other. By using commercial crossplotting tools available in many interpretation packages provides the color bar shown in Figure 5.10 and 5.11. Such representation has greatly enhanced the interpretation of SOM facies.

Comparing Figure 5.10 and 5.11, note observe that both SOM facies maps are able to delineate the karst, faults, and fractures equally well. This observation verifies the assumption that I made earlier, that adding a penalty weight does not significantly alter the curvature and similarity anomaly contributions. The polygonal karst regions are characterized by purple and cyan facies, where purple corresponds to anticlinal components and cyan synclinal components. Compared to the co-rendered structural curvatures, both SOM facies maps provide details about smaller scaled karst caves that are not identifiable on structural curvatures, most of which correspond to fracture joints (blue arrows). We are also able to identify the major faults (red arrows) close to the polygonal karst regions, suggesting a tectonic control of the karst development (Khatiwada et al., 2013, Qi at al., 2014). The main difference between Figure 5.10 and 5.11 comes from regions marked with yellow and orange arrows. In Figure 5.11, the yellow arrow regions are in a lime green facies, where the orange arrow regions are in an orange facies. In contrast, these regions look nearly identical in Figure 5.10, all being

brownish cellular textures that somehow follow the trend on the curvature attributes. The lime green versus orange facies change in Figure 5.11 reflects the frequency variation found in Figure 5.7, where low peak frequency regions are in lime green facies (yellow arrows), and high frequency regions are in orange facies (orange arrows). The peak frequency provides information on tuning thickness, which adds another dimension besides surface morphology. The SOM facies map from equally weighted attributes, on the other hand, does not distinct such frequency variation clearly. After z-score normalization, the contrast in some attributes (such as curvature and similarity in this example) are much higher than the others (such as spectral decomposition attributes) and dominates the distance calculation. SOM is biased towards finding clusters in those attributes. After stretching and squeezing the distance along each attribute dimension using the precomputed weights, I am able to recover the variation in peak frequency that is otherwise buried in the strong curvature contrast. Figure 5.12 provides a SOM result generated from a subset of equally weighted attributes. In this example, I only use structural curvatures, peak spectral frequency and magnitude, and GLCM homogeneity as input attributes, with the objective to better define the extent of polygonal karst while reducing the dominance of edge attributes that we previously observed when there were three more edge attributes. By eliminating the overprint from amplitude curvature and similarity, we recognize a cleaner silhouette of the polygonal karst; however, the change in frequency in the less deformed regions is still difficult to identify. This facies map is excellent in delineating structural deformations of karst, faults, and fractures, but fails to properly represent frequency information.

To look for evidence of the facies discovered in Figure 5.11, I take four composite vertical sections (a) – (d) from seismic amplitude, and co-render with the SOM facies (Figure 13). In order to allow the background seismic amplitude to come through, I modulate the opacity of the SOM facies, which leads to faded color on the vertical sections compared to on the map along Horizon A. Horizon A lies vertically at the center of the SOM analysis window, so we expect to see the same facies in Figure 5.11 to appear at the center (vertically) of the colored zone on vertical sections. In section (b), I use a red-white-blue colorbar for seismic amplitude to better visualize the thickness change in reflectors, where the black curve marks Horizon A. We see a change in layer thickness in the two red ovals, which verifies the facies change that I previously interpreted. The karst features identified on Horizon A nicely match the synclinal events on seismic amplitude profile (marked as red curves), with the perimeter in purple (which fades to magenta on the vertical sections) delineating the extent. We identify a large dome shaped karst developed in the center of a collapse in section (d), which may related to the compression from the two reverse faults on the east and west sides (blue curves). Most polygonal karst features develop along or in the vicinity of faults, and many of the smaller scale, isolated karst caves appear at joints of folds. The location of karst indicates that fractures of different scales might have accelerated the weathering and dissolution process of the Ellenburger formation.

DISCUSSION

By now I have demonstrated that by adaptively weighting the input attribute, information from input attributes are more adequately presented in the SOM facies map. To further investigate the effectiveness of the proposed attribute weighting scheme, I

provide a second example from the Barnett Shale, in which the main target is to map the interbedded layers within the Barnett Shale formation overlying the Ellenburger formation. In this example, I use structural attributes and inversion attributes as inputs to SOM, where the weights calculated using the proposed approach are provided in Table 5.1. Inversion attributes are closely related to the elastic properties and lithology of the rocks, and therefore are effectively used to characterize shale reservoirs. On the other hand, structural attributes are useful when mapping faults and folds in the shale formation. Figure 5.14 shows SOM facies maps from both equally weighted (Figure 5.14a) and data-adaptively weighted (Figure 5.14b) input attributes. We observe that the result from adaptively weighted attributes provides better defined layers, contain less "noise" associated with the structural attributes (black ovals), and still shows the major fault in the west of the survey. Furthermore, while increasing the weights of the inversion attributes, we also note the weights are consistent with the hierarchy of attribute dependence. P and S impedances are two independent attributes directly inverted from seismic inversion and have the highest weights; Lambda/Mu and V_p/V_s are calculated from P and S impedances and have more modest weights; Poisson's Ratio is calculated from V_p/V_s and has the lowest weight. Such weight values indirectly verifies the dependency among input attributes.

CONCLUSIONS

In this study, I augment the attribute selection process with an attribute weighting process for seismic facies analysis. I define the attribute weight to be a function of both SOM response and interpreter's preference. By using such weights, information in the input attributes are more adequately represented in the SOM facies map, and less

dominated by attributes with very high contrast. The weights also provides information on attribute inter-dependency, which may allow interpreters to understand the origin and behavior of different attributes, and in turn to select the most appropriate group of input attributes for seismic facies analysis. Being a heuristic method in nature, the proposed approach is one of the many possible ways for attribute selection, however I believe it is the very first attempt to using weighting in place of simple selection. I hope that the proposed approach will inspire others to add further metrics to the quantitative attribute selection problems.

CHAPTER 5 TABLES

Attribute	Weight
Coherent energy	0.52
Dip magnitude	0.69
Energy ratio similarity	0.52
P impedance	1.66
S impedance	1.73
Lambda/Mu	0.99
V_p/V_s	0.94
Poisson's Ratio	0.72

Table 5.1. Input attributes and their corresponding weights used in the second Barnett Shale example. Note the weight values of inversion attributes are consistent with attribute dependency.

CHAPTER 5 FIGURES

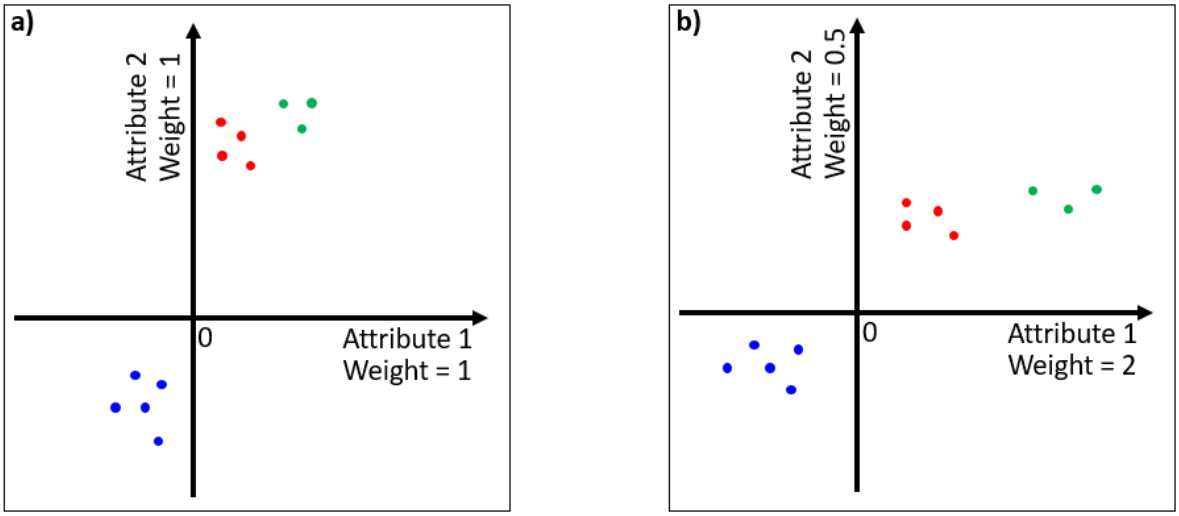
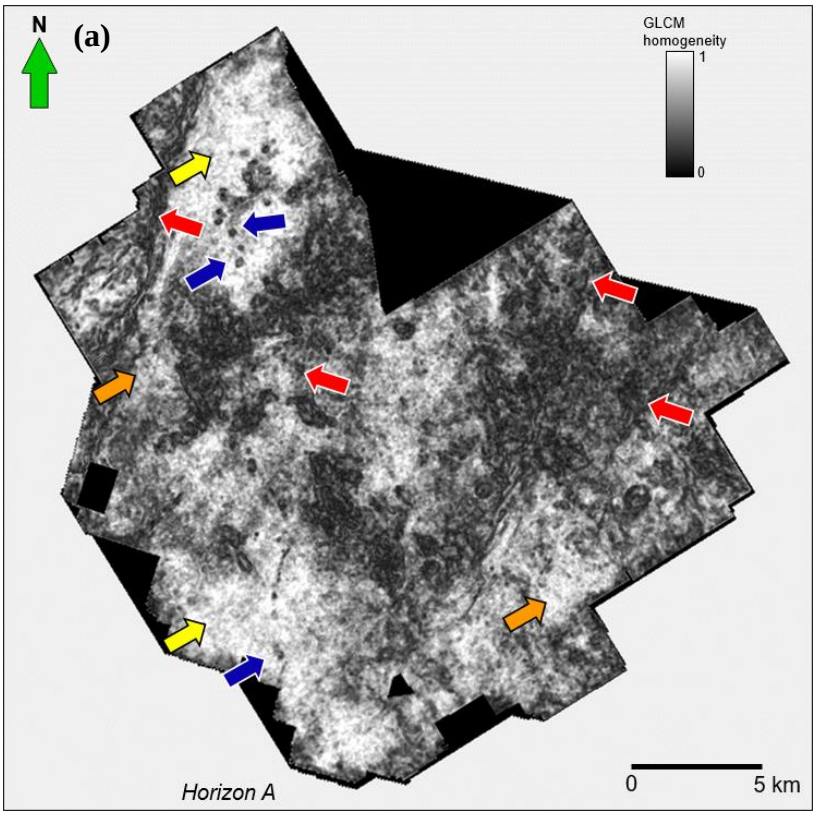
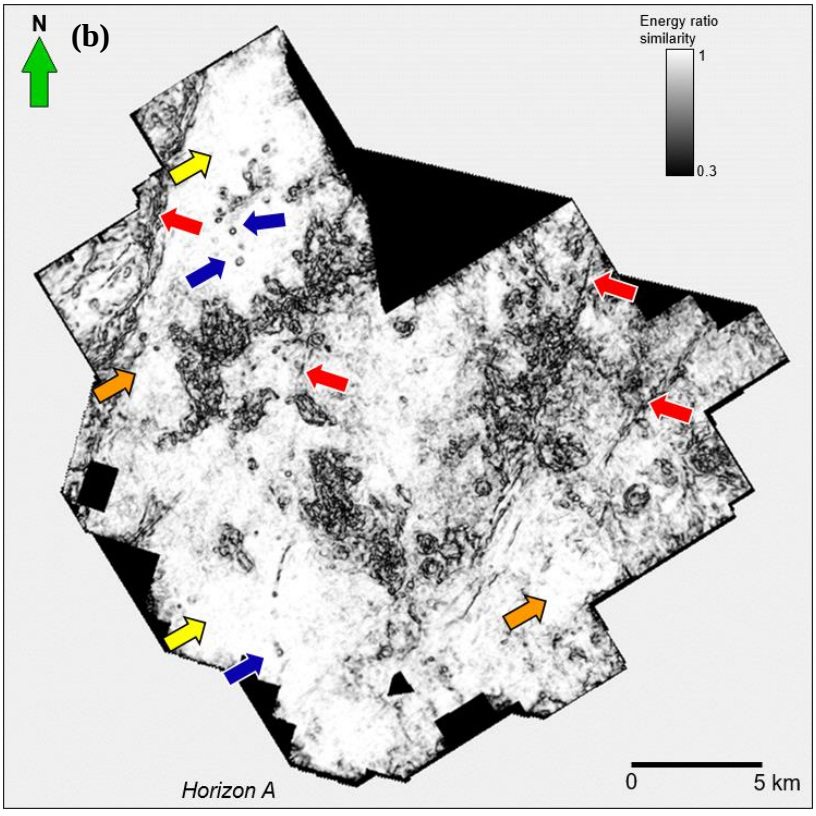


Figure 5.1. A schematic drawing to show the effect of weighting attributes. Before analysis, the attributes are transformed to have equal contribution to clustering using a z-score algorithm. (a) Two equally weighted attributes with three clusters. The red and green clusters are very close such that a distance-based clustering algorithm may only find two instead of three clusters, placing red and green samples into a single cluster. (b) The same data samples but now with different weights applied to the two attributes changing their distance from the origin. The red and green clusters are now better separated.


N
(a)
GLCM
homogeneity
1
0
Horizon A
0
5 km


N
(b)
Energy ratio
similarity
1
0.3
Horizon A
0
5 km

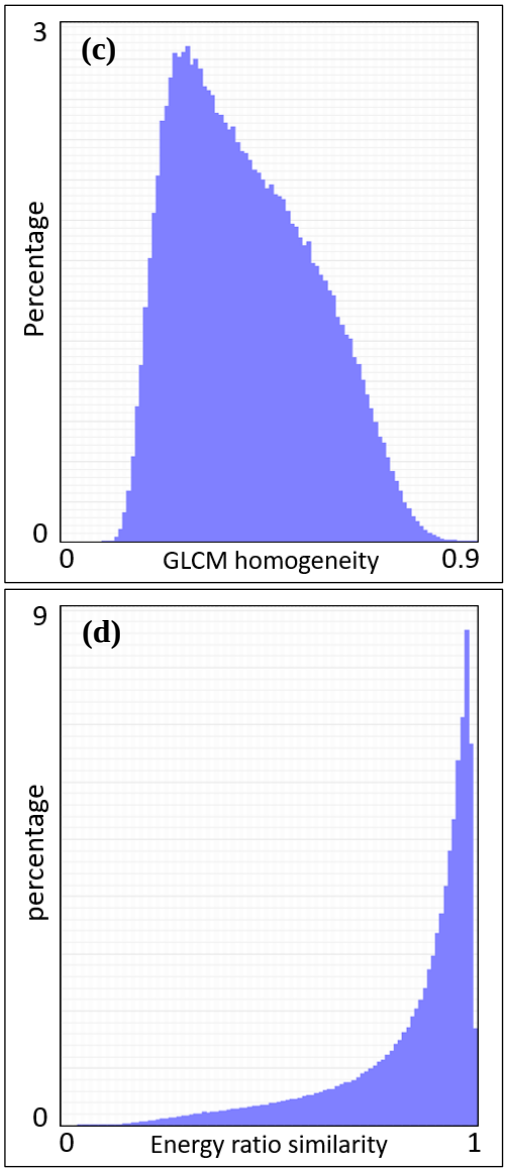


Figure 5.2. A phantom Horizon A *25 ms* below the top of Ellenburger formation through the (a) GLCM homogeneity and (b) Energy ratio similarity volumes. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. I loosely define the GLCM homogeneity as a "body" attribute as it represents information within a window along structural dip. It is useful for mapping the extent of the highly deformed regions. In contrast, the energy ratio similarity is an edge attribute that highlights the boundaries among different features. (c) The histogram of GLCM homogeneity within the analysis window. (d) The histogram of energy ratio similarity within the analysis window.

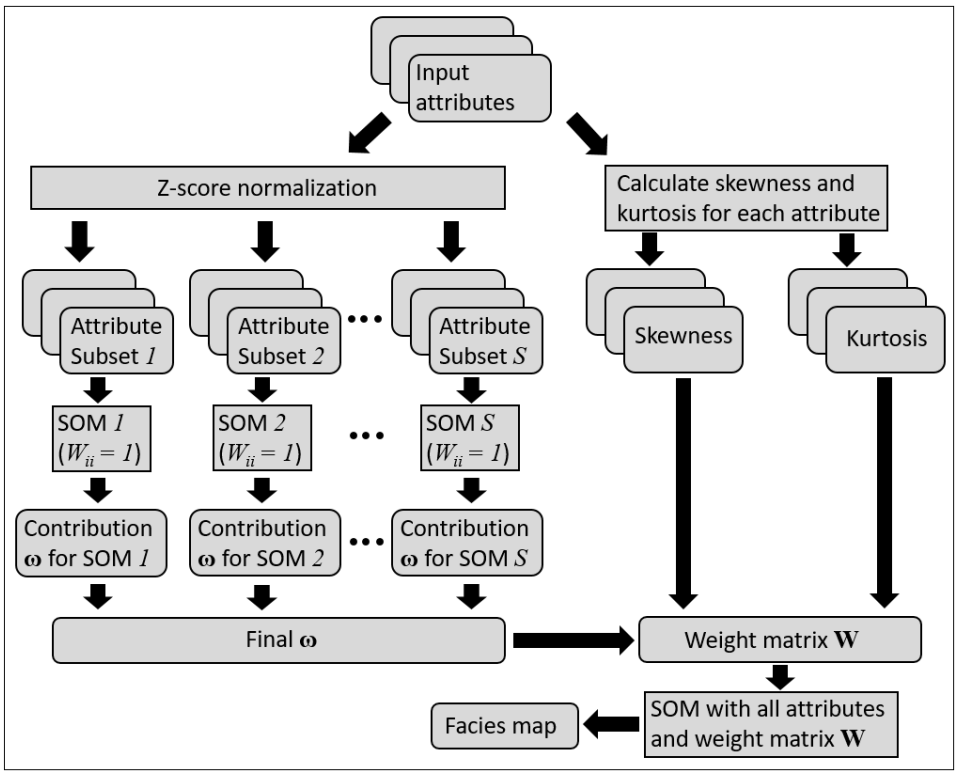


Figure 5.3. The proposed attribute weighting workflow.

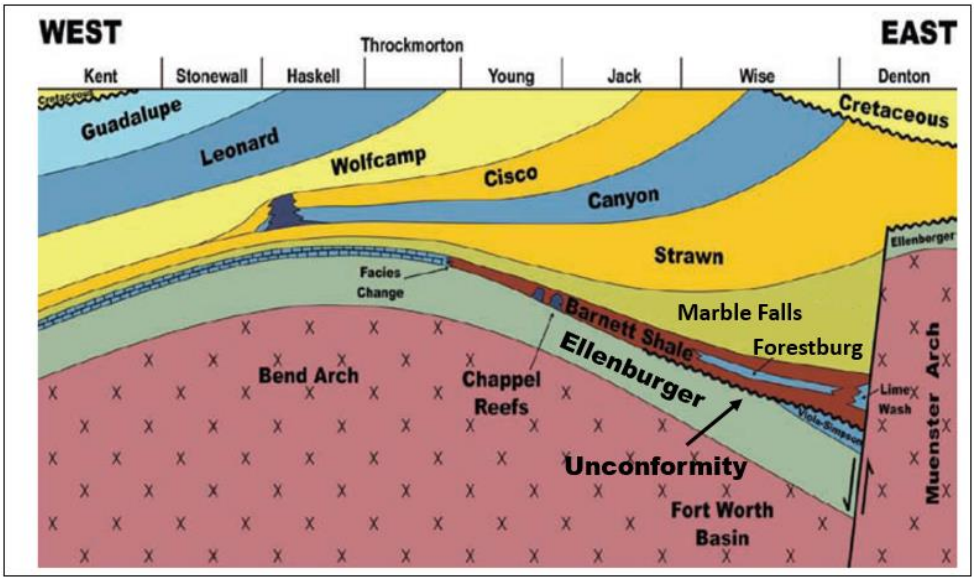


Figure 5.4. Stratigraphic cross section of the Fort Worth Basin. In the study area, the Barnett Shale lies unconformably on top of the dolomitic Ellenburger formation. The seismic survey showing the Ellenburger karst is to the south of Young County in this plot (after Pollastro et al., 2007).

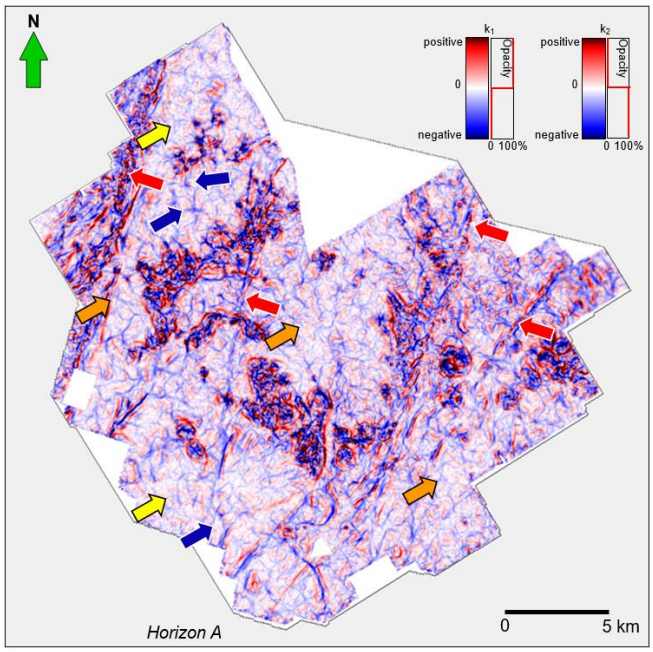


Figure 5.5. Co-rendered structural curvatures k_1 and k_2 along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers.

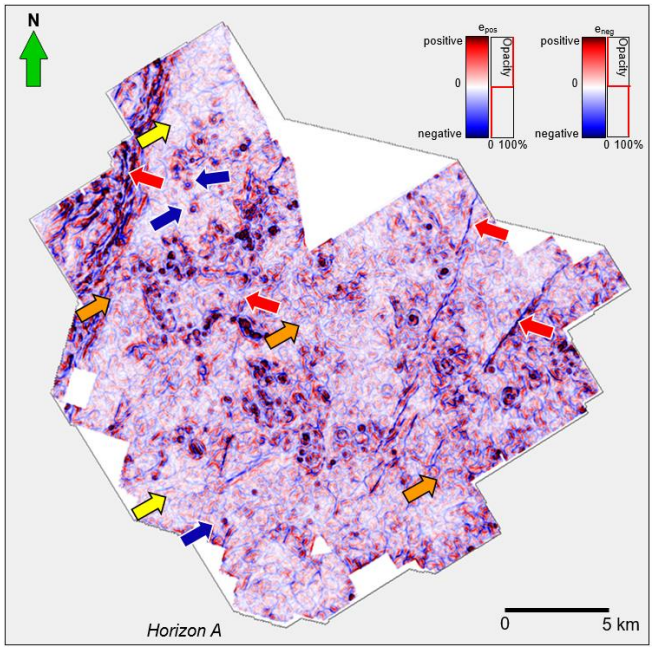


Figure 5.6. Co-rendered amplitude curvatures e_{pos} and e_{neg} along a phantom Horizon A *25 ms* below the top of the Ellenburger formation Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. Amplitude curvature shows more details about the small scale karst.

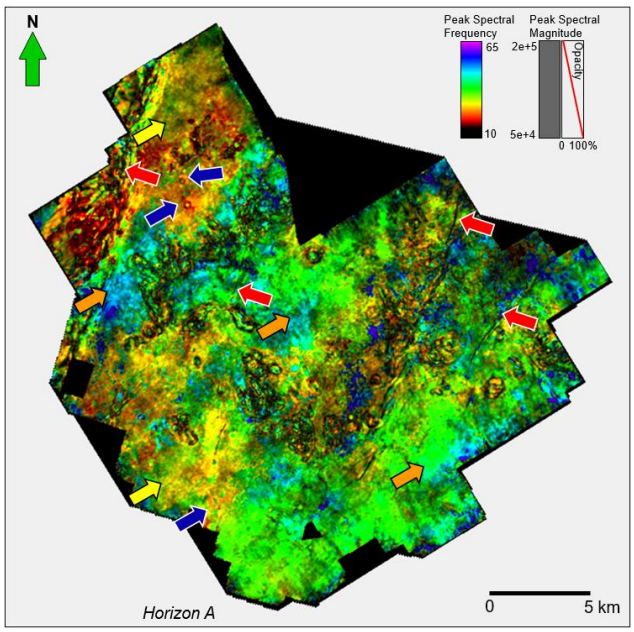


Figure 5.7. Peak spectral frequency modulated by peak spectral magnitude along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. The polygonal karst regions are of low frequency and low magnitude, possibly due to the non-specular reflection from the highly deformed reflectors.

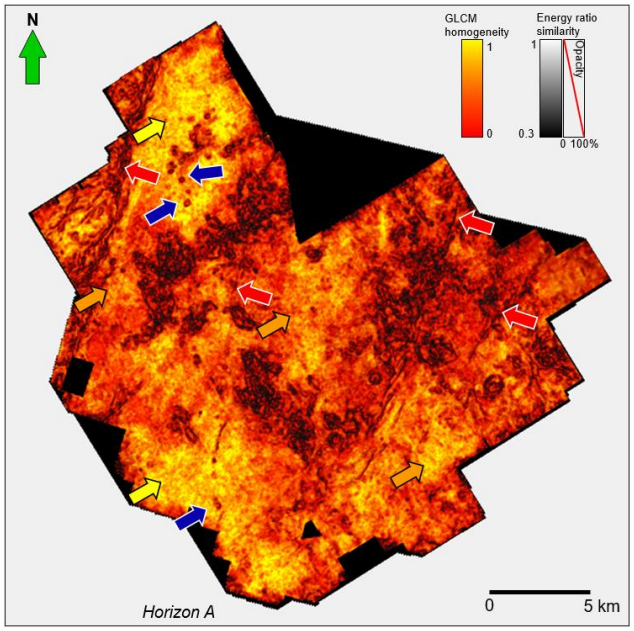


Figure 5.8. Co-rendered GLCM homogeneity and energy ratio similarity along a phantom Horizon A *25 ms* below the top of Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers.

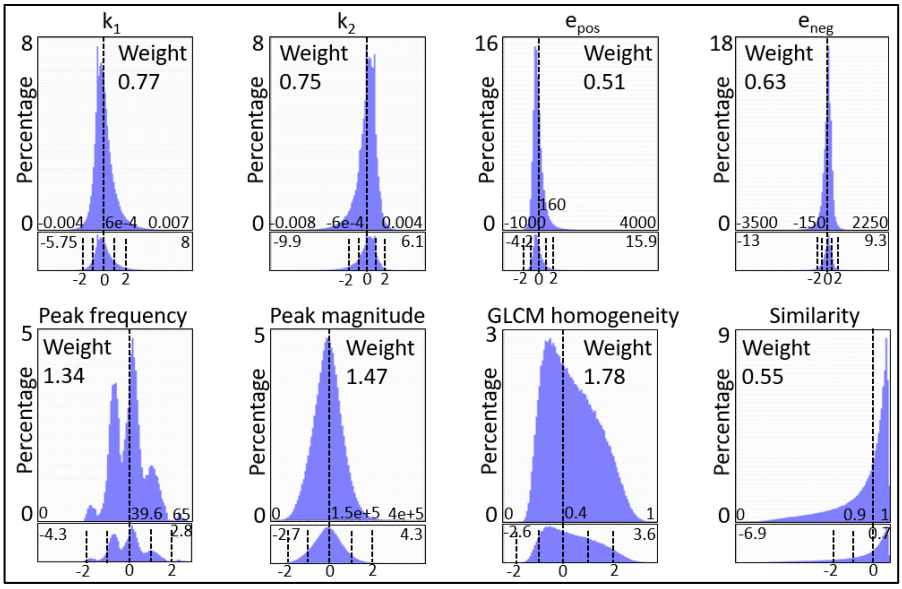


Figure 5.9. Histograms of the eight input attributes within the analysis window. Weights are computed using the method described in this paper. Note that histograms exhibiting high kurtosis and/or skewness are assigned lower weights, while those that rea broader and more symmetric are assigned higher weights. Values at the bottom of each histogram mark the z-scores, and dashed lines mark values at zero, one, and two standard deviations. Note curvature and coherence attributes span several standard deviations.

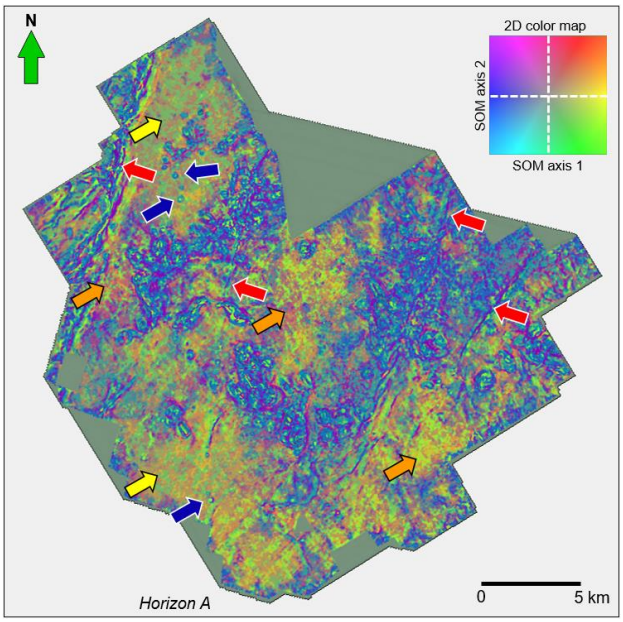


Figure 5.10. SOM facies map from equally weighted attributes along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. The dashed lines in the 2D color maps denotes the 1D colorbar used for each projection along a SOM axis in order to approximate the 2D color map.

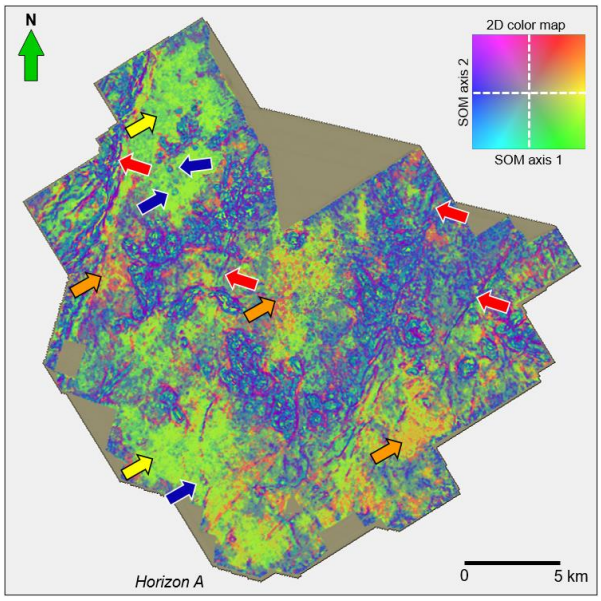


Figure 5.11. SOM facies map from adaptively weighted attributes along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. The dashed lines in the 2D color maps denotes the 1D colorbar used for each projection along a SOM axis in order to approximate the 2D color map.

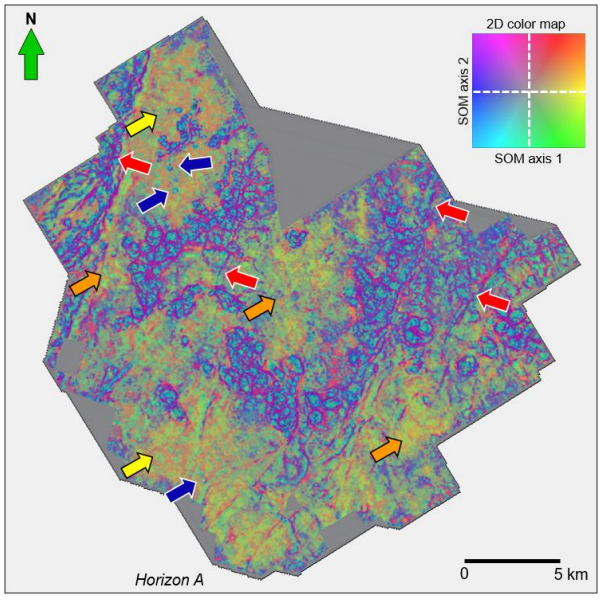


Figure 5.12. SOM facies map from a subset of equally weighted attributes along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. The dashed lines in the 2D color maps denotes the 1D colorbar used for each projection along a SOM axis in order to approximate the 2D color map.

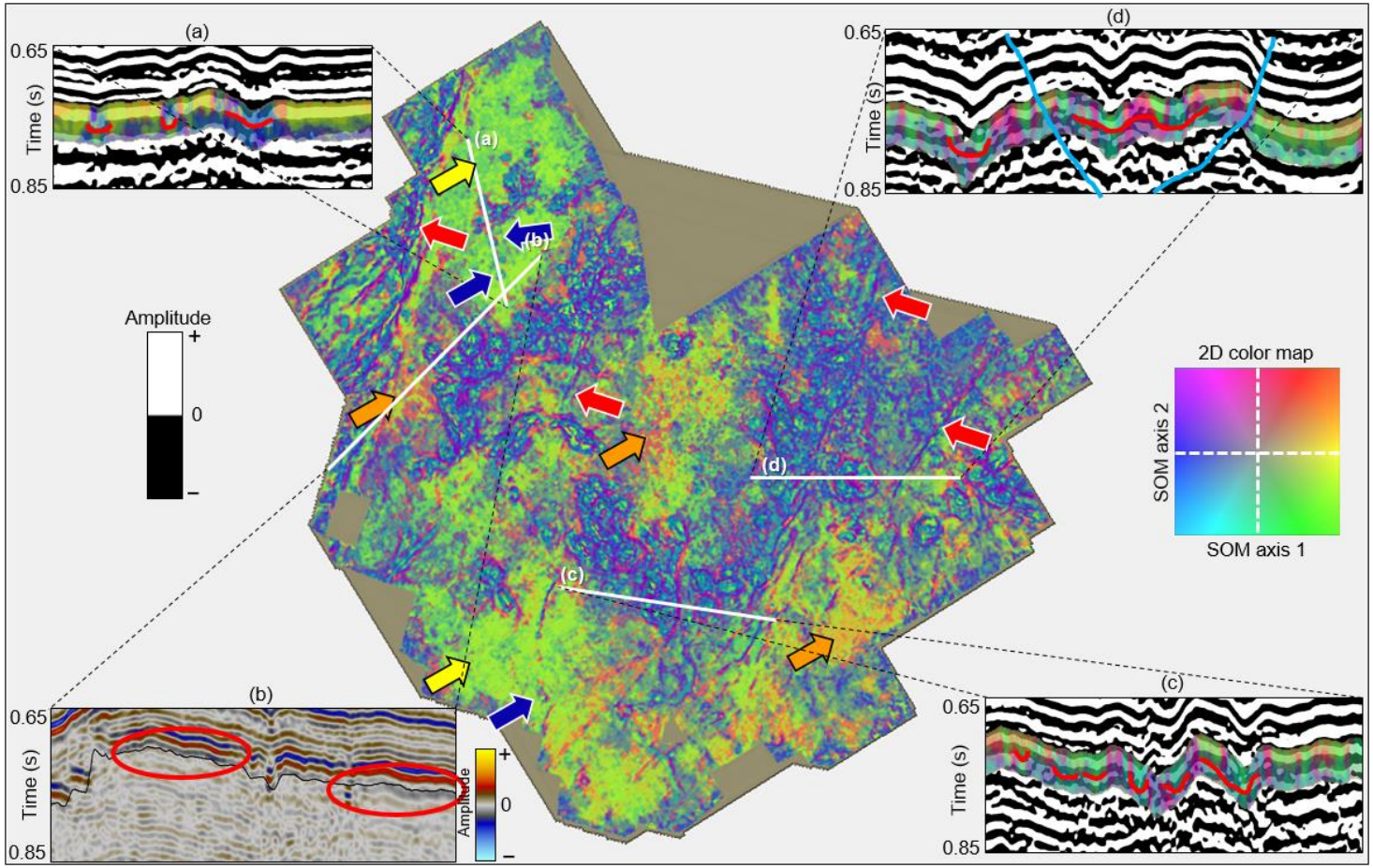


Figure 5.13. SOM facies map from adaptively weighted attributes along a phantom Horizon A *25 ms* below the top of the Ellenburger formation. Red arrows denote locations of large regional faults. Blue arrows denote small scale, isolated karst collapse features. Yellow arrows denote less deformed regions of relatively thick layers, while orange arrows denote less deformed regions of relatively thin layers. The dashed lines in the 2D color maps denotes the 1D colorbar used for each projection along a SOM axis in order to approximate the 2D color map. Inserts (a) to (d) shows seismic evidence of the interpreted facies.

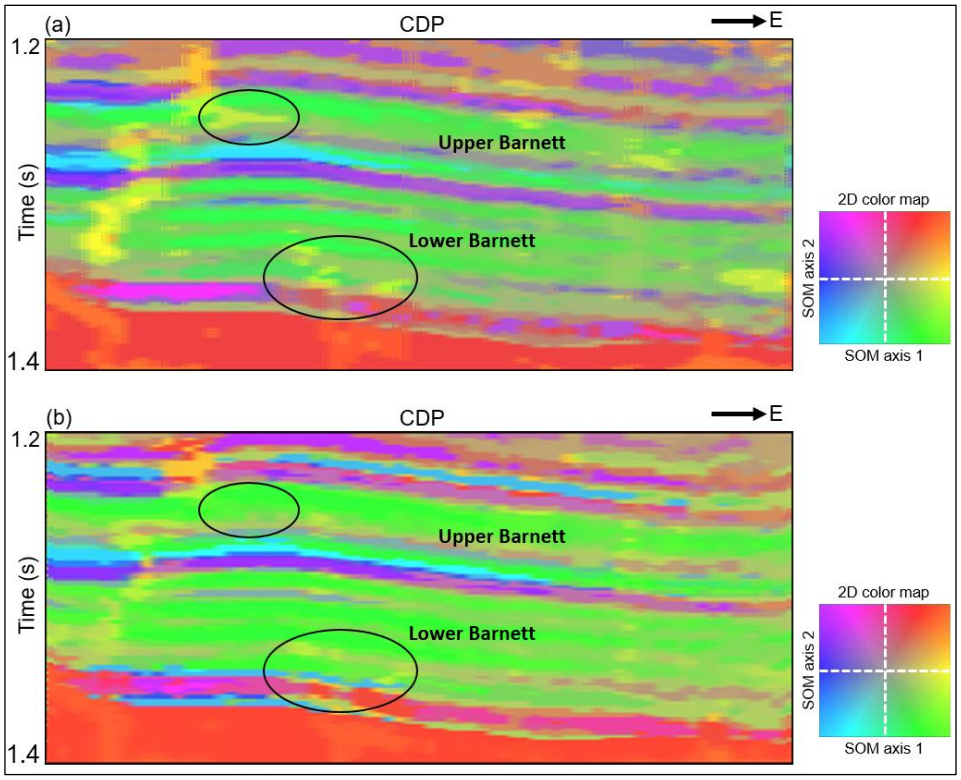


Figure 5.14. Vertical sections from a second Barnett Shale seismic survey. (a) SOM facies map with equally weighted input attributes; (b) SOM facies map from adaptively weighed input attributes. Weights for each attribute are provided in Table 1. We notice the layers within the shale formations are better defined in (b), and with less noise associated with structural attributes (black ovals).

REFERENCES

Altamar, R. P., and K. J. Marfurt, 2015, Identification of brittle/ductile areas in unconventional reservoirs using seismic and microseismic data: Application to the Barnett Shale: Interpretation, **3**, T233-T243.

Barnes, A. E., 2007, Redundant and useless seismic attributes: Geophysics, **72**, P33-P38.

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221-2224.

Benabdeslem, K., and M. Lebbah, 2007, Feature selection for self-organizing map: IEEE 29th International Conference on Information Technology Interfaces, 45-50.

Canter, K. L., D. B. Stearns, R. C. Geesaman, and J. L. Wilson, 1993, Paleostructural and related paleokarst controls on reservoir development in the Lower Ordovician Ellenburger Group, Val Verde basin, in R. D. Fritz, J. L. Wilson, and D. A. Yurewicz, eds., Paleokarst Related Hydrocarbon Reservoirs: SEPM Core Workshop 18, 61-99.

Chen, Q., and S. Sidney, 1997, Seismic attribute technology for reservoir forecasting and monitoring: The Leading Edge, **16**, 445-448.

Coburn, T. C., J. M. Yarus, and R. L. Chambers, 2005, Stochastic modeling and geostatistics: principles, methods, and case studies, vol. II, AAPG computer applications in geology 5: AAPG.

Coléou, 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.

Dorrington, K. P., and C. A. Link, 2004, Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction: Geophysics, **69**, 212-221.

Hampson, D. P., J. S. Schuelke, and J. A. Quirein, 2001, Use of multiattribute transforms to predict log properties from seismic data, Geophysics, **66**, 220-236.

Hart, B. S., and R. S. Balch, 2000, Approaches to defining reservoir physical properties from 3-D seismic attributes with limited well control: An example from the Jurassic Smackover Formation, Alabama: Geophysics, **65**, 368-376.

Hu, C., W. Lu, and Y. Zhang, 2014, Seismic facies classification using visualization-induced self-organizing mapping: 76th EAGE Conference and Exhibition 2014.

Kalkomey, C. T., 1997, Potential risks when using seismic attributes as predictors of reservoir properties: The Leading Edge, **16**, 247-251.

Khatiwada, M., G. R. Keller, and K. J. Marfurt, 2013, A window into the Proterozoic: Integrating 3D seismic, gravity and magnetic data to image subbasement structures in the southeast Fort Worth basin: Interpretation, **1**, T125–T141.

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

Kupecz, J.A., and L. S. Land, 1991, Late-stage dolomitization of the Lower Ordovician Ellenburger Group, West Texas: Journal of Sedimentary Petrology, **61**, 551-574.

Mahalanobis, P. C., 1936, On the generalized distance in statistics: Proceedings of the National Institute of Sciences of India, **2**, 49–55.

Matos, M. C., K. J. Marfurt, and P. R. S. Johann, 2009, Seismic color self-organizing maps: Presented at 11th International Congress of the Brazilian Geophysical Society, Extended Abstracts.

Pollastro, R. M., D. M. Jarvie, R. J. Hill, and C. W. Adams, 2007, Geologic framework of the Mississippian Barnett shale, Barnett-paleozoic total petroleum system, Bend arch — Fort Worth Basin, Texas: AAPG Bulletin, **91**, 405–436.

Poupon, M., K. Azbel, and G. Palmer, 1999, A new methodology based on seismic facies analysis and litho-seismic modeling: The Elkhorn Slough field pilot project, Solano County, California: 69th Annual International Meeting, SEG, Expanded Abstracts, 927–930.

Qi, J., B. Zhang, H. Zhou, and K. Marfurt, 2014, Attribute expression of fault-controlled karst — Fort Worth Basin, Texas: A tutorial: Interpretation, **2**, SF91-SF110.

Roden, R., T. Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **4**, SAE59-SAE83.

Roy, A., B. L. Dowdell, and K. J. Marfurt, 2013, Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma: Interpretation, **1**, SB109–SB124.

Schuelke, J. S., and J. A. Quirein, 1998, Validation: A technique for selecting seismic attributes and verifying results: 68th Annual International Meeting, SEG, Expanded Abstracts, 936-939.

Song, C., Z. Liu, Y. Wang, X. Li, and G. Hu, 2017, Multi-waveform classification for seismic facies analysis: Computers & Geosciences, **101**, 1-9.

Strecker, U., and R. Uden, 2002, Data mining of 3D post- stack attribute volumes using Kohonen self-organizing maps: The Leading Edge, **21**, 1032–1037.

Sullivan, E. C., K. J. Marfurt, A. Lacazette, and M. Ammerman, 2006, Application of new seismic attributes to collapse chimneys in the Fort Worth Basin: Geophysics, **71**, B111-B119.

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

Zhao, T., J. Zhang, F. Li, and K. J. Marfurt, 2016, Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps: Interpretation, **4**, SB79-SB89.

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

CHAPTER 6

CONCLUSIONS

In this dissertation, I explored the feasibility of machine learning/pattern recognition techniques on seismic interpretation, and focused on one technique, the self-organizing map (SOM), improving it to better address multiattribute seismic interpretation problems. Machine learning/ pattern recognition techniques can effectively reduce the human labor involved in the traditional seismic interpretation, while reducing the interpretation uncertainty by combining information from multiple seismic attributes. Such technology innovation happens in every aspect of exploration geophysics. For example, while seismic imagers no longer use hand drawing on papers for migration and rely heavily on computers to perform the operation, there are still thousands of seismic imagers in the industry. Similarly, replacing human interpreters by automatic systems is not what I advocate and is deemed impossible. Interpreters will outsource the labor intensive steps to computers, oversee the processes, quality control the final products, and make decisions based on their experience. With the development of autopickers, seismic attributes, and seismic facies analysis techniques, I anticipate interpreters will be able to generate more complete, more quantitative products, testing and evaluating more hypotheses.

Being an emerging topic, there are numerous developments and improvements that await us. I focus on SOM as it is the most popular and commonly available unsupervised learning method for the seismic interpretation community. Traditional SOM was not developed specificly for seismic interpretation problems. With the improvements I made, I believe that the seismic facies represent the information from

input attributes more appropriately than the traditional SOM. While the distance-preserving improvement (discussed in Chapter 3) is unique to SOM, the stratigraphic constraint (Chapter 4) and attribute weighting (Chapter 5) are equally valid for other unsupervised seismic facies analysis methods. In this sense, SOM is just a proxy that demonstrates the value of such improvements; the broader message I wish to convey is a general solution on how to tailor an unsupervised learning technique to properly handle multiattribute seismic facies analysis. I hope that this dissertation contributes to building a "Turing machine" that built for seismic interpretation.