UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

LATENT SPACE CLASSIFICATION OF SEISMIC FACIES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ATISH ROY
Norman, Oklahoma
2013

LATENT SPACE CLASSIFICATION OF SEISMIC FACIES

A DISSERTATION APPROVED FOR THE
CONOCOPHILLIPS SCHOOL OF GEOLOGY AND GEOPHYSICS

BY

Dr. Kurt J. Marfurt, Chair

Dr. Roger M. Slatt

Dr. G. Randy Keller

Dr. Deepak Devegowda

Dr. Vikram Jayaram

To my beloved late dida

ACKNOWLEDGMENTS

To earn a PhD in Geophysics is like a dream come true for me. It has been an eventful three and half years of my life at University of Oklahoma. I have reached this goal with the support and guidance of a lot of people.

I received a lot of guidance, knowledge and encouragement on my research from my supervisor Dr. Kurt Marfurt. Initially when I started under Dr. Marfurt I was not very comfortable in computer programming. However, once he wrote the PCA program, wrote my name as the author of the program, and made everybody understand that I was the author of the program. This gave me tremendous encouragement to understand how the code works and it was how I started to write programs for my research. I think my self very lucky to be his student. He was never strict with me and always encouraged me visit home frequently and spend time with my family back in India. This gave me further encouragement to work hard when I returned. Dr. Marfurt is one of the most wonderful persons I have ever met in my life.

I also want to express my gratitude to Dr. Vikram Jayram who is one of my committee members. In his busy schedule he used guided me how to proceed in my research, and gave new ideas to my research. He helped me to understand the complicated part of my research (the GTM technique), and because of this quick understanding of the GTM theory, formulation of GTM into AASPI took me just a couple of months. He also sat with me for long to make my research writings more scientific and easy to read.

I would also like to express my sincere thanks to Dr. J. Tim Kwiatkowski. He helped whenever I faced problems in the codes that I was unable to solve. He has been very helpful to me. He gave me some most important new ideas on how to explore the probabilistic GTM technique, which is one of the important parts of my PhD research.

Also, I want to thank Marcilio Matos who was one of my initial mentors to understand the SOM algorithm. My PhD research is based on his initial works.

I would also like to thank Dr. Roger Slatt who is also one of my committee members for his important comments and feedbacks on my research. I learnt a lot of reservoir characterization from his courses. I also thank my other committee members Dr. Deepak Devegowda for giving me the chance to work with his dataset and provide me the opportunity to present my works to one of his industry sponsor. I would also like to thank Dr. Randy Keller for being one my committee member. He is always nice to me and encouraged my research.

I would like to thank the entire staff of ConocoPhillips School of Geology and Geophysics: Donna S. Mullins, Nancy Leonard, Adrianne Fox, Teresa Hackney and Jocelyn Cook for their help and support.

I also want to thank the AASPI Consortium sponsors, all the student members at the AASPI consortium and my fellow graduate students at the ConocoPhillips School of Geology and Geophysics for their help and support. I also like to thank my friends: Roderick, Ben, Alfredo, Toan, Oswaldo, Brad, and Araceli at who helped me in various ways in my research. I would also like to thank my close friends Avinash who has been here since I came to OU, helped me in different ways and is my very good partner for computer games. I also like to thank Sumit for his *bhokali* help in various ways. I would like to acknowledge the support of some of my other friends Supratikda, Nabanitadi and Bijit.

I will not be able to express in few words how the love, affections and support of my *ma* and *baba*. Constant prayer and never-ending love of *ma* for me helped me to get success in my life.

I would also like to thank deeply my love *Tora* for staying beside me and have faith in me always. We had several ups and downs in our relationship, however we endured everything together and hopefully we will be together in the coming years. Her sweet smile and love gives me constant encouragement for my works.

I also want to express my love to all my family members, *Tintin, bhaiti, chotomoni, bondhu, puchke, bonti, mejomoni* and *meso* for their everyday wish, love and blessings of *dadubati* and *thama,* that made me succeed and feel that I am never alone. Finally, I would like to thank my late *dida* without whose love and blessings I think I would not have progressed so far in my life.

TABLE OF CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

ABSTRACT

Supervised and unsupervised seismic facies classification methods are slowly gaining popularity in hydrocarbon exploration and production workflows. Unsupervised clustering is data driven, unbiased by the interpreter beyond the choice of input data and brings out the natural clusters present in the data. There are several competing unsupervised clustering techniques, each with advantages and disadvantages. In this dissertation, I demonstrate the use of various classification techniques on real 3D seismic data from various depositional environments. Initially, I use the popular unsupervised Kohonen self-organizing maps (SOMs) algorithms and apply it to a deep-water Gulf of Mexico 3D dataset to identify various deep-water depositional facies including basin floor fans, mass transport complexes and feeder channels. I then extend this algorithm to characterize a heterogeneous Mississippian Chert reservoir from Oklahoma and map the locations of the tight/non-porous chert and limestone vs. more prospective porous tripolitic chert and fractured chert zones. The tight chert and dense limestone can be highly fractured, giving rise to an additional seismic facies. In both the case studies, a large number of potential classes are fed into the SOM algorithm. These “prototype vectors” are clustered and colors are assigned to them using a 2D gradational RGB color-scale for visual aid in interpretation.

Kohonen SOM suffers from the absence of any proper convergence criterion and rules for parameter selection. These shortcomings are addressed by the more recent development of generative topographic mapping (GTM) algorithm. GTM is based on a probabilistic unsupervised classification technique and “generates” a PDF to map the data about a lower dimensional “topographic” surface residing in high dimensional attribute space. GTM predicts not only which cluster best represents the data, but also how well it is predicted by all other clusters. For this reason, GTM interfaces neatly with modern risk analysis workflows. I apply the GTM technique to classify 15 sets of horizontal well parameters in one of the recent unconventional shale plays, correlating the results with normalized estimated ultimate recovery (EURs), allowing an estimation of EUR based on the most relevant parameters.

I extend the GTM workflow to consider multi-attribute inversion volumes and do seismic facies classification for a Barnett shale survey. With the aid of microseismic data, the clusters from GTM analysis are interpreted as brittle or ductile. I also apply the GTM technique to the P-impedance (Z_P), lambda-rho ($\lambda\rho$), mu-rho ($\mu\rho$) and the V_P/V_S volumes from a Veracruz Basin survey in Southern Mexico that was acquired over a heterogeneous conglomerate reservoir.

Finally, I introduce limited supervision into both the SOM and GTM algorithms. The target vectors for both SOM and GTM are the average attribute vector about the different facies identified from the well logs. This supervision introduced user-defined clusters. In the preliminary supervision, I use multiattribute minimum

Euclidean distance measures, comparing the results with the unsupervised SOM results. For GTM, I calculate the probability of occurrence of the well facies in the survey.

Given the appropriate 3D seismic attribute volumes, SOM and GTM workflows will not only accelerate seismic facies identification, but also with GTM, quantify the identification of different petrotypes or heterogeneities present in the reservoir zone. The final product of my dissertation is a suite of algorithms, workflows, user interfaces and user documentation allowing others to build upon and extend this research.

CHAPTER 1

MOTIVATION AND OBJECTIVE

Seismic facies classification is a critical step in understanding the depositional history and properties of a reservoir and therefore extremely useful in hydrocarbon exploration and development. Traditionally skilled interpreters delineate seismic facies on 2D lines by visually scanning the waveforms, frequency, amplitude, phase and geometric relationship to their neighbors on the vertical sections, producing a 2D output map of seismic facies. Seismic attributes provide quantitative estimates of these properties including not only frequency, amplitude and phase, but also dip magnitude, dip azimuth, similarity, curvature and reflector convergence. With the adoption of 3D technology and increasing survey size, such manual techniques became extremely time consuming. The number of seismic attributes has also increased dramatically, providing increasingly accurate measurements of reflector morphology, but also greatly expanding the amount of data to be analyzed.

For these reasons, considerable effort has been devoted to more efficient 3D volume based seismic facies estimation, resulting in two types of algorithms and workflows. In supervised seismic facies classification the interpreter provides picks or target vectors around wells or interesting features in the seismic dataset and the dataset is classified accordingly. In the unsupervised classification the interpreter feeds in multi-attribute or seismic amplitude volume and the classification is data driven.

Supervised interpretation is well established in the exploration industry with several commercial products available. However, unsupervised classification is less mature, with external advances being pushed by applications in consumer purchasing, public health and national security needs. This dissertation focuses on hybrid algorithms with both unsupervised and supervised classification for seismic facies exploration. There are several such unsupervised seismic facies analysis techniques based on K-means clustering, principal component analysis (PCA), neural networks and Kohonen Self-organizing Maps (SOM). K-means clustering requires *a priori* knowledge of the number of clusters before classification. However with the noisy seismic data, with their continuity and low dimensionality does not hold the perquisite for a good K-means clustering (Coleou et al., 2003). Principal component analysis is widely used to reduce the redundancy and excess dimension of the data. The PCA is commonly used to reduce the data-dimensionality to make the classification more accurate and efficient. Each of the techniques solves different objective functions to form the classification, with each having their merits and demerits.

Kohonen SOM (Kohonen, 1982) originally developed for gene pattern recognition, is one of the most popular classification techniques, and it has been implemented in at least three commercial software packages for seismic facies classification. Mathematically the major advantage SOM has over K-means classification techniques is that the clusters in the SOM grid space are ‘topologically” related to the input dataset through the means of an intermediate latent space. 2D and 3D SOM grid spaces can be mapped to 2D and 3D colorbars to show the data classes (Matos 2006). I applied SOM technique on various dataset for multi-attribute classification.

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. Apart from general seismic facies classification, GTM can be used to predict the probability of occurrence of a particular seismic facies. The GTM can also be used for estimating physical parameters such as porosity, permeability and EUR estimations from the training parameters. For this reason, GTM not only can be used for seismic facies classification but also applied into modern risk analysis and used in decision making.

In this dissertation, I implement and apply Kohonen SOM and GTM to real 3D multi-attribute seismic datasets. I also introduce supervision to SOM and GTM using well control, resulting in a hybrid algorithm that finds user-defined clusters and uses the data alone to generate the remaining clusters.

Further, with the supervised GTM I output the probabilistic estimation of a seismic facies which can be included in the modern risk analysis workflow.

REFERENCES

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, 10, No. 1, 215-234.

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

Matos, M. C., P. L. Osorio, and P. Johann, 2007, Unsupervised seismic facies analysis using wavelet transform and self-organizing maps: Geophysics, v. 72, p. P9-P21

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

CHAPTER 2

INTRODUCTION

Classification is the arrangement of a dataset according to their observed similarities. Designing of a perfect classifier is often impossible; designing the probability of occurrence for each category make more sense (Duda and Hart, 2001). A seismic facies volume represents the variability in lithology, rock-properties and/or other physical properties of the reservoir rocks, and is an important part of seismic exploration workflow. Originally, highly skilled interpreters generated the seismic facies map. However, with the ever-increasing 3D seismic data volumes and the requirement to have fast and accurate results, we need to automate the process of seismic facies classification.

K-means and PCA were some of the earliest attempts at computer assisted seismic facies analysis. Supervised neural networks based on the back-propagation method have been proven much more effective allowing the computer to imitate the interpreter. The back-propagation method is a supervised classification that requires vectors of input attributes and a suite of user defined seismic facies target vectors (picks). Such supervised analysis in seismic add user insight and/or bias to the classification. When the geology is unknown or when there are very few wells in a large seismic survey, a supervised classification may reveal only a subset of the natural classes present in the data. In this situation, unsupervised classification techniques should be utilized.

DISSERTATION STRUCTURE

This dissertation consists of seven chapters, including motivation/objectives (chapter 1), introduction (chapter 2) and a conclusion (chapter 7). Chapters 3-6 are in the form of scientific papers and have been presented at international conferences or submitted to peer review journals.

Chapter 3 applies Kohonen SOM on a deep-water 3D seismic survey in the northern Gulf of Mexico. The objective of the paper is to show how this unsupervised seismic facies analysis can be used to distinguish different deep-water depositional features including basin floor fans, mass transport complexes, debris flows and feeder channels. Proper selection of the input seismic attribute volumes is the key to effective classification differentiate the different depositional features. After the SOM training is complete, the trained prototype vectors are color-coded by using a 2D gradational colorscale. Those traces with similar seismic nature are assigned the same colors, resulting in a 3D seismic facies volume. This chapter was published and presented in the Gulf Coast Section SEPM Annual Convention and Exhibition in Houston in 2011.

Chapter 4 applies SOM unsupervised workflow to characterize heterogeneous Mississippian chert reservoirs in Osage County Oklahoma. The heterogeneity in the Mississippian cherty limestone is formed by different stages of diagenesis of the Mississippian limestone. There are two major types of chert: tight/non-porous chert and the highly porous tripolitic chert. The dense chert and the dense limestone facies can be highly fractured, giving rise to two additional seismic facies. I evaluate two workflows – one based on structural attributes (coherency, dip magnitude, reflector convergence and coherent energy) and a second with the texture, frequency and amplitude attributes (GLCM energy, GLCM entropy, GLCM homogeneity, spectral bandwidth and coherent energy). A third workflow based on supervised facies classification was also done where the facies similarity was based upon the minimum Euclidean distance measures (MED). Three average wavelets, based on the borehole image logs from two horizontal wells in the survey, formed the target vectors in this supervised analysis. Further the results are compared with the post-stack P-impedance results within the chert reservoir. The works from this chapter has been submitted to “Interpretation” jointly published by the SEG and AAPG.

Chapter 5 is an introduction to the generative topographic mapping (GTM) algorithm, which takes care of the limitations of the Kohonen SOM process. This is more of a tutorial paper on GTM. In the first application GTM technique is applied to classify 15 sets of horizontal well parameters in an unconventional shale survey, correlating the results with normalized EURs, allowing an estimation of EUR based on the most relevant parameters. In addition, I have re-coded GTM to handle a much larger seismic dataset as input and generated a new workflow for unsupervised seismic facies analysis. I have considered multiattribute seismic inversion volumes (P-impedance, lambda-rho and mu-rho) from a Barnett shale survey as input to the GTM algorithm. The output GTM projections volumes are cross-plotted in interpretation software. By this method, the posterior probability values of the data are projected onto the 2D latent space forming different clusters. These clusters are interpreted and different user-defined lithofacies are defined. These lithofacies within the Barnett Shale are first correlated with the microseismic events at the well locations. Based on the concentration of the microseismic events the lithofacies are interpreted as brittle or ductile. Four zones within the Barnett Shale formation are identified and the seismic in each of the zones are then colored consistent with the different facies types interpreted from the microseismic events. This study does a 3D regional mapping of the "fracable" and ductile shale. The works from this chapter will submitted in the "Geophysics" journal.

Chapter 6 applies the unsupervised multi-attribute GTM algorithm to a carbonate conglomerate oil field in the Veracruz Basin of southern Mexico. The multi-attribute analysis is done with the different seismic inversion volumes (P-impedance, lambda-rho ($\mu\rho$), mu-rho ($\lambda\rho$) and Vp/Vs). After GTM modeling the dataset is projected as a mean posterior probability on a 2D latent space. Different clusters formed in the 2D latent space are color-coded by the user to identify different seismic facies present in the dataset. The average EURs are plotted with the unsupervised seismic facies volume to classify the good reservoir and the bad reservoir facies. In the next stage the average multi-attribute waveforms are calculated from three different wells in the survey and a PDF is created for each of the three well data vector. Then each of these three PDF is compared with all the data vectors to quantitative estimate probability of occurrence of a well vector within the reservoir units. This study helps to understand the reservoir compartmentalization, which will provide insight to the variable production. The probabilistic facies distribution from supervised GTM analysis can be used to study the geologic risk analysis related to this reservoir. The woks from this chapter will be submitted to “Interpretation” jointly published by the SEG and AAPG.

Each of the above papers is followed by the appropriate references. Chapter 7 includes my conclusions and recommendations based on my research. The appendix includes the documentation for running geometric attribute as a batch mode, 2D SOM, 3D SOM and the GTM at the, allowing the reader to replicate my unsupervised and supervised classification workflows using my software.

CHAPTER 3

Application of 3D clustering analysis for deep marine seismic facies classification – an example from deep water northern Gulf of Mexico

Atish Roy ^*, *Marcilio M. Matos*# *and Kurt J. Marfurt*^
^ *The University of Oklahoma*
#Independent Consultant

ABSTRACT

The most popular seismic attributes fall into three broad categories – those that are sensitive to lateral changes in waveform and structure such as coherence and curvature, those sensitive to thin bed tuning and stratigraphy, such as spectral components, and those sensitive to lithology and fluid properties – such as AVO and impedance inversion. We present a workflow that mimics multiattribute clustering routinely done by human interpreters that can differentiate depositional packages characterized by subtle changes in the stratigraphic column as well as lateral changes in texture. The best input attributes are those that are mathematically independent and rotationally invariant sensitive to the seismic facies of interest.

Supervised and artificial neural networks form the popular clustering algorithms in the seismic industry. Among the artificial neural network techniques Kohonen self-organizing maps form a popular clustering technique. The clusters in unsupervised

data analysis are defined by the data themselves, without any *a priori* information. In supervised training, a subset of clusters is pre-defined by the interpreter. The input data volumes are compared to these clusters; some are assigned to the predefined clusters, while others form clusters outside the area of supervision. The self-organizing map (SOM) is one of the most effective unsupervised pattern recognition techniques, because it identifies natural clusters in the dataset and the SOM grid points are topographically attached.

To avoid guessing at the number of clusters necessary to represent the data, We have over-defined the number of initial clusters through the use of a large number of "prototype vectors" (PVs), which after subsequent iterations tend to converge to the lesser number of "natural clusters" using a Kohonen SOM neighborhood training rule. After the training is complete the modified PVs are then color-coded by using a 2D gradational colorscale. Those traces with similar seismic nature are assigned the same colors, resulting in a 3D seismic facies volume. Calibration is done *a posteriori* by co-rendering the colored PVs with the original seismic amplitude data, input attributes, and well control. We apply this clustering workflow to a deep-water 3D seismic survey in the northern Gulf of Mexico and calibrate it to a previous interpretation made using traditional methods.

INTRODUCTION

Seismic facies are groups of different seismic reflections based on the amplitude, frequency, reflection, geometry and reflection continuity. In the past, seismic facies mapping and classification required time-consuming manual interpretation by a skilled interpreter. The interpreter examines successive vertical sections through seismic volumes to determine the seismic character, which is then correlated to a seismic facies using well data. Given large 3D seismic data volumes, there is a need for robust automated seismic facies classification algorithms which can help segment the large data volume into seismic facies components. Through visualization the interpreter links different architectural elements through the understanding of depositional processes, resulting in an integrated image of the depositional environment and reservoir heterogeneity.

Our work is built upon that done by many others who have worked on volumetric seismic facies mapping. Coleou *et al.* (2003) showed how by treating seismic amplitude samples within a vertical analysis window as “attributes”, one could generate a 2D map of seismic facies having similar waveforms. Each waveform was plotted against a 1D latent space and a corresponding 1D color bar giving rise to the popular algorithm known as “waveform classification”. A latent space is a lower-dimensional manifold embedded in attribute space that approximately contains the vast majority of the probability mass (Wallet *et al.*, 2009). In an ideal situation, virtually all the waveforms (or alteratively, suite of attributes at a given voxel) should be mapped onto a 1D, 2D or 3D subspace. Strecker and Uden (2002) were perhaps the first to use 2D latent spaces with geophysical data, using multidimensional attribute volumes to form N-dimensional vectors at each voxel. 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) used GLCM texture attributes, principal component analysis and SOM using a 1D latent space to map seismic facies offshore Angola. Using supervised artificial neural networks rather than unsupervised classification using SOM to cluster GLCM texture attributes, Corradi *et al.*, (2009) mapped sand and sealing vs. non-sealing shale facies offshore west Africa. There are many different supervised and unsupervised clustering algorithms. K-means clustering works well where the data are noise free and have a relatively high dimensionality (Coleou *et al*, 2003). Statistically, seismic data have high continuity, and therefore a relatively low dimensionality. Coleou *et al.*, (2003) found that the addition of noise further violates the assumptions of K-means clustering, at least for trace-shape analysis.

Another commonly used commercial product is based on Kohonen self-organizing maps (SOM). Coleou *et al.*, (2003) were among the first to apply Kohonen's SOMs to seismic waveform classification using a 1D latent space. More recently Matos *et al.* (2009) and Roy and Marfurt (2010) built on work by Strecker and Uden (2002) to show the advantage of extending the SOM grid space to 2D and

3D, with corresponding 2D and 3D colorbars used to delineate the subsurface depositional environment. In this study, we show the value of automated seismic facies classification to the study of a shallow sedimentary column within a salt minibasin in the Northern Gulf of Mexico.

The Kohonen Self-Organizing Map

Kohonen (1982) presented his clustering and dimensionality reduction methods as a “self-organizing process” whereby a “simple network of adaptive physical elements” is made to resonate in a particular way with externally provided signals (a “primary event space”) and tried to link these ideas with the functionality of the human brain (Murtagh, 1995). This popular neural network method was initially used in biology and computer science for data mining purposes, has since been used in speech recognition, and automated pattern-recognition in seismic exploration.

SOM (Kohonen, 2001) clusters data such that the statistical relationship between multidimensional data is converted into a much lower dimensional SOM grid space that preserves the geometrical relationship among the data points. Mathematically each SOM unit within the SOM grid space preserves the metric relationships and topologies of the multidimensional input data. SOM prototype vectors (PVs) or neurons have the same dimension as the input data (e.g. 12 dimensions for 12 input attribute values). These PVs are arranged in a regular low-dimensional grid or map (2D or 3D for our work), thereby topologically connecting it to its neighbors. By construction, SOM preserves the original topological structure within this *N*-dimensional attribute space, making it amenable for seismic facies analysis.

SOM Clustering Analysis

SOM is closely related to vector quantization methods (Haykin, 1999). Initially We assume that the input are represented by *J* vectors in a *N*-dimensional vector space R^n, $\mathbf{x}_j = [x_{j1}, x_{j2}, x_{j3} \ldots\ldots x_{jN}]$ where *N* is the number of input attributes (or amplitude samples for “waveform” classification) and $j=1,2,\ldots,J$ is the number of vectors analyzed. The objective of the algorithm is to organize the dataset of input seismic attributes into a geometric structure called the SOM. SOM consists of neurons or prototype vectors (PVs) organized by a lower-dimension grid, usually 2D, which are representative of the input data that lies in the same *N*-dimensional space as the input seismic attributes. PVs are also termed as SOM units and typically arranged in 2D hexagonal or rectangular structure maps that preserve the neighborhood relationship among the PVs. In this manner PVs close to each other are associated with input seismic attribute vectors that are similar to each other. The number of these PVs in the 2D map determines the effectiveness and generalization of the algorithm.

Let’s consider a 2D SOM represented by *P* prototype vectors $\mathbf{m}_i$, $\mathbf{m}_i = [m_{i1,} m_{i2} \ldots\ldots m_{iN}]$, where $i=1, 2, \ldots, P$ and *N* is the dimension of these vectors defined by the number of input attributes (or samples for waveform classification). The 2D SOM can be understood as a 2D sheet upon which the interconnected PVs are imbedded. After assigning the initial location of the PVs to form either a hexagonal (Figure 3.1B) or rectangular grid, we “train” them (i.e., changing their direction and moving them around laterally in their SOM grid space) to best represent the input data. After SOM neighborhood training these PVs become cluster centers representing the different classes of the input dataset.

To illustrate this somewhat complicated algorithm let us imagine a fanciful example from Matos et al. (2009). Figure 3.2A show some fruits that have different properties. Let us consider two of their properties: their aspect ratio (fruit shape), peak frequency (fruit color), and vitamin C content of each fruit. After training, the fruits are arranged into three major groups as shown in Figure 3.2B. Now if we draw an analogy with a 2D SOM grid map, the PVs represent the fruits, the shape and the color of the fruits are the attributes of the PVs and these different attributes helps to arrange them into different groups or clusters. Figure 3.1A shows 300 samples belonging to three distinct Gaussian distributions having the same standard deviation but different means. These 300 input vectors have three attributes or dimensions (N=3). Each sample vector in 3D space is cross-correlated with itself and all other vectors, resulting in a 3 by 3 covariance matrix. To begin, we choose three times the square root of each of the first two eigenvalues (three standard deviations) of this covariance matrix to define our initial 2D SOM grid space, which we sample with an 11x7 regular hexagonal grid (Figure 3.1B). This definition results in the SOM grid space representing approximately 99.7% of the input dataset. Thus the total PVs or the number of over-defined classes is 77 (P=77). Each of these individual PVs is denoted by a vector $\mathbf{m_i}$ of dimensionality 3 (the same dimensionality as the input data vectors).

The weight of the PVs associated with each lattice element i lying inside the neighborhood radius $\sigma(t)$ of the 2D map is updated with successive training.

During the SOM training process, an input vector is initialized and is compared with all *N*-dimensional PVs on the 2D grid, or latent space. The prototype vector with the best match (the winning PV) will be updated as a part of SOM neighborhood training. A small neighborhood of PVs around the "winner" prototype vector are also updated (Figure 3.3A) such that the PVs lying within a distance of "$\sigma(t)$" is excited and updated. With successive iteration, this neighborhood radius $\sigma(t)$ decreases. Thus, in each iteration, the winning prototype vector is brought closer to the input vectors.

Given this background, Kohonen (2001) defines the SOM training algorithm using the following five steps:

Step 1: Consider an input vector, which is randomly chosen from the set of input vectors.

Step 2: Compute the Euclidean distance between this vector $\boldsymbol{x}$ and all PVs $\mathbf{m}_{\boldsymbol{i}}$,*i=1, 2,…p*. The prototype vector $\boldsymbol{m_b}$, which has the minimum distance to the input vector $\mathbf{x}$, is defined to be the "winner" or the Best Matching Unit, $\mathbf{m}_{\boldsymbol{b}}$:

$$\left|\left|\mathbf{x}-\mathbf{m}_{\boldsymbol{b}}\right|\right| = \boldsymbol{MIN}\left\{\left|\left|\mathbf{x}-\mathbf{m}_{\boldsymbol{i}}\right|\right|\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

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

$$\mathbf{m}_{\boldsymbol{i}}(t+1) = \mathbf{m}_{\boldsymbol{i}}(t) + \alpha(t)h_{bi}(t)[\mathbf{x}-\mathbf{m}_{\boldsymbol{i}}(t)] \quad if \ \left|\left|\mathbf{r}_{\boldsymbol{i}}-\mathbf{r}_{\boldsymbol{b}}\right|\right| \leq \sigma(t) \qquad \text{(2a)}$$

$$= \mathbf{m}_{\boldsymbol{i}}(t) \qquad if \ \left|\left|\mathbf{r}_{\boldsymbol{i}}-\mathbf{r}_{\boldsymbol{b}}\right|\right| > \sigma(t), \quad \text{(2b)}$$

where the neighborhood radius defined as $\sigma(t)$ is predefined for a problem and decreases with each iteration. t. $\mathbf{r}_{\boldsymbol{b}}$ and $\mathbf{r}_{\boldsymbol{i}}$ are the position vectors of the winner PV $\mathbf{m}_{\boldsymbol{b}}$ and the i^{th} PV $\mathbf{m}_{\boldsymbol{i}}$ respectively. We also define $h_{bi}(t)$ as the neighborhood function, $\alpha(t)$ as the exponential learning function and T as the length of training. $h_{bi}(t)$ and $\alpha(t)$ decrease with each iteration in the learning process and they are defined as

$$h_{bi}(t) = e^{-(||\mathbf{r}_b - \mathbf{r}_i||^2/2\sigma^2(t)} \quad \text{, and} \quad \text{..............................} \; (3)$$

$$\alpha(t) = \alpha_0 (\frac{0.005}{\alpha_0})^{t/T} \quad . \quad \text{..............................} \; (4)$$

The training of the PVs for the above dataset is illustrated in Figure 3.3. The green hexagon represents a neighborhood $\sigma(t)$ at an early iteration, while the smaller blue hexagon represents a neighborhood at a later stage. Although 77 prototype vectors are used, after training they group themselves into three separate clusters in the 2D latent space.

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 PVs in the SOM grid space) is reached.

Step 5: Color-code the trained PVs using 2D or 3D gradational colors (Matos *et al.* 2009). We will use an HSV model with for 2D spaces will be defined as hue, $\mathcal{H}$,

$$\mathcal{H} = tan^{-1}\left(\frac{v-{}^{1}/_{2}}{u-{}^{1}/_{2}}\right) \quad \text{..............................} \; (5)$$

and saturation, $\mathcal{S}$, as

$$\mathcal{S} = \left[\left(u - {}^{1}\!/_{2}\right)^{2} + \left(v - {}^{1}\!/_{2}\right)^{2}\right]^{1/2} \text{ ....} (6)$$

where u and v are the projected components onto the 2D SOM grid space defined by the eigenvectors $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$. The input data in this example visually contain three separate classes which gives rise to three separate clusters in this 2D SOM grid space (Figure 3.4A). These new set of PVs are colored using the 2D HSV color palette with equations 5 and 6 (Figure 3.4B). This coloring scheme will be used for generating the seismic facies volumes discussed in the remainder of the paper.

APPLICATION

Application of SOM on synthetic data

To verify the effectiveness of the above workflow for waveform classification, we apply it to the 3D synthetic model shown in Figure 3.5. The 3D volume is composed of four different "waveforms" shown in Figure 3.5C. These 16-sample waveforms can represent seismic amplitude samples or impedance values. The location of the different waveforms A, B, C and D are highlighted both on the map view and crossline view of the synthetic model (Figures 3.5A and B). The data are sampled at 2ms, and consist of 50 inline by 26 crossline 30ms (16 sample) long traces.

The 2D SOM grid space containing the prototype vectors is defined as before by considering three standard deviations of the variability (square root of the eigenvalues λ_1 and λ_2) along the two principal component directions (eigenvectors $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$). Thus this SOM grid space represents approximately 99.7% of the input dataset. We sample this SOM grid space using a regular grid of more than 300, 16-dimensional prototype vectors similar to that in Figure 3.3A. The minimum Euclidean distance between the 16-dimensional input data with the prototype vectors, defines a winner for each data vector as part of the SOM training algorithm (Steps 2 and 3).

Although more than 300 prototype vectors were used, the resulting facies map clusters into four different colors corresponding to the four different waveforms in the input dataset (Figure 3.6). The waveforms are superposed on the 2D map (Figure 3.7B) to show how the different colors are assigned to the different waveforms. The synthetic seismic is colored based on the similarity of the waveform (Figure 3.7A). The waveforms with their respective colors are explicitly shown in Figure 3.7C.

Application of SOM to generate 3D seismic facies volume

Our 3D SOM algorithm is an extension of the 2D SOM algorithm described above. The input consists of several mathematically independent volumetric attributes where the number of input attributes determines the mathematical dimensionality of the data. Due to the limitation of our visualization software which provides only 256 colors, we have limited our over-defined prototype vectors to a maximum of J=256. In this application, we normalize our input data vectors using a Z-score algorithm. Thus our input data has a vector assigned to each of the (x, y, z) location in our volume (which are actually the normalized input attribute values at that location). We call this new volume the normalized multi-attribute volume and project it onto a 2D SOM grid space by Principal Component Analysis. The 2D SOM grid space is defined as explained earlier. If there are five input attribute volumes, each of the PVs in the 2D SOM grid space is 5-dimensional. This 2D SOM grid space is sampled uniformly by 256 PVs. The PVs are trained in the 2D SOM grid space and their positions updated after each iteration, resulting in the new updated position of the PVs. When the updating slows down the training process stops. With an increasing number of iterations, the PVs move closer to each other and to the data points within the SOM grid space. The HSV colors are assigned to the PVs according to their distance from their center of mass and their azimuth (equations 5 and 6). Once trained, the distance is computed between each PV, $\mathbf{p_i}'$, and the multiattribute data vector, $\mathbf{x}$, at each voxel using

$$\left|\left|\mathbf{x} - \mathbf{p}'_{\boldsymbol{b}}\right|\right| = \boldsymbol{min}\left\{\left|\left|\mathbf{x} - \mathbf{p}'_{\boldsymbol{i}}\right|\right|\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

where $\mathbf{p}_{\boldsymbol{b}}{}'$ is the nearest trained PV to the input data sample vector $\mathbf{x}$. Each voxel is then assigned the color of $\mathbf{p}'_{\boldsymbol{b}}$. In this manner, two dissimilar neighboring samples in the seismic volume will be far apart in the SOM grid space and have different colors. Conversely, two similar samples in the seismic volume will have nearly the same color. Each color represents a seismic facies, most of which are geologic facies, but some which may be seismic ‘noise’ facies.

SOM on a 3D seismic dataset from the Northern Gulf of Mexico

Our study area is within the tabular salt mini-basin, tectono-stratigraphic province (Diegel *et al*, 1995) of the Northern Gulf of Mexico (GOM) margin (Figure 3.8A). There are primarily four depositional environments for deep water systems: channels, thin bed-levees, basin floor fan lobes or sheets and mass transport complexes (MTCs) (Figure 3.8C). The sediments in the salt minibasins follow a “fill and spill” depositional process (e.g., Winker, 1996) preserving the basin floor fan and channel-levee sediments.

A 100 km^2 prestack depth-migrated 3D seismic data volume (Figure 3.8D) covers almost this entire salt minibasin. Sarkar *et al.*, (2010) describes a suite of different mass transport complexes (MTCs) preserving the basin floor fan in the system. They co-render strata-slices of coherent energy and inline amplitude gradient to highlight the basin floor fans and their feeder channels.

To illustrate the effectiveness of this new workflow, we consider zones identified by Ferrero, (2007) and Sarkar *et al.*’s (2010), which highlights the MTCs and Basin floor fan (BFF). They did the interpretation with the well logs and sparse biostratigraphic data. We considered four zones for analysis, highlighted in an arbitrary vertical section across the survey (Figure 3.9). Zone 1 is the deepest MTC with a less chaotic flow A. Above it zone 2 has the second MTC flow B which is more erosive in nature and is more related to salt tectonics. This mass transport complex contains some small slump bocks along with some debrites and undifferentiated MTC deposits. The MTC deposits are also characterized by the presence of a buried channel-submarine fan unit within the chaotic sediments. The depositional patterns of the upper and the lower MTCs are quite random (Sarkar *et al.*, 2010). Overlying this is the zone 3 having strong continuous seismic amplitude, which forms the basin floor fan sequence. Zone 4 is the uppermost zone in our analysis, which is a lower energy MTC and is a basin wide depositional feature. Some of the submarine channel systems are also identified in this zone.

The input to our clustering algorithm consists of five volumetric attributes, which help discriminate different seismic facies distribution: the coherent energy, eigenstructure coherence, two GLCM texture attributes (entropy and variance) and the magnitude of reflector convergence. Table 1 summarizes the seismic characteristics in different deep-water depositional environment, which determined the subsequent choice of seismic attributes. Note that these attributes are not sensitive to the direction of the survey or rotation of the sediments with compaction. We note that attributes such as phase and azimuth are inappropriate for this cluster analysis since they are mathematically discontinuous. These five attribute values are first normalized using a z-score algorithm for each sample to remove any bias prior to clustering with the Kohonen SOM algorithm.

DISCUSSIONS

Table 1, summarizes different seismic amplitude and attribute patterns associated with different deep water deposits. Figure 3.10 shows an arbitrary vertical section through the facies volume co-rendered with the seismic amplitude section. The interpretation is done in a top-down sequence and the zones from 1-4 are interpreted independently. The continuous high amplitude depositional features are colored light to dark blue. The random chaotic MTCs are colored in magenta. The background marine pelagic shale is colored in lighter yellowish green color. Figures 3.11A to E show different phantom horizon slices through the data volumes and highlights different depositional features. The location of the phantom horizon slice is marked in the seismic section below each of the figures.

Table 1. Summary of 3D attribute assisted cluster analysis for deep-water deposits

Deep water deposits	**Seismic amplitude pattern and internal configuration**	**Attribute anomalies**	**Facies color in our analysis**
Mass Transport Complexes	Moderate to high amplitude, discontinuous, chaotic, hummocky, rotated blocks	Low coherence, high energy. Rotated blocks and pinch out patterns resulting in high reflector convergence. Irregular bed thickness gives rise to non-uniform tuning frequencies. High values of GLCM entropy and variance.	Different hues of blue and cyan

Basin Floor Fans	High amplitude, continuous, isolated or connected features within the fan. Sub-parallel reflectors.	High coherence, high energy. Subparallel bedding resulting in low reflector convergence. Thicker bedding gives rise to lower tuning frequencies. Moderate values of GLCM entropy and variance.	Yellowish green and green
Marine Pelagic Shale	Moderate to low amplitude, continuous, very thin and separated from the MTC	Moderate coherence, lower energy. Subparallel bedding resulting in low reflector convergence. Thin bedding gives rise to higher tuning frequencies. Moderate values of GLCM entropy and variance.	Pink to cyan

Figure 3.11A is a phantom horizon slice through the MTC complex in zone 1. The blue color facies highlights the relatively continuous flow A. It flows from the northeast and the more chaotic, incoherent flow B (magenta facies) encroaches the flow A from the north and south. Figure 3.11B is another phantom slice, which highlights the chaotic flow B from zone 2. The flow is mostly from the northeast marked by a yellow dotted line. The Figure 3.11C shows some confined high amplitude sand bodies (dark blue) with strong seismic amplitude, which may be gas

charged. The flow B being chaotic has more characteristic of a MTC caused by slope failure due to regional salt-tectonics.

The basin floor fan sequence identified in Ferrero, (2008) and Sarkar *et al*, 2010 is properly highlighted with this facies classification volume. Figure 3.11D shows the complex depositional features causing a fan deposit (blue color facies) in zone 3. The lighter green color forms a thin deposit of marine pelagic background shale. A stratal slice above this basin fan sequence in Figure 3.11E highlights number of channels which feeds the fan deposits. As we go above the zone other basin floor fan can interpreted across the survey. The continuous high amplitude in the west and in the east of the survey is interpreted as fan deposits (Figure 3.11F). Some feeder channels are also highlighted in this zone.

The last zone analyzed is the zone 4, which is also a MTC but is more regional in nature. Figure 3.12 shows a stratal slice in the middle of the zone. The younger submarine channel systems are interpreted in the north east of the survey. It forms a distributary system as this channel (highlighted in blue arrow) moves south and forms fan deposits at the mouth. The edges of this feeder channel system are defined by the magenta color highlighted in the zone 4 across the vertical seismic section AA'.

Figures 3.13a and b show the geobodies formed corresponding to the basin floor fan. Although we are using a 2D SOM grid space, plotted against hue and lightness to color SOM unsupervised seismic volume, most commercial seismic interpretation packages allow only a 1D color bar. To the strata shown in Figure 3.11E, we apply transparency to the desired blue facies, which is easily achieved using the 1D opacity table below where the blue facies are set to be opaque and all others to be transparent. Figure 3.13A shows the volume probe with only the blue facies. Automatic geobodies can be extracted in most of the commercial software. With this opaque blue facies an automatic extraction of geobodies forms the different geo-blob (Figure 3.13B). However in order to properly generate a geobody, we need to apply opacity directly to the colorbar in two dimensions by defining a polygon, painting, or some other process.

CONCLUSIONS

Seismic stratigraphy forms an integral part of many if not most seismic interpretation workflows. Seismic attributes provide a means of quantitatively measuring amplitude and morphologic features such as onlap, coherence, and reflection strength. SOM is a well-established tool in seismic facies mapping, typically applied to a window of seismic amplitude data about a given horizon, giving rise to “waveform classification” algorithms. Waveform classifications have several disadvantages. Waveforms are not sensitive to the behavior of adjacent traces, unlike attributes such as dip magnitude and coherence. Waveforms are also quite sensitive to the source wavelet and thus cannot readily be applied to stratal slices. Our research is in very early stages, and we do not claim to have used the ‘best attributes’ for our analysis. Other attributes such as impedance inversion at different angles, spectral components, and anisotropy may further help us discriminate between different seismic packages. Barnes and Laughlin (2002) found using mathematically independent attributes is critical to classification. We feel intuitively that the attributes should be azimuthally invariant unless there are strong reasons to prefer given directions.

The linkage of seismic facies classification to the generation and extraction of geobodies is obvious, but difficult given today’s commercial visualization software based on 1D colorbars rather than 2D or 3D color tables. The use of pastel vs. dark colors allows co-rendering the seismic facies classifications with seismic amplitude on vertical slices and coherence and other edge-sensitive attributes on time and horizon slices plotted against a gray scale.

This initial application is unsupervised. In this workflow, the results should be validated against well logs and conventional seismic stratigraphic interpretation. Supervision can be introduced by explicitly defining, and then fixing prototype vectors to represent interpreter-defined clusters, either of a classic facies seen on a vertical slice, or of a less well-defined facies centered about a producing well or dry hole. More sophisticated supervision will form a continuum between SOM as shown here and traditional applied neural network technologies, blurring the differences between the two methods.

By using a large number of PVs and colors, we avoid the need to know the number of clusters in the data. Instead, the interpreter lumps PVs of similar colors into clusters and associates them with a particular facies as part of the interpretation process. As with multiattribute blending, 24-bit color will provide much greater depth and detail in images than the current 8-bit color limitations on many interpretation workstations. A limitation to this 2D color mapping is the folding of the PVs when plotted in the 2D SOM grid space, such that multiple PVs of different nature can have the same color. A partial solution is to project the data in a 3D SOM grid space and apply a 3D gradational colorbar to the PVs.

Our initial attempts at supervision are limited to predefining a suite of prototype vectors formed by averaging the input data vectors within a user-defined subvolume.

ACKOWLEDGEMENTS

We acknowledge Hess Inc. for the use of its deepwater GOM data volume for research and Schlumberger for providing the Petrel software, in which interpretation and visualization was done. The interpretation was done based on earlier studies by Ferrero in her MS thesis and discussions with Dr. Roger M. Slatt of OU. We would also like to thank the financial support of the industry members for our Attribute-Assisted Seismic Processing and Interpretation (AASPI) consortium at the University of Oklahoma.

LIST OF FIGURES

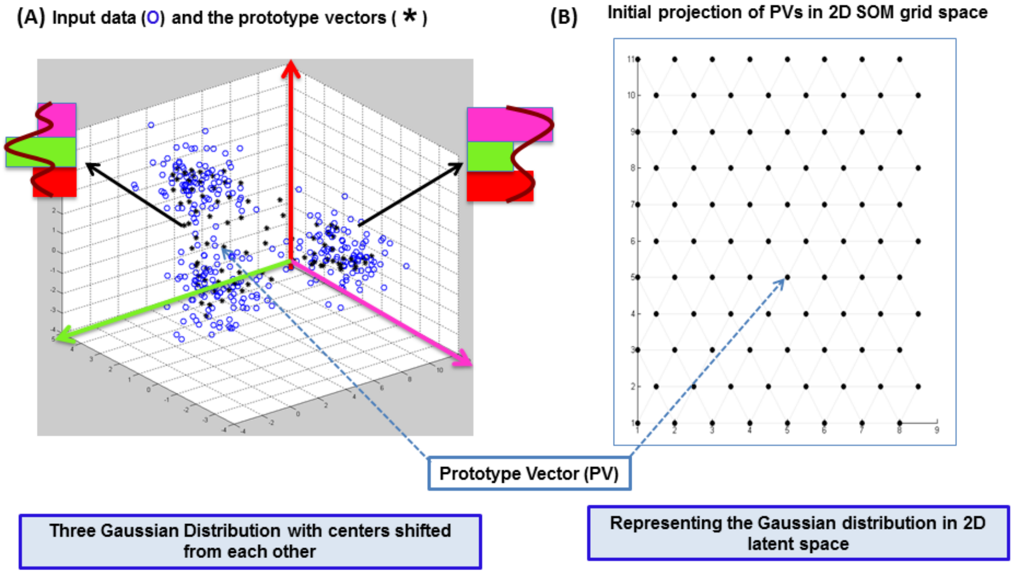


Figure 3.1. (A) A plot of samples corresponding to three normal Gaussian distributions with the same standard deviation but different means. The red plus signs represents the original data points and the blue dots the Prototype Vectors in 3D; (B) the initial PVs with the same dimensionality as the input is plotted in the latent space by Principal Component Analysis.

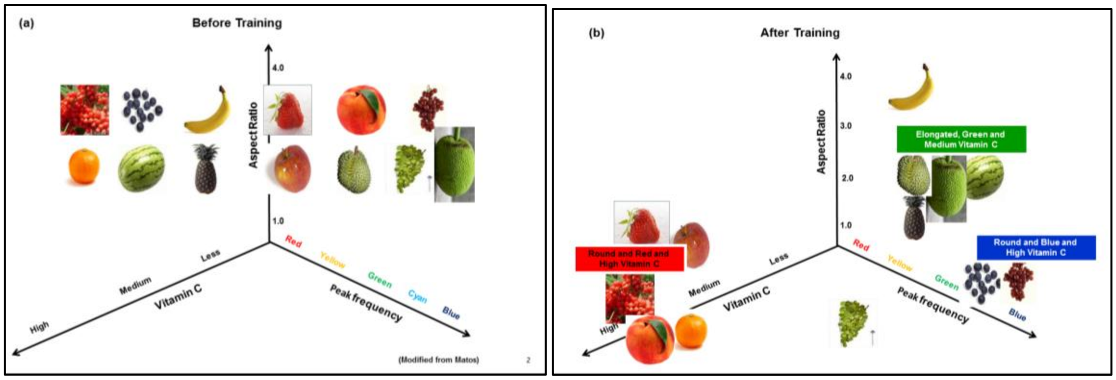


Figure 3.2. Analogy of classification of different fruits illustrating unsupervised SOM clustering analysis. **(a)** The unorganized fruits before training; **(b)** Clustering of the fruits after training using three attributes (color, aspect ratio and Vitamin C content of the individual fruits).

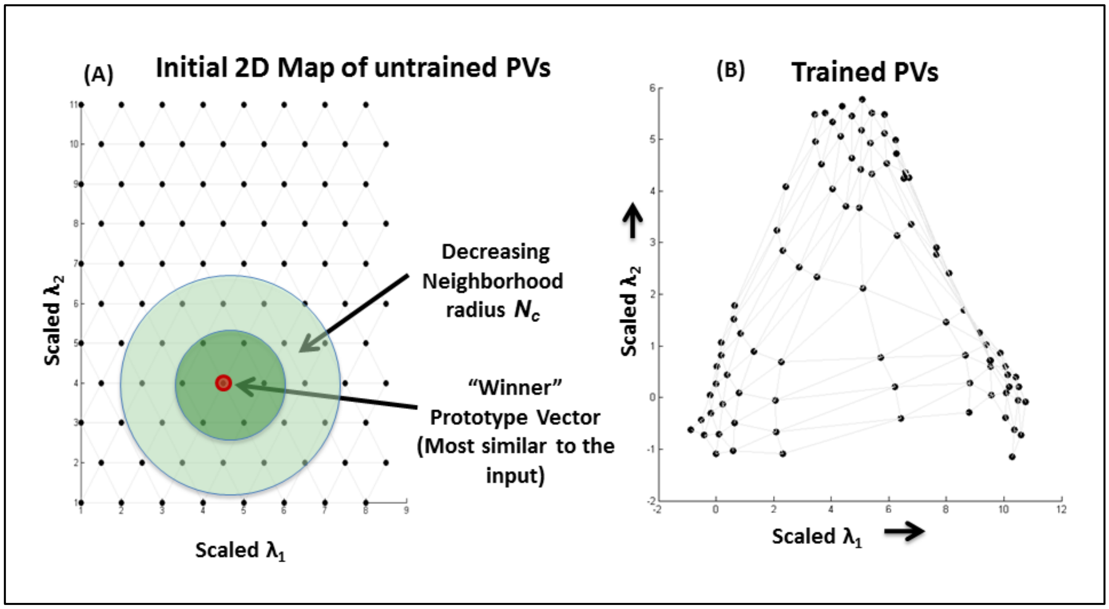


Figure 3.3. 2D map of the PVs in latent space (A) before and (B) after the training. Around the “Winner PV” the larger green circle represents a neighborhood at an early iteration while the smaller green circle represents a neighborhood at a later iteration. After the training the PVs form three clusters, thus classifying the three different types of input data properly.

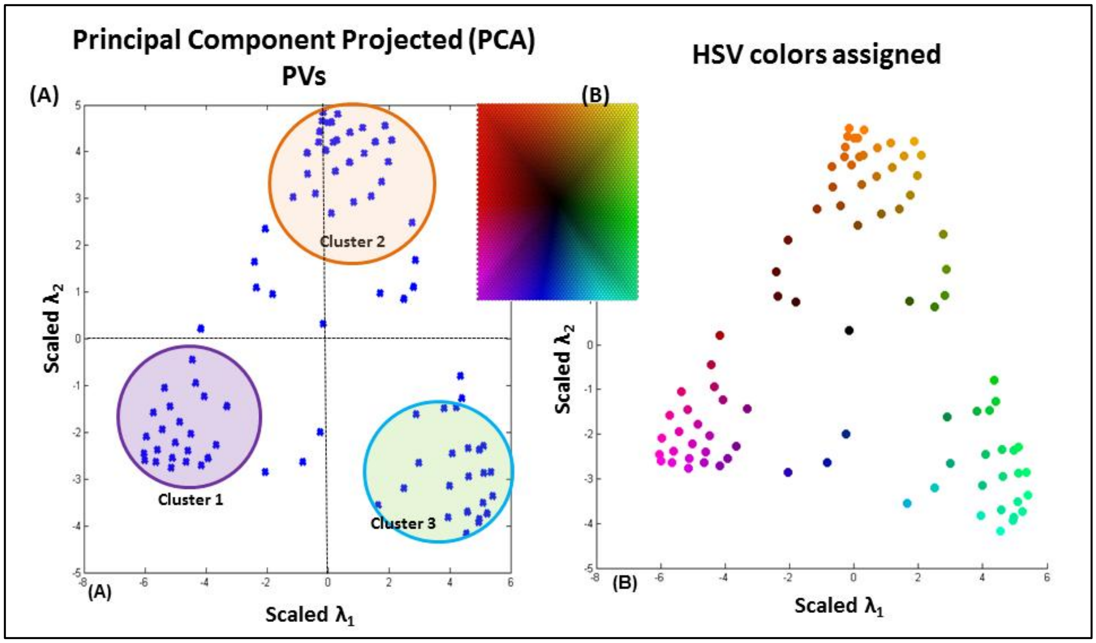


Figure 3.4. Interpreter-driven clustering by mapping a 2D latent space against a 2D colorbar. (A) PVs in the 2D latent space before assigning colors, (B) colored by the 2D HSV color table using the Equations 5 and 6 for Hue and Saturation.

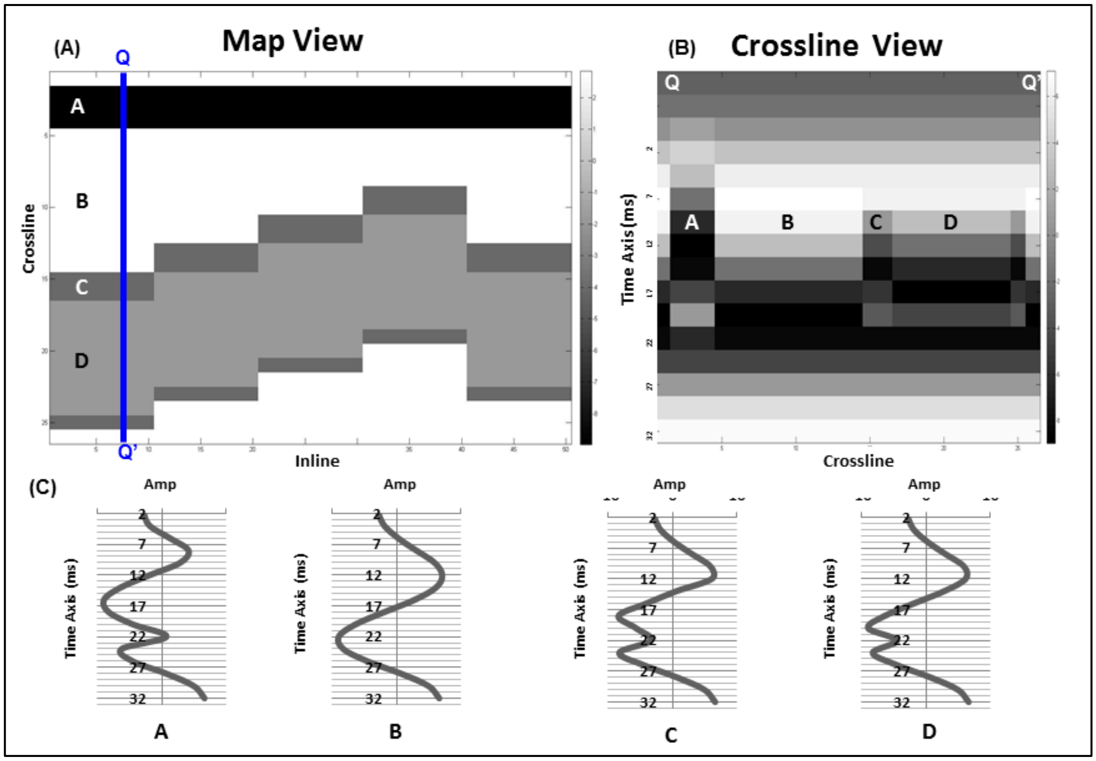


Figure 3.5. A synthetic data volume 30ms (16 – 2ms samples) having four different waveforms consisting of 50 inlines by 26 crosslines. Noise level is 0.1 of the RMS amplitude of the signal. (A) Map view and (B) Vertical slice along line QQ'. (C) The four noise-free waveforms used to generate the synthetic volume.

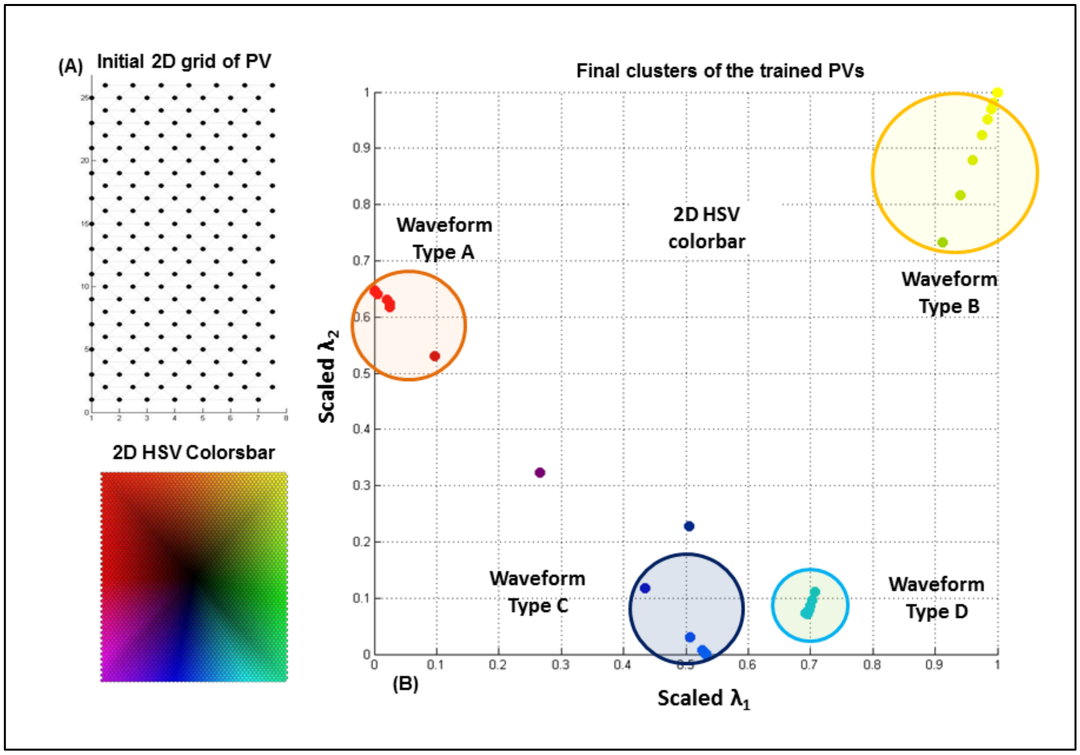


Figure 3.6. (A) Initial set of PVs. **(B)** Training and HSV coloring of the projected Prototype Vectors (PVs) in latent space. Although more than 300 prototype vectors were used, the resulting facies map clusters into four different colors corresponding to the four different waveforms in the input dataset. Arrows indicate the location of each of the noise-free synthetic waveforms A, B, C, and D. Almost all of the clusters fall on top of each other. Scatter away from the clusters such as the dark blue ‘cluster’ are associated with stretch in the latent space. After clustering, the PVs are colored with the HSV colorscale (left) described by Equations 5 and 6.

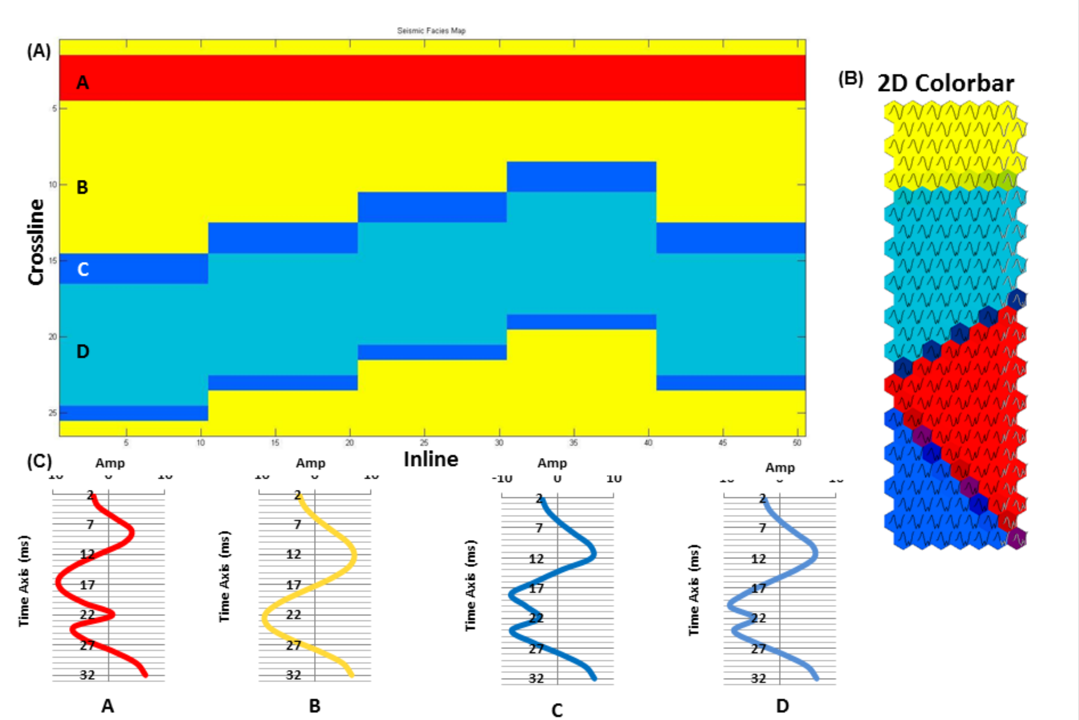
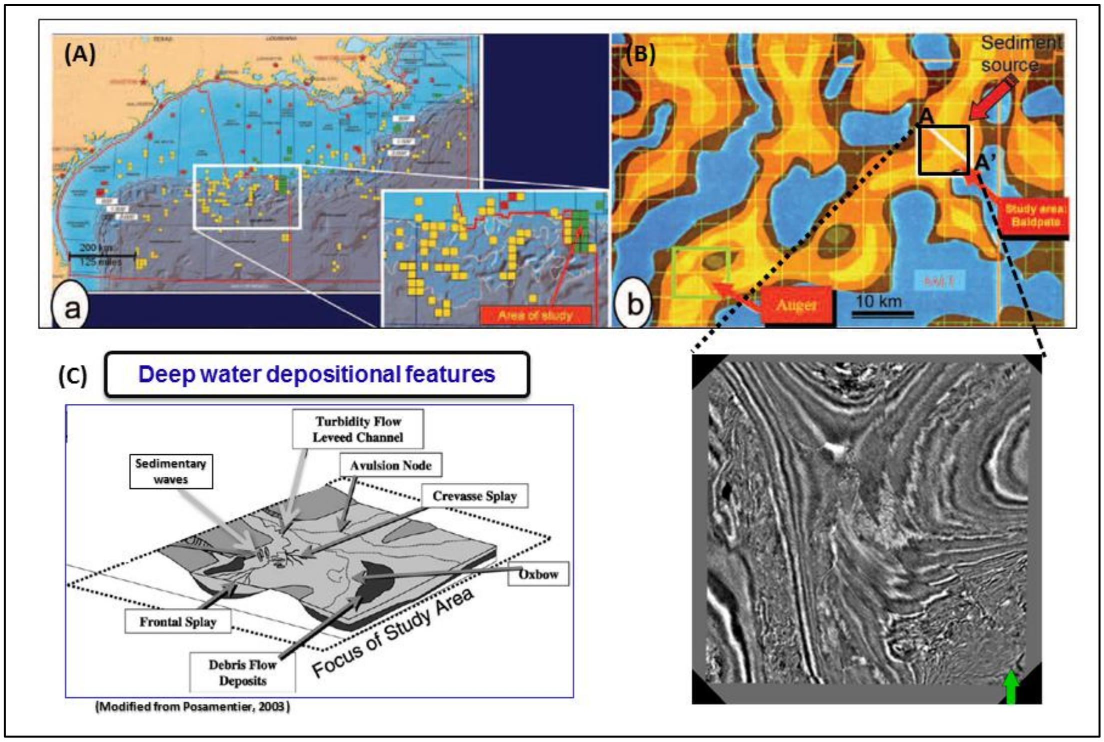


Figure 3.7. (A) The seismic facies map showing four different colors corresponding to the four different waveforms that constitute the synthetic dataset. **(B)** The waveforms are superposed on the 2D colormap showing the color assignment to the different waveforms. **(C)** These four different waveforms are explicitly shown with corresponding colors assigned to it. Note the transitional waveforms along the boundaries which do not represent any data, but provide a continuum in the latent space. Rather than choosing the final number of clusters (in this case, four) as described by Matos et al. (2007), we simply let the interpreter visually cluster according to their color perception (Matos et al, 2009).



(Modified from Posamentier, 2003)

Figure 3.8. (A) The location of the study area within the northern salt minibasin province of Gulf of Mexico (Okuma *et al*, 2000). **(B)** Regional sedimentary fairway map with the survey location. (C) Some of the deep-water depositional features highlighted from Posamentier *et al*, 2003.

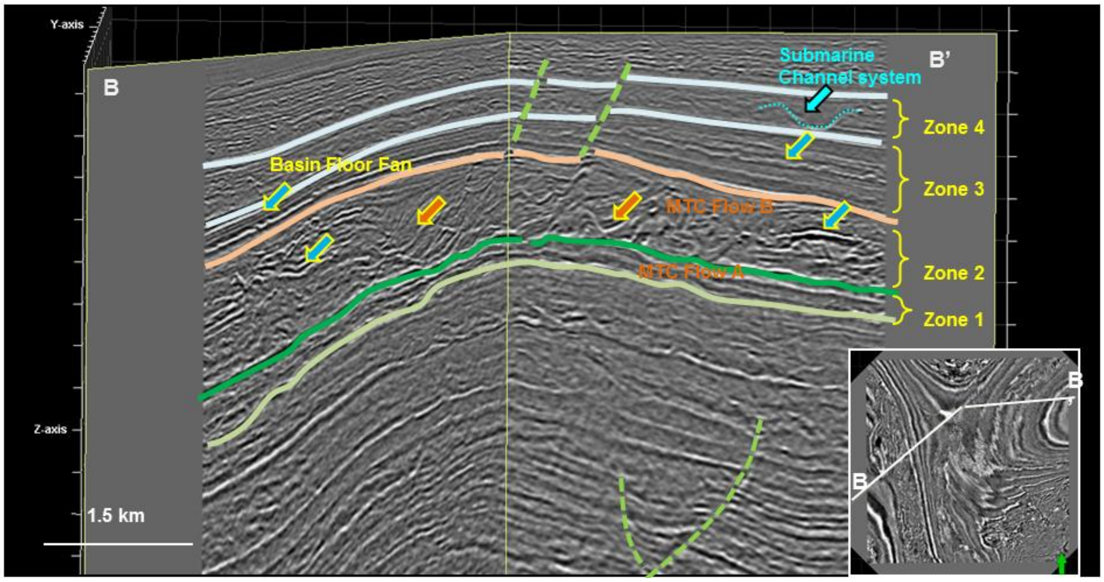


Figure 3.9. An arbitrary seismic section across the survey showing the different zones (identified by Ferrero, 2007) used for analysis with the SOM clustering. Zone 1 has Mass transport complex (MTC) flow-A. A more chaotic MTC flow-B is in zone 2. A Basin floor fan sequence is highlighted in the zone 3. Zone 4 is also a regional MTC with some submarine channel complex visible.

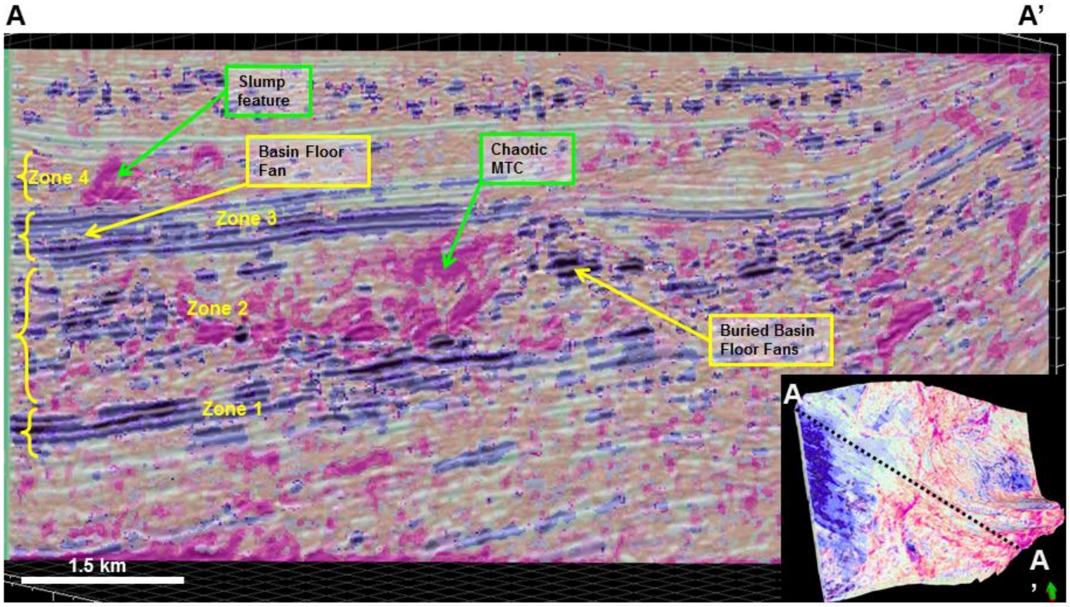
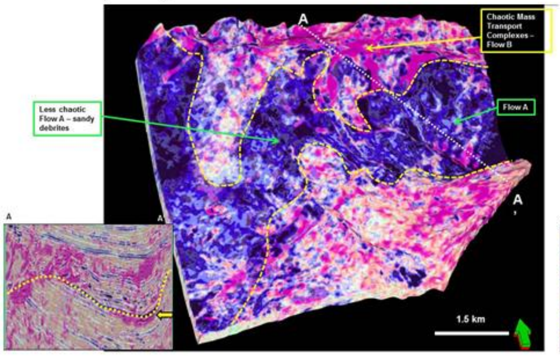
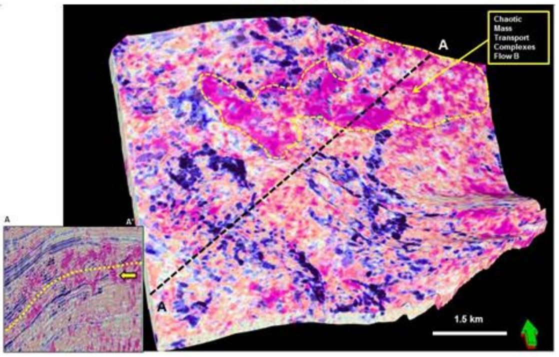


Figure 3.10. An arbitrary unsupervised seismic facies section co-rendered with the seismic amplitude section across AA'. The four zones are analyzed independently. The MTC flow in the zone 1 is more continuous (blue facies). The zone 2 consists of chaotic MTC (magenta), shale (yellowish green) and buried sand bodies (dark blue facies). Zone 3 is a basin floor sequence formed by a complex set of distributaries. Zone 4 is another regional MTC with some submarine channel complex visible in the north east corner of the survey.

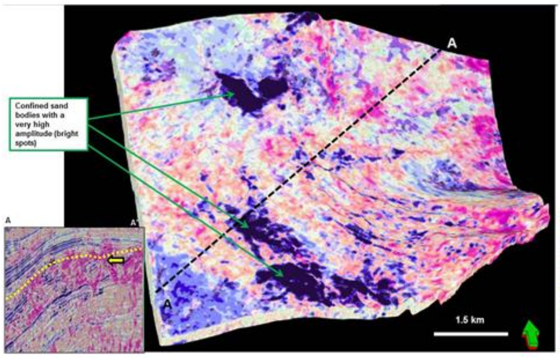
(A) Strata slices within the volume probe of MTC flow A (zone 1) – Erosive and Chaotic MTC Flow B



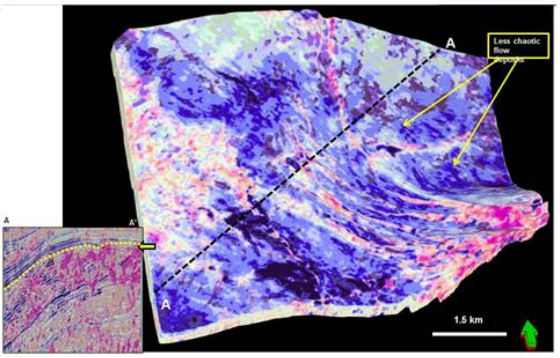
(B) Strata slices within the volume probe of analysis zone 2



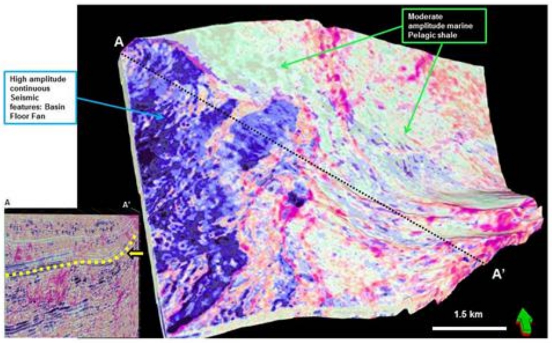
(C) Stratal Slice within volume probe of flow B (zone 2)- Confined sand bodies



(D) Stratal Slice within volume probe of flow B (zone 2) – later stage of Flow B: less chaotic



(E) Stratal Slice within volume probe of Zone3



(F) Stratal Slice within volume probe of Zone3

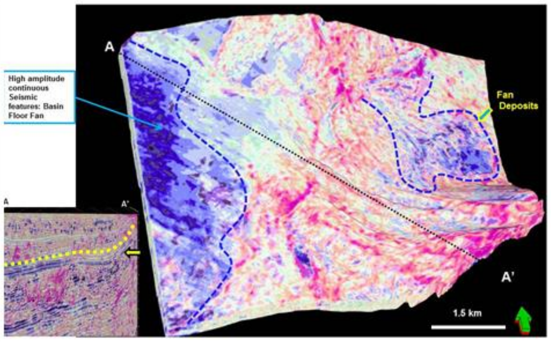


Figure 3.11. Unsupervised seismic facies horizon probe showing different stratal slices in the four different zones starting from the lowest sequence. In the below corner the location of the stratal slice is marked on the vertical section. **(A)** The strata in zone 1 showing a relatively continuous Flow-A (blue facies) from north east of debris and MTC being encroached by a more chaotic flow-B (magenta) in the north and south. The flow-B is more erosive in nature and is more related to slope failure to salt tectonics. **(B)** The strata in zone 2 within the chaotic MTC flow-B. The chaotic nature of the MTC can be seen in this zone. **(C)** The chaotic MTC in zone 2 caused due to slope failure and salt tectonics have eroded large sand bodies within itself as shown by the dark blue colored facies. With their confined nature and high seismic amplitude these sand bodies may have hydrocarbon charged. **(D)** Higher up in the zone there is relatively some less chaotic MTC flow. **(E)** In the zone 3 the basin floor fan sequence can be properly visualized from this facies volume. The strata shows a complex set of distributaries and basin fans in the west (blue facies) with pale green background of pelagic shale. **(F)** Moving up in the zone 3 the later deposited basin floor fan sequence and other set of distributaries are interpreted in the east.

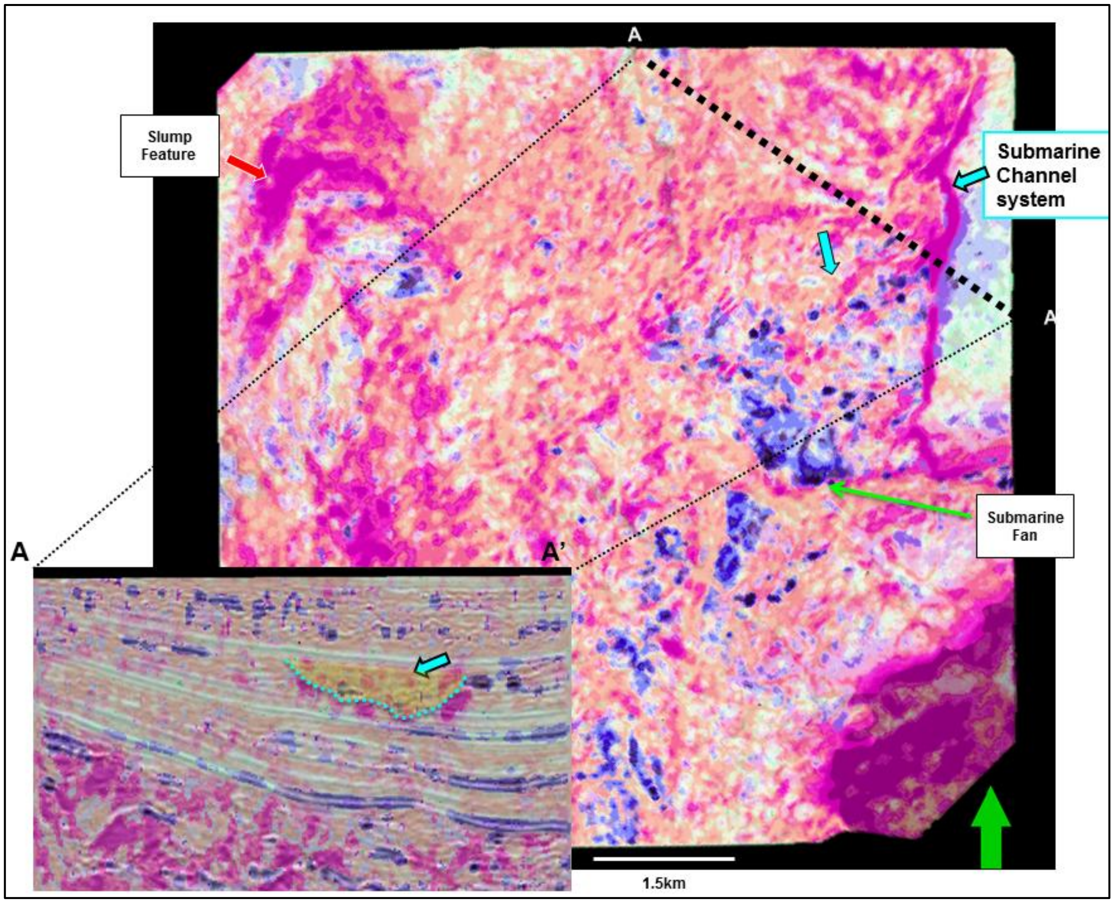
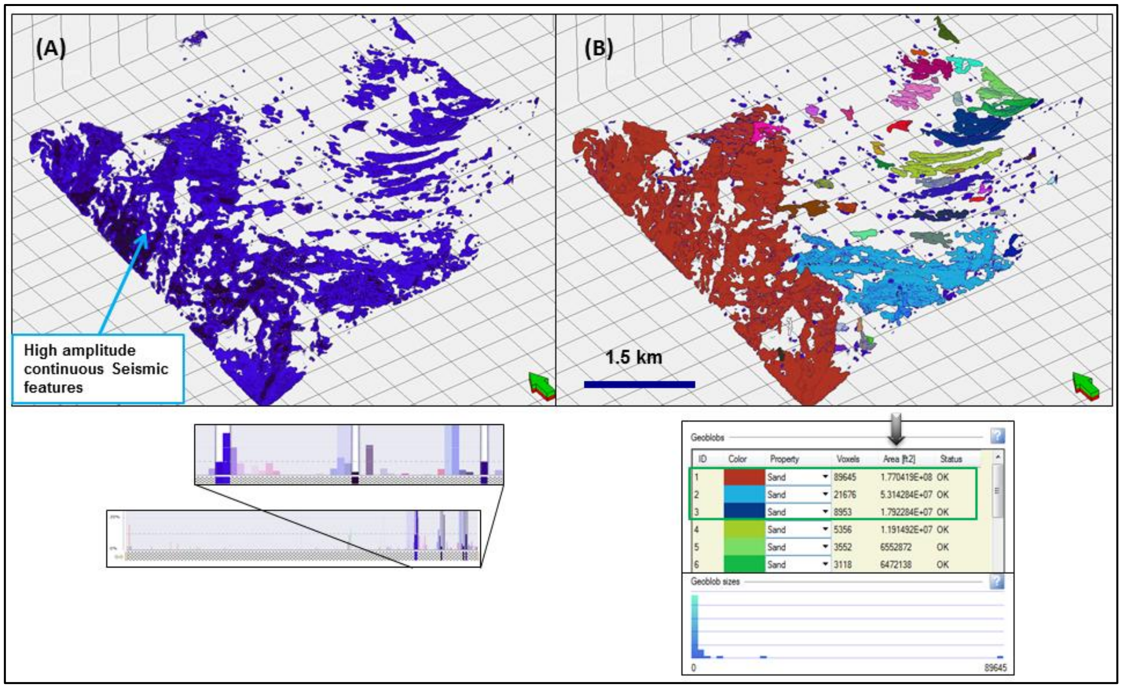


Figure 3.12. The zone 4 forms another MTC complex but is more regional in nature. A younger feeder channel in the north east corner can be easily interpreted in this strata from the unsupervised classified volume. AA' is a vertical section of the unsupervised volume overlayed on the seismic amplitude. The feeder distributary channel complex is highlighted in this vertical section with a blue arrow.



ID	Color	Property	Voxels	Area [ft2]	Status
1	🟥	Sand	89645	1.770419E+08	OK
2	🟦	Sand	21676	5.314284E+07	OK
3	🟦	Sand	8953	1.792284E+07	OK
4	🟩	Sand	5356	1.191492E+07	OK
5	🟩	Sand	3552	6552872	OK
6	🟩	Sand	3118	6472138	OK

Figure 3.13. Extracting geobodies from the unsupervised classified volume. Although coloring of the unsupervised classified volume was done in the SOM grid space using a 2D colorbar, the coloring in most of the interpretation software is done by a 1D colorscale. Modifying the transparency we can highlight certain seismic facies. The blue facies from the Figure 3.11E showing mostly the BFF and other similar depositional seismic facies. **(A)** The blue facies is made opaque and the rest transparent in the 1D colorbar shown below. **(B)** With this single colored facies an automatic geobody extraction is done to create different geobodies which can be quantified as highlighted in the table.

RFERENCES

Barnes, A. E., and K. J. Laughlin, , 2002, Investigation of methods for unsupervised classification of seismic data: 73rd Annual International Meeting Society of Exploration Geophysicists, Expanded Abstracts, p. 2221-2224.

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

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

Diegel, F. A., J. F. Karlo, D. C. Schuster, R. C. Shoup, and P.R. Tauves, 1995, Cenozoic structural evolution and tectono-stratigraphic framework of the northern Gulf coast continental margin: AAPG Memoir, v. 65, p. 109-151.

Ferrero, B.H., 2007, An integrated 3D seismic sequence-stratigraphic analysis of the Pleistocene strata above Baldpate Field, Garden Banks, deepwater Gulf of Mexico, MS thesis The University of Oklahoma.

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, v. 91, p. 1665-1683.

Haykin, S., 1999, Neural Networks: A Comprehensive Foundation, Second edition, Prentice Hall.

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

Kohonen,T., 2001, Self-organizingSelf-organizing Maps, 3rd ed.: Springer- Verlag.

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.

Matos, M. C., P. L. Osorio, and P. Johann, 2007, Unsupervised seismic facies analysis using wavelet transform and self-organizing maps: Geophysics, v. 72, p. P9-P21.

Murtagh, F., and M. H. Pajares, 1995, The Kohonen self-organizing map method: An assessment: Journal of Classification , v. 12, p. 165-190.

Roy, A., and K. J. Marfurt, 2010, Applying self-organizing maps of multiattributes, an example from the Red-Fork Formation, Anadarko Basin: 81st Annual International Meeting Society of Exploration Geophysicists, Expanded Abstracts, p. 1591-1595.

Sarkar, S., K. J. Marfurt., and R. Slatt, , 2010, Generation of sea-level curves from depositional pattern as seen through seismic attributes-seismic geomorphology analysis of an MTC-rich shallow sediment column, northern Gulf of Mexico, The Leading Edge, v. 29, p. 1084- 1091.

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

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

Winker, C. D., 1996, High resolution seismic stratigraphy of a late Pleistocene submarine fan ponded by salt-withdrawal min-basins on the Gulf of Mexico Continental slope: Proceedings from the 1996 Offshore Technology Conference, paper OTC 8024, p.619-628.

CHAPTER 4

Characterizing a Mississippian Tripolitic Chert Reservoir using 3D Unsupervised and Supervised Multi-attribute Seismic Facies Analysis: An example from Osage County, Oklahoma

*Atish Roy *, Benjamin L. Dowell and Kurt J. Marfurt, The University of Oklahoma*

ABSTRACT

Seismic stratigraphy is based on reflector configuration, with coherent reflectors having a distinct amplitude, frequency, and phase. Typically, reflector configurations described as parallel, converging, truncated, hummocky, will be interpreted within a sequence stratigraphic framework. Skilled interpreters use their expertise to identify stratigraphic packages separated by erosional unconformities, maximum flooding surfaces, and other surfaces. In principal, a given pattern can be explicitly defined as a combination of waveform and reflector configuration properties, although such "clustering" is often done subconsciously.

Computer-assisted classification of seismic attribute volumes builds on the same concepts. Seismic attributes quantify characteristics not only of the seismic reflection events, but also measure aspects of reflector configurations.

The Mississippi Lime resource play of northern Oklahoma and southern Kansas provides a particularly challenging problem. Here the seismic facies are diagenetic (tight limestone, stratified limestone and non-porous chert, and highly porous tripolitic chert) or structural (fractured vs. not-fractured chert and limestone) rather than stratigraphic. Skilled interpreters can recognize the different facies even though they may not able to accurately define the exact diagenetic or tectonic history. During the past several years, over 4700 wells have been drilled in the Mississippi Lime play. Such dense drilling provides means of calibrating the computer-generated "natural" clusters that represent the data. Using a 3D seismic survey acquired in Osage County Oklahoma, we use Kohonen self-organizing maps (SOM) classification technique to represent different diagenetically altered facies of the Mississippi Lime play. The 256 prototype vectors (potential clusters) reduce to only three or four distinct "natural" clusters. We use ground truth of seismic facies seen on horizontal image logs to fix three average attribute data-vectors near the well locations, resulting in three "known" facies, and do a minimum Euclidean distance (MED) supervised classification. The predicted clusters correlate well to the post-stack impedance inversion result.

INTRODUCTION

Seismic interpreters are experts at pattern recognition and as fast as an eight-year old in finding Waldo (Handford, 2012). Older, experienced interpreters are usually more proficient at such pattern recognition than younger, perhaps more technically savvy interpreters. Nevertheless, 3D seismic surveys are not only growing larger, but also are being acquired by small companies and even smaller partnerships in resource plays like the Mississippi Lime of Oklahoma and Kansas, companies that often do not have seismic stratigraphers on their staff. Our workflow will help seismic interpreters with an easy and quick way to estimate the variation of seismic facies in the survey with the help of various multiattribute data as input.

Seismic attributes quantify not only reflector amplitude, frequency, and phase but also through estimates of dip magnitude, dip azimuth, reflector convergence, reflector rotation, and coherence, they also quantify reflector configurations, which forms the basis of seismic stratigraphy. These physically independent (but statistically correlating) attribute volumes add a "dimension" to our analysis. Two or three attributes (corresponding to two or three dimensions) are effectively analyzed by interactive crossplot tool. The analyses of more than three attributes require a different workflow along with higher computational complexity.

There are competing techniques for coping with the classification of data with excessive dimensionality. One of the approaches is to reduce the data dimension by combining features. Principal component analysis (PCA) is a classical approach to reduce dimensionality and provides an orderly suite of projections that best represents the data in a least-squares sense. We use Kohonen SOM classification technique to classify our multi-attribute dataset. Before SOM training, we need to reduce the data dimension into a low dimensional space through PCA analysis of the data. This lower dimensional space represents the majority probability mass of the data (Wallet et al., 2009). The vector components of the latent space are hidden and the SOM algorithm clusters these vectors with information from the data-vectors.

Barnes and Laughlin (2002) reviewed several unsupervised clustering 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 the clustering algorithms, they favored SOM since there is ordered mapping of the clusters (topologically ordered); with similar clusters lying adjacent to each other in the latent space. Coleou et al.'s (2003) seismic "waveform classification" algorithm is implemented using SOM, where the "attributes" are seismic amplitudes on a suite of 16 phantom horizon slices, where the mean in 16-dimensional space when plotted one element after the other can be envisioned to be a waveform. Plotting each cluster against a 1D color bar results in a 2D map of seismic facies having similar waveforms. They generalize 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 latent spaces with geophysical data, using multidimensional attribute volumes to form *N*-dimensional vectors at each voxel. 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) clustered GLCM texture attributes based on their Euclidean distance in the texture attribute space and 1D SOM to map seismic facies offshore Angola. He used 256 prototype vectors 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. (2011) built on these concepts and developed a classification workflow of multi-seismic attributes computed over a deep-water depositional system, using SOM. 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 non-linear Sammon space (Sammon, 1969). This projection was then colored using a gradational 2D colorscale like that of Matos, (2008) thus facilitating the interpretation.

The other classification technique is supervised classification. When there is significant well control or when the interpreter has significant expertise, one can perform supervised classification. The most popular means of supervised classification are based on artificial neural networks (ANN). Meldahl et al. (1999) used seismic energy and coherence attributes coupled with interpreter control (picked locations) to train a neural network to identify hydrocarbon chimneys. West et al. (2002) used a similar workflow but where the objective was seismic facies analysis and the input attributes were textures. Corradi et al. (2009) used GLCM textures and ANN, with controls based on wells and skilled interpretation of some key 2D vertical slices and mapped sand and sealing vs. non-sealing shale facies offshore west Africa.

In this paper we utilize Kohonen self-organizing maps (SOM) for unsupervised classification of seismic facies. We further incorporate supervision by use of minimum Euclidean distance (MED) measure. We begin with an overview of SOM. Next, we introduce the Mississippi Lime play and how it is expressed by seismic attributes. We then cluster our data in an unsupervised manner, finding the natural clusters. We then extracted three average attribute data-vectors from two horizontal wells in our survey based on the fracture density calculated by White (2013) with the borehole image logs. This a posteriori supervision is simple – clusters (seismic facies) that correspond to those penetrated by the poor well are considered to be higher risk, and should be avoided, while seismic facies that correspond to those penetrated by the good well serve as higher priority targets. We conclude with a comparison of our clusters against a more conventional post-stack inversion volume.

METHODOLOGY

Kohonen’s (1982) original SOM algorithm is based on localized (neighborhood) training of prototype vectors (e.g a seismic waveform or a vector of attributes) on a 2D SOM grid. While the vectors can move to better represent the nearby data, the spatial relationship between the prototype vectors and its neighbors on the SOM grid is preserved, or topologically ordered. In our case, we will use multi-attribute data vectors as input into the SOM clustering algorithm.

To avoid guessing at the number of clusters necessary to represent the data, we over-define the number of initial clusters through the use of a large number of “prototype vectors” (PVs). As experienced by Gao (2007), subsequent iterations using a Kohonen SOM neighborhood training function results in the large number of PVs “clumping” into a smaller number of actual clusters that represent the true variability in the data. After the training is complete, the modified PVs are then color-coded by using a 2D gradational colorscale. Clumped prototype vectors have nearly identical colors. Each *N*-dimensional data vector of attributes is compared to these 256 trained prototype vectors. The data voxel is then assigned the color associated with the prototype vector that most closely matches the corresponding data vector, resulting in a 3D seismic facies volume.

Kohonen SOM Clustering Analysis

The SOM was first developed by Kohonen in the biological sciences, but is now commonly used in speech recognition, economics, and of course, geophysical data analysis. Excellent implementations of SOM can be found in Matlab and in at least two commercial interpretation packages (one of which was used by Coleou et al., 2002, and Gao, 2007). However, the commercial packages today appear to use a 1D latent space, while the Matlab implementations are not amenable to handling large 3D seismic data volumes. An early implementation described by Strecker and Uden (2002) does use a 2D latent space but may no longer be commercially available. We describe the algorithm as we have implemented it so that others can duplicate or improve upon our effort.

The Kohonen SOM algorithm assumes that the input is represented by J vectors in an N-dimensional vector space $\mathcal{R}^n$, $\mathbf{x}_j= [x_{j1}, x_{j2}, x_{j3} .... x_{jN}]$ where N is the number of input attributes (or amplitude samples for "waveform" classification) and $j=1,2,\ldots,J$ is the number of vectors (one vector per voxel in 3D, one vector for map location in 2D). The input attributes have different units of measurements, resulting in radically different numerical ranges. We chose a simpler approach of choosing physically independent attributes and computing their Z-score.

Our SOM implementation defines mapping from the $\mathcal{R}^n$, input data space to a 2D SOM grid space where the PVs assigned to each grid point are topologically ordered. The 2D SOM grid space can be understood as a 2D sheet upon which the interconnected imbedded PVs lie. The prototype vectors are represented by $\mathbf{m}_i$, $\mathbf{m}_i$= [m_{i1}, m_{i2}.... m_{iN}], where $i=1, 2, ..., P$ where P is the number of PVs. We project the PVs onto a rectangular structure map that preserves the neighborhood relationship among the PVs. A projected PV in the 2D SOM grid space occupies a single grid point. We initialize the PVs using PCA with the 2D SOM being defined by the first two eigenvectors with amplitudes ranging between $\pm 3\sqrt{\lambda_1}$ and $\pm 3\sqrt{\lambda_2}$ (three standard deviations of the variability in the data as defined by the eigenvalues λ_1 and λ_2. After training, these PVs deform and move out of the 2D plane to move closer to the data-vector such that the 2D SOM grid better represents the natural clusters present in the input data. After several iterations, the PVs continue to move (organize), clumping into subsets. PVs that are close in the SOM grid space will represent attribute vectors that are similar to each other. The number of these clumped PVs determines the effectiveness and generalization of the algorithm.

For example Figure 4.1a shows 300 samples belonging to three distinct Gaussian distributions having the same standard deviation but different means. These 300 input vectors have three attributes or dimensions (N=3). Each sample vector in 3D space is cross-correlated with itself and all other vectors, resulting in a 3 by 3 covariance matrix. To begin, we choose three times the square root of each of the first two eigenvalues (three standard deviations) of this covariance matrix to define our initial 2D SOM grid space, which we sample with an 11x7 regular hexagonal grid (Figure 4.1b). This definition results in the SOM grid space representing approximately 99.7% variance of the input dataset. There are 77 PVs, much more than needed to represent the three natural clusters. Each of these individual PVs is denoted by a vector $\mathbf{m_i}$ of dimensionality three (the same dimensionality as the input data vectors).

During the SOM training process, an input vector is initialized and is compared with all N-dimensional PVs. The prototype vector with the best match “winner” and its surrounding PVs will be updated, thereby “training” that neighborhood of the SOM. A Gaussian neighborhood function is defined with about the “winner” PV as its center and $\sigma(t)$ as its variance. With each subsequent iteration, the neighborhood radius $\sigma(t)$ (variance) decreases. Thus, in each iteration, the winning prototype vector is brought closer to the data vectors in the input data space while its corresponding node organizes (or clumps) into one of the clusters formed in the 2D SOM space. After 100 iterations, the node SOM grid space clumps into the 3 classes present in the input data (Figure 4.1c). The trained PVs are color-coded using 2D gradational colors (Matos *et al.* 2009). We will use an HSV model with for 2D spaces defined as hue, $\mathcal{H}$,

$$\mathcal{H} = tan^{-1}\left(\frac{v-{}^{1}\!/_{2}}{u-{}^{1}\!/_{2}}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

and saturation, $\mathcal{S}$, as

$$\mathcal{S} = [\left(u - {}^{1}\!/_{2}\right)^{2} + \left(v - {}^{1}\!/_{2}\right)^{2}]^{1/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where u and v are the projected components onto the 2D SOM grid space. The input data in this example visually contain three separate classes which gives rise to three separate clusters and they are color-coded using the 2D HSV color palette with equations 1 and 2 (Figure 4.1d). This coloring scheme will be used for generating the seismic facies volumes discussed in the remainder of the paper. A more mathematical explanation of all the above steps are given in Kohonen (2001) and Roy et al. (2011).

We start the training by first projecting this normalized multi-attribute data onto a 2D SOM grid space defined by using the eigenvalues and eigenvectors obtained from PCA. The SOM grid space is thus uniformly sampled with grid points that are projections of PVs having the same dimensions as the number of attribute volumes taken as input. Due to the limitation of our visualization software, which provides only 256 colors, we have limited our over-defined PVs to be 256. We apply the SOM training rule to cluster these vectors in the SOM grid space. The PVs are updated after each iteration and they slowly move towards the data-vector in the input space, resulting in an updated projection of the PVs onto the SOM grid space. As the updating slows down, the training process stops. The SOM manifold can be far from planar and can even unfold itself, linearly projecting the multi-dimensional PVs onto a 2D SOM grid space causes some overlap in the projections. Thus after the final iteration we do a non-linear projection of the PVs using a Sammon projection (Sammon, 1969). This algorithm is based upon a point mapping of a set of *N*-dimensional vectors to a lower dimensional space such that the inherent structure of the data is approximately preserved. Sammon mapping helps in reducing some of the overlap of the projection of the PVs (nodes), by maintaining inter-PV distance measures corresponding to inter-grid point distance measures in the 2D SOM grid space. We use a 2D HSV color model to assign continuous color to the PVs according to the distance from the center of mass and the azimuth of their projections. Once trained, the Euclidean distance is computed between each trained PV, $\mathbf{m}_i'$ and the multi-attribute input data vector, $\mathbf{x}_n$ at each voxel using

$$||\,\mathbf{x}_n - \mathbf{m}'_b\,|| = min\,\{||\,\mathbf{x}_n - \mathbf{m}'_i||\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

where, $\mathbf{m}_b'$ is the nearest trained PV to the input data sample vector $\mathbf{x}_n$. Each voxel in the 3D data space is then assigned the color corresponding to $\mathbf{m}_b'$. In this manner, two dissimilar neighboring samples in the seismic volume that are far apart in the SOM grid space will have different colors (Roy et al., 2011). Conversely, two similar samples in the seismic volume will have nearly the same color. Each color represents a seismic facies, most of which are geologic facies, but some which may be seismic 'noise' facies. Figure 4.2 shows a flowchart explaining the multi-attribute SOM process.

APPLICATION

The Mississippi "Lime" – Tight Limestone, Tight layered Chert and lime, Porous Tripolite, and Fractured chert

The general stratigraphic column (Figure 4.3a) shows the Mississippian tripolitic chert interval is below the Mississippian - Pennsylvanian unconformity. Other tripolite targets are within the Mississippian limestone. These weathered and/or detrital intervals of highly porous rock are present in north-central Oklahoma and south-central Kansas. The Mississippi Lime and tripolitic chert reservoirs have been producing hydrocarbons for more than 50 years. Although the tripolite is widespread across the region, it is not continuous and is highly heterogeneous (Rogers, 2001). Our dataset is from Osage County, Oklahoma, (Figure 4.3b) which sits within the Cherokee Platform which is bounded to the west by the Nemaha Uplift and to the east by the Ozark Uplift (Johnson, 2008).

Matson (2013) subdivides the Mississippian in this study area into the tight St. Joe limestone, and the Osage A and Osage B levels. Osage B consists of alternating thinly layered limestone and nonporous chert which does not undergo subsequent diagenetic alteration (but often shattered chert). Osage A can be diagenetically altered through meteoric processes and consists of siliceous limestone and high porosity tripolitic chert.

Rogers (2001) suggested most Mississippian tripolitic chert developed primarily from weathered or eroded Mississippian Lime that was deposited as muddy debris flows. Figures 4.4a and b show the two schematic depositional models for the chert and the formation of high-porosity and low-density chert.

In the reef model, the reef is eroded by wave action and material is deposited downslope as debris flows. Silica then replaces some of the limestone (stage-one diagenesis) to form high-density, low-porosity chert. Later, sea level drops and meteoric water dissolves much of the remaining calcite (stage-two diagenesis) to form low-density, high-porosity tripolite. In the breccia model, karst breccias formed in a subaerial environment are submerged and the silica replaces some of the limestone (stage-one diagenesis) to form chert. When the sea level drops the meteoric water dissolves the remaining calcite (stage-two diagenesis) to form tripolite.

The tripolitic chert is widespread, but unlike the limestone, it is discontinuous throughout the Mississippian section (Rogers, 2001). Much of the limestone has also been altered to a dense, non-porous chert that exhibits fractures due to shrinkage and subsequent tectonic activities. The tripolitic chert reservoirs are heterogeneous and have high-porosity and low-permeability forming sweet spots in the reservoir. The limestone and non-porous chert are highly fractured. Ideally, horizontal wells are drilled perpendicular to the fractures in the limestone and tight chert parts of the formation, hydraulically fractured and/or acidized. This provides the necessary plumbing to produce multiple sweet spots. The log responses of tripolitic chert zones show low-resistivity, low-density and high porosity (25-30%). With the proposed seismic facies analysis, we try to visualize and map the heterogeneous chert reservoirs to optimize the well locations to economically exploit the sweet spots.

Multiattribute seismic Facies Analysis

In our 3D SOM algorithm, the input consists of several physically independent volumetric attributes where the number of input attributes determines the intrinsic dimensionality of the data. In this application, we normalize our input data vectors using a Z-score algorithm. Thus our input data consists of a vector assigned at each voxel or (x, y, z) location in our 3D survey. We evaluate three different workflows to estimate seismic facies. The first workflow applies unsupervised Kohonen SOM analysis to structural attributes, while the second workflow applies the same unsupervised Kohonen SOM analysis to texture attributes. The third workflow we derived a supervised multi-attribute seismic facies analysis in which three average multi-attribute wavelets near the wells are compared to the multi-attribute dataset, based on the MED measure. The case studies based on the three workflows provide a qualitative analysis of the heterogeneous chert reservoir in the survey.

Attribute Selection and Unsupervised Classification

We evaluated two different hypotheses corresponding to two different sets of input volumetric attributes to our clustering algorithm. This is done in order to find the ‘best attribute’ and to do a proper analysis of the heterogeneous chert reservoir.

Structural Attributes

Our first hypothesis (workflow 1) assumes that the magnitude of reflector convergence, coherence, coherent energy, and dip magnitude (Figure 4.5) better map the structural heterogeneity of the chert layers. Coherent energy will bring out amplitude changes associated with tripolitic and tight limestone. Coherence will highlight the discontinuities within the reservoir. Dip magnitude will capture deformation, while reflector convergence will map unconformities. The shattered chert and the Tripolitic chert will be more discontinuous compared to the tight limestone regions. Using these attributes as input data to SOM algorithm will result in a “structural” classification that emphasizes differences in reflector orientation, continuity and configuration in the data volume. Figure 4.6 show different features in the 3D seismic facies volumes generated as output after the first analysis.

Texture Attributes

Our second hypothesis (workflow 2) assumes that the tripolitic chert, silica-rich limestone and the St. Joe tight limestone have different textures (Figure 4.7). Coherent energy, the spectral bandwidth computed from spectral decomposition (Zhang et al., 2008), and three GLCM (Gray level co-occurrence matrix) texture attributes are used in the analysis. A common everyday example of textures is the association of the grain seen on furniture and the tree from which it was made. Matos et al. (2011) used texture to identify Osage-age channel features in a different Osage Co. seismic survey. For GLCM analysis an ensemble of traces are examined as an image and the distribution of pixel values within a sub-region of data are described mathematically, effectively quantifying the spatial organization of seismic reflection (West et al., 2002). The GLCM energy is the sum of the squared values of the pixels defined this Grey Level Co-occurrence Matrix, the GLCM entropy is a statistical measure of randomness of the seismic amplitude and the GLCM homogeneity highlights regions having strict stationary statistics (invariant mean and variance). Dense cherty limestone and tight chert exhibit high coherent energy, thicker continuous reflectors, higher amplitude, more homogeneous and exhibit less entropy (Figure 4.7). Layered chert and lime will exhibit comparatively less coherent energy, thin reflectors, less homogeneity and more entropy. Tripolitic chert with high porosity and low-density, will have lower amplitude, be most discontinuous, least homogeneous and exhibit high entropy (be more disorganized) (Figure 4.7). We use these differences to classifying the dataset on the basis of amplitude and texture variations. Figures 4.8 and 4.9 show different features in the 3D seismic facies volumes generated as output after unsupervised classification.

Supervised Classification

Waveform similarity based on distance metrics have long been used in seismic interpretation. Typically the interpreter compares a vector of samples (e.g. Johnson, 2000) or attributes (e.g. Michelena et al., 1998) extracted from productive or non-productive wells to every trace along the horizon. Here we perform a similar exercise but do the comparison volumetrically. Furthermore if the MED between the vector at a given well lies beyond a user-defined threshold distance the voxel is assigned to an unclassified cluster. This workflow was introduced by Poupon et al. (1999) in correlating wells to seismic waveforms, where the supervision was not only the actual seismic about the well but also a suite of synthetic seismic traces generated through petrophysical modeling and fluid substitution.

The supervised seismic facies classification techniques described in this paper is based upon distance metrics similarity characteristics. These metrics compare two multi-attribute data vectors and return a scalar value based on some notion of similarity. We begin by training some of the dataset near the wells and turned their average into target vectors into target or training vectors. These are then compared with the rest of the data vectors. The MED is measured between the training vectors and the data vectors. This measure is utilized for providing quantitative measures of similarity between two seismic facies.

In our supervised classification workflow, three facies are selected based on the borehole image log interpretation from White, (2013) (Figure 4.10). Three small zones are taken, two along Well-A and one along Well-B to perform a supervised classification about the Mississippian chert reservoir zone (Figure 4.9). The texture attributes used in unsupervised workflow 2 are also utilized in the supervised classification. Three sets of average waveform are calculated from the three zones and are these training data-vectors called facies-1, facies-2, and facies-3 as shown in Figure 4.9.

Well A was drilled in the north east corner of the survey and encountered mostly the St. Joe limestone and dense cherty limestone. However the dense chert and the dense limestone facies can be highly fractured, giving rise to two additional seismic facies. It is observed from the fracture density calculations along the well bore (White, 2013) that there are some natural fractures in the dense chert of the Well-A (Figure 4.9a). Thus two zones are identified, one in the region of dense cherty limestone and the other in the region of dense chert with natural fractures (layered chert and limestone). The average data-vector of the facies-1 (light blue-violet color) is extracted from Well A around the vertical and horizontal zone, having the tight, least fractured cherty limestone. The average data-vector of facies-2 (yellow color) is extracted from the horizontal section of Well A, in which tight fractured chert are observed. The average data-vector of both facies-1 and facies-2 have similar nature with high homogeneity, high energy and low entropy (measure of disorder). However, these two facies type have different amplitudes for each attribute value as observed in Figures 4.10a and b.

Well B was drilled in the southern edge of the survey. It is a producing well and has been drilled mostly in the tripolitic chert formation. Facies-3 (light green color) is the average data-vector extracted from the horizontal section of the Well B that has more tripolitic chert facies present (Figure 4.9b). The average data-vector from facies-3 shows less coherent energy, lower bandwidth spectrum, higher disorder (entropy), low homogeneity and moderate energy (Figure 4.10c). These average data-vectors are compared to the same z-score normalized dataset. We measure the Euclidean distance between the facies vectors and each data-vector, where the most similar data-vector to the facies vector will have the least distance. The confidence values of the well data-vectors are fixed to 80%. The data-vector in the reservoir zone is identified and the corresponding data voxel is color-coded according to its closest facies type. A cutoff value is defined which controls the unclassified class facies-4 (gray color) that does not fall in either of the above three groups.

DISCUSSIONS

Although we started our analysis with an over-defined number of 256 PVs, the unsupervised SOM clumps (clusters) them into a smaller number of “natural clusters” present in the multi-attribute dataset. From Figures 4.6b and 4.8b we see that the projections of these PVs form four/five clusters after the final iteration.

The unsupervised seismic facies analysis from the structural attributes (workflow 1) helps to map the discontinuity and the overall distribution of rock types within the Mississippian chert reservoir (Figure 4.6a). The harder higher-density chert will have greater amplitude and a higher coherence compared to either the fractured or the low-density tripolitic cherts. Analyzing Well A, we find the magenta, pink and orange facies correlates to tight limestone and fractured tight chert (corresponds to St Joes Limestone and Osage B) as shown in Figure 4.6a. From a posteriori analysis of Well B we interpret the yellow and green colors correlates to the tripolitic chert rich facies (corresponding to Osage A formation) also shown in Figure 4.6a. In addition, the combined effects of the dip-magnitude, coherency and the reflector convergence attributes helps in distinguishing and the subtle variations, fractures, faults and karst features which corresponds to the blue and cyan color. Some of the structural features are concentrated in the fractured chert zones but most of the structural features are concentrated in the tripolitic chert rich zones

The second workflow using unsupervised seismic facies analysis from the texture attributes (workflow 2) appears to define a greater number of seismic facies in the survey area (Figure 4.8a). After the final iteration five major clusters are formed which are interpreted by analyzing Wells A and B. The dark red zones correspond to the dark red cluster formed in the outliers, interpreted as tight cherty limestone (St. Joe limestone). Figure 4.9a shows the horizontal well overlayed on the vertical section of the seismic facies volume. Overlaying the vertical section with the fracture density plot (White, 2013) we correlate, that the lighter red areas with high-density fractured chert rich or the layered chert and limestone. From Figure 4.9b we infer that the lighter colored (light green and yellow) facies are tripolitic chert rich zones (Osage A).

The supervised seismic facies volume (Figure 4.11) is quite similar to the previous unsupervised SOM classified volumes. The blue-violet facies corresponds to St Joe's limestone rich areas. The high-density fractured cherts corresponding to layered chert and limestone (Osage B formation) and corresponds to yellow facies (facies-2). The tripolitic chert rich areas corresponding to Osage A formation, (facies-3) and are in green. These green areas correspond to the diagenetically altered chert with low-density and high-porosity values. The tripolitic chert facies is not uniform across the survey and occur in patches. The remaining data those does not fall in either of the above three groups are color-coded gray.

Using Wells A and B, a post-stack P-impedance was done (Dowdell, 2012), and the results are compared with the multi-attribute unsupervised seismic facies analysis (Figure 4.12). High impedance (purple) corresponds to the regions interpreted in the facies volumes as the high-density chert or having St. Joe's limestone. Intermediate (green) impedance occurs in the regions having fractured non-porous chert or layered chert. Finally, low-impedance regions (yellow and red) correspond to the rocks of low density and high porosity and correspond to the tripolitic chert-rich zones (Figure 4.12a). The unsupervised seismic facies volume from the texture attributes (Workflow 2) is similar to Figure 4.12a. The zones interpreted as dense limestone correspond to the regions having high impedance values. Similarly, the low-impedance zones are interpreted as tripolitic chert-rich zones in the unsupervised facies volume from the "texture" attributes. In this case we can replace coherent energy in workflow 2 with acoustic impedance resulting in the improved classification. The impedance volume co-rendered with the k_1 principal curvature

(Figure 4.12b) suggests that there are more fractures in the low-impedance regions compared to the limestone-rich areas (White, 2013). The unsupervised seismic facies volume from the structural attributes (Workflow 1) shows structural features in blue and shows similar distribution pattern of the k_1 principal curvature as in Figure 4.12b.

CONCLUSIONS

Seismic stratigraphy and seismic facies analysis is routinely used in both clastic, and carbonate sedimentary deposits. However, facies analysis of different diagenetic altered rocks like chert is not a common workflow. Proper selection of the input seismic attribute volumes is the key to effective classification to differentiate the various diagenetically altered rocks. The proposed unsupervised SOM algorithm generates natural clusters that are formed from the overdefined classes. We have used two different sets of attribute volumes to identify different expressions of diagenesis and structural deformation present in the Mississippi chert reservoir.

The three facies (St. Joe’s tight limestone, fractured and layered Osage B chert, and high-porosity Osage A tripolitic chert) are identified from the borehole image logs and are used as training vectors for supervised classification. This supervised classification produced results consistent with the unsupervised classification. While identifying porous tripolitic chert as the sweet spots, the tight layered cherts with the natural fractures are identified as areas that may be more effectively stimulated by hydraulic fracturing. The post stack inversion result combined with the different seismic facies volumes helps understanding the prospective zones of this Mississippian chert reservoir.

With all the three workflows, we infer the “best” attributes for classifying the heterogeneous Mississippi Chert reservoir can be GLCM entropy, GLCM homogeneity, spectral bandwidth, coherence and P-impedance.

This technique is less well established in mapping diagenetically altered strata like the Mississippi Lime. More conventional techniques like P-impedance inversion requires better well control for the area and good well ties that require much effort from geophysicists and is a time consuming job. The proposed unsupervised multi-attribute workflow requires comparatively much less effort and can be done before the wells are available. This analysis can be done as a preliminary workflow to understand the reservoir. This classification is a step forward where we choose the best attributes that can represent the reservoir and perform our clustering on the chosen set of multi-attribute volumes. For our study, we show that SOM multi-attribute clustering successfully maps the various chert facies within the reservoir zone. These results are consistent with the multi-attribute supervised classification volume, P-impedance volume and the borehole image logs. Thus the proposed workflow can be confidently and easily performed when there is not much well information available in the survey.

LIST OF FIGURES

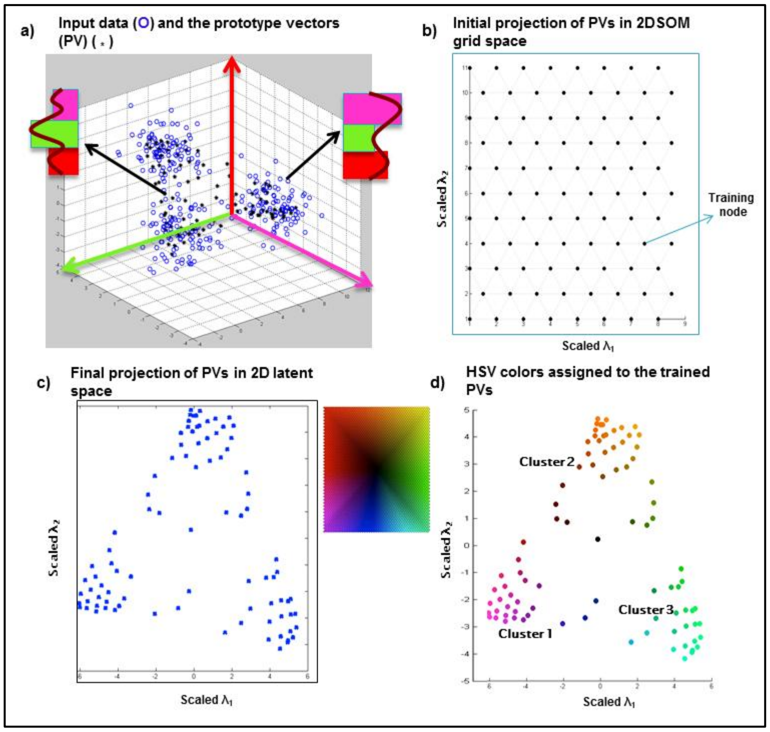


Figure 4.1. Kohonen Self-organizing maps training algorithm and 2D gradational HSV coloring. **(a)** Three distinct Gaussian distribution with same standard deviations but different means (blue circles). To begin with we choose a representation of the data-vectors (called the prototype vectors) having same three dimensions as the data-vectors. A PV in the input space (black dots in (a)) is assigned to each grid point in the 2D SOM grid space. **(b)** The initial projections of the PVs in the 2D SOM grid space. **(c)** Final projection of the trained PVs in the 2D SOM grid space forming three classes as present in the input dataset. The trained PVs are plotted in the 3D data-space (black dots). **(d)** A 2D gradational HSV color scaling of the trained PVs are done for better visualization (after Matos, 2009).

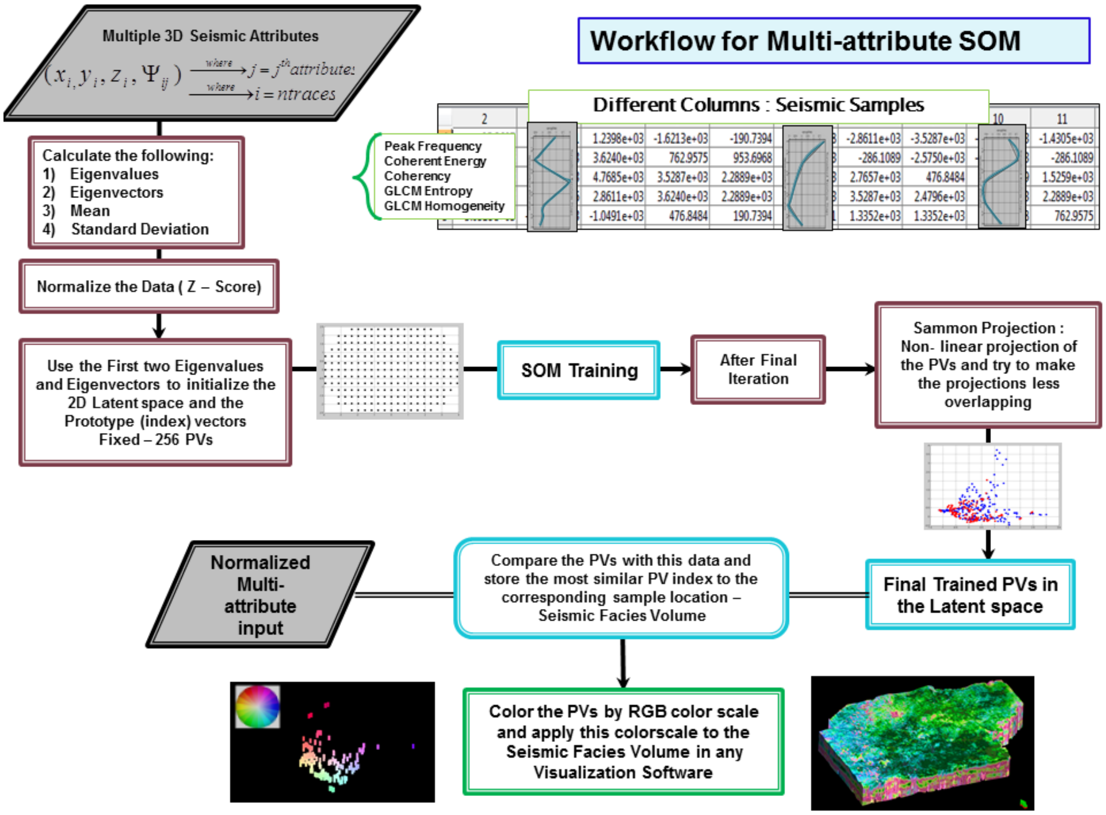


Figure 4.2. Workflow for the unsupervised SOM classification of the multi-attribute dataset.

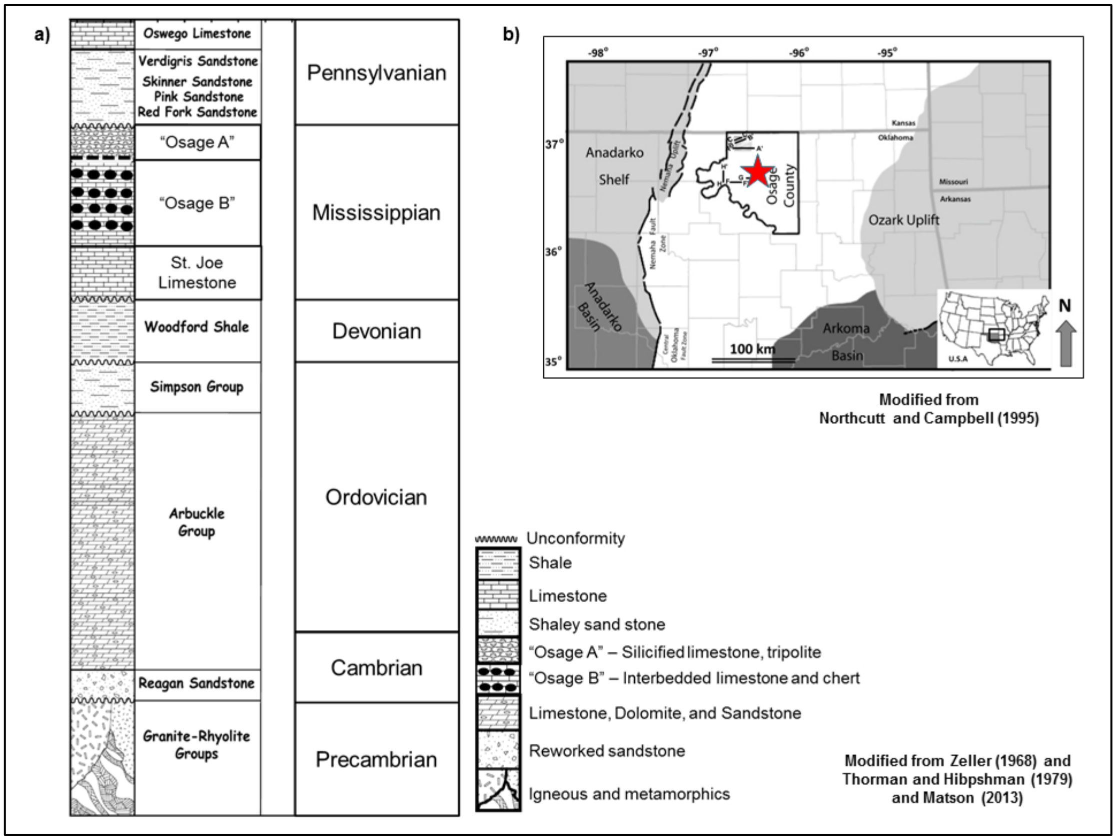
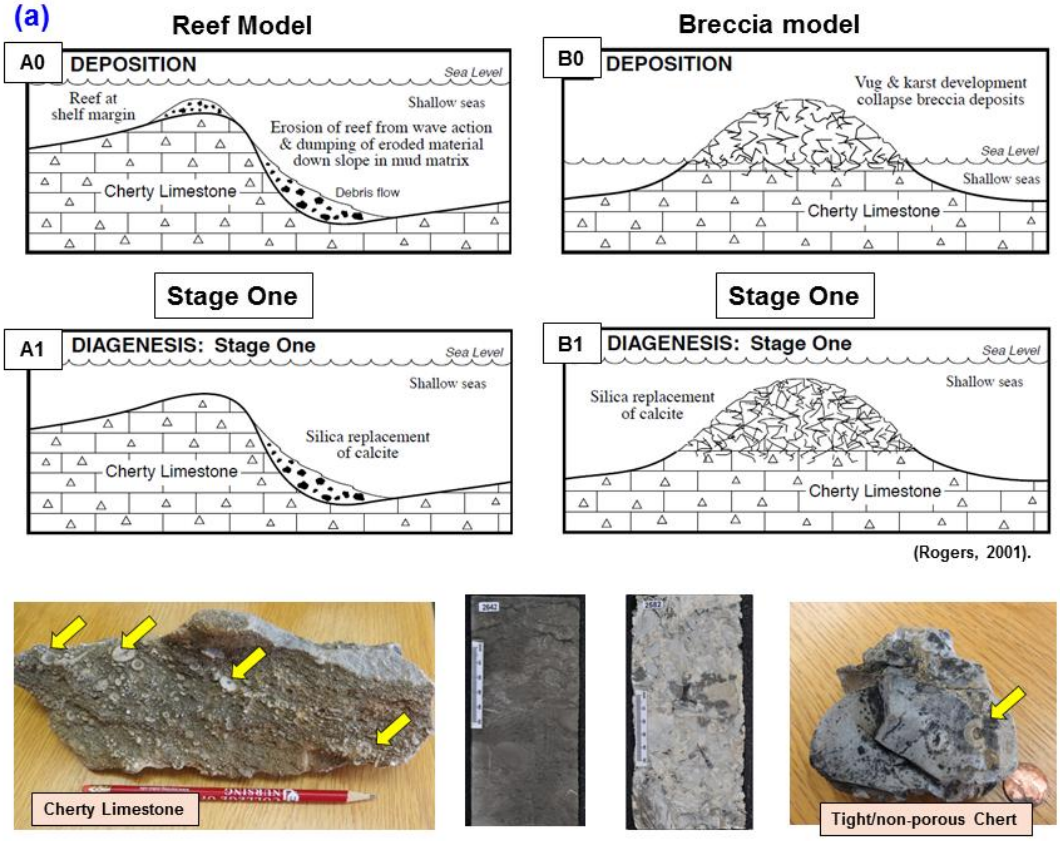


Figure 4.3. (a) The general stratigraphic column with the Mississippian tripolitic chert interval is present at the unconformity between the Pennsylvanian and Mississippian age. (**b)** The location of the seismic survey from Osage County, Oklahoma, within the Cherokee Platform province. The Cherokee Platform is bounded on the west by the Nemaha Uplift and to the east by the Ozark Uplift. Matson (2013) subdivides the Mississippian in this study area into the tight St. Joe limestone, and the Osage A and Osage B levels.


(a)
Reef Model
Breccia model
A0
DEPOSITION
Sea Level
Reef at shelf margin
Shallow seas
Erosion of reef from wave action & dumping of eroded material down slope in mud matrix
Cherty Limestone
Debris flow
B0
DEPOSITION
Vug & karst development collapse breccia deposits
Sea Level
Shallow seas
Cherty Limestone
Stage One
Stage One
A1
DIAGENESIS: Stage One
Sea Level
Shallow seas
Silica replacement of calcite
Cherty Limestone
B1
DIAGENESIS: Stage One
Sea Level
Silica replacement of calcite
Shallow seas
Cherty Limestone
(Rogers, 2001).
Cherty Limestone
2642
2582
Tight/non-porous Chert

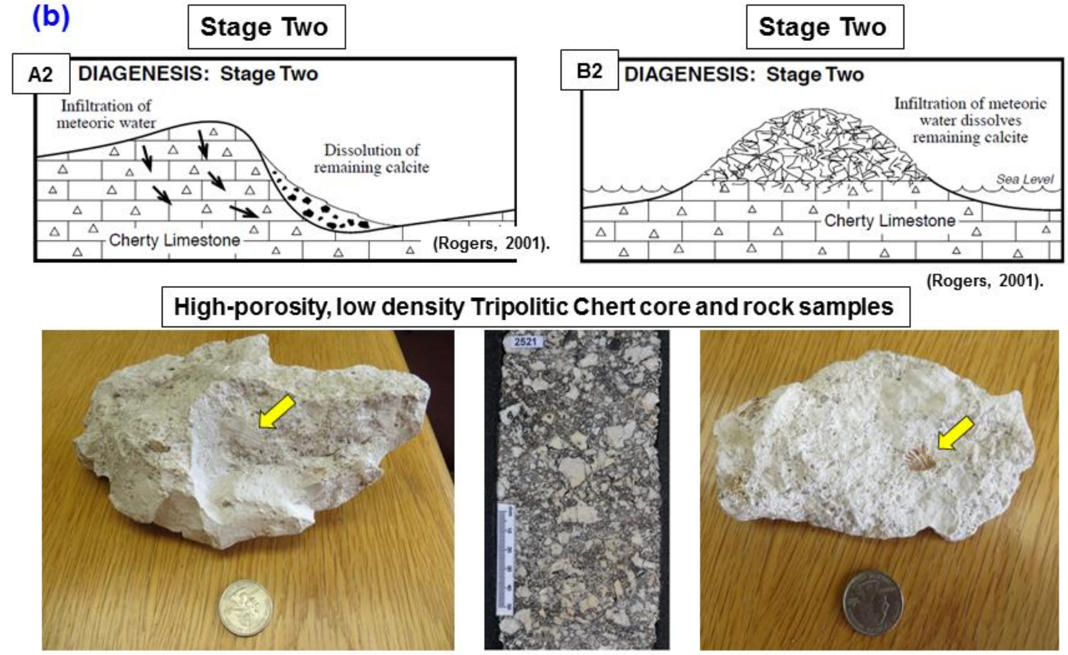

(b)
Stage Two
Stage Two
A2
DIAGENESIS: Stage Two
Infiltration of meteoric water
Dissolution of remaining calcite
Cherty Limestone
(Rogers, 2001).
B2
DIAGENESIS: Stage Two
Infiltration of meteoric water dissolves remaining calcite
Sea Level
Cherty Limestone
(Rogers, 2001).
High-porosity, low density Tripolitic Chert core and rock samples
2521

Figure 4.4. Two different models for the diagenesis the Mississippian Chert developed from weathered and/or eroded limestone (after Rogers, 2001 using nomenclature of Matson, 2013). In the reef model the chert formation at the reef margin. In the Breccia model the chert forms as sub-aerially exposed breccia deposits. **(a)** Stage-one for both the settings is the silica replacement of calcite in a submarine environment. The siliceous limestone and the layered chert and limestone (corresponding to Osage B formation) are formed in this environment. **(b):** Stage-two digenesis is a result of erosion and uplift, infiltration by meteoric water. This results in flushing of the rock and dissolution of the remaining calcite by low-pH fluid and absence of no new silica precipitation. This results in moldic porosity and vuggy porosity (yellow arrows) that is common in the low-density high-porosity tripolitic chert (corresponding to Osage A).

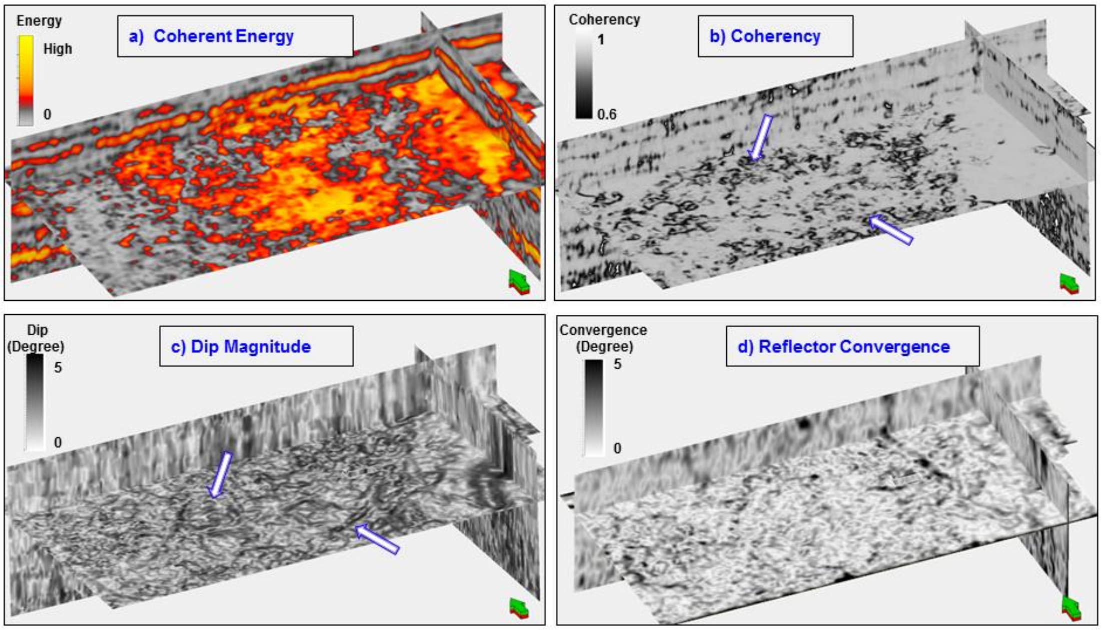


Figure 4.5. Structural attributes used in unsupervised analysis (workflow 1). **(a)** Coherent energy, **(b)** Coherence, **(c)** dip magnitude and **(d)** reflector convergence. Blue arrows indicate discontinuities and structural features that will form their own clusters in Figure 4.6.

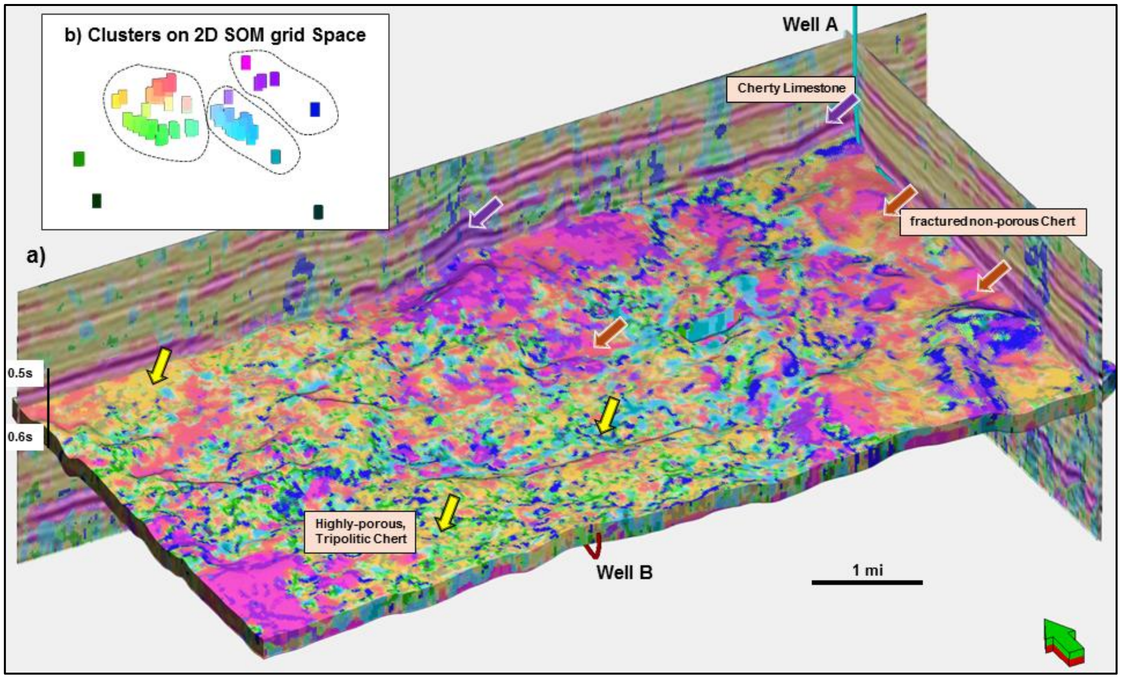


Figure 4.6. (a) The result of multi-attribute unsupervised seismic facies classification (workflow 1) within the Mississippi lime using the four “structural” attribute volumes shown in Figure 4.5 as input. (b) The projections of the 256 prototype vectors (clusters) plotted in the 2D SOM grid space clump into four main clusters magenta/dark pink, red/orange, yellow/green or cyan in color. The blue cluster is one of the outliers. A posteriori analysis from the borehole image log and fracture density diagram of Well A shows magenta or dark pink color correlate to the tight limestone facies (St. Joe Limestone) (indicated by the magenta arrows) while orange and red correlate to the dense layered chert and limestone facies (Osage B) (indicated by the red arrows). Similar a posteriori analysis from Well B shows that light green and yellow correlate to tripolitic chert facies (Osage A) (indicated by the yellow arrow). The forth cluster blue and cyan correlates to areas of low coherence and high dip

magnitude, which we interpret to be areas of faults, fractures and karst (blue arrows as in Figure 4.5).

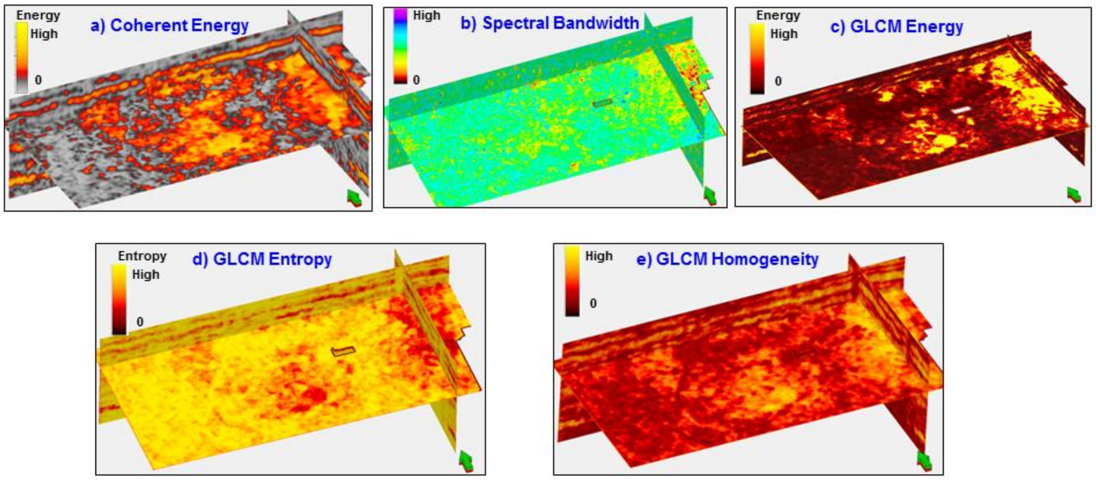


Figure 4.7. The input attribute volumes used in workflow 2: **(a)** coherent energy, **(b)** spectral bandwidth, (**c)** GLCM energy, **(d)** The GLCM entropy, and **(e)** GLCM homogeneity. Our implementation of GLCM measures lateral variations of reflectivity along structure. In contrast we use bandwidth over spectral attributes to measure the vertical variations in texture. The GLCM energy is a measure of the GLCM matrix energy, not the square of the seismic amplitude. GLCM energy increases if amplitude values repeat (e.g. as a homogeneous, striped, or checkerboard pattern). GLCM entropy is high if the amplitude values are random. GLCM homogeneity is high if amplitude values are smoothly varying.

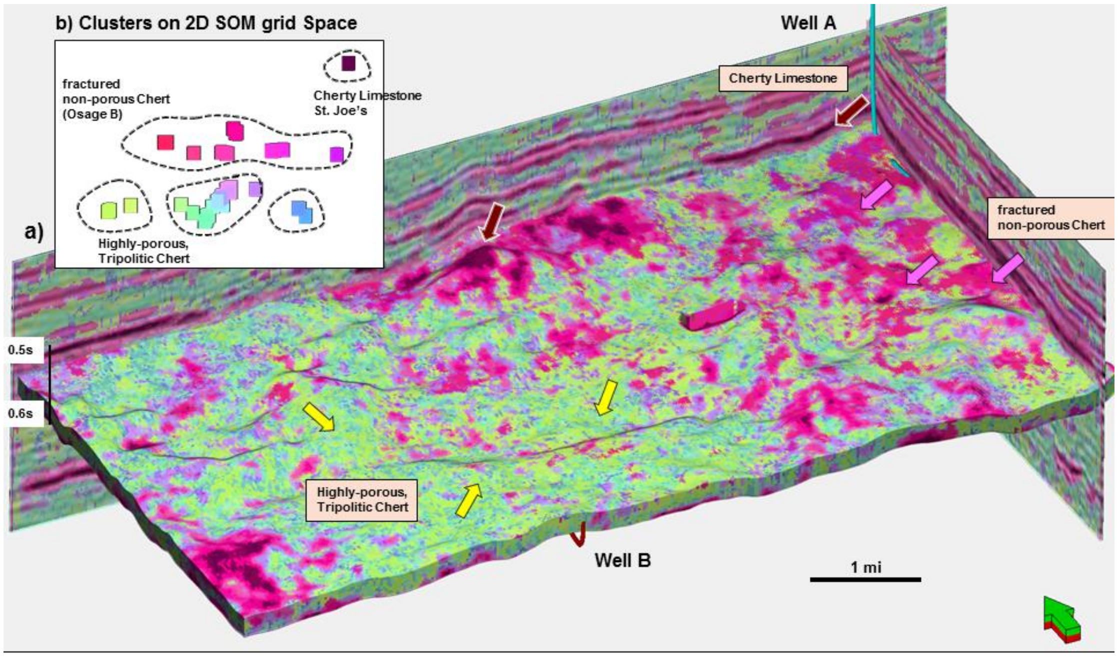


Figure 4.8. The result of multi-attribute unsupervised seismic facies classification from workflow 2 within the Mississippian lime zone using the "texture" attribute volumes shown in Figure 4.7. (b) The projections of the 256 prototype vectors (clusters) plotted in the 2D SOM grid space clump into mainly into four different clusters. The dark red/brown projection of the PV (as an outlier) forms the fifth facies type. A posteriori analysis from the borehole image log and fracture density diagram of Well A shows dark red and brown correlate to the tight cherty St. Joe's limestone facies (indicated by the brown arrows) while light pink and violet correlate to the layered chert and lime facies (Osage B formation) (indicated by the pink arrows). Similar a posteriori analysis from Well B shows that light green and yellow correlate to tripolitic chert facies (Osage A formation) (indicated by the yellow arrow).

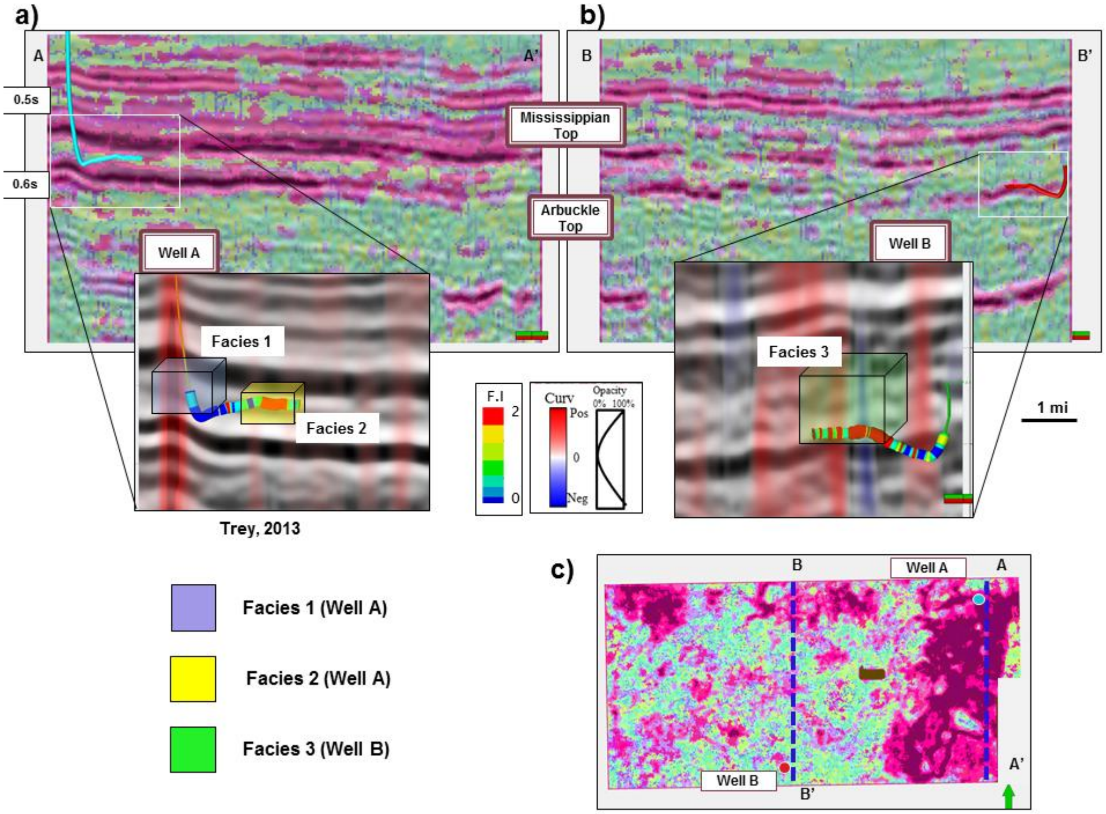
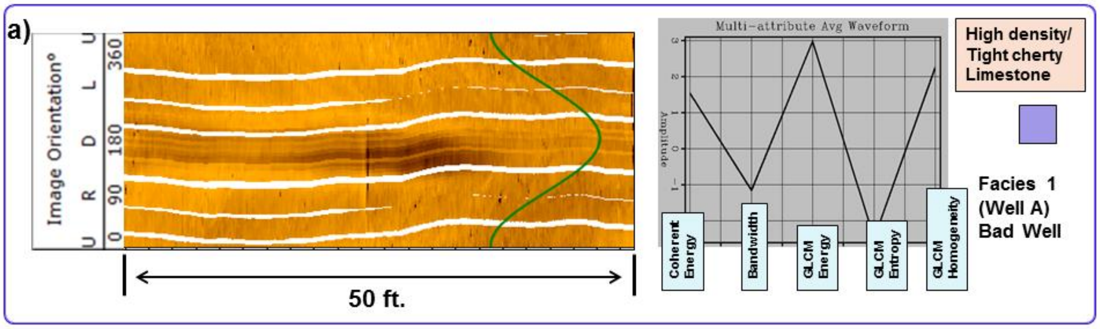
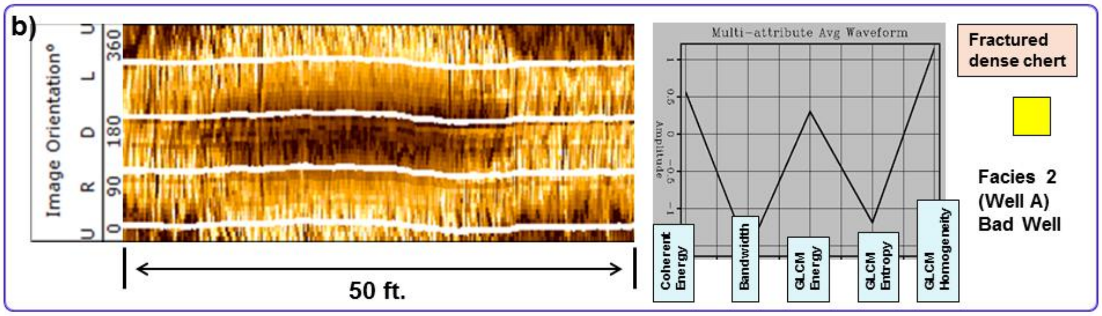


Figure 4.9. Subvolumes used to extract three average data-vectors, which will be used in supervised multi-attribute seismic facies classification. The three sub-volumes have been chosen on the basis of the image log interpretation from White (2013). Hot colors indicate high fracture densities based on the two wells. **(a)** Two sub-volumes are considered around Well A. The blue-violet Facies 1 corresponds to the tight/non-porous limestone (St Joe's Limestone). The yellow Facies 2 corresponds to the fractured dense chert (Osage B). **(b)** The sub-volume considered around Well B. The light green facies 3 corresponds to tripolitic chert facies and the region having the highest fracture density (Osage A). **(c)** Map view of the location of the two wells in the survey.


a)
Image Orientation°
U R D L U
0 90 180 360
50 ft.
Multi-attribute Avg Waveform
Amplitude
Coherent Energy
Bandwidth
GLCM Energy
GLCM Entropy
GLCM Homogeneity
High density/ Tight cherty Limestone
Facies 1 (Well A) Bad Well


b)
Image Orientation°
U R D L U
0 90 180 360
50 ft.
Multi-attribute Avg Waveform
Amplitude
Coherent Energy
Bandwidth
GLCM Energy
GLCM Entropy
GLCM Homogeneity
Fractured dense chert
Facies 2 (Well A) Bad Well

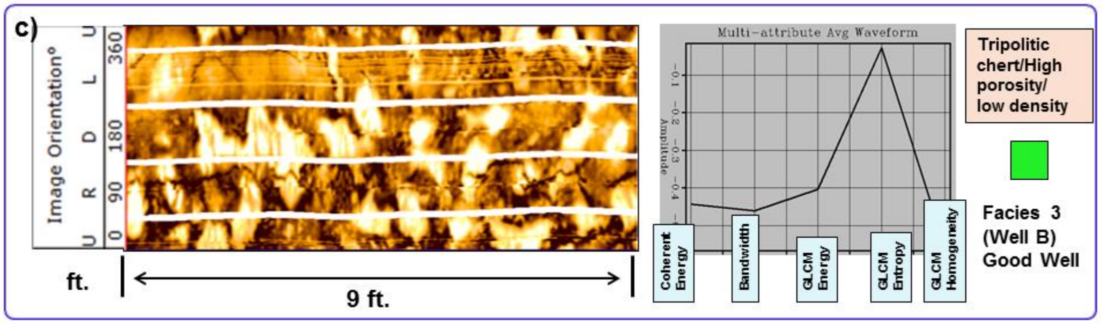

c)
Image Orientation°
U R D L U
0 90 180 360
ft.
9 ft.
Multi-attribute Avg Waveform
Amplitude
Coherent Energy
Bandwidth
GLCM Energy
GLCM Entropy
GLCM Homogeneity
Tripolitic chert/High porosity/ low density
Facies 3 (Well B) Good Well

Figure 4.10. The three average waveforms extracted from the three sub-volumes around the wells with the images of the borehole image logs for each of the facies type. (a) The borehole image log of Well A corresponds to the tight St. Joe limestone formation. The corresponding facies type is defined to be Facies-1 (violet color) (b) The borehole image log within the Osage B formation shows many natural fractures and interbeded chert and lime. The corresponding facies type is defined to be Facies-2 (yellow color). The average data-vectors selected from the Well A have similar patterns with high coherent energy, high energy and high GLCM homogeneity, narrow bandwidth and low GLCM entropy. However, the amplitude values of the attributes are less in Facies-2 indicating that this facies is fractured and layered chert not as tight and dense as the Facies 1. (c) The image log shows the presence of low density diagenetically altered tripolitic chert (corresponding to the Osage A formation). The average multi-attribute data-vectors selected from the Well B (Facies 3) has low coherent energy, spectral amplitude, GLCM homogeneity and GLCM energy values and high GLCM entropy value, which is a seismic signature of the low density, high porous tripolitic chert.

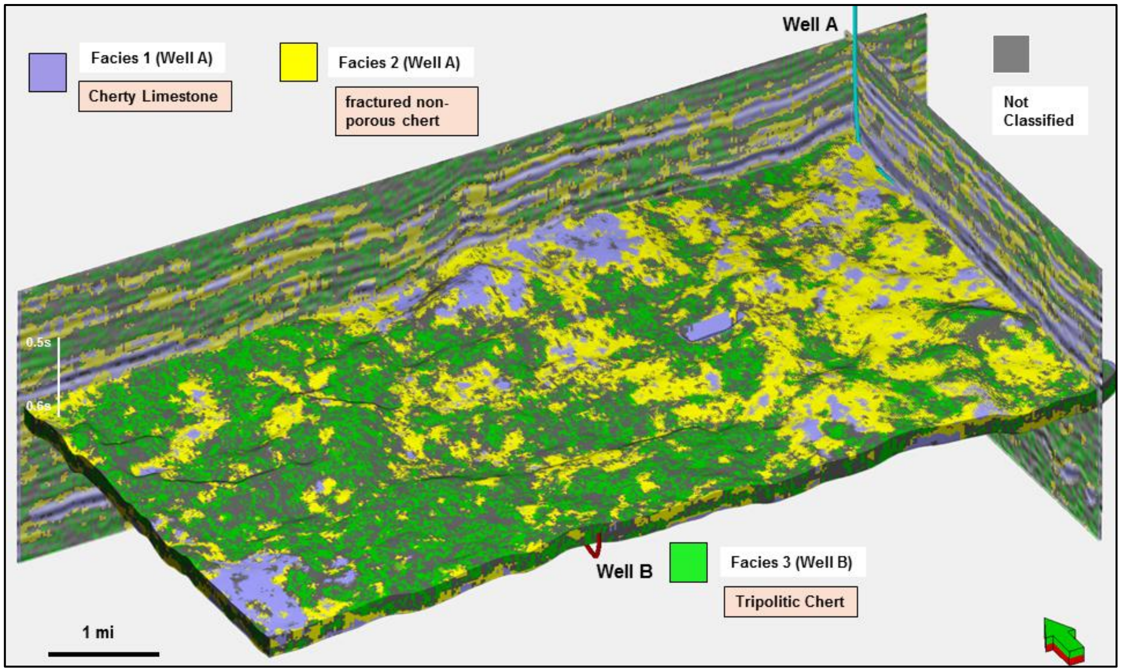


Figure 4.11. The multi-attribute supervised seismic facies volume, with the facies defined in Figure 4.10. The violet seismic facies (Facies-1) corresponds to the tight St. Joe's limestone, the yellow seismic facies (Facies 2) corresponds to fractured chert with interbeded chert and lime regions and the green seismic facies (Facies 3) corresponds to the tripolitic chert rich zones. The facies, which are not similar to these three facies, are color-coded gray.

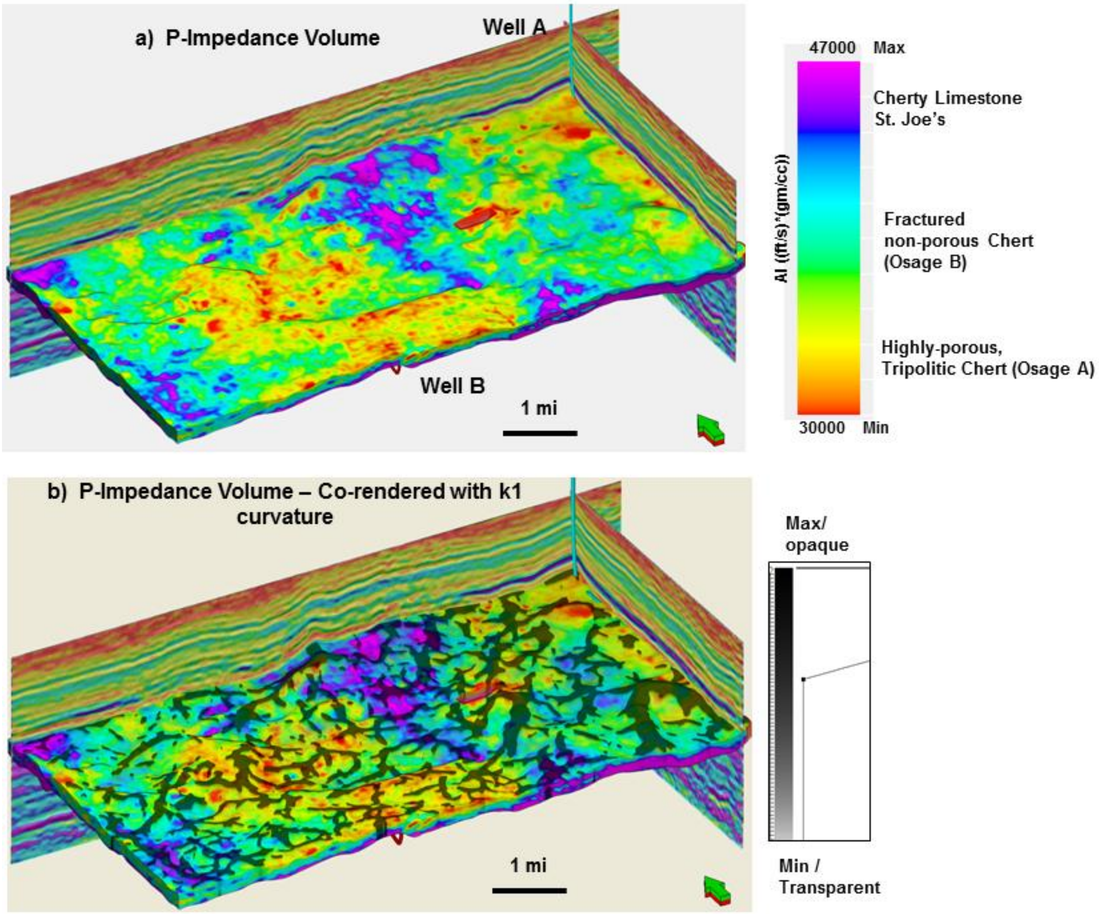


Figure 4.12. The results of post-stack P-impedance inversion. **(a)** The less dense highly porous tripolitic chert regions have low impedance (red and yellow colors). The green areas correspond to the fractured or layered dense chert and lime regions. The high impedance regions (cyan-blue violet) correspond to the dense chert and cherty limestone. This result when compared with unsupervised seismic facies analysis workflow 2 shows that the dense limestone and chert rich zones correspond to high impedance regions and the low impedance corresponds to the tripolitic chert rich areas. Similarly, for the fractured layered chert and lime corresponds to the regions with medium to low impedance. **(b)** The P-impedance volume is co-rendered with the

positive principal curvature k1. When compared with the multi-attribute analysis with the structural attributes (workflow 1) it shows similar discontinuities/fractures, faults and karst features accompanied with the different chert facies.

REFERENCES

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 73rd Annual International Meeting Society of Exploration Geophysicists, Expanded Abstracts, 2221-2224.

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 enviroment described by a seis- mic facies classification volume: Marine and Petroleum Geology , v. 26, p. 495-503.

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.

Handford, M., 2012, Where's Waldo?: The 25th Anniversary Edition: Candlewick Press, 32 p.

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

Johnson, K.S., 2008, Geologic History of Oklahoma, *in* K.S.Johnson and K.V. Luza, eds., Earth sciences and mineral resources of Oklahoma: Oklahoma Geological Survey Educational Publication 9, p. 3-5.

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.

Matos, M., M. Yenugu, and K.J. Marfurt, 2011, Integrated seismic texture segmentation and cluster analysis applied to channel delineation and chert reservoir characterization: Geophysics, p. 11-21.

Matson, S., 2013, Mississippi lime play: reservoir definition, production examples and considerations for stimulations, Mississippi Lime play Forum, January, 2013, Oklahoma City.

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 Society of Exploration Geophysicists, Expanded Abstracts, 931-934.

Michelena, J. R., E. S. M. Gonzalez, and M. Capello, 1998, Similarity analysis: A new tool to summarize seismic attribute information, The Leading Edge, April 1998.

Northcutt, R. A., and J. A. Campbell, 1995, Geologic provinces of Oklahoma: Oklahoma Geological Survey Open- File Report 5-95, scale 1:750,000, 1 sheet.

Poupon, M., Gil, J., D. Vannaxay, and B. Cortiula, Tracking Tertiary delta sands (Urdaneta West, Lake Maracaibo, Venezuela): An integrated seismic facies classification workflow, The Leading Edge, September 2004, 909-912.

Rogers, S.M., 2001, Deposition and diagenesis of Mississippian chat reservoirs, north-central Oklahoma: AAPG Bulletin, **85**, 115-129.

Roy, A., M. Matos, and K. J. Marfurt, 2011, Application of 3D clustering analysis for deep marine seismic facies classification – an example from deep water northern Gulf of Mexico: GCSSEPM 31st Annual Bob. F. Perkins Research Conference, 410-439.

Sammon, W. J., 1969, A nonlinear mapping for data structure analysis, IEEE Transaction on Computers, Vol. C-18. No. 5.

Snyder, R., A case history of the East Hardy Unit, Mississippian Highway 60 trend, Osage County, Oklahoma, Mississippi Lime play Forum, January, 2013, Oklahoma City.

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.

Thorman, C. H., and M. H. Hibpshman, 1979, Status of mineral resource information for the Osage Indian Reservation, Oklahoma: U.S. Geological Survey and Bureau of Mines, Administrative Report BIA-47, p. 1–60.

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.

Watney, W.L., W.J. Guy, and A.P. Byrnes, 2001, Characterization of the Mississippian chat in south-central Kansas: AAPG Bulletin, **85**, 85-113.

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.

White, G.H., 2013, Calibration of surface seismic attributes to natural fractures using horizontal image logs, M.S. Thesis, The University of Oklahoma

Zeller, D. E., 1968, The stratigraphic succession in Kansas: Kansas Geological Survey Bulletin, v. 189, p. 81.

Zhang, K., K. J. Marfurt, M. C. Matos, and J. T. Kwiatkowski, 2008, Time-frequency domain spectral balancing and phase dispersion compensation, 78th Annual International Meeting of the SEG, Expanded Abstracts.

CHAPTER 5

Seismic facies classification of unconventional reservoirs using Generative Topographic Mapping

ABSTRACT

Classification methods fall into two major classes, the supervised and unsupervised classification. An ideal unsupervised classification method is one which does not require direct human intervention and subdivides the data volume into its “natural clusters”. However classification in the actual higher dimensional data space is difficult. Principal component analysis, self-organizing maps, and generative topographic mapping provide dimensionality reduction from a higher-dimensional input data-space to a lower dimensional model space. Among these techniques neural net and Kohonen self-organizing maps (a sub-group of artificial neural network) are the most popular classification methods routinely done for Generally well log prediction or analysis and seismic facies modeling. Although they have been successful in many hydrocarbon exploration projects, they have some inherent limitations. Here we explore one of the recent techniques known as the Generative Topographic Maps (GTM), which takes care of the shortcomings of the Kohonen Self organizing maps. We explore the formulation of this technique and applied it on two different datasets to demonstrate probabilistic clustering. Firstly, we applied the GTM technique to classify 15 sets of horizontal well parameters in a one of the new shale

plays, correlating the results with normalized EURs, allowing an estimation of EUR based on the most relevant parameters.

Secondly we applied GTM on the inverted 3D seismic volumes of a Barnett shale dataset for a multi-attribute unsupervised clustering. The final seismic facies volume was a result of user defined *a posteriori* supervision to classify the higher probability density of the data in the 2D latent space. Different clusters formed from the GTM analysis of the Barnett shale survey are interpreted from different horizontal wells with micro-seismic data. The study shows the effectiveness of GTM technique as an unsupervised probabilistic approach for data classification.

INTRODUCTION

In the past, due to computational inefficiency, the dimensionality reduction from a higher-dimensional input data-space to a lower dimensional model space were only linear structures (hyper-planes) in the data space. The technique described in this paper illustrates a non-linear relationship between *D*-dimensional data-space and a lesser *L*-dimensional model (D>L). Generally the *L*-dimensional model is called a latent space, which is a lower-dimensional manifold embedded in attribute space that approximately contains the vast majority of the probability mass of the data (Sevensen, 1998). It is assumed that a smaller number of hidden or latent variables in this latent space generate the large number of input dataset. Every latent variable is mapped to a corresponding point in the data space to form *L*-dimensional manifold in the *D*-dimensional data-space. Therefore, a distribution defined in the latent space corresponds to a distribution in the data-space thus establishing a probabilistic relationship between the two spaces (Sevensen, 1998). After proper parameter estimation, we can relate the points in the *D*-dimensional data-space to grid points in *L*-dimensional curved manifold, which in turn relates to points in *L*-dimensional latent space. Thus for data points in the high dimensional space one can find some representation in the lower dimensional latent space. The above technique formulates a generative model where vectors in the data space are generated by grid points in the latent/model space. Bishop et al, 1998 proposed such a generative model called the Generative Topographic Model (GTM). Data visualization after GTM modeling is the key for using GTM as a probabilistic classifier.

This GTM classification algorithm has been rarely used in hydrocarbon exploration. Wallet et al., (2009) were the first to apply the GTM technique to a suite of phantom horizon slices through a seismic volume generating a “waveform classification”. We have implemented GTM efficiently to handle much larger seismic dataset as input and generated a new workflow for unsupervised seismic facies analysis.

The GTM Algorithm

The GTM is a non-linear dimensional reduction technique that provides a probabilistic representation of the data-vectors in a corresponding latent space. In this paper our input data will consists of D-dimensional data vectors. In our first example the data are multiple engineering measurements forming vectors corresponding to discrete wells. In our second example, the data are a suite of seismic attributes forming a vector at each 3D voxel. A set of non-linear continuous and differentiable basis functions are used to map points $\mathbf{u} = u_1, u_2, \ldots, u_L$ in the L-dimensional latent space into corresponding points in the data space. This mapping generates points defining L-dimensional non-Euclidean manifold $\mathcal{S}$ embedded within the data-space (Figure 5.1). A Gaussian noise model (PDF) for a data-vector $\mathbf{x}_n$ is then defined centered on the mapped grid points on the non-Euclidean manifold $\mathcal{S}$. This Gaussian noise model defines the space in which the data vector lies. In this manner, the probability distribution of a data-vector $\mathbf{x}_n$ is obtained by summing the individual PDFs associated with each of the mapped grid points. This initial stage is called a constrained Gaussian mixture because the Gaussian centers are constrained by the grid points in the latent space. As we iterate each component of the mixture model is moved towards the data-vector that it best represents. For each iteration the mapping parameters are determined using an Expectation Maximization (EM) algorithm (Dempstar *et al*, 1977). Since the points on the manifold $\mathcal{S}$ relates to the points on the latent space by a probability distribution, a posterior probability projection is then found for each of the data-vectors in the L-dimensional latent space. This posterior probability projection of the data onto the latent space is used for data visualization. Detailed implementation of this algorithm is given below.

GTM Theory

In our implementation of GTM, we have fixed the latent space to be 2-dimensional (L=2) to facilitate subsequent analysis using a 2-component commercial cross-plotting tool, which will be later used for visualization. An array of regularly spaced latent space variable points (nodes) is arranged in this 2D latent space, labeled k=1,2,3….,K and are denoted by $\mathbf{u}_k$. In addition, a set of j =1,2,3…., J non-linear basis functions are introduced in the mapping function (Figure 5.1). Following Bishop et al., (1998) we use Gaussian functions $\phi_j(\mathbf{u}_k)$ as non-linear basis functions, with their centers j also uniformly arranged in regular grid spacing. These Gaussian basis functions are defined as

$$\phi_j(\mathbf{u}_k) = e^{-\frac{(\mathbf{r}_j - \boldsymbol{u}_k)}{2\sigma^2}}, \tag{1}$$

where, $\mathbf{r}_j$ is the position vector of the center of the j^{th} basis function, $\mathbf{u}_k$ is the position vector of the k^{th} node, and σ is the common width (standard deviation) of these Gaussian basis functions.

In order to determine the PDF in the data-space, we need to map the pre-defined latent space variables $\mathbf{u}_k$ onto the D-dimensional data space to a corresponding set of reference vectors, $\mathbf{m}_k$. These reference vectors $\mathbf{m}_k$ lie on the 2D non-Euclidean manifold $\boldsymbol{\mathcal{S}}$. This non-linear transformation is given by,

$$\mathbf{m}_k = \sum_{j=1}^{J} \mathbf{W}_{kj}\phi_j(\mathbf{u}_k), \tag{2}$$

where, k=1,2,3….,K are the indices for latent space variables, and J is the set of non-linear Gaussian basis functions. $\mathbf{W}$ is a D x J matrix that denotes the weights of the mapping function. Figure 5.1 depicts an overview of the GTM mapping from the latent space to the data space. Specifically, nine ($K=9$) latent space variables are mapped to nine reference vectors in the 2D non-Euclidean manifold with four ($J=4$) non-linear Gaussian basis functions.

It is highly improbable that the data lies only on the 2D manifold $\boldsymbol{\mathcal{S}}$ in the data-space. We assume these deviations are normally distributed and can be represented by a weighted sum of N-dimensional Gaussian probability density functions (PDF).

For simplicity, the variance σ^2=1/β is assumed to be same for all nodes, however, these nodes can move, such that the noise model is defined for a data-vector $\mathbf{x}_n$ with a radially symmetric Gaussian functions with centers at $\mathbf{m}_k$ (Figure 5.2) is given by

$$p(\mathbf{x}_n|k,\mathbf{W},\beta) = (\frac{\beta}{2\pi})^{\frac{D}{2}} e^{-\frac{\beta}{2}||\mathbf{m}_k-\mathbf{x}_n||^2}, \quad (3)$$

Since the centers of the Gaussian mixture components are dependent on points in the latent space (given by equation 2) the distribution $p(\mathbf{x}_n|k,\mathbf{W},\beta)$ corresponds to a constrained Gaussian mixture model (Hinton et al, 1992). The probability distribution in the data-space for the GTM model is obtained by summing over all K Gaussian components for a given $\mathbf{W}$ and β:

$$p(\mathbf{x}_n|\mathbf{W},\beta) = \sum_{k=1}^{K} P(k)\, p(\mathbf{x}_n|k,\mathbf{W},\beta) = \sum_{k=1}^{K} \frac{1}{K} (\frac{\beta}{2\pi})^{\frac{D}{2}} e^{-\frac{\beta}{2}||\mathbf{m}_k - \mathbf{x}_n||^2} \quad (4)$$

To make the above problem mathematically tractable, we require that the latent space points are constrained to a grid. Thus the prior probabilities *P(k)* of each of these latent space variables are given by the sum of delta functions centered on the nodes $\mathbf{u}_k$, where the indices of the nodes runs from *k=1,2,….., K.* We initialize *P(k)=1/K* factor is the normalization factor such that the total probability sums to unity.

The parameterization of the GTM model **W** (weight matrix) and β (inverse of the noise variance) is done by maximum likelihood estimation. This maximization is well suited for solution using an Expectation Maximum (EM) algorithm (Dempstar et al, 1977). The EM algorithm consists of two major steps. The first or expectation step is to find the expected value for a data-vector using the current estimated parameters. The second, or maximization step uses the expectation value to provide a new estimate of the parameters. The EM algorithm is a good choice for estimation problems like GTM because of its generality and guaranteed convergence (Bishop, 1995).

Consider for a certain iteration the weight matrix is given by $\mathbf{W}_{\text{old}}$ and the inverse of the noise variance given by β_{old}. In the “E-step” of the EM algorithm we calculate the posterior probability or *responsibility,* (R_{nk}) for each of the *K* components in latent space for every data-vector $\mathbf{x}_n$ using these GTM model parameters $\mathbf{W}_{\text{old}}$ and β_{old}. Using Bayes’ theorem we obtain the posterior probability as,

$$R_{nk}(\mathbf{W}_{old}\,,\beta_{old}) = \frac{e^{-\frac{\beta}{2}||\mathbf{m}_k - \mathbf{x}_n||^2}}{\sum_k e^{-\frac{\beta}{2}||\mathbf{m}_k - \mathbf{x}_n||^2}}, \tag{5}$$

where the dataset are represented by $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \ldots\ldots \mathbf{x}_N$. However the value of $e^{-\frac{\beta}{2}||\boldsymbol{m}_k - \boldsymbol{x}_n||^2}$ becomes very small for most of the values of K causing a numerical underflow in the calculation which results in the denominator of equation 5 becoming zero. This is avoided by subtracting a constant factor from the argument of the exponential, $e^{-\frac{\beta}{2}||\mathbf{m}_k - \mathbf{x}_n||^2}$, which is equivalent to multiplying the numerator and denominator of equation 5 by a constant. Next, in the Maximization or "M-step", we use these responsibilities to update the model to compute a new weight matrix $\mathbf{W}_{\text{new}}$ by solving a set of linear equations

$$\mathbf{\Phi}^T \mathbf{G}_{old} \mathbf{\Phi} \mathbf{W}^T{}_{new} = \mathbf{\Phi}^T \mathbf{R}_{old} \mathbf{X} \tag{6}$$

where **G** is a K x K diagonal matrix with elements $G_{kk} = \sum_{n=1}^{N} R_{nk}$, $\mathbf{\Phi}$ is a K x J Matrix with elements $\mathbf{\Phi} = \phi_j(\mathbf{u}_k)$ and $\mathbf{R}_{\text{old}}$ is a K x N matrix with elements R_{nk} given by equation (5). The updated value of the inverse of the noise variance is given by β_{new} where

$$\frac{1}{\beta_{new}} = \frac{1}{ND} \sum_{n=1}^{N} \sum_{k=1}^{K} R_{nk_{old}} \,|\left|\mathbf{W}_{kj_{new}} \phi_j(\mathbf{u}_k) - \mathbf{x}_n\right||^2. \tag{7}$$

To regularize the matrix inversion we add a constant λ and rewrite the equation (6) to be

$$(\boldsymbol{\Phi}^T \mathbf{G}_{old} \boldsymbol{\Phi} + \lambda \mathbf{I}) \mathbf{W}^T{}_{new} = \boldsymbol{\Phi}^T \mathbf{R}_{old} \mathbf{X} \qquad (8)$$

where $\mathbf{I}$ is the J x J identity matrix and $\lambda = \alpha/\beta$, α being a real constant parameter. This constant α makes a prior distribution over $\mathbf{W}$ and scales the 2D manifold.

After GTM training the 2D manifold $\boldsymbol{\mathcal{S}}$ is stretched to represent regions of lower data density and is squeezed to represent regions with greater data density. To construct a GTM model we require choosing a number of parameters which is discussed next.

Initialization and Parameter selection in GTM

The selection of the input data vector (e.g. which seismic attributes best represent the desired facies) is the most critical parameter selection of GTM and to clustering in general (Barnes and Laughlin, 2002). Parameter selection specific to GTM defines facies resolution/discrimination and runtime. The number of points in the grid points (nodes) K in the latent space should be such that it is dense enough to approximate a continuous distribution of the data. By construction, the number of grid points is equal to the number of Gaussian mixtures in the data-space. Also to note, choosing a very large number of grid points increases the computation time and memory usage. Thus a trade-off should be maintained between selecting the number of grid points and computational time.

We also define a set of non-linear Gaussian basis function centers *J*, on this 2D latent space grid and take care to set $J < K$ to avoid rank deficiency of the $\mathbf{\Phi}$ matrix. The common width of these basis functions ε is also set prior to the training. ε controls the smoothness of the 2D manifold. A smooth manifold in the data space facilitates fitting the data-vectors during training. This parameter ε remains constant for the whole process. The matrix consisting of the non-linear basis functions (Gaussian functions in our case) $\mathbf{\Phi} = \phi_j(\mathbf{u}_k)$ is calculated at initialization and remains constant for all the subsequent iterations.

We initialize the weight matrix $\mathbf{W}$ such that the initial GTM model approximates the principal component of the dataset. The common value of the inverse variance of the Gaussian PDFs β is initialized to be the inverse of the $(L+1)^{\text{th}}$ eigenvalue from PCA where L is the dimension of the latent space. Since in our case $L=2$ we initialize β to be the inverse of the third eigenvalue.

Data Visualization in GTM

Visualization is key to effective clustering. After we have estimated the parameters $\mathbf{W}_{\text{new}}$ and β_{new}, we are able to define a new posterior probability distribution of the data by the latent space grid points:

$$R_{nk}(\mathbf{W}_{new}, \beta_{new}) = \frac{e^{-\frac{\beta}{2}||\mathbf{m}_k-\mathbf{x}_n||^2}}{\sum_k e^{-\frac{\beta}{2}||\mathbf{m}_k-\mathbf{x}_n||^2}} \,. \tag{9}$$

These posterior probabilities or the “responsibility”, R_{nk} values in equation 9 can be mapped to the entire grid points $\mathbf{u}_k$ (Figure 5.3). Such an explicit projection for all the data-vectors will in general results in too much redundant information in the latent space. For this reason Bishop et al., (1998) proposed projecting the mean or the mode posterior probability projections onto the latent space.

The mean of the posterior probability distribution of a data-vector $\mathbf{x_n}$ is obtained by projecting onto all the grid points (nodes) the responsibilities of each node, thus computing the average location in **u**:

$$\mathbf{U}_{mean}(\mathbf{x}_n) = \sum_{k=1}^{K} R_{kn}\mathbf{u}_k. \tag{10}$$

The mode of the posterior probability distribution is similar:

$$\mathbf{U}_{mode}(\mathbf{x}_n)|_k = MAX(R_{kn}). \tag{11}$$

Figure 5.3 illustrates such a calculation of the posterior probability for one data-vector. The mode of the discrete K nodes occurs at the the magenta point. The mean of the posterior probability projection of the same data-vector $\mathbf{x_n}$ occurs at the green point and in general falls between the grid points.

Summary of the GTM workflow for facies classification

The GTM workflow is summarized in the following steps (Figure 5.4):

Initialization

- Choose an appropriate suite of attributes to differentiate the different reservoir performance results or seismic facies.

- Define the number of latent variables K, the basis function J, the relative width of the basis functions, ε, and the scaling factor, α,
- Generate the latent space grid $\mathbf{u}_k$, where $k=1,2,\ldots\ldots,K$,
- Generate the grid for the Gaussian basis function centers $\mathbf{r}_j$, where $j=1,2,\ldots\ldots,J$,
- Compute the set of Gaussian basis function $\mathbf{\Phi}$ (from equation 1),
- Initialization of $\mathbf{W}$ and β from PCA analysis of the data $\mathbf{X}$,
- Compute the reference vectors, $\mathbf{m}_k$ (from equation 2),
- Compute $||\mathbf{m}_k - \mathbf{x}_n||^2$ for the Gaussian PDFs (equation 3), and
- Calculate the responsibility $R_{nk}(\mathbf{W}, \beta)$ (equation 5).

  Training

- Update the weight matrix $\mathbf{W}_{\text{new}}$ (equations 6 or 8),
- Update $\mathbf{m}_k$ and calculate the new $||\mathbf{m}_k - \mathbf{x}_n||^2$,
- Update the inverse variance β_{new} (equation 7),
- Calculate the new responsibility $R_{nk}(\mathbf{W}_{new}, \beta_{new})$, and
- Compute posterior mean or mode projection for QC.

  Training continues until the model converges (i.e. the value of the inverse variance β_{new} *stabilizes)*

APPLICATIONS

GTM example 1: a reservoir engineering application

In our first example we wish to determine which combination of completion process and reservoir properties are correlated with high and low expected ultimate recovery (EUR). The test dataset consists of 137 horizontal wells from shale survey displayed in the map view in Figure 5.5. Each well is color-coded by its scaled EURs. In contrast, the rapidly varying EUR in the center of the map is less likely to be a reservoir property and more dependent on the specific engineering parameters. Geo-hazards such as faults will also exhibit anomalous engineering behaviors.

Our input data consists of 15 horizontal well parameters, which we hypothesize to affect the EUR of the wells. We train on 137 and validate with 8 wells. The 15 horizontal well parameters considered for analysis are:

- Total clean volume of sand,
- Total proppant volume,
- Total 100 mesh sand,
- Total non-100 mesh sand,
- Daily peak rate,
- Cluster spacing,
- Number of fracture stages,

- Total perforations,
- Total perforation cluster,
- Total perforation length,
- Contour permeability,
- Average treating rate,
- Formation thickness,
- Porosity,
- Average proppant concentration,

All the variables are normalized using a z-score algorithm to minimize any bias due to units of measure. The 2D latent space is uniformly sampled using 144 points and forming a square grid of 12 x 12 points. These 144 latent space variables are mapped into the *N=15* dimensional data-space using four Gaussian basis functions with equal variance. After GTM training the manifold in the data space will be stretched in regions of low-data density and compressed in the region of high data density. With the trained GTM model parameters, the posterior probabilities $R_{nk}(\mathbf{W}_{new}, \beta_{new})$ of the data-vectors are calculated using Bayes' theorem (equation 9). We then use these posterior probabilities and project them onto the 2D latent space to form either a mean and or a mode distribution map using equations 10 and 11. Finally projected points are colored by the scaled EUR values. Once trained we validate the clustering using the estimated EURs from the GTM property map and compare with the true EURs for the 8 wells not used in training.

The mean (Figure 5.6) or the mode (Figure 5.7) of the posterior probability (*responsibilities*) distribution map of every data-vector is plotted in the 2D latent space before and after 100 iterations. The color bar represents scaled EURs. Note that the data-vectors from wells with high EUR projection map onto the upper right corner of the latent space.

Some of the wells from the mean posterior probability distribution map are analyzed as highlighted in Figure 5.8a. The upper set correspond to the set of wells having high EURs and the below set corresponds to the wells having low EURs. The average of the normalized well parameters for these two set of wells are plotted in Figure 5.8b. The average data-vector from wells with high EURs (in red) are having different characteristics compared to the wells with low EURs. The wells with good EURs have higher proppant, sand volume, less cluster spacing, higher fracture stages, more perforations, and higher porosity, whereas the wells with bad EURs have opposite characters.

These posterior probabilities are crucial to our application of the GTM for EUR predictions. To begin, we assume that we have a property E_n (which is nothing but the EURs) that is associated with the n^{th} well. We can multiply this E_n to the posterior probabilities of the n^{th} well data-vector which are projected on the 2D latent space (Figure 5.9a). From this we get the EUR map for 1 well (Figure 5.9b). Then, we can formulate a weighted sum of the EUR at each grid point k in the latent space for all the N wells given by Equation (9).

$$\tilde{q}_k = \frac{\sum_n^N R_{nk} E_n}{\sum_n^N R_{nk}} \qquad (9)$$

This will give us a most likely EUR map from all the wells over the latent space (Figure 5.9c). We first should analyze this EUR map to qualitatively determine if there is significant correlation between the latent space and the physical property. Notice for our EUR map, the latent space highly correlates with the EUR values. In addition, we can also use this new $\tilde{q}_k$ to predict the EUR at another location in the latent space $\hat{E}_m$, which is not present in the training data set. This is done by calculating the posterior probability V_{km} of the new m^{th} data-vector and multiplying the posterior probability value at each grid point k with the weighted sum of the EUR values. Thus, for predicting the EURs $\hat{E}_m$ we use,

$$\hat{E}_m = \sum_{k=1}^{K} V_{km} \, \tilde{q}_k \qquad (10)$$

where, the indices *m* representing a validating well.

Figure 5.9d is the plot of the predicted EURs from the GTM model vs. the true given EURs for the validating sets of 8 wells with a very good linear least-square fit.

This study shows how different horizontal well parameters affect the average EUR for horizontal wells. The horizontal well parameters like the sand used for fracturing, the proppant volume, the number of perforations, their cluster spacing, the porosity mostly affects which wells have the good or bad EURs. The model parameters used in our GTM model fit the dataset properly and correlates the latent space to the EURs of each well. A very good least square fit of the predicted EURs to the actual EURs gives confidence to our model. This workflow can be conveniently extended to estimate which of the parameters most affects the EURs of the horizontal wells in a shale play.

Next, we will apply this GTM algorithm to do multi-attribute seismic facies classification for the Mississippian Barnett Shale of the Fort Worth Basin in Texas.

GTM example 2: Probabilistic Multi-attribute seismic facies analysis of the Barnett Shale from the Fort Worth Basin

The Fort Worth Basin is a foreland basin, located in north-central Texas and is associated with the late Paleozoic Ouachita orogeny. This basin is bounded by the Muenster Arch to the northeast, the Ouachita Thrust Front to the east, the Bend Arch to the west, the Red River Arch to the north, and the Llano Uplift to the south (Figure 5.10a). In this study area the Barnett sits on an angular unconformity above the Cambrian to upper-Ordovician-age carbonates of the Ellenberger Group and Viola Formation and overlying Pennsylvanian-age Marble Falls Limestone (Figure 5.10b). In between, the Forestburg Limestone divides the Barnett formation into Upper and Lower Barnett zone (Figure 5.10c). The Barnett Shale is not homogeneous, but rather can be subdivided into siliceous shale, argillaceous shale, calcareous shale, and limestone layers, with minor amounts of dolomite (Singh, 2008).

The Fort Worth Barnett shale gas play is traditionally more of an engineering driven play. It requires hydraulic fracturing for gas production. Our 3D seismic survey consists of a 200 square mile survey in the North East Fort Worth Basin. The data are sampled at 110ft by 110ft by 2ms. However, the above survey was acquired after numerous vertical and horizontal wells have been drilled and hydraulically fractured. For effective well placement within the survey in future drilling, care should be taken to identify the brittle zones by mapping the geomechanical rock type of the Barnett shale.

GTM Workflow for Seismic Facies Estimation

The inputs to our GTM algorithm are different seismic inversion volumes (P-impedance, lambda-rho, mu-rho) which help in understanding the highly heterogeneous nature of the Barnett shale. For the above attribute generations the seismic data between the Marble faults horizon and the Viola limestone is considered. The impedance volumes better reflect a heterogeneous shale reservoir based rock type. The Lamé parameters of seismic inversion such as lambda-rho ($\lambda\rho$) and mu-rho ($\mu\rho$) correlates to "fracability" and different elastic properties of rocks. Simple cross-plot between two such elastic properties from the wells sometimes help in segregating different rock types. However, it is very difficult to separate between classes when we cross-plot any of these two seismic volumes. Figure 5.11 shows the cross-plot of the lambda-rho ($\lambda\rho$) vs. mu-Rho ($\mu\rho$) for the Upper Barnett shale (Figure 5.11a) and Lower Barnett shale (Figure 5.11b). Both the cross-plots form one single cluster of data-points and we are unable differentiate this cluster. The same is true when we cross-plot with the P-impedance volume. Thus it is not possible to identify the heterogeneity in shale just by simple cross-plot.

The above three volumes: P-impedance, $\mu\rho$ and $\lambda\rho$, between the Lower Marble Falls and the Viola limestone, are considered as input for the GTM unsupervised clustering analysis. Thus our data become 3-dimensional. However this data with different amplitude values needs to be normalized first, which is done by subtracting off the mean and then dividing it by the standard deviation for each of the three volumes (z-score algorithm).

Initially a 2D latent space is uniformly sampled with $K = 1600$ number of regular grid points (nodes) and square grid is defined with 40 grid points in each side. An additional square grid of $J = 144$ defined the centers of the non-linear basis functions in the same 2D space as the mapping function. These non-linear Gaussians basis functions have the same standard deviation equals to half the separation of the neighboring basis function centers. A multi-attribute PCA analysis is done to initialize the weight matrix $\mathbf{W}$ and the inverse of the noise variance β of the GTM model. The training dataset to the GTM consists of the decimated P-impedance, $\mu\rho$ and $\lambda\rho$ volumes. $\mathbf{W}$ and β, are updated with each iteration through expectation-maximization (EM) algorithm. The EM algorithm is guaranteed to converge to local maxima and possibly a global maxima (Dempstar *et al*, 1977).

We stopped the training after 100 iterations when the value of the inverse variance β stabilized (Figure 5.12a). After training the latent space the final GTM model is applied to the original normalized dataset, which is then projected onto this 2D latent space (Figure 5.12b). This projected mean posterior probability distribution in the 2D latent is then typically used for user defined cluster analysis. An innovative and easy workflow to do this is to store the values of the posterior probability projections for each of the 2D latent space axis into two separate volumes, which we call as GTM projection volumes. Thus, the GTM output consists of two separate volumes. Four different zones are identified in the gamma-ray log from a well within the survey. The GTM projection volumes constrained in the four zones within the Barnett formation are cross-plotted in commercial interpretation software. The mean posterior projections of these four zones form separate clusters onto the latent space (Figure 5.15). These different clusters formed are then interpreted by qualitatively by visualization and the micro-seismic dataset from the wells separated them into "fracable" or ductile lithofacies. Finally, we came up with 9 clusters within the Barnett shale.

Discussions of Seismic Facies Analysis on Barnett Shale

For identifying different clusters, we have analyzed several well-probes within the Barnett shale interval combining them with the microseismic dataset. Microseismic is one of the most important fracture diagnostic techniques to map the effectiveness of fractures due to hydraulic fracturing. Microseismic records the slippage events that accompany a hydraulic fracturing and then locates the origin of the events in the subsurface.

The well locations are given in Figure 5.16. Microseismic events from Well A and a suite of Wells B (B1, B2, B3, and B4) are discussed in this study (Figures 5.13 and 5.14). Well-probes are created by cross plotting the two projection volumes from GTM. Next, we identify, draw polygons around the clusters and assign them with different colors. Figure 5.13 shows the well-probe for well A. The 2D histogram (Figure 5.13a) shows the occurrence of various clusters over which eight user-defined polygons as shown (Figure 5.13b). The well-probe in Figure 5.13d is colored accordingly. The Upper and the Lower Barnett Shale exhibit different cluster composition and in turn different from the intervening Forestburg limestone in gray color. The Upper Barnett is mostly identified in three clusters which are colored blue, cyan and yellow. The most variability is found in the Lower Barnett Shale with 4.5 different clusters formed in the 2D latent space. Figure 5.13d shows the well-probe overlaid with the microseismic events. Slatt et al, (2011) identified 1st order, 2nd order and 3rd order ductile-brittle couplets within the Barnett Shale (Figure 5.13c). It is observed that more microseismic events are confined in the light green and the red lithofacies of the Lower Barnett Shale and less microseismic events penetrate into the brown facies. The well-probe corresponds to the 2nd order brittle ductile couplets. The arrow highlights the ductile and brittle zone in the well probe and in the 2D histogram plot.

The suite of Wells B forms similar well-probe after we do *a posteriori* coloring of the 2D latent space consistent with Well A. The Upper Barnett consists of the similar three facies as before and the Forestburg limestone is colored gray. Analyzing the 2D histogram (Figure 5.14a) we can add one more class (pink) as shown in Figure

5.14b. The well-probe (Figure 5.14c) is colored with these nine user-defined clusters. The microseismic data are added for all the four wells (Figure 5.14d) and it shows the presence of most of the microseismic events confined in the light green, red and pink colored lithofacies. Again, it is observed the microseismic events miss the brown lithofacies. Thus light green, red and pink lithofacies are brittle in nature compared to the brown colored facies, which is more ductile. This is consistent with the 2nd order brittle ductile couplets (Slatt, et al, 2011) and the studies done by Perez (2013). Similarly the arrow highlights the ductile and brittle zone in the well probe and in the 2D histogram plot.

After analyzing the well-probes we extend our interpreted clusters to the seismic volume. The GTM projection volumes are cross-plotted for four different zones with their boundaries defined above and below by stratal slices corresponding to the zones defined in the gamma-ray log from Figure 5.15a. Figure 5.15 b, c, d and e show the mean posterior probability projections of these four zones (B, C, D, and E) within the Barnett shale formation. Different clusters are recognized in the cross-plot and are colored, consistent with the clusters and colors of the well-probes.

A horizon-probe within the upper Barnett shale (zone B) shown in Figure 5.16c. The 2D histogram and the user defined clusters of the data from zone B are given in Figure 5.16a and b. Three clusters are formed in the upper half of the 2D latent space. Which are then colored consistent with the well-probes. Most of the facies of the Upper Barnett Shale falls into the blue, green and the yellow clusters.

For the Lower Barnett Shale three different horizon-probes (Figure 5.17, 5.18 and 5.19) are created based on the different zones C, D and E identified in the gamma ray signature of the lower Barnett shale (Figure 5.15 a). Figure 5.17c shows the top section of the Lower Barnett Shale (zone C). It is observed in the mean posterior probability projections (Figure 5.15c and Figure 5.17a) that it has some similar nature as the Upper Barnett Shale but also have different rock type since a group of clusters are identified in the lower part of the latent space.

Figure 5.18c is the horizon-probe for the middle zone D of the Lower Barnett Shale. The cluster occurs mostly in the lower half of the latent space (Figure 5.15d and Figure 5.18a). User-defined polygons are drawn around these clusters (Figure 5.18b) to color the seismic horizon-probe. Earlier we discussed from the well-probes that the microseismic events are mostly concentrated in the pink, light green and the red color facies. Also this zone mostly has siliceous non-calcareous shale lithofacies (Singh, 2008). Thus we infer that zone D with mostly pink, light green and the red color lithofacies indicates good fracturing brittle shale.

The lowermost zone (zone E) of the Lower Barnett Shale is the zone of hot shale (Pollastro et al, 2007) with very high gamma ray (Figure 5.15a). Figure 5.19c is a horizon-probe from the lower most zones. The brown facies is the most abundant facies in this zone. From the high gamma values in the well logs and the fewer occurrences of the microseismic events we interpret the brown facies as the most ductile shale which may be due to high TOC concentration in this zone (Singh, 2008 and Perez, 2013).

CONCLUSIONS

The two case studies, first on the horizontal well dataset in an unconventional shale survey and the second on the 3D seismic data from Barnett shale discusses the various workflows for clustering with the GTM algorithm. For the reservoir engineering problem we had a high correlation of the latent space model with the horizontal well parameters. The horizontal well parameters like the sand used for fracturing, the proppant volume, the number of perforations, their cluster spacing, and the porosity mostly affects EURs of the wells. Finally we got a very good prediction model predicting EURs for our blind test wells. This unsupervised clustering workflow can be conveniently extended for analyzing horizontal well parameters in other shale plays.

Secondly the GTM application in the Barnett shale with the P-impedance, $\lambda\rho$ and $\mu\rho$, well characterizes the brittle and the ductile zones present in the shale. Each of the shale lithofacies is tied to the microseismic data set from the wells of the survey. From these microseismic events we can interpret clusters and differentiate between the more "fracable" and ductile shale. Nine set of lithofacies have been identified within the Barnett formation from the GTM unsupervised cluster analysis. Within the four/five lithofacies identified in the Lower Barnett Shale we conclude that the brown facies is more ductile in nature and probably more TOC rich. With larger concentration of the microseismic events the red, light green and the pink lithofacies are identified as more brittle in nature. This study will help in identifying the sweet spots within the survey to identify zones for placing new horizontal well effectively.

Because the GTM theory is deep rooted in probability, other than seismic facies classification it can also be applied into modern risk analysis and used in decision making. We can extend the GTM application in seismic exploration by projecting the mean posterior probabilities of a target data in the 2D latent space and then calculate the probability estimates of the data falling in the category of the target vector. We thus have a probabilistic estimate of the facies type and allows add another dimension to the geologic risk analysis workflow using probabilistic facies classification output.

LIST OF FIGURES

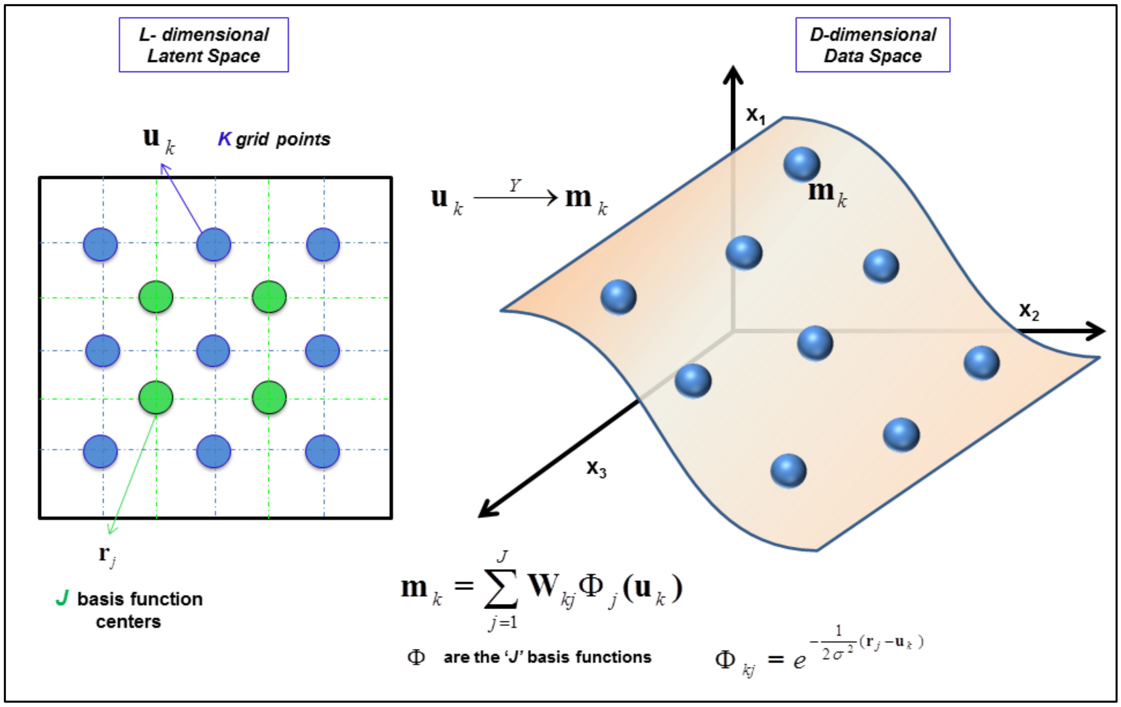


Figure 5.1. Non-linear mapping of the latent space to the data space. The prior distribution consists of latent space variables (K) ordered on a regular grid (blue circles) residing in an L-dimensional latent space. In this figure L=2. $\boldsymbol{\Phi}$ consists of a regular array of J non-linear basis functions. With the linear combination of these basis functions the latent space (blue circles) are mapped to the data-space (blue spheres) where they form a 2D non-Euclidean manifold, S, such that each node $\mathbf{u}_k$ is mapped to a corresponding vector $\mathbf{m}_k$ in data-space, given by $\mathbf{m}_k = \sum_{j=1}^{J} \mathbf{W}_{kj} \boldsymbol{\phi}_j(\mathbf{u}_k)$ (equation 2).

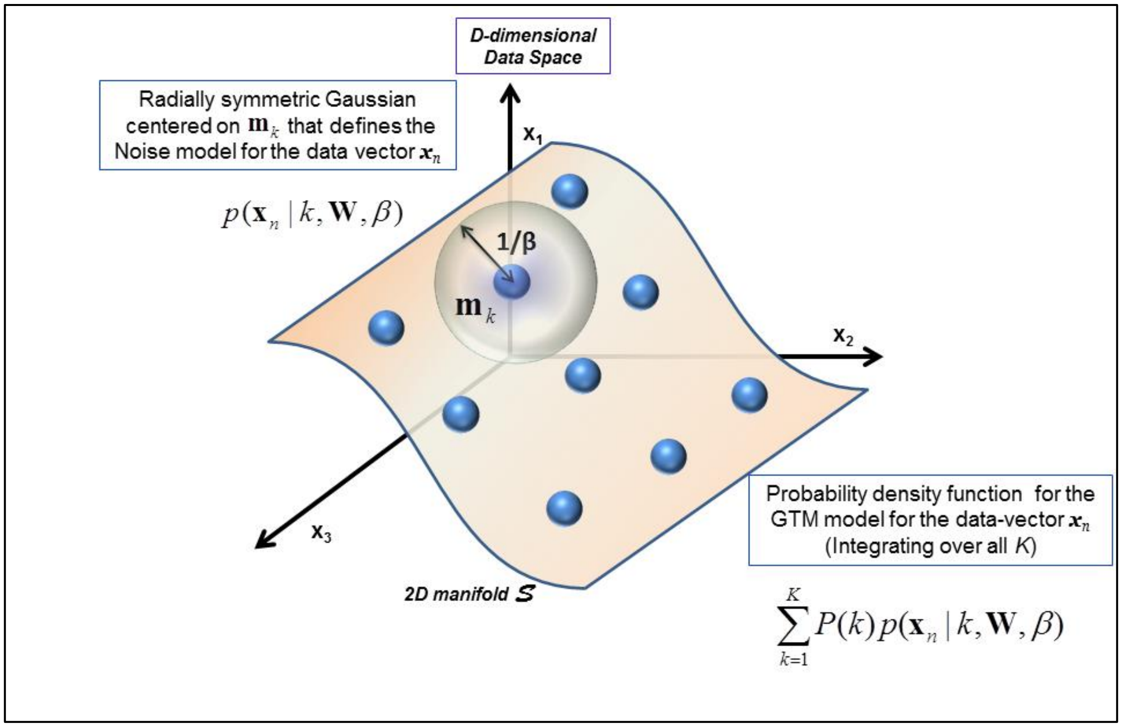


Figure 5.2. Generating Gaussian probability density functions (PDF) to represent the data vectors. In general data-vectors are scattered about $\mathcal{S}$. We assume the misfit represented by a suite of PDFs. Specially we assume an isotropic Gaussian noise distribution $p(\mathbf{x}_n|k, \mathbf{W}, \beta)$, centered at $\mathbf{m}_k$ and having a variance of 1/β (equation 3), defines each data-vector $\mathbf{x}_n$. The final probability density function in the data space is obtained by integrating the Gaussian PDFs for all $\mathbf{m}_k$ where $k=1, 2, \ldots K$ grid points $\sum_{k=1}^{K} P(k)\, p(\mathbf{x}_n|k, \mathbf{W}, \beta)$ (equation 4), where $P(k)$ is the prior distribution at each node in the latent space.

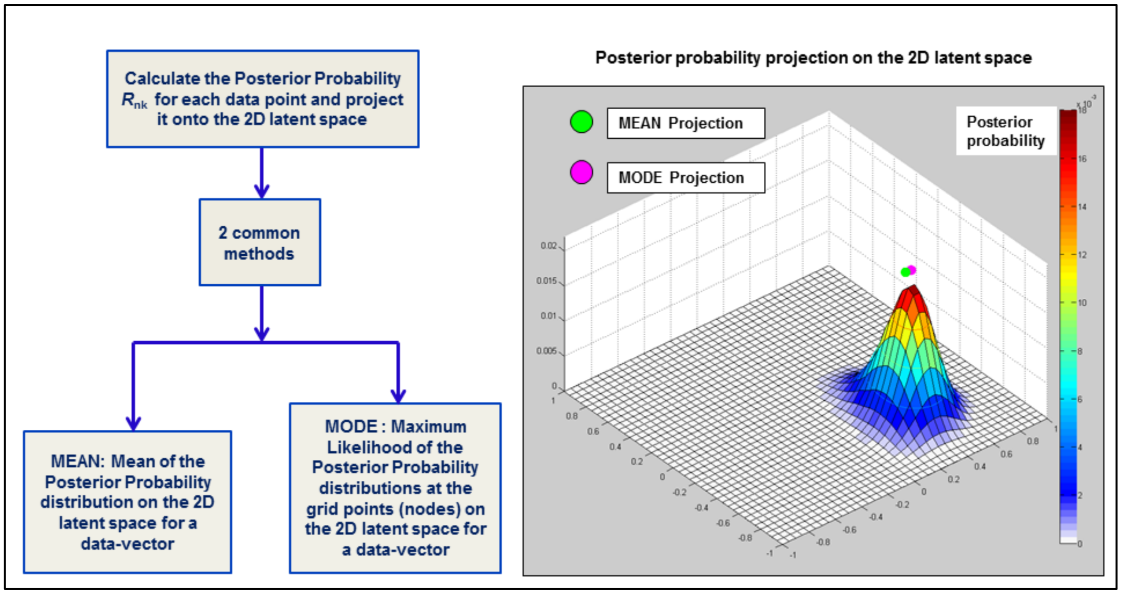


Figure 5.3. Workflow for data visualization in GTM. After training the new estimated parameters $\mathbf{W}_{\text{new}}$ and β_{new}, the new posterior probabilities representing the data-vectors can be obtained using Bayes' theorem and is given by $R_{nk}(\mathbf{W}_{new}, \beta_{new})$. These posterior distribution "responsibility" values can be plotted for all the grid points, $\mathbf{u}_k$ in the 2D latent space. The mean location will assign the value $\mathbf{U}_{mean}(\mathbf{x}_n)$ to be the weighted average of the posterior distribution values and will in general fall between neighboring values of $\mathbf{u}_k$. The mode will assign the value $\mathbf{U}_{mode}(\mathbf{x}_n)$ to be the location of the greatest posterior distribution value in the 2D latent space and will always correspond to a discrete gridded value of $\mathbf{u}_k$.

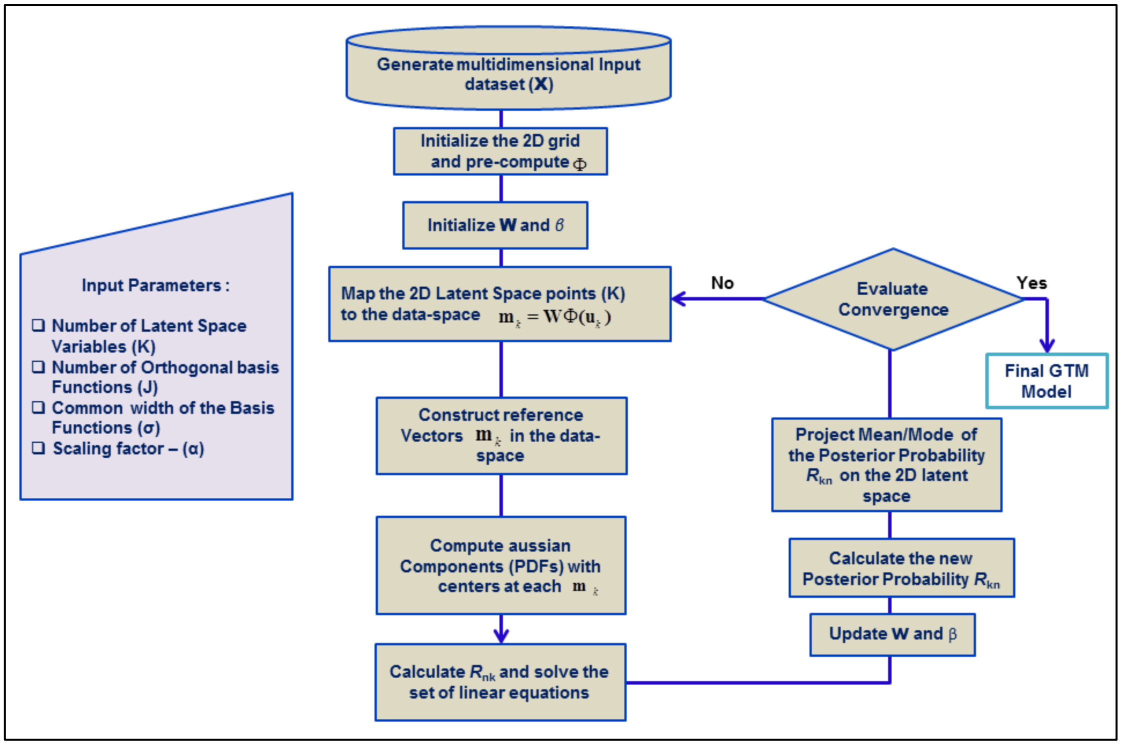


Figure 5.4. The flow chart for generative topographic mapping workflow.

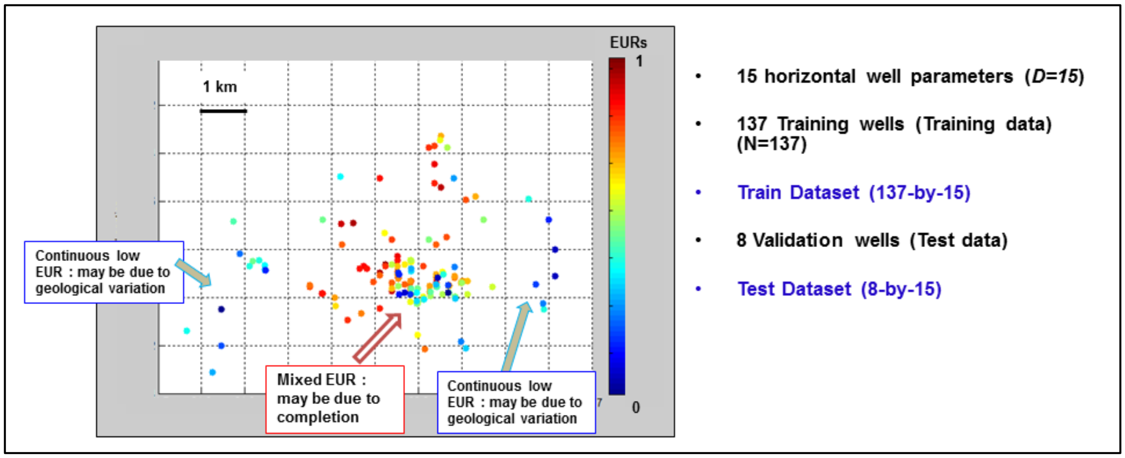


Figure 5.5. The Well Locations for the unconventional shale play of an area roughly 1000 km^2. Colors correlate to scaled EUR ranging between 0 and 1. 137 wells are used to train and eight wells used to validate the GTM. In general, the reservoir properties of the shale are smoothly varying laterally. However, near the center of the survey

note the proximity of high- and low-EUR wells. We attribute these variations to difference in completion practices, including stimulations of geohazard.

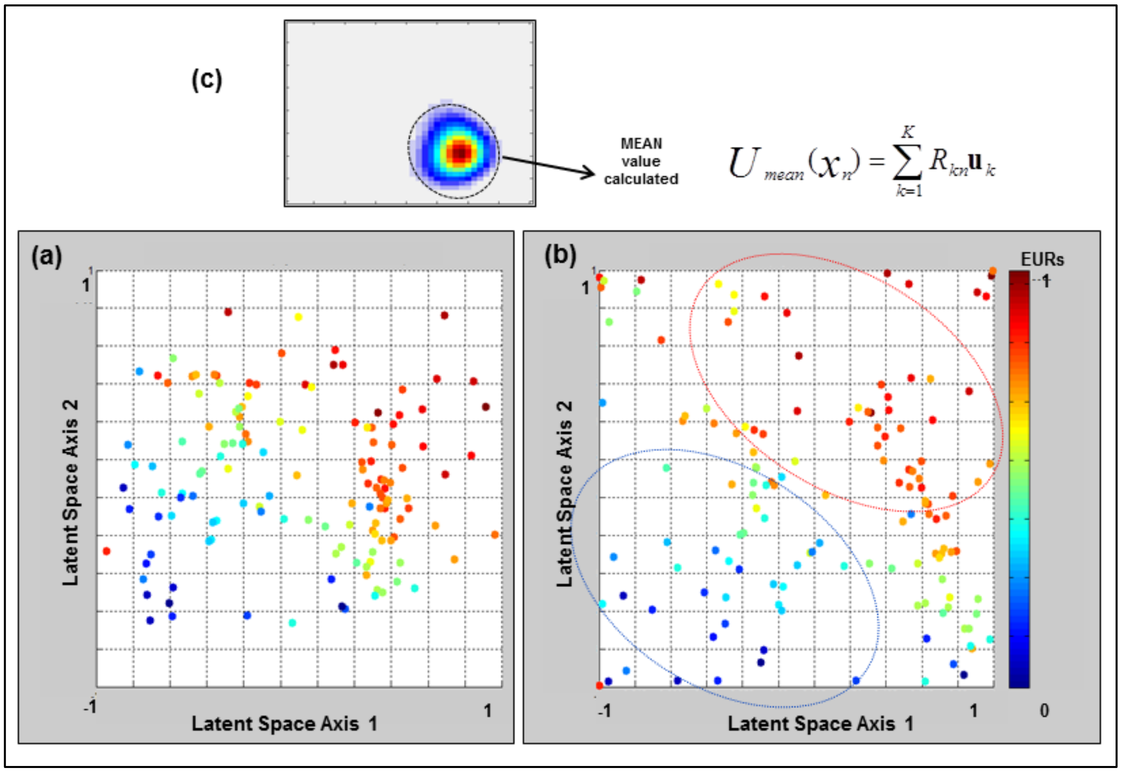


Figure 5.6. The mean posterior distribution map of the “*responsibilities*” of the data in the 2D Latent space. The mean location of a data-vector is the weighted average of the posterior distribution values and will in general fall in between neighboring values of $\boldsymbol{u}_k$. (a) Initial and (b) final distribution of the posterior mean projections of all 137 well data onto the latent space after 100 iterations. (c) The PDF of a representative well vector projected onto the latent space. While only the mean value is plotted in (a) and (b), the full distribution can be used to better quantify risk. The plot is color-coded

by the scaled EURs. Note the two different clusters corresponding to the good (red) and the bad EURs (blue) in the final mean distribution map.

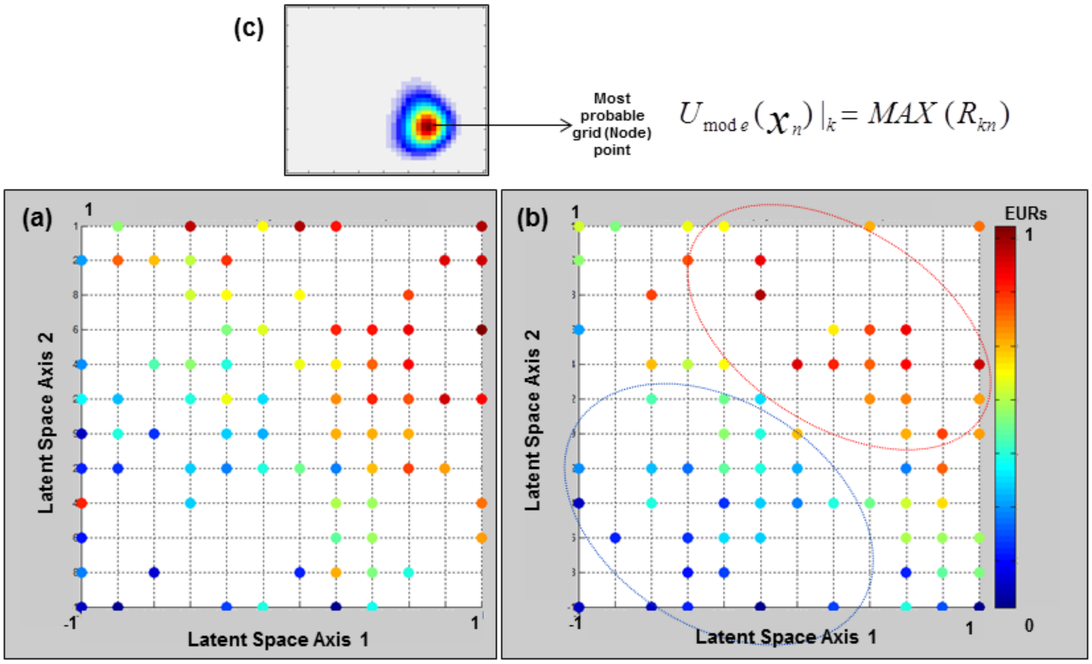


Figure 5.7. The same GTM classification shown in Figure 5.6, but now using the mode of the posterior probability projections of the data. (a) Initial and (b) final distribution of the posterior mode projections of all 137 well data onto the latent space after 100 iterations. (c) The data-vectors are projected onto the most likely grid points (grid points with the most posterior probability value) and will always correspond to a discrete gridded value of $\mathbf{u}_k$. Note that in this mode projection the posterior probability values are assigned only to the grid locations. The plot is color-coded from low to high EUR values. The latent space shows a more orderly separation between the good, moderate and the bad EURs for the final iteration in the mode distribution map.

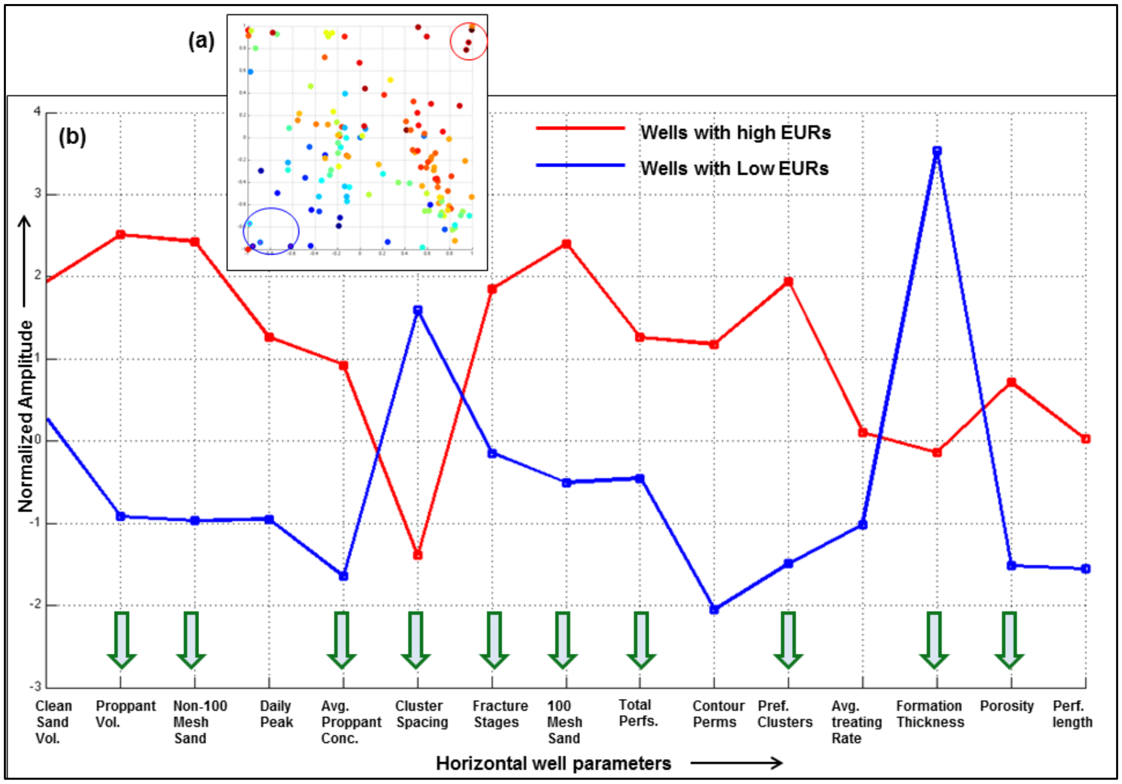


Figure 5.8. (a) Some of the wells from the mean posterior probability distribution map are analyzed as highlighted. The upper corner corresponds to wells with high EUR and the bottom corner corresponds to wells with low EUR. (b) The 15 normalized well parameters for each of the two set of wells are averaged and are plotted forming two sets of averaged data-vectors. The red corresponds to the average data-vector for wells with good EURs and blue is the average data-vector for wells with low EURs. Note that mostly the well parameters differ radically for the two cases. The wells with good EURs have higher proppant, sand volume, less cluster spacing, higher fracture stages, more perforations, and higher porosity, whereas the wells with bad EURs have opposite characters (highlighted with arrows).

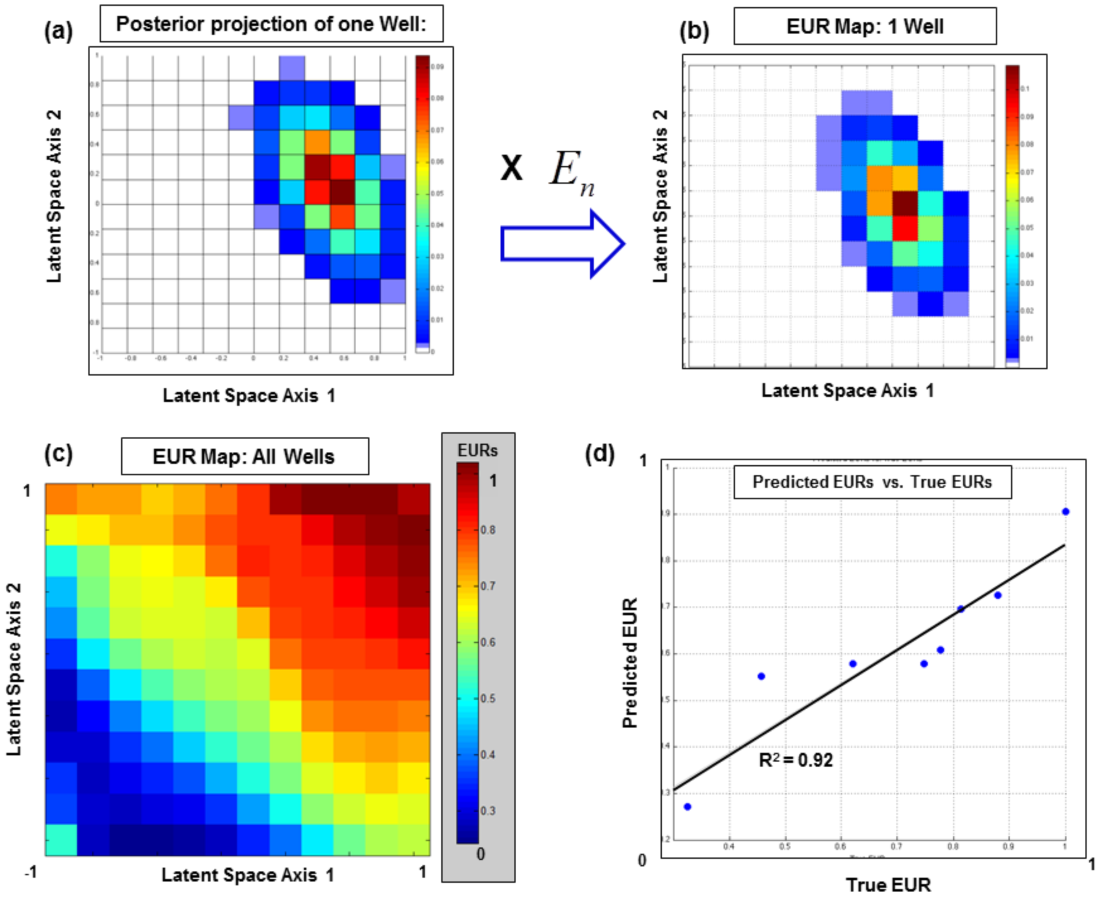
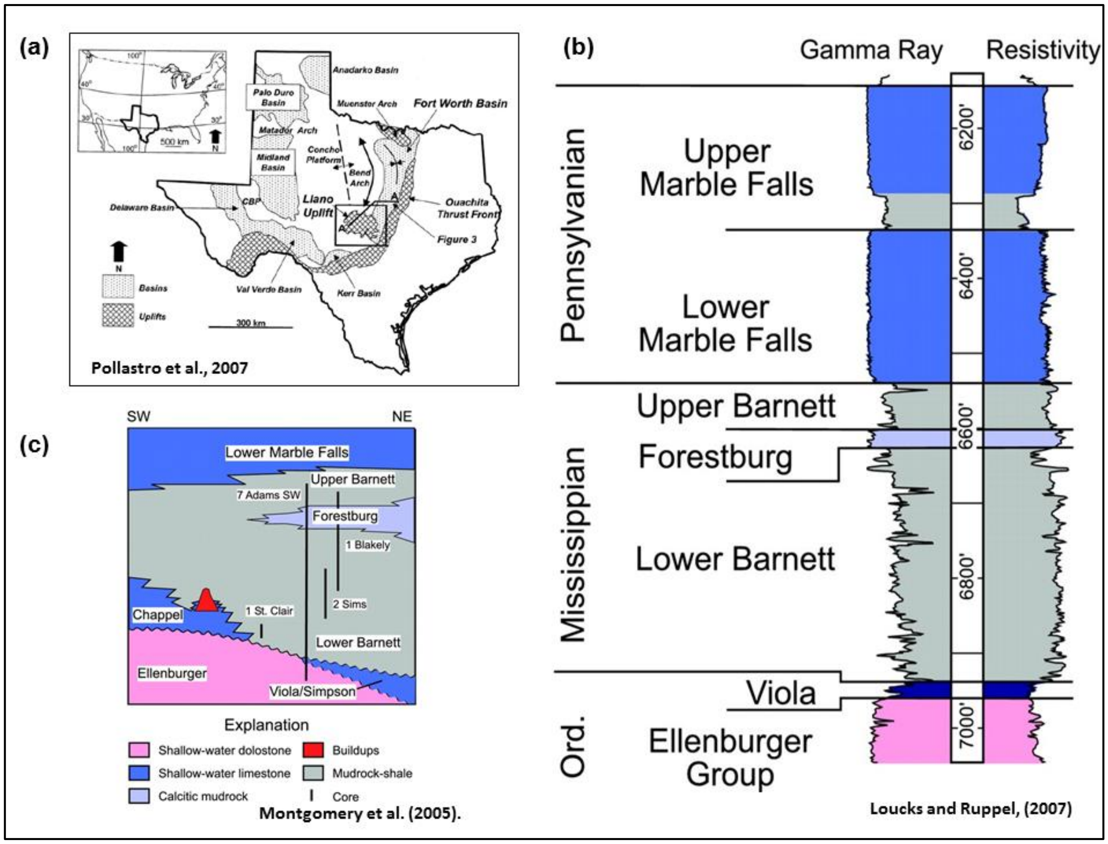


Figure 5.9. EUR prediction through GTM modeling. (a) The posterior probability of the data-vector from the n^{th} well. (b) The EUR for the n^{th} well, E_n is multiplied with the posterior projection values onto 2D latent space in (a). The result gives an EUR map for 1 well (Figure 5.9b). Then, we can formulate a weighted sum of the EUR at each grid point k in the latent space for all the wells (given by Equation 9) and form the EUR “map” over the latent space. Note the high correlation of the latent space with the EURs. (b) Plot showing the predicted EUR from the EUR property map in (a) vs. the true EUR for the 8 validation wells not used in training the GTM. The least square fit shows and excellent correlation.



Pollastro et al., 2007

Montgomery et al. (2005).

Loucks and Ruppel, (2007)

Figure 5.10. (a) The map of Texas highlighting the Ft. Worth basin and other major basins and uplifts. The Ft. Worth basin is bounded by the Muenster Arch to the northeast, the Ouachita Thrust Front to the east, the Bend Arch to the west, the Red River Arch to the north, and the Llano Uplift to the south. **(b)** Stratigraphic section including the gamma ray and resistivity logs showing the major units. The Barnett Shale sits on an angular unconformity above the Cambrian to upper-Ordovician-age carbonates of the Ellenburger Group and Viola Formation and is overlain by the Pennsylvanian-age Marble Falls Limestone, and is divided into Upper and Lower units

by the Forestburg Limestone (c) Cross-section of the stratigraphy of Ft. Worth basin from Montgomery et al. (2005).

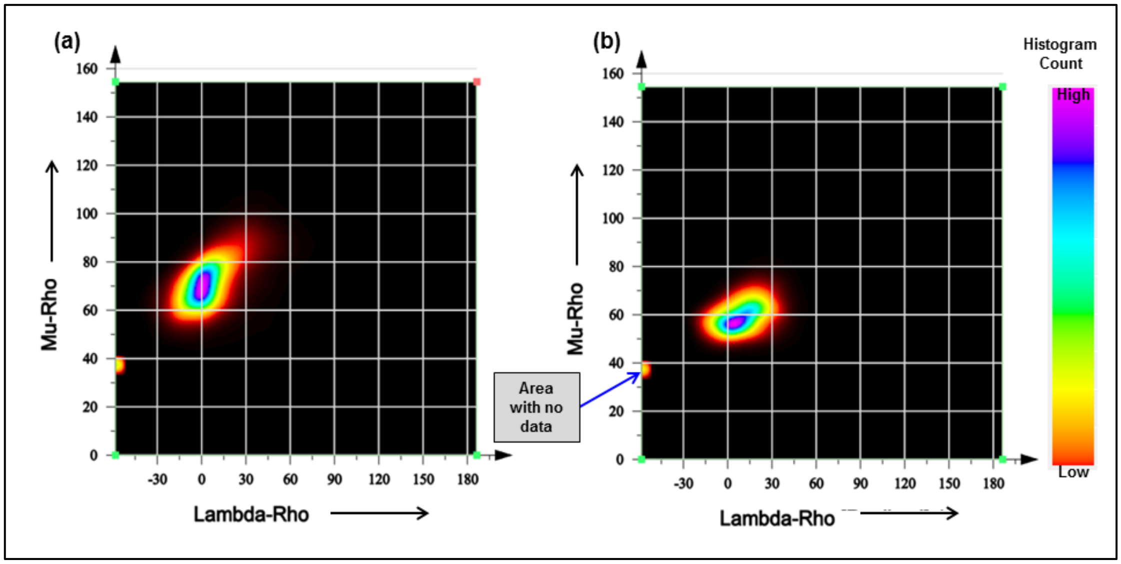


Figure 5.11. 2D histogram obtained by cross-plotting the mu-rho vs. lambda-rho volumes. The cross-plot shows a single cluster for **(a)** the Upper Barnett Shale **(b)** Lower Barnett Shale. A small cluster (marked with an arrow) at the edge of the two plots corresponds to a no permit zone of the data.

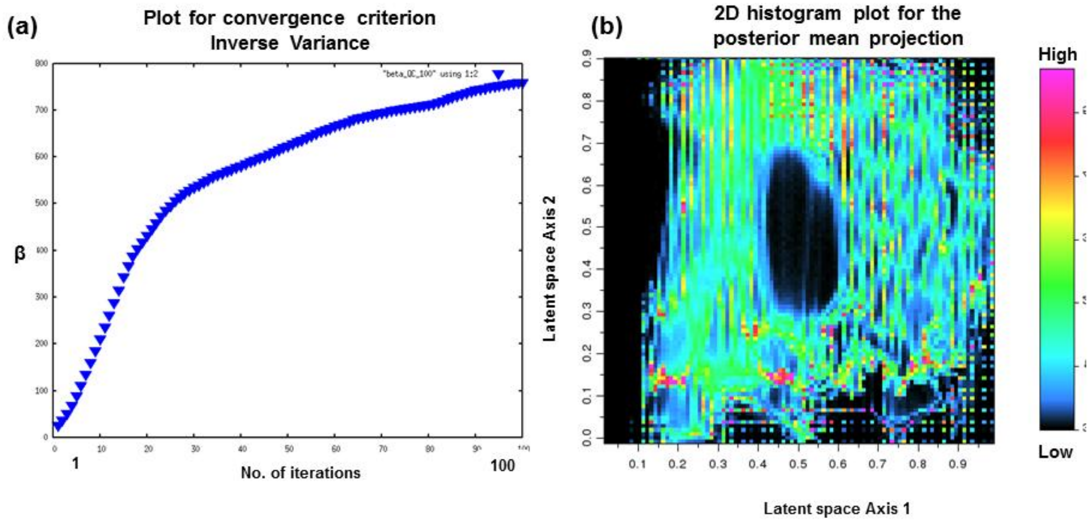


Figure 5.12. QC of the GTM analysis after performing 100 iterations. **(a)** The plot of the inverse variance (β), stabilizing as the number of iterations increases. **(b)** A 2D histogram of the mean projections of posterior probability of the dataset defined above and below by the Lower Marble Falls and the Viola limestone respectively

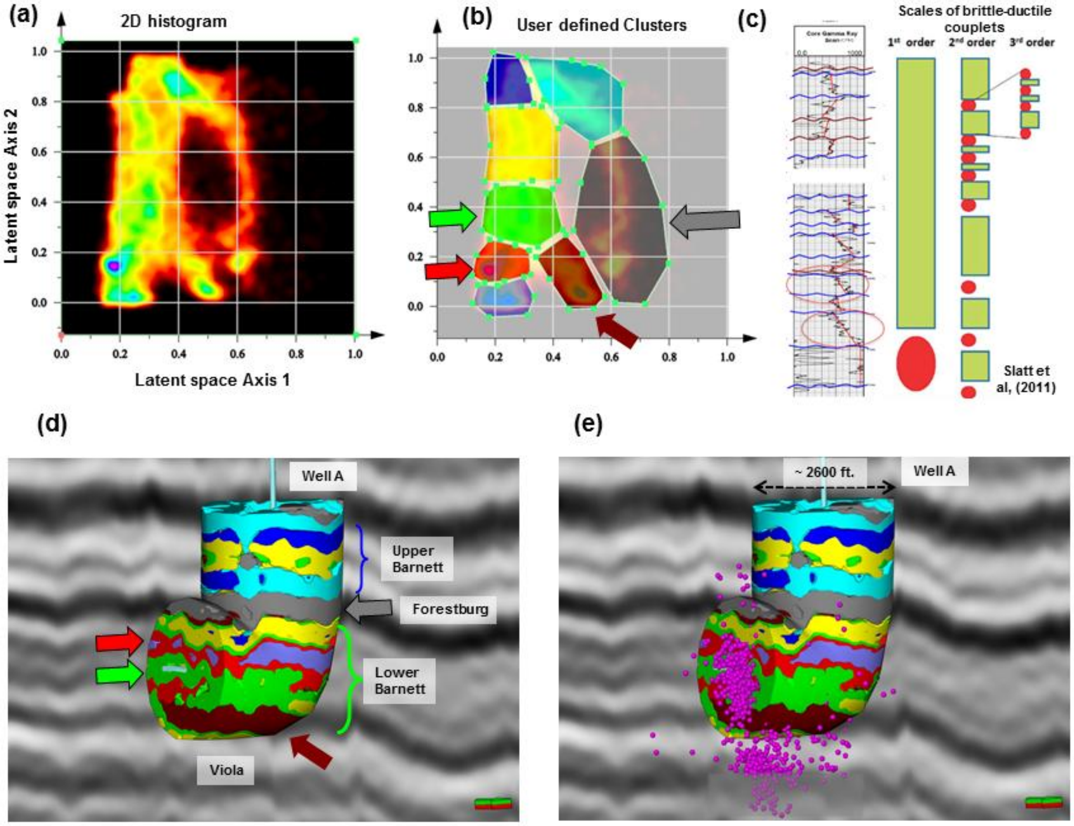


Figure 5.13. Well-probes for Well A generated by cross-plotting the two GTM projection volumes. **(a)** The 2D histogram generated from the cross-plot of the two GTM projection volumes. **(b)** Eight user-defined polygons drawn around the clusters seen in (a). **(c)** Brittle-ductile couplets proposed by Slatt et al, (2011). **(d)** Well-probe data colored by the clusters selected in (b). The Upper Barnett, the Lower Barnett exhibit a different cluster composition and are in turn different from the intervening Forestburg Limestone (in gray). **(e)** The microseismic events from this well are plotted along with the well-probe. Note the microseismic events are more localized in the red and light green facies and misses the brown facies, thus the red and light green facies

are interpreted as brittle and brown facies to be ductile. The results are consistent with the 2nd order brittle-ductile couplets proposed by Slatt et al, (2011).

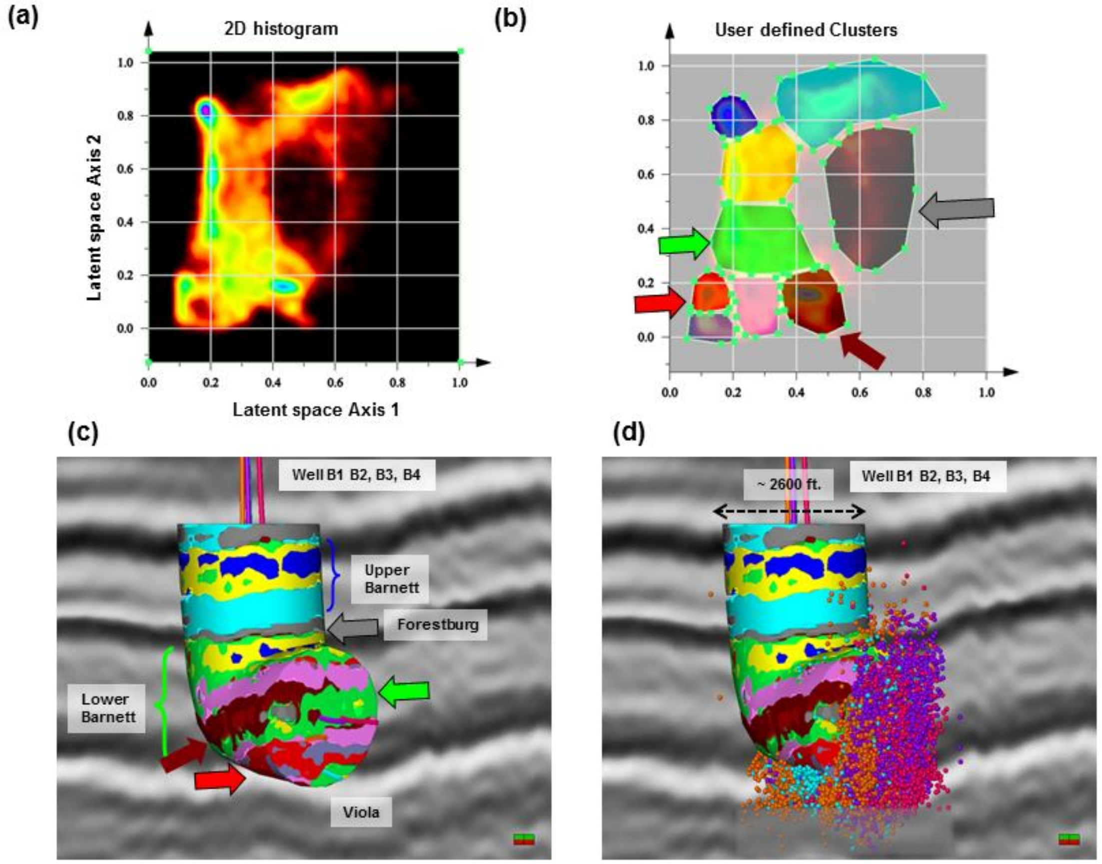


Figure 5.14. Well-probes generated for a suite of horizontal wells B1, B2, B3, and B4, generated cross-plotting the two GTM projection volumes. **(a)** The 2D histogram generated from the cross-plot of the two GTM projection volumes. **(b)** Nine user-defined polygons drawn about clusters seen in (a). **(c)** The well-probe data colored by the nine clusters. The Forestburg Limestone appears as gray and divides the Upper Barnett from the Lower Barnett Shale. **(d)** The microseismic events from these wells are plotted along with the well-probe. Similar to Well A the microseismic events are more localized in the red, pink, and the light green facies and are interpreted as brittle

facies and the brown facies has less microseismic events and are interpreted as ductile facies. Here also the results are consistent with the 2nd order brittle-ductile couplets (Slatt et al., 2011).

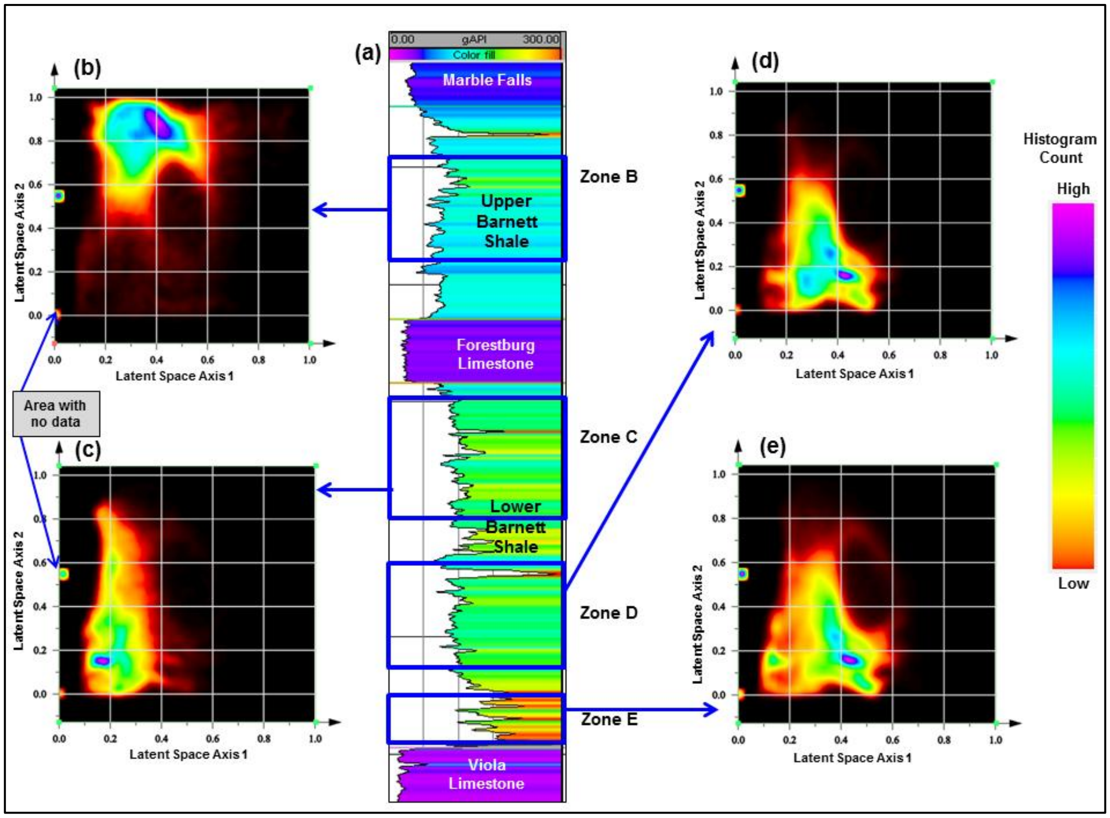


Figure 5.15. (a) Four different zones within the Barnett Shale selected using a gamma ray log from a well within the survey. The corresponding 2D histograms of the mean posterior probability projections for **(b)** the Upper Barnett (zone B), **(c)** the top of the Lower Barnett (zone C), **(d)** the middle of the Lower Barnett (zone D) and **(e)** the bottom of the Lower Barnett Shale (zone E) are shown.

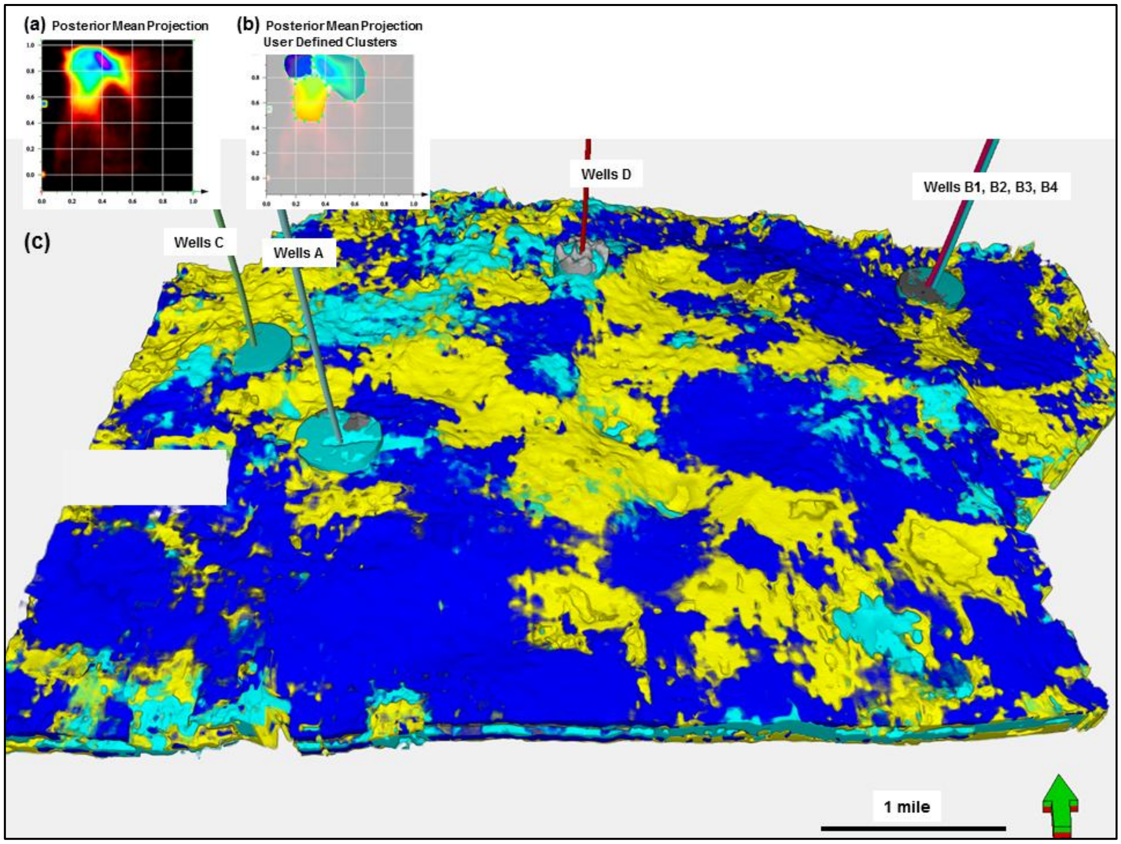


Figure 5.16. (a) 2d histogram of zone B defined above and below by stratal slices corresponding to the well log picks in Figure 5.15a. **(b)** User defined polygons are created and are colored consistent with the well-probes in Figures 5.13 and 5.14. **(c)** The facies volume probe of the Upper Barnett Shale zone B visualized along with the well-probes. The seismic dataset is colored accordingly to the clusters selected. Most of the facies of the Upper Barnett Shale falls into the blue, cyan and the yellow clusters.

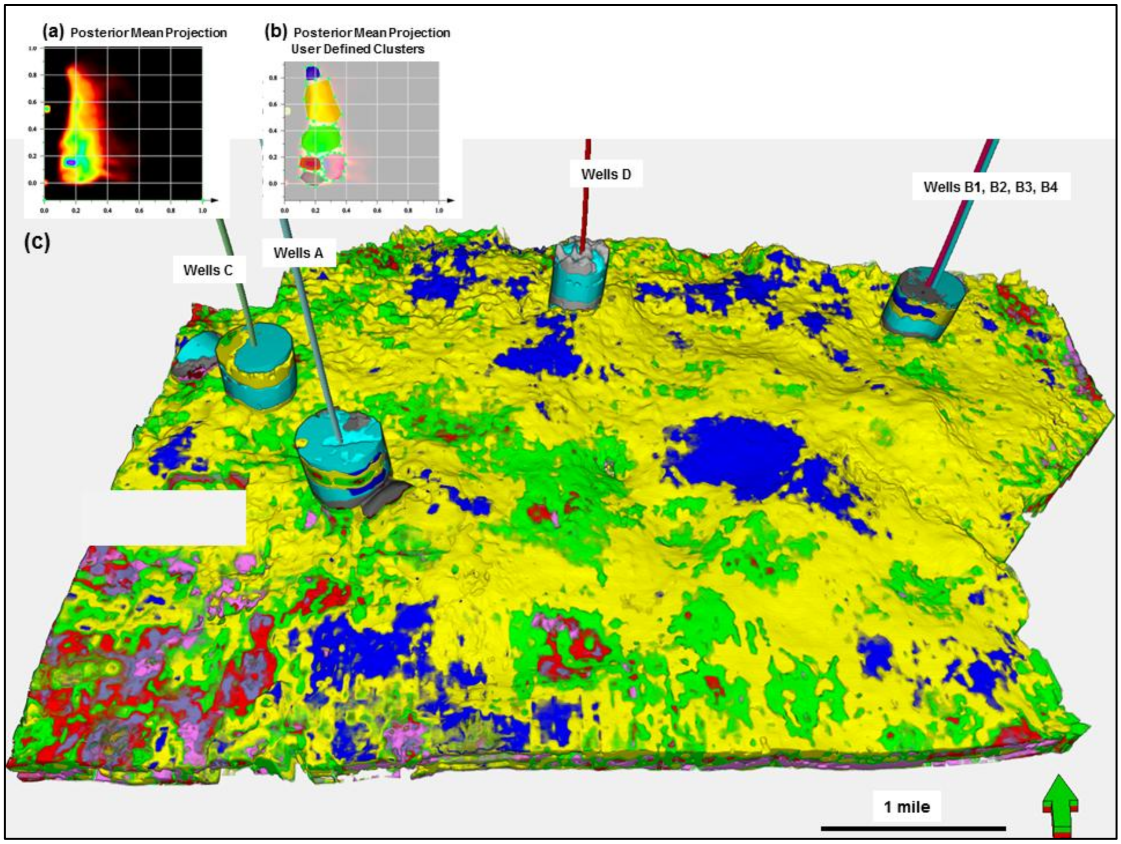


Figure 5.17. (a) 2d histogram of zone C defined above and below by stratal slices corresponding to the well log picks in Figure 5.15a. **(b)** User defined polygons are created and are colored consistent with the well-probes in Figures 5.13 and 5.14. **(c)** The facies volume probe of top section of the Lower Barnett Shale zone C visualized along with the well-probes. The seismic dataset is colored accordingly to these clusters selected. The yellow and blue facies is common in the Upper Barnett Shale is also found abundantly in this zone.

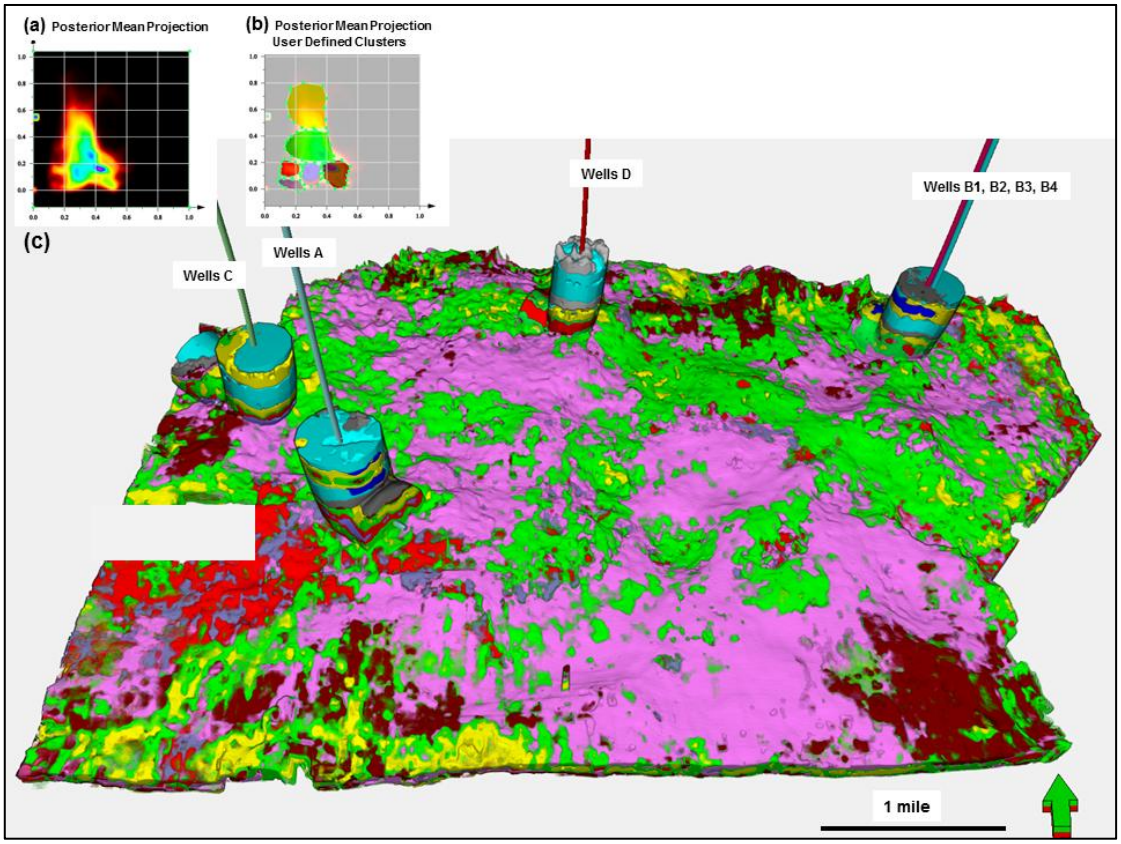


Figure 5.18. (a) 2d histogram of zone D defined above and below by stratal slices corresponding to the well log picks in Figure 5.15a. **(b)** User defined polygons are created and are colored consistent with the well-probes in Figures 5.13 and 5.14. **(c)** The facies volume probe of middle section of the Lower Barnett Shale zone D visualized along with the well-probes with the colors selected according to the clusters in (b). Note that this zone has least similarity to the Upper Barnett Shale. With more of the microseismic events concentrated in the pink, light green and red facies as seen in the well-probes, and the dominance of siliceous non-calcareous shale lithofacies (Singh, 2008), this zone 3 is interpreted as brittle with results consistent with Perez (2013).

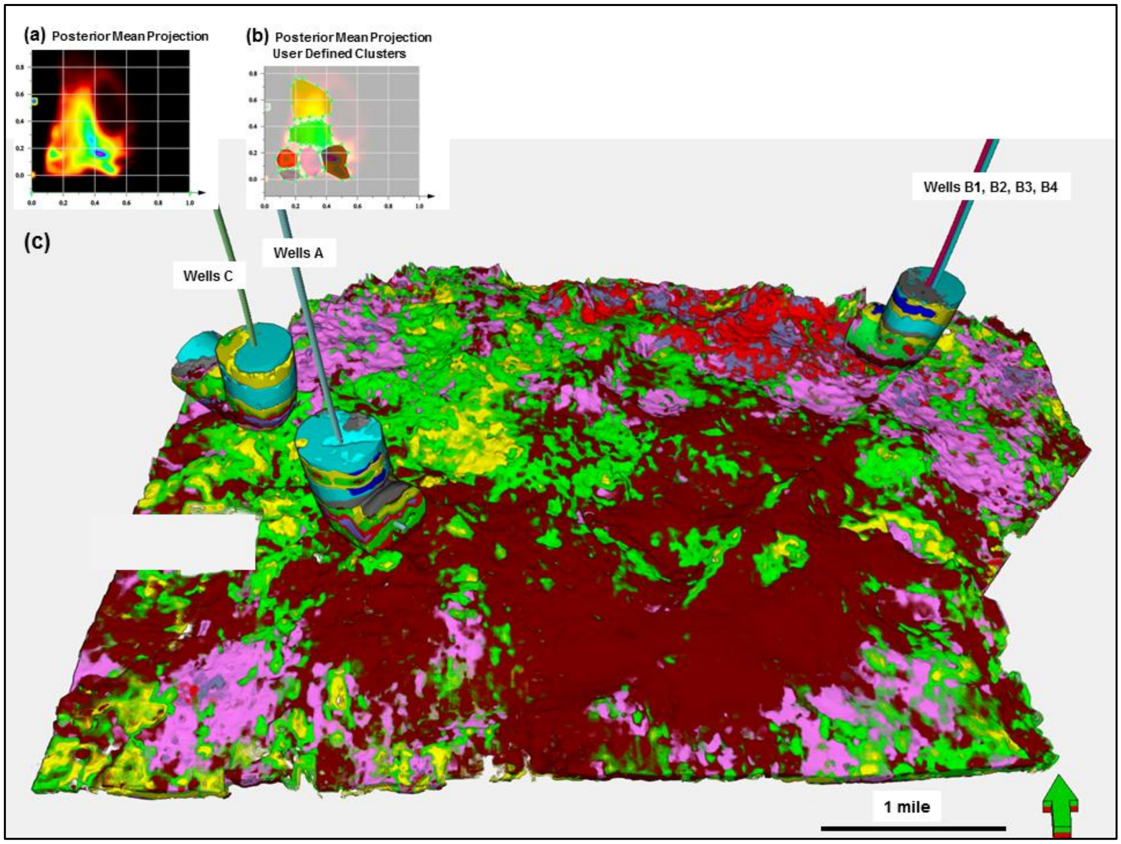


Figure 5.19. (a) 2d histogram of zone E defined above and below by stratal slices corresponding to the well log picks in Figure 5.15a. **(b)** User defined polygons are created and are colored consistent with the well-probes in Figures 5.13 and 5.14. **(c)** The facies volume probe of bottom section of the Lower Barnett Shale zone E visualized along with the well-probes with the colors selected according to the clusters in (b). This zone corresponds to the hot gamma ray zone (Pollastro et al., 2007). Six clusters are also identified from the mean posterior probability projections (in the top inset) are polygons are drawn and are colored consistent with the well-probes. With very few of the microseismic events in the brown colored facies we interpret from (Singh, 2008 and Perez 2013) the brown colored rock to be ductile and high in TOC content. The pink, light green and red facies are the regions with brittle shale.

REFERENCES

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 73rd Annual International Meeting Society of Exploration Geophysicists, Expanded Abstracts, p. 2221-2224.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, 10, No. 1, 215-234.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, Development of the generative topographic mapping: Neurocomputing, 21, No. 1, 203-224.

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, B, **39** no. 1:1 38.

Duda, R.O., P. E. Hart, and D. G. Stork, Pattern Classification, 2001, 2nd Edition, Wiley Publications.

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.

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

The GTM Matlab Toolbox v1.0 beta, M. Svensen, 1996.

Perez, Altamar, R., 2013, Correlation of surface seismic measurements to completion quality: Application to the Barnett Shale, PhD dissertation.

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.

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

Singh, P., 2008, Lithofacies and Sequence Stratigraphic Framework of the Barnett Shale, Northeast Texas: Ph.D. dissertation, University of Oklahoma.

Svensen, M., 1998, GTM: The Generative Topographic Mapping: PhD Dissertation, Aston University.

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

CHAPTER 6

Mulitattribute Generative Topographic Mapping for facies estimation of a carbonate wash, Veracruz Basin, Southern Mexico

Atish Roy, Araceli S. Romero, Tim J. Kwiatkowski, and Kurt J. Marfurt*

ABSTRACT

Seismic facies estimation is a critical component in understanding the stratigraphy and lithology of hydrocarbon reservoirs. With the adoption of 3D technology and increasing survey size, manual techniques of facies classification have become increasingly time consuming. Besides, the numbers of seismic attributes have also increased dramatically, providing increasingly accurate measurements of reflector morphology, but also greatly expanding the amount of data to be analyzed. This in turn leads to add a “dimension” to the data. Principal component analysis, self-organizing maps, and Generative Topographic Mapping (GTM) reduce such dimensionality by projecting the data onto a lower order space in which clusters can be more readily identified and interpreted.

In our case study we apply GTM to perform multi-attribute seismic facies classification of a carbonate conglomerate oil field in the Veracruz Basin of southern Mexico. The presence of conglomerate carbonates makes the reservoirs both laterally

and vertically highly heterogeneous, which is observed at well log, core slab and thin section scales. We apply unsupervised GTM classification to determine the “natural” clusters in the dataset. Finally we introduce supervision into GTM and calculate the probability of occurrence of seismic facies seen at the wells over the reservoir units. In this manner, we are able to assign a level of confidence (or risk) to encountering facies that correspond to good and poor production.

INTRODUCTION

Kohonen self-organizing maps (SOM) is currently the most popular unsupervised clustering technique in seismic exploration. Unlike K-means, SOM clusters are ordered, with similar clusters adjacent to each other. This similarity is visualized by mapping the clusters against 1D (Poupon et al. 1999, Coleou et al. 2003) or 2D (Strecker et al. 2002) color bars. However, there are several limitations to the SOM algorithm. First, there is no definite rule for selecting the neighborhood function, training radius, and learning rate because these parameters are data dependent. Second, there is no quantitative measure of “confidence” in the final clustering results because of the absence of any defined cost function to indicate the convergence in the final iteration.

For these reasons, Bishop et al. (1998) developed the generative topographic mapping (GTM) algorithm as an alternative to SOM. GTM finds a model based on a probability density function (PDF) that describes the distribution of the *D*-dimensional data in terms of a smaller number of latent variables, or cluster nodes that approximate the vast majority of the probability mass of the data (Sevensen, 1998). After iterative parameter estimation, this manifold (curved surface in the data space) allows us to relate the points in the D-dimensional data-space to grid points in the lower L-dimensional latent space. In general, GTM estimates the probability that any given data vector is represented by each and every cluster node, providing a direct link to risk analysis. To visualize the PDFs of large 3D seismic data volumes for interpretation, we can approximate each data vector by the mean or the mode of the PDF projected onto the latent space.

Wallet et al. (2009) were 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”. Roy et al. (2013) later expanded GTM implementation to cluster reservoir engineering completion parameters from a set of 137 horizontal wells from the Haynesville Shale. In this study mapping EUR in the latent space shows the correlation of good vs. poor wells to variability in proppant used, fluid injected, and porosity. These original MATLAB codes were computationally demanding in terms of memory and processing requirements (Wallet et al, 2009). For this reason Roy et al. (2013) re-implemented GTM to handle much larger input data volumes including P-impedance, lambda-rho ($\lambda\rho$), mu-rho ($\mu\rho$) and other volumes as input. Using a 2D latent space, we define polygons around the clusters to classify different lithofacies found within the Barnett Shale. Using the location of microseismic events, we classified the different lithofacies as being brittle or ductile shale (Slatt et al, 2011).

We begin with an overview of the GTM algorithm, parameterization and visualization. We then apply GTM in an unsupervised manner to map the natural (inherent) clusters in two variable producing reservoir units within a heterogeneous “carbonate wash” reservoir in the Veracruz Basin of Mexico and correlate them to production. Then we incorporate supervision into the GTM algorithm by using the wells to define attribute vectors that represent good and poor production. Finally, we use Bhattacharya (1943) measure to derive similarity between the probability density function to produce mapped regions corresponding to low and high risk of production.

METHODOLOGY

The GTM Algorithm

GTM is a non-linear dimensional reduction technique that provides a probabilistic representation of the data-vectors in a corresponding latent space. A set of non-linear continuous and differentiable basis functions is used to map points $\mathbf{u} = u_1,u_2,\dots,u_L$ in the L-dimensional latent space into an L-dimensional non-Euclidean manifold $\boldsymbol{\mathcal{S}}$ embedded within the D-dimensional data-space (Figure 6.1). Data vectors $\mathbf{x}_n$ are modelled by a suite of Gaussian PDFs centered on the mapped grid points on the non-Euclidean manifold $\boldsymbol{\mathcal{S}}$, thereby defining the space in which the data vector lies. The initial stage is called a constrained Gaussian mixture because the Gaussian centers are constrained by the grid points in the latent space. As we iterate each component of the mixture model is moved towards the data-vectors that it best represents. For each iteration the mapping parameters are determined using an Expectation Maximization (EM) algorithm (Dempstar et al, 1977). Relating the points on the manifold $\boldsymbol{\mathcal{S}}$ relates to the points on the latent space by a probability distribution provides a posterior probability projection for each data-vector in the L-dimensional latent space. This posterior probability projection of the data onto the latent space is used not only for data visualization but also to estimate the confidence in which each data point belongs to a given cluster. Detailed implementation of this algorithm is given in Chapter 5.

GTM Initialization and Parameter Selection

The selection of the input data vector (e.g. which seismic attributes best represent the desired facies) is the most critical parameter selection of GTM and to clustering in general (Barnes and Laughlin, 2002). Parameter selection specific to GTM defines facies resolution/discrimination and runtime. The number of points, K, forming the grid in the latent space should be dense enough to approximate a continuous distribution of the data and differentiate the target facies. By construction, the number of grid points is equal to the number of Gaussian mixtures in the data-space. Choosing a very large number of grid points increases the computation time and memory usage.

We also define a set of non-linear Gaussian basis function centers J, on this 2D latent space grid and take care to set $J < K$ to avoid rank deficiency of the $\boldsymbol{\Phi}$ matrix.

The common width of these basis functions ε is also set prior to the training. ε controls the smoothness of the 2D manifold. A smooth manifold in the data space facilitates fitting the data-vectors during training. The parameter ε remains constant for the whole process. The matrix consisting of the non-linear basis functions (Gaussian functions in our case) $\mathbf{\Phi} = \phi_j(\mathbf{u}_k)$ is calculated at initialization and remains constant for all subsequent iterations.

We initialize the weight matrix $\mathbf{W}$ such that the initial GTM model approximates the principal component analysis (PCA) of the dataset. The common value of the inverse variance of the Gaussian PDFs β is initialized to be the inverse of the $(L+1)^{\text{th}}$ eigenvalue from PCA where L is the dimension of the latent space. Since in our case $L=2$ we initialize β to be the inverse of the third eigenvalue.

GTM cluster visualization

Visualization is key to effective clustering. After we have estimated the parameters $\mathbf{W}_{\text{new}}$ and β_{new}, we are able to define a new posterior probability distribution of the data by the latent space grid points. Using Bayes' theorem we obtain,

$$R_{nk}(\mathbf{W}_{new}\,,\beta_{new}) = \frac{e^{-\frac{\beta}{2}||\mathbf{m}_k-\mathbf{x}_n||^2}}{\sum_k e^{-\frac{\beta}{2}||\mathbf{m}_k-\mathbf{x}_n||^2}}\,. \qquad (1)$$

These posterior probabilities or the “responsibility”, R_{nk} , values in equation 1 can be mapped to the latent space grid points $\mathbf{u}_k$ (Figure 6.2). We will use such an explicit projection of average data vectors about good and bad wells when we introduce supervision. However, such an explicit projection for 100s of millions of seismic attribute data vectors (one per voxel) in general results in too much information in the latent space. We therefore follow Bishop et al. (1998) and project the mean or the mode posterior probability projections onto the latent space.

The mean of the posterior probability distribution of a data-vector $\mathbf{x_n}$ is obtained by projecting onto all the grid points (nodes) the responsibilities of each node, thereby computing the average location in **u**

$$\mathbf{U}_{mean}(\mathbf{x}_n) = \sum_{k=1}^{K} R_{kn}\mathbf{u}_k. \qquad (2)$$

This mean projection of the posterior probability of the data vector is used in generating the clusters in the 2D latent space.

Summary of the GTM workflow

The GTM workflow is summarized in the following steps (Figure 6.3):

Initialization

- Choose an appropriate suite of attributes to differentiate the different reservoir performance results or seismic facies.

- Define the number of latent variables *K*, the basis function *J*, the relative width of the basis functions, ε, and the scaling factor, α,
- Generate the latent space grid $\mathbf{u}_k$, where *k=1,2,….., K,*
- Generate the grid for the Gaussian basis function centers $\mathbf{r}_j$, where *j=1,2,….., J,*
- Compute the set of Gaussian basis function $\boldsymbol{\Phi}$,
- Initialization of **W** and β from PCA analysis of the data **X,**
- Compute the reference vectors, $\mathbf{m}_k$,
- Compute $||\mathbf{m}_k - \mathbf{x}_n||^2$ for the Gaussian PDFs, and
- Calculate the responsibility $R_{nk}(\mathbf{W}\ ,\beta\ )$.

  Training

- Update the weight matrix $\mathbf{W}_{\text{new}}$,
- Update $\mathbf{m}_k$ and calculate the new $||\mathbf{m}_k - \mathbf{x}_n||^2$,
- Update the inverse variance β_{new} ,
- Calculate the new responsibility $R_{nk}(\mathbf{W}_{new}\ ,\beta_{new})$ (equation 1), and
- Compute posterior mean projection for QC (equation 2).

  Training continues until the model converges (i.e. the value of the inverse variance β_{new} stabilizes)

APPLICATION

Geological setting of Veracruz Basin

The Veracruz Tertiary Basin (VTB) (Figure 6.4a) is a foreland basin (Prost and Aranda, 2001) developed at the foothills of the buried tectonic front, filled with sequences of sandstone, shale and conglomerates deposited from Paleocene through recent (Cruz-Helu et al., 1977). The sediments come from a variety of sources: igneous complexes (such as the Santa Anna high), metamorphic complexes (La Mixtequita, the Sierra de Juarez and Macizo de Chiapas), and carbonates from the Plataforma de Córdoba (PEMEX, 2010; Cruz-Helú et al., 1977).

The field is composed of five reservoirs vertically separated by impermeable shale layers, of which the older are Middle Eocene in age (EOC–3, EOC–10, EOC–20 and EOC–30) that produce an average 22^0 API oil. EOC–10, EOC–20 and EOC–3 produce in the top of the structure (Figure 6.4b). In contrast, EOC–30 and EOC–50 are present only in the eastern flank of the anticline, where they produce 500 m down dip with respect to the crest of the structure (Figure 6.4b).

The seismic survey consists of a 100 km^2 of high quality prestack time migrated gathers and a stacked volume cut out from the much larger mega-merge survey over the Tertiary Veracruz Basin. The prestack impedance inversion volumes used in the classification algorithm are derived from this seismic dataset.

GTM workflow for multi-attribute seismic facies classification

The presence of conglomerate carbonates makes the reservoirs both laterally and vertically highly heterogeneous. This heterogeneity is observed at well log, core slab and thin section scales (Romero-Pelaez, 2012). The facies varies from breccias to conglomerates interbedded with poorly sorted calcareous sandstones and shale re-deposited in slope and basin-floor environments. The conglomerates and sandstone are rich in carbonate grains. These carbonate conglomerates are diagenetically altered providing secondary porosity, which is much more prevalent than the primary porosity. The porosity for the pay zones ranges from 11 to 17%. However, the reservoir may be under-sampled, such that the recovered samples may not be fully representative of all reservoir facies. Thus through our unsupervised clustering workflow we have tried to understand this heterogeneity present in the reservoir. A seismic facies model of the EOC-30 and the EOC-10 reservoir units were created and the results were then correlated with the productivity of the wells.

The inputs to our unsupervised GTM algorithm are different seismic inversion volumes (P-impedance, lambda-rho, mu-rho and V_p/V_s), which help in understanding the highly heterogeneous conglomerate reservoir of the Veracruz Basin. The impedance volumes better reflect a heterogeneous reservoir based on the variation of the porosity and the rock type. They also help to determine the relationship between the desired rock property such as lithology, porosity or clay volume. The other products of seismic inversion such as V_P to V_S ratio (V_P/V_S), Lambda-Rho ($\lambda\rho$) and

Mu-Rho ($\mu\rho$) helps in understanding the rock-type and different elastic properties of rocks. Cross plotting between two such elastic properties from the wells sometimes helps in segregating different rock types. However, for this reservoir, two-component crossplots do not separate the classes. Figure 6.5 shows the crossplot of the seismic volumes $\lambda\rho$ vs. $\mu\rho$ (Figure 6.5a) and P-wave impedance (Z_p) vs. V_P to V_S ratio (V_P/V_S) (Figure 6.5b).

With the above four volumes (V_P/V_S, Z_P, $\lambda\rho$ and $\mu\rho$) our input data dimension becomes 4-dimensional, that is, each voxel in 3D physical space is associated with an associated 4-attribute or 4-dimensional vector. PCA and GTM both attempt to minimize the error between a projected space and the data. Without normalization, small errors in fitting Z_P or Z_S impedances (with values on the order of 10^9 Pa) would overwhelm the signal in V_P/V_S (with dimensionless values ranging between 1.5 and 3). Data vectors are therefore normalized by subtracting off the mean and dividing by the standard deviation for each attribute component or dimension using a z-score algorithm.

Initially a 2D latent space is uniformly sampled with K = 30x30=900 square grid of nodes (the blue circles in Figure 6.1a). A slightly coarser square grid of J = 16x16=256 nodes define the centers of the non-linear Gaussian basis functions in the same 2D space (the green circles in Figure 6.1a). The separation between the neighboring Gaussian Basis function centers is set to half the standard deviation of the Basis function centers and the grid points on the 2D latent space. A multi-attribute PCA analysis initializes the weight matrix **W** and the inverse of the Gaussian (noise) variance β of the GTM model. We randomly select 0.1% of the data vectors to train the latent space, updating **W** and β with each iteration using an expectation-maximization (EM) algorithm. The EM algorithm is guaranteed to converge to possible local maxima onto the likelihood surface. We stop the iteration when the value of the inverse variance, β, stabilizes (Figure 6.6a). After training the latent space, the final GTM model parameters are applied to all 100% of the data vectors, which are then projected onto the 2D latent space (Figure 6.6b). The interpreter then interactively analyzes, or “clusters” the projected posterior probability distributions in the 2D latent space and visualizes the results in 3D as geobodies.

Unsupervised facies model of the EOC-10 and EOC-30 reservoir units

We focused our analysis on the resulting GTM seismic facies volume of the EOC-10 and the EOC-30 reservoir units. Both reservoir units are dated as Middle Eocene, produce 22^0 API oil, are composed of similar mineralogy content, and exhibit a similar porosity distribution. Figures 6.7a and b show the mean projection after GTM training of the dataset for reservoir units lying only within EOC-10 and EOC-30 unit respectively. The posterior probability values of the data vectors from each of the voxel locations of the probe are projected onto all the grid locations and their mean locations calculated and projected onto the 2D latent space forming a 2D histogram. Hernández-Martínez (2009) proposed the regional conceptual sedimentary model (Figure 6.7c) to be a slope and basin-floor fan, with distributary channels oriented northwest-southeast, very similar to the coarse-grained, sand-rich turbidite system described by Bouma (2000). We define clusters in this histogram by interactively drawing polygons about the high-density areas of the plot, thereby hypothesizing similar seismic facies for reservoir units EOC-30 and EOC-10. The mineralogy content, the porosity and reservoir type of these two reservoir units are similar, which appear as similar location of the clusters onto the 2D latent space (Figure 6.7). The dataset falling within these polygons are highlighted for these two reservoir units. This creates the 3D seismic facies volumes (or geobodies) of the two reservoir units shown in Figure 6.8.

Romero-Paleaz (2012) analyzed these same reservoir units using a commercial implementation probabilistic neural network (PNN) in order to predict effective porosity and clay volumes. To validate our clustering, we performed an independent facies estimation of the GTM seismic facies volume and visually correlated with reservoir property volumes calculated by supervised probabilistic neural network (PNN). Figure 6.9 compares predicted effective porosity and clay volume to seismic classes obtained from GTM for reservoir unit EOC-10 and the EOC-30 reservoir units. The orange- and green-colored GTM facies approximately correlate to the conglomerate sandstone with minimum clay content. The pink- and purple-facies GTM volumes approximately correlate to the low effective porosity clay-rich facies seen in the EOC-30 reservoir unit and along the normal faults.

Correlating the production data with the seismic facies the facies model

The unsupervised GTM classified volume is correlated visually to the seven-month average well production. The wells in these two zones are along the structural highs.

Figure 6.10a shows the volume probe within the EOC-10 reservoir unit with the well locations. Figure 6.10b shows a phantom horizon 10ms below the EOC-10 top. In general the light green and the orange color facies are better reservoir rock. The green facies in the north are good producers (Wells J and I). Well W on the left tested brine water and falls into the orange GTM-facies, which we associate with good reservoir quality, but lies in the structural lows. Well X is a dry well with logs and core showing hard-cemented facies and falls in the brown-colored GTM facies making it a poor reservoir quality rock. The pink- and purple- facies correspond to facies rich in clay content.

A similar analysis was done for the EOC-30 reservoir unit. Figure 6.11a shows the GTM facies volume and the well locations in this reservoir unit. Figure 6.11b shows productivity pie charts of the wells on a phantom slice 10ms below the EOC-30 top. Wells A, B, C, D and E exhibit mediocre production and lie in the orange facies. Well F exhibits the highest oil production, although for a short term before producing water. Inspecting the GTM facies volume (Figure 6.11a) Well F is located adjacent to the light green-GTM facies. In general, the orange facies, and light green facies are associated with good reservoir quality rock for this reservoir unit. However since the structure is dipping south, water productivity of the wells increases southwards.

Now we take average seismic attribute data-vector around three of the well locations, calculate the PDF of each well data-vector and compare the overlap of the PDF with other seismic attribute data-vectors. This will give the most likely facies volume corresponding to a Well facies.

Supervised GTM Classification based on Bhattacharya measure

Since GTM is based on probability measures and statistics, we can use the probability distribution functions to measure similarities or dissimilarities between two PDFs. Let R_{wk} be the posterior probability distribution corresponding to a well data-vector as shown in Figure 6.12a. Let R_{nk} be the posterior probability of any other seismic attribute data-vector, n as shown in Figure 6.12b. Then by Bhattacharya (1943) measure, we can find the similarities (Figure 6.12c) between the two PDFs by

$$d_n = \sum_{k=1}^{K} \sqrt{R_{wk} R_{nk}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3)}$$

where, k are the grid points of the 2D latent space. Thus when two distributions are identical ($R_{nk} = R_{wk}$), we have a coefficient of 1 $d_n = \sum_{k=1}^{K} R_{wk} = 1$. In contrast when there is no overlap between the PDFs $d_n = 0$. Thus, this coefficient ranges from$0 \leq d_n \leq 1$.

In this manner, one computes the value of d_n for all the data-vectors in the survey resulting in a supervised facies "similarity" volume, quantitatively comparing each voxel to good and poor wells.

The average data-vectors around three wells are calculated: poor producer Well X, good producer Well K and moderate producer Well E. The same four inversion "attribute" volumes (V_P/V_S, Z_P, $\lambda\rho$ and $\mu\rho$) are used as input, with the average vectors calculated for sub-volumes around each of the three well locations. Figure 6.13 shows the average vector for each of the three well locations. The following three well data-vectors have the following characteristics. The average vector for the dry Well X has a lower $\mu\rho$ value, which is different from rest of the two wells. The average data-vector for Well K shows a larger $\mu\rho$ value. The average data-vector for Well E shows a higher $\lambda\rho$, $\mu\rho$ and V_P/V_S value compared to the rest. Finally, we calculate the overlap of the posterior probability distribution of each of these well data-vectors and the remaining data-vectors by calculating the coefficient from the Bhattacharya (1943) measure, for each of the data-vectors. The output "similarity" volumes for each well type are shown in Figures 6.14 to 6.16.

Probabilistic estimation of facies for the EOC-10 and EOC-30 reservoir units

Three volumes are created after the similarities between the PDF of the well-vector and the data-vectors are calculated with Bhattacharya measure. Each volume highlights the likelihood that a given voxel corresponds to those about a target well. Figure 6.14 shows a geobody within the EOC-10 and EOC-30 reservoir unit through Bhattacharya measure showing the likelihood of facies corresponding to the dry Well X. The most probable regions appear as hot colors and the least probable regions appear as cold colors. The regions surrounding well X show the occurrence of bad reservoir rock. This facies type is also present along the faults mapped using conventional interpretation techniques.

Figure 6.15 shows the geobody after PDF similarity is calculated between the data and the good producer Well K. Hot areas correspond to regions with high likely that the facies is similar to Well K. Note that there is only limited zones in the two reservoir units similar to Well K. Most of the reservoir exhibits a low probability of occurrence (blue and cyan regions) of Well K facies.

Figure 6.16 is the geobody through the Bhattacharya measure between the moderate producer Well E data-vector and all the data-vectors. Most of the regions within the EOC-10 and EOC-30 reservoir units have high likelihood of facies being similar to well E. Note that this facies is least likely to occur around well X and around the faults.

DISCUSSION

GTM facies generated from input V_P/V_S, P-impedance, $\mu\rho$ and $\lambda\rho$ volumes provide similar facies to those generated using a supervised PNN to estimate V_{clay}, and helps in identifying the different rock-types of the reservoir units. Geobodies can be conveniently calculated for each facies, which will quantify the amount of good or bad reservoir rock present.

Analysis of the clusters was done from the regional geological information and from the unsupervised GTM facies analysis results. Figure 6.18 summarizes the results of unsupervised and supervised GTM cluster analysis. In the unsupervised case different clusters are identified by their separation in latent space. The meanings of these clusters are evaluated a posteriori through the use of well logs, core, production data, the ANN-generated V_{clay} volume, and the original impedance volumes. The violet and the pink facies along the faults are clay- rich, the brown facies corresponding to the dry well X is tight, the orange and the light green facies comprise good reservoir rocks made up of porous limestone conglomerate “wash”. The dark-green facies in the unsupervised analysis corresponds to moderate reservoir rocks made up of lower porosity, harder limestone conglomerate.

The three PDF similarity volumes, shown in Figures 6.14 to 6.16, give the most likely occurrence of the facies type within the reservoir units. Around Well X and along the faults (Figure 6.14) GTM predicts bad reservoir rocks. Thus for the arrow marked in Figure 6.14 it is 90-100% more likely that the facies is similar to that of Well X. The moderate reservoir rock corresponding to Well E (Figure 6.16) is abundant and most likely facies type to be found in the reservoir units. Thus this probabilistic estimate of the reservoir facies can add a factor in the modern day risk analysis workflows. However, the most of the productive wells are in the north with water production increasing southward. Faults running north-west to south-east the structure also play an important role in the well production (Romero-Paleaz, 2012).

CONCLUSIONS

In contrast to Kohonen SOM which is based on heuristics and empirical arguments, GTM is based on probability and statistics (Sevensen, 1998). GTM provides a method for modeling continuous, low-dimensional, non-linear structures in high dimensional data space. In this paper, where each dimension corresponds to an input seismic attribute volume, we successfully model a 4D attribute space with a 2D manifold $\mathcal{S}$ which in turn is constrained by the K grid points defined in the 2D latent space. This latent space contains the mean locations representing probability density of the data-vectors.

We chose our lower dimensional space to be 2D since the results take the form of a histogram, allowing us to directly link the GTM results to the 2D crossplot utilities found in most commercial software. Within the commercial crossplot software, the interpreter recognizes (or hypothesizes) clusters, defines them with colored polygons, and displays the corresponding data vectors by coloring their voxel locations as geobodies. Thus, unlike K-means and SOM, the proposed unsupervised GTM workflow facilitates a posteriori supervision to interactively define and color the seismic clusters/facies.

Since GTM is rooted in probability theory, it provides a direct link to supervised classification. In this limestone conglomerate/wash oil field, we provide a very simple, but versatile workflow by assigning each data vector to the well data vector that it best represents. Unlike earlier similarity workflows used by Johnson et al. (2001) and Michelena et al. (1998), all well data are used simultaneously to define the latent space distribution. Therefore, we know not only which well-vector matches a data-vector but also a probabilistic estimate as to how well each attribute data-vector represents a given well. This capability allows us to develop and apply a supervised GTM workflow that generates a suite of seismic "similarity" volumes with the probabilistic estimates of the facies types being similar to the well vectors used to train the GTM model. This provides an estimate of how likely we should find a facies if we drill in a certain location. We anticipate that such probabilistic estimates will fall neatly into modern risk analysis evaluation of drill locations and reservoir valuation.

LIST OF FIGURES

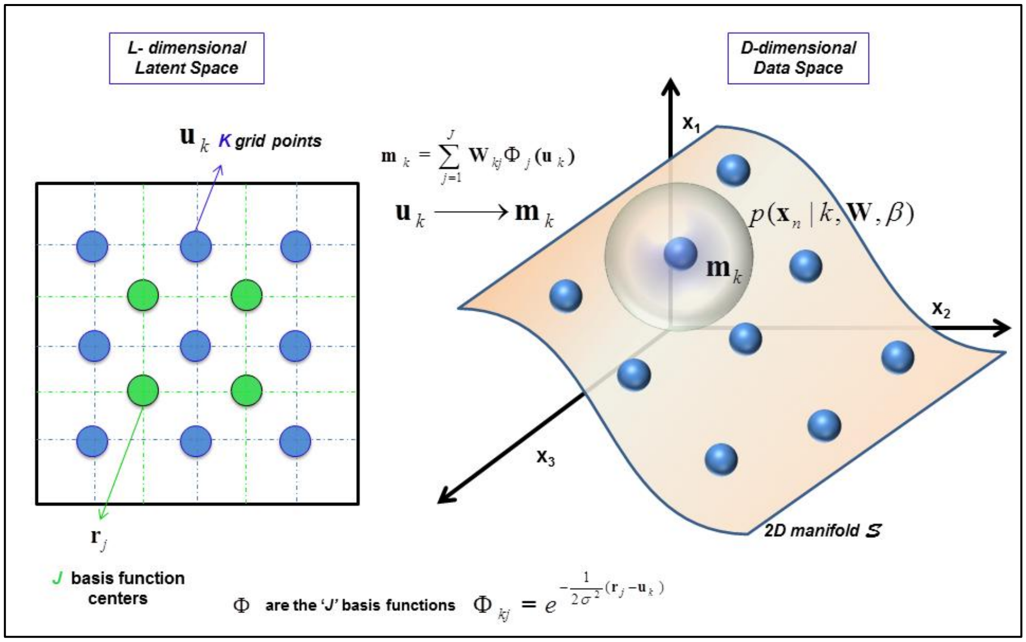


Figure 6.1. Non-linear mapping of the latent space to the data space. The prior distribution consists of latent space variables (*K*) ordered on a regular grid (blue circles) residing in an *2*-dimensional latent space. **Φ** consists of a regular array of *J* non-linear basis functions. With the linear combination of these basis functions the latent space (blue circles) are mapped to the data-space (blue spheres) where they form a 2D non-Euclidean manifold, S, such that each node $\mathbf{u}_k$ is mapped to a corresponding vector $\mathbf{m}_k$ in data-space, given by $\mathbf{m}_k = \sum_{j=1}^{J} \mathbf{W}_{kj}\boldsymbol{\phi}_j(\mathbf{u}_k)$. In general data-vectors are scattered about $\mathcal{S}$. Specially we assume an isotropic Gaussian noise distribution $p(\mathbf{x}_n|k, \mathbf{W}, \beta)$, centered at $\mathbf{m}_k$ and having a variance of 1/β. The final probability density function in the data space is obtained by integrating the Gaussian PDFs for all $\mathbf{m}_k$ where *k=1, 2, ….K* grid points.

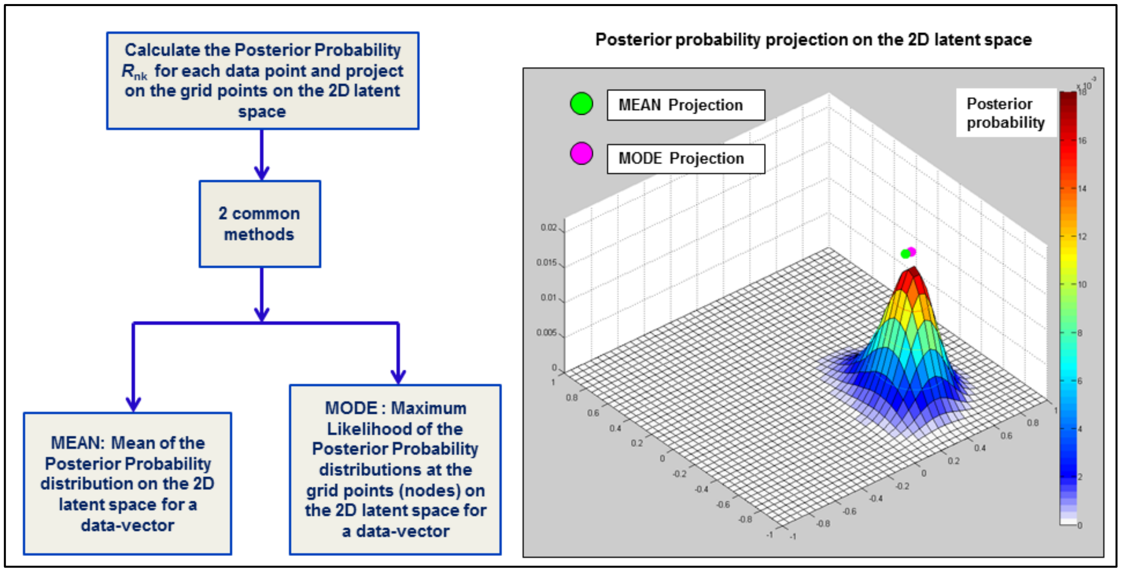


Figure 6.2. Workflow for data visualization in GTM. After training the new estimated parameters $\mathbf{W}_{\text{new}}$ and β_{new}, the new posterior probabilities representing the data-vectors can be obtained using Bayes' theorem and is given by $R_{nk}(\mathbf{W}_{new}, \beta_{new})$. These posterior distribution "responsibility" values can be plotted for all the grid points, $\mathbf{u}_k$ in the 2D latent space. The mean location will assign the value $\mathbf{U}_{mean}(\mathbf{x}_n)$ to be the weighted average of the posterior distribution values and will in general fall between neighboring values of $\mathbf{u}_k$. The mode will assign the value $\mathbf{U}_{mode}(\mathbf{x}_n)$ to be the location of the greatest posterior distribution value in the 2D latent space and will always correspond to a discrete gridded value of $\mathbf{u}_k$.

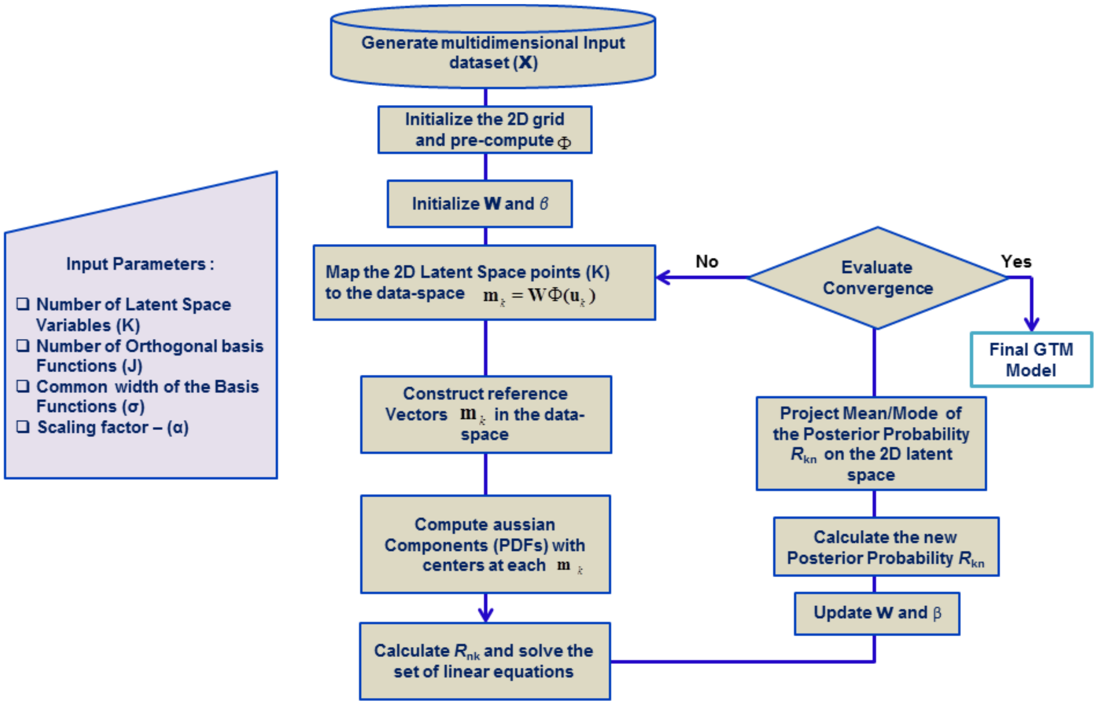


Figure 6.3. The flow chart for generative topographic mapping workflow.

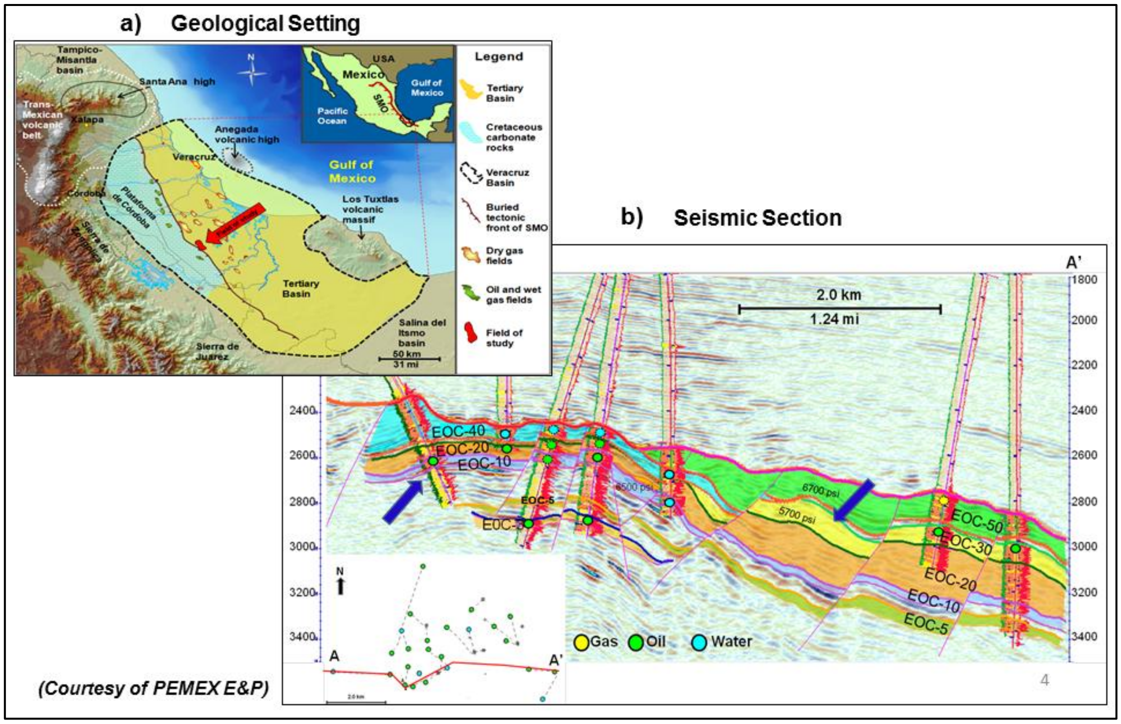


Figure 6.4. (a) Detail of the Veracruz Basin showing its boundaries and its two geological subdivisions: The Plataforma de Córdoba, which is the buried tectonic front of the Sierra de Zongolica, and the Veracruz Tertiary Basin. The field of study (indicated by the red arrow) is located in the western margin of the buried tectonic front in Upper and Middle Eocene age Tertiary sediments. (Map courtesy of PEMEX E&P based on previous work by Prost and Aranda, 2001; Romero, 2012). **(b)** Structural cross section through the field. EOC–10 and EOC–20 produce in the western portion of the structure; EOC–40 is water bearing and is distributed over the western portion of the anticline; EOC–30 and EOC–50 produce in the eastern flank (Hernández-Martínez, 2009). EOC-10 and EOC-30 are the reservoir units, which we analyzed through GTM clustering. (Image is courtesy of PEMEX E & P).

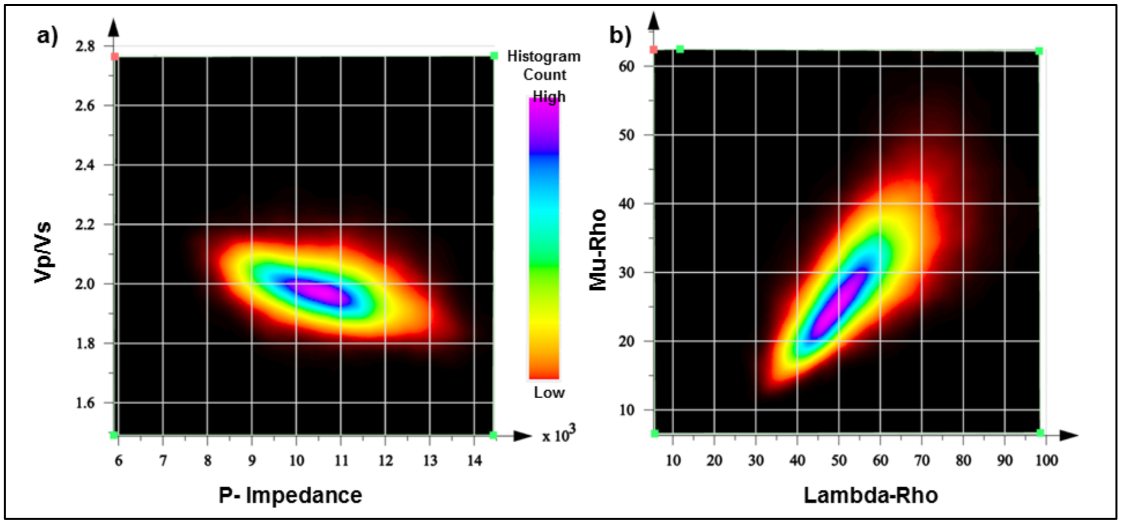


Figure 6.5. 2D histogram plot generated from cross-plotting the different volumes taken as input to GTM of the EOC-30 zone. **(a)** Cross-plot from seismic volumes Vp/Vs vs. P-impedance. **(b)** Cross-plot from seismic: Mu-Rho vs. Lambda –Rho. Note that the crossplot generated forms a single cluster and does not help in distinguishing different facies

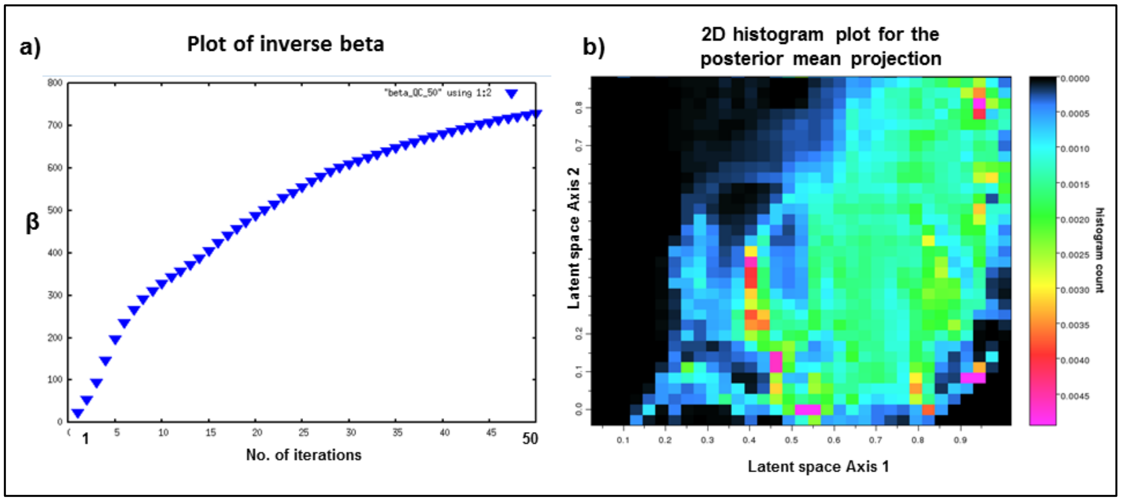
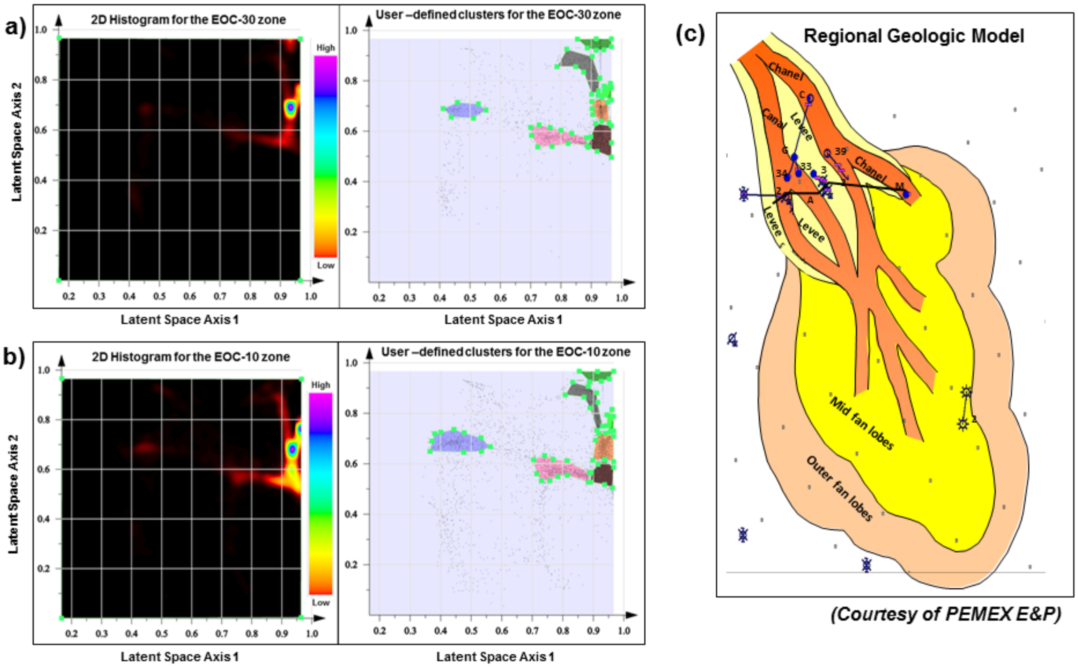


Figure 6.6. QC of the GTM analysis after performing 50 iterations. **(a)** The plot of the inverse variance (β), stabilizing as the number of iterations increases. **(b)** A 2D

histogram of the mean projections of posterior probability of the dataset within the reservoir zone between EOC-50 and EOC-5 (from Figure 6.4b).



(Courtesy of PEMEX E&P)

Figure 6.7. The 2D cross- plot of the mean posterior distribution map of the "*responsibilities*" of the data onto the 2D Latent space for the reservoir units EOC-30 and EOC-10. The cross plot is generated by cross plotting to GTM projection volumes. **(a)** The projection of the mean "*responsibilities*" of EOC-30 unit. The 2D histogram is on the right and the scatter crossplot is on the left. Seven clusters are visible on the latent pace corresponding to the high-density points. These clusters are delineated by polygons with different colors and in the subsequent figure will help to visualize the different classes in the seismic data. **(b)** The projection of the mean "responsibilities" of EOC-10 unit. The mineralogy content and porosity distribution for the EOC-10 and the EOC-30 reservoir units being similar the clusters for both

these reservoir units lie on the same location in the 2D latent space. They are also color-coded similarly since both reservoir units have similar rock type. **(c)** Regional conceptual sedimentary model (Courtesy of Petróleos Mexicanos (PEMEX) E & P. Hernández-Martínez, 2009)

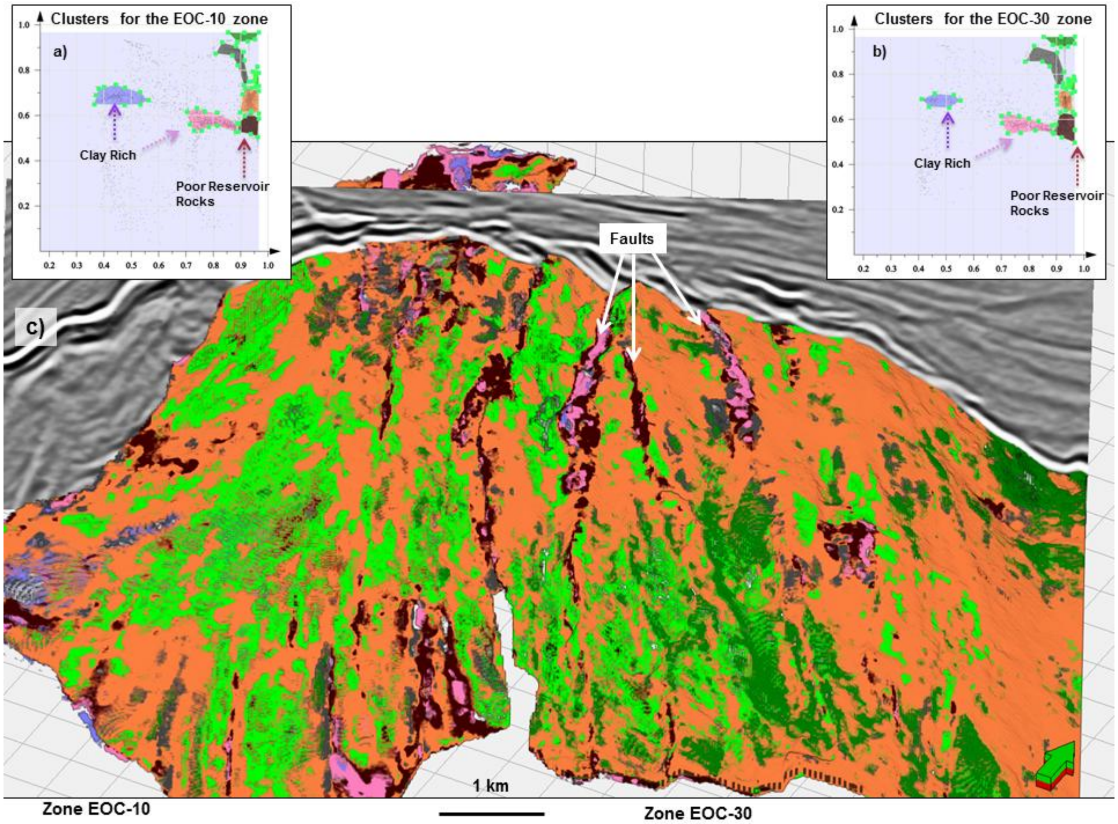


Figure 6.8. Generating the seismic facies volume (geobodies) from GTM clustering within the reservoir units EOC-10 and EOC-30, considering the input seismic volumes - lambda-rho ($\lambda\rho$) vs. mu-rho ($\mu\rho$) and P-wave impedance (Zp) vs. V_p to V_s ratio (V_p/V_s). Different polygons around classes signify rock types for reservoir units **(a)** EOC-10 and **(b)** EOC-30. Seven different facies class have been identified from the

clusters in the latent space and are delineated by polygons of different colors. **(c)** The horizon probe generated for the EOC -10 and the EOC-30 reservoir units after the unsupervised GTM analysis. The white arrows highlight the faults. The most abundant facies are the orange facies.

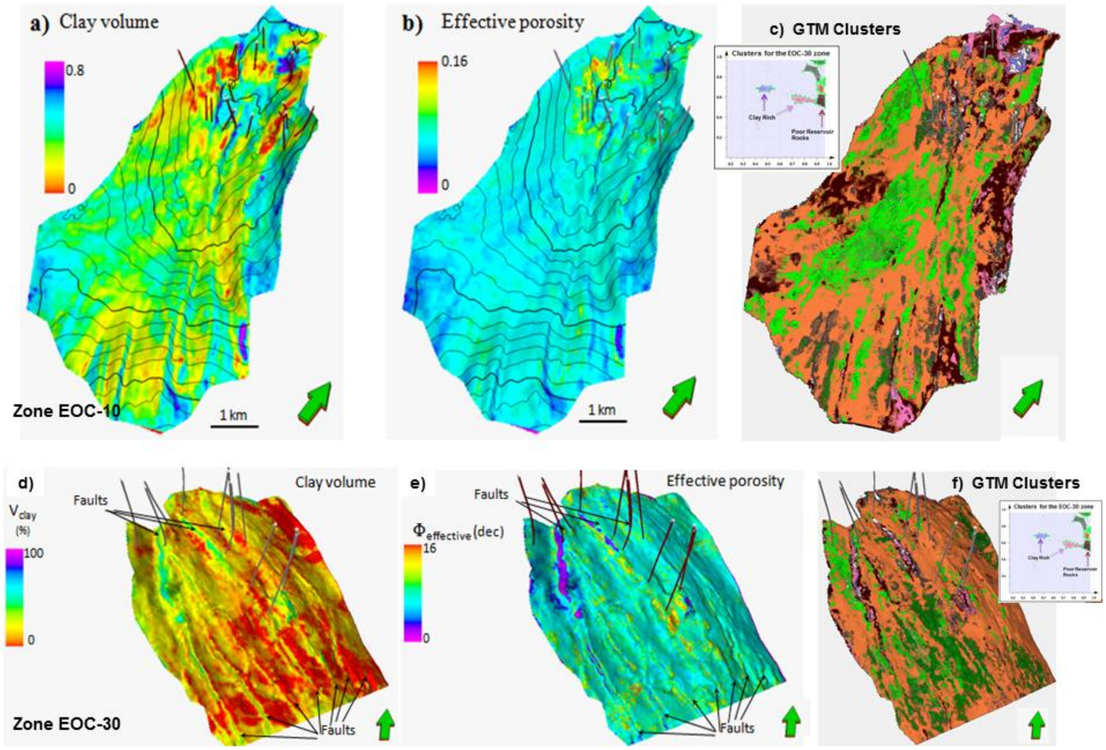


Figure 6.9. Horizon phantom slice 10ms below the top of EOC-10 and the EOC-30 reservoir units, through **(a)** clay volume, **(b)** effective porosity predicted from supervised probabilistic neural networks (PNN), and **(c)** unsupervised seismic facies volume from GTM for the of EOC-10 reservoir unit. **(d)** The clay volume and **(e)** effective porosity predicted from supervised probabilistic neural network (PNN), and **(f)** seismic facies volume from GTM of EOC-30 reservoir unit.

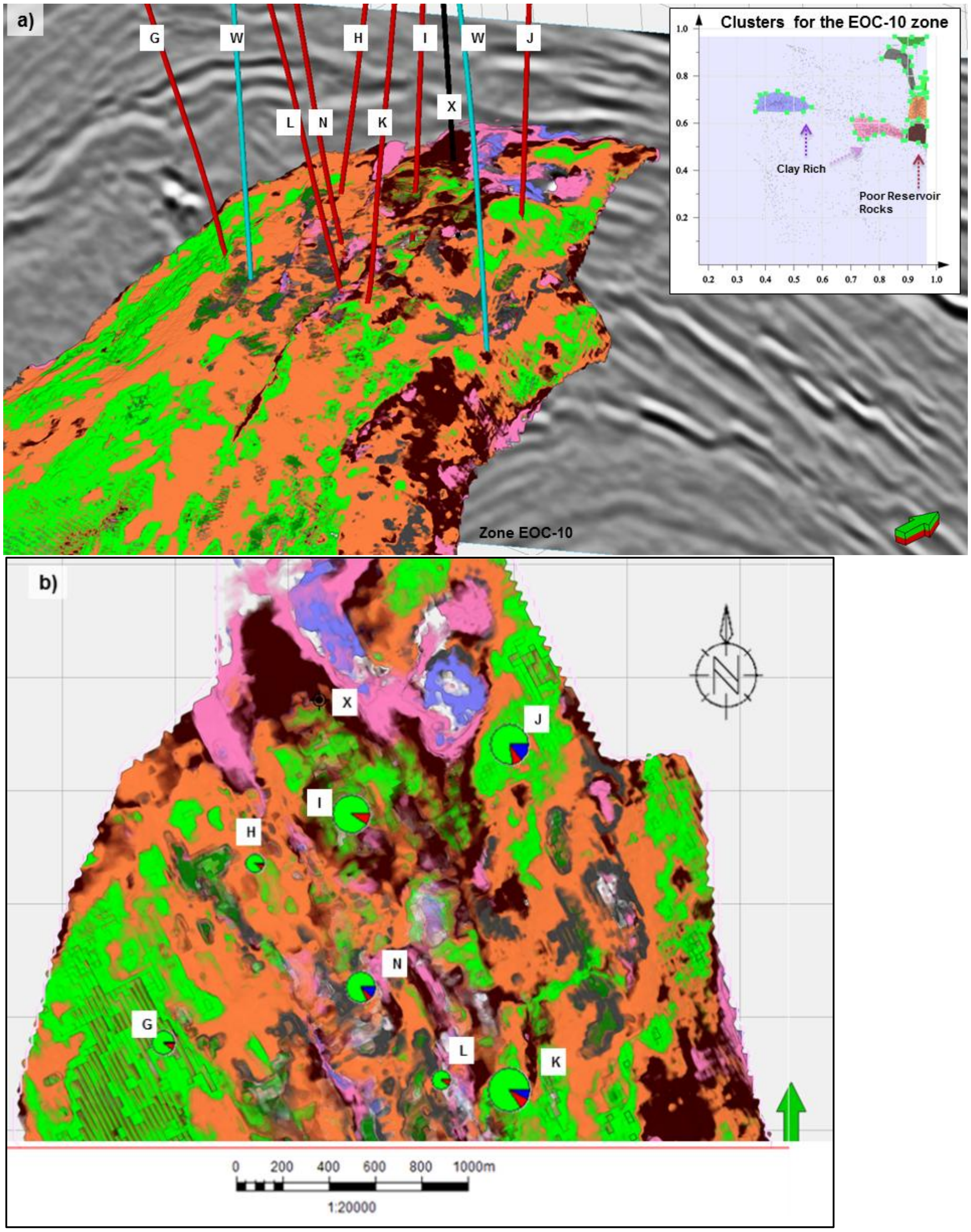


Figure 6.10. (a) The GTM seismic facies volume with the well locations for the EOC-10 reservoir unit. The red wells are the producing wells and the blue the injector well. The well X is a dry well and falls within the brown facies region. These brown colored

faces are probably bad reservoir quality rocks making X a dry well. In this unit also the wells are located at the structural highs. Most of the producing wells are along the greyish-green and light green facies. Correlating with the clay volume the pink-purple colored facies corresponds to the high clay content facies from. **(b)** Map view of the top of the EOC-10 reservoir unit. The pie charts at the well locations show the average production of the well for a seven-month period. In the pie chart green is for oil, red for gas and blue for water. Structural high is in the North. The green facies in the north are good producers (Wells J and I). Well H in orange facies is a moderate producer. Well G in the west lies in the structural low making it a moderate producer. In the south the well K lying in the light green facies is the most productive whereas the wells in the greyish facies are moderate-low producers. Note the pink and purple facies correlates with clay-rich areas.

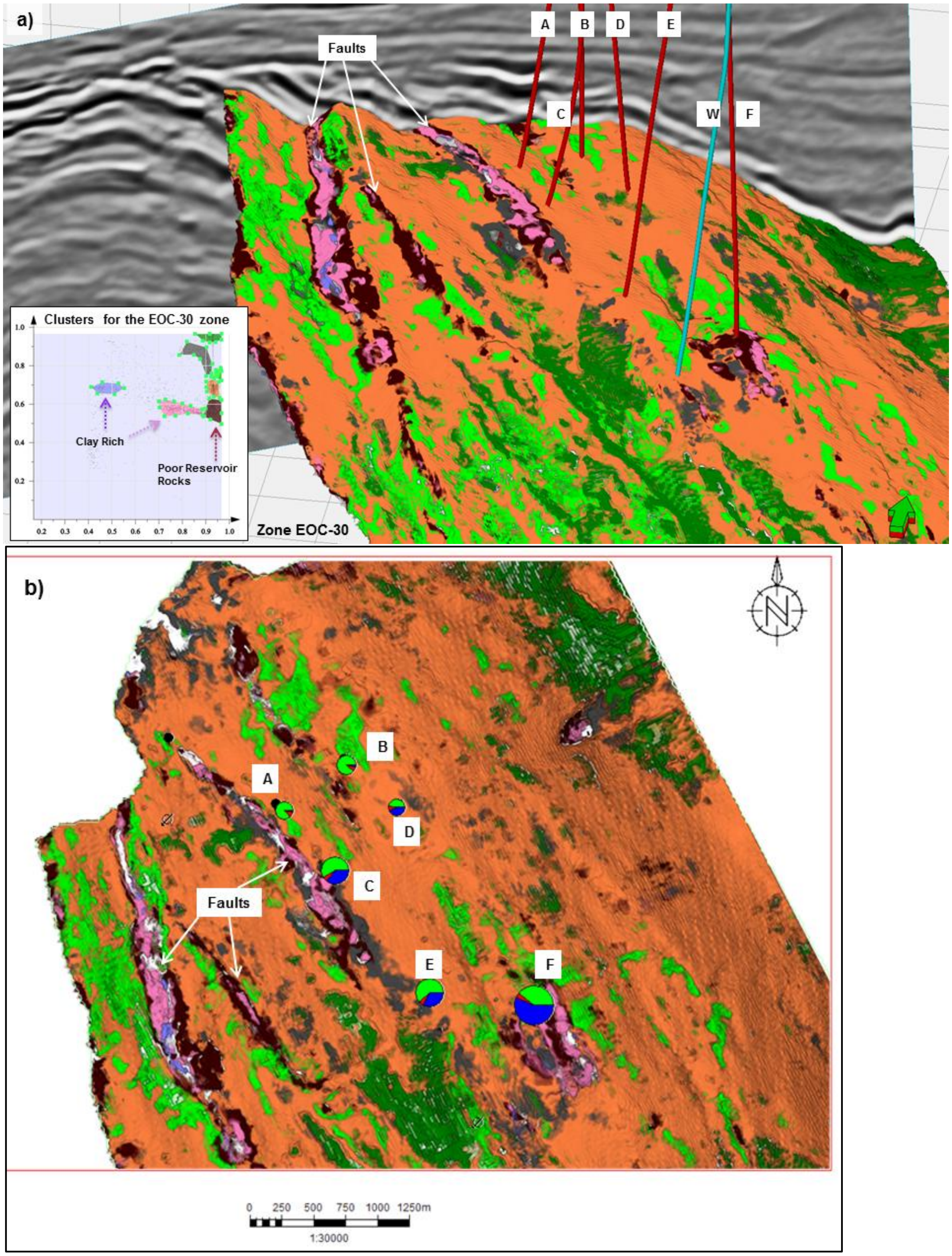
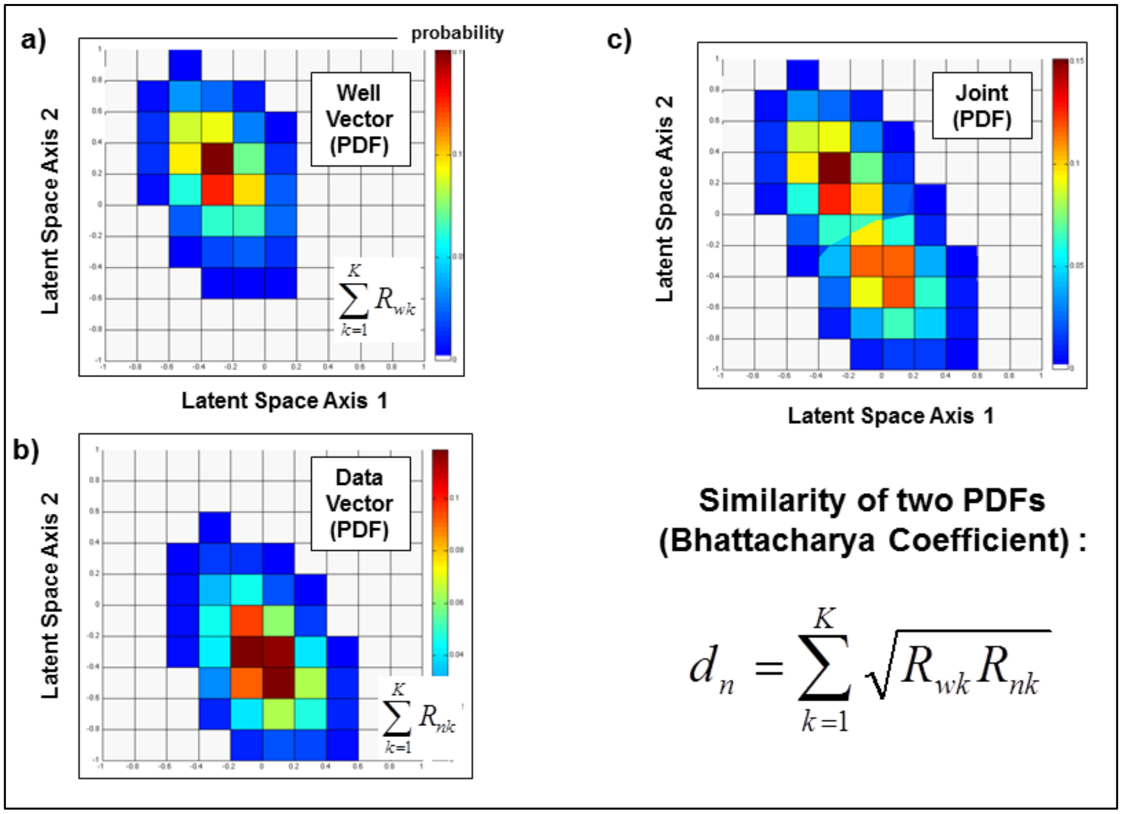


Figure 6.11. (a) The GTM seismic facies volume with the well locations for the EOC-30 reservoir unit. The red wells are the producing wells and the blue the injector well. The wells are mostly situated at the structural highs. The most abundant rock type

within this zone is the one with the orange color. Comparing with the well information and with the clay and effective porosity volumes from probabilistic neural network (PNN) these corresponds to the conglomerate sandstones with moderate porosity. The brown and pink color facies are mostly along the faults and probably shows the depositions close to normal faults along hanging walls. These regions also correspond low porosity and relatively rich in clay. The dark green corresponds the highest impedance regions, with moderate effective porosity, which is interpreted as the hard conglomerate deposits having least clay content. **(b)** The map view of the top of the EOC-30 reservoir unit. The pie charts at the well locations show the average production of the well for a seven-month period. In the pie chart green is for oil, red for gas and blue for water. Structural high is in the North. Wells A, B, C, D and E are low-moderate producing wells. Thus the orange rock is a moderate quality reservoir rock with moderate production. The water productivity increases as we go south (wells E and F). The well F is having the largest production in terms of both oil and water is situated in the mixed grey/light green and pink facies. The grey/light green facies are better reservoir quality rocks.



Similarity of two PDFs (Bhattacharya Coefficient) :

$$d_n = \sum_{k=1}^{K} \sqrt{R_{wk} R_{nk}}$$

Figure 6.12. A schematic representation of the supervised GTM analysis workflow: (a) The PDF representing a data vector computed from the seismic attributes about any well. (b) The PDF of a data vector corresponding to voxel *n* in the seismic attribute volume. (c) The joint PDF of the average well data vector and the data vector. The coefficient d_n from the Bhattacharya measure is the measure of the similarity (overlap) between the two PDFs.

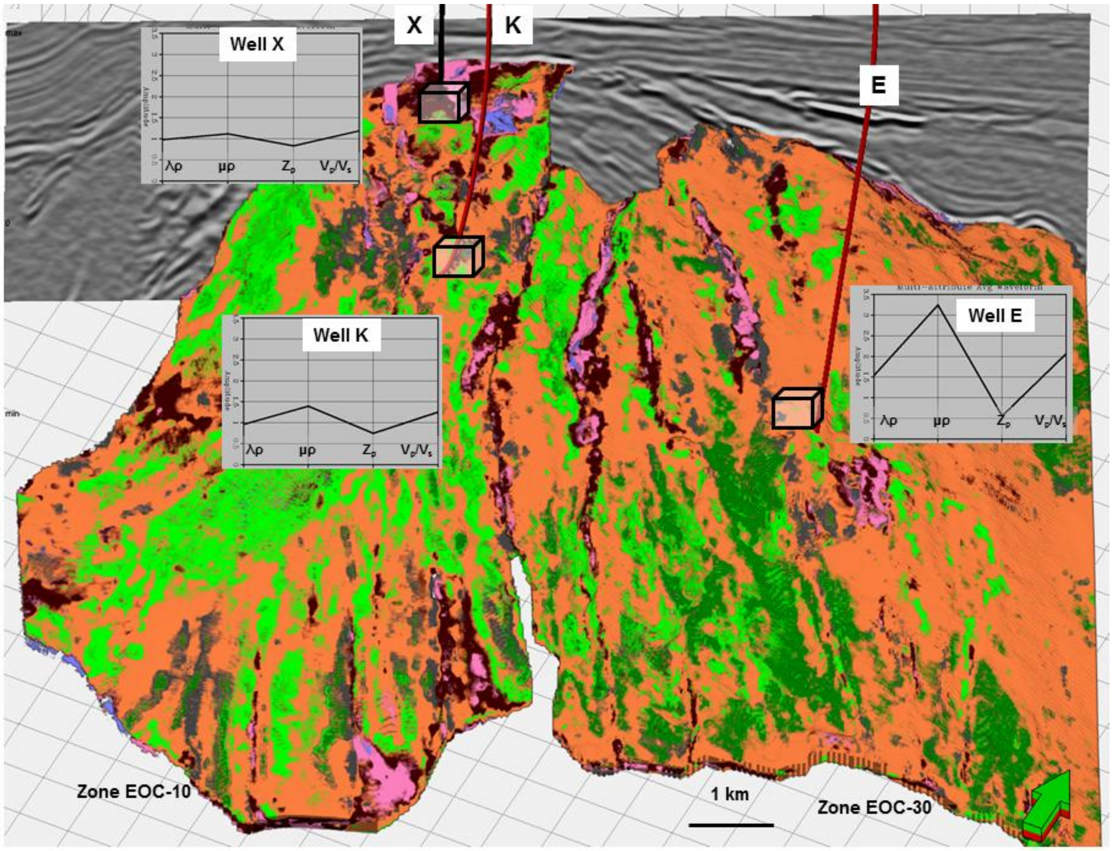


Figure 6.13. The average of the data-vectors calculated from a sub-volume (displayed as small cubes) around the each well location. Each input attribute has been previously normalized using a Z-score algorithm. These three average (target) vector around the wells are used to train the GTM, resulting in a supervised analysis.

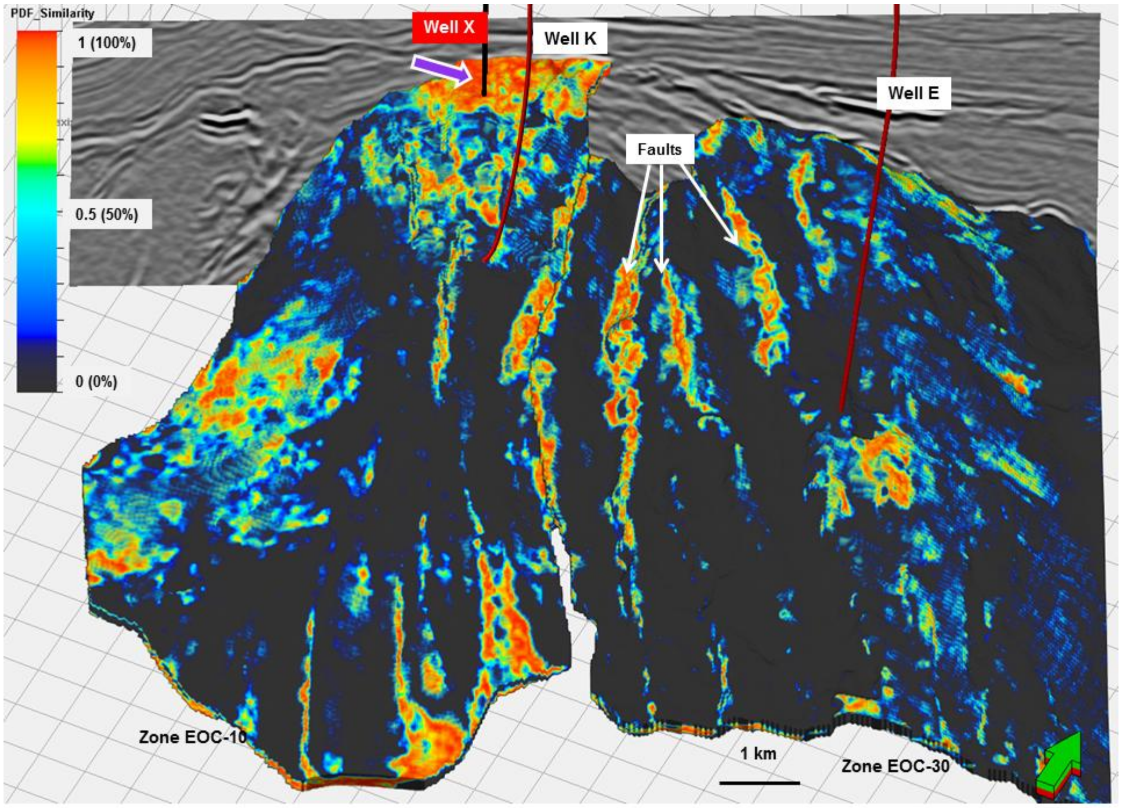


Figure 6.14. The results of the most likely occurrence of the seismic facies corresponding to dry well X within the EOC-10 and the EOC-30 reservoir units. The regions, which are most similar to Well X facies appears as hot colors. The least similar regions appear in cold colors. Note the occurrence of Well X Facies is confined mostly along the faults and around Well X.

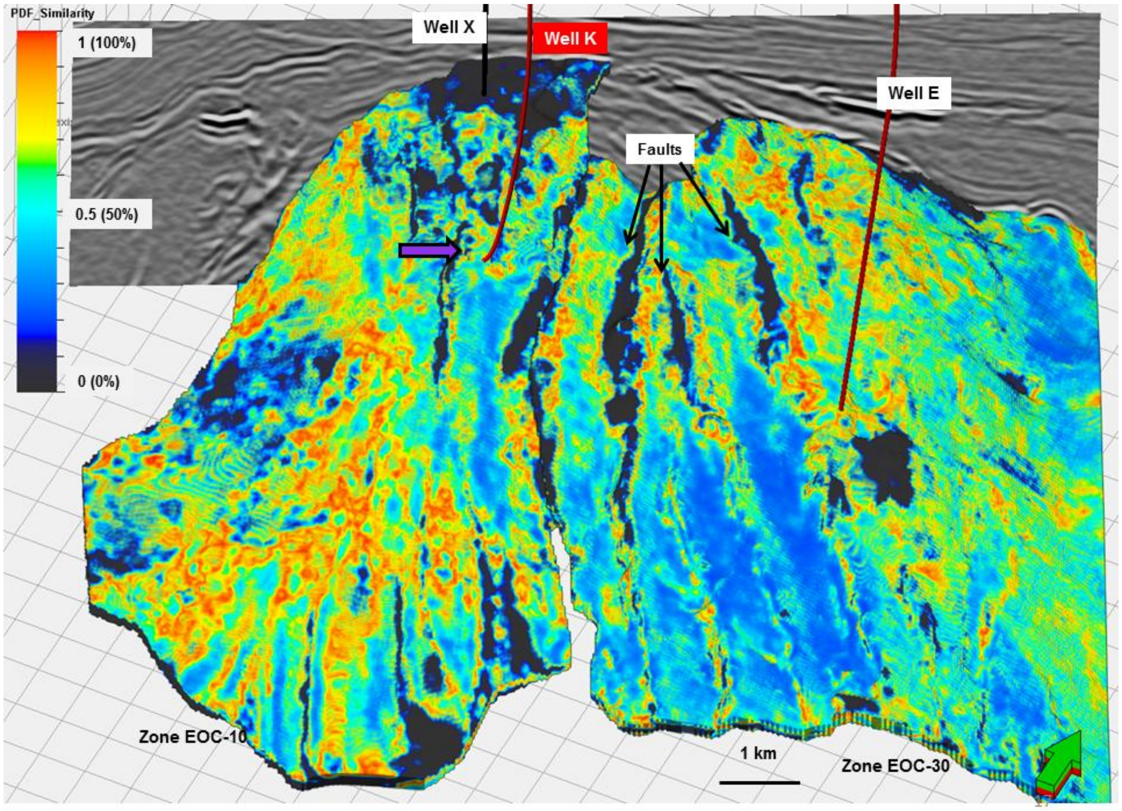


Figure 6.15. A geobody within the EOC-10 and the EOC-30 reservoir units which shows the most likely occurrence of the facies type corresponding to good producer well K. The regions, which are most similar to the facies type at Well K, appear in hot colors. The least similar regions appear in cold colors. Note that this facies is not so abundant in the reservoir units and the likelihood of occurrence ranges from 40-70%.

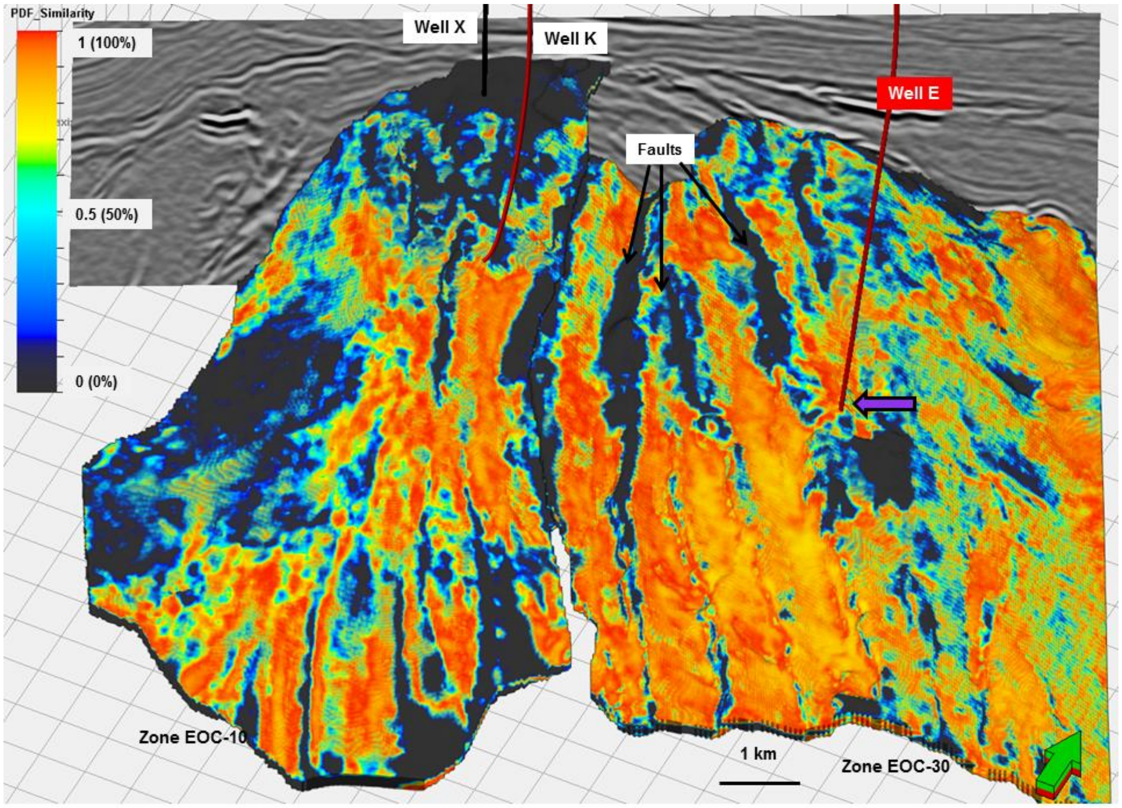


Figure 6.16. A geobody within the EOC-10 and the EOC-30 reservoir units which shows the most likely occurrence of the facies type corresponding to moderate producer well E. The regions, which are most similar to the facies type at Well E, appear as hot colors and the least similar regions appear as cold colors. Note that this type of facies is very abundant in the reservoir units and the likelihood of occurrence mostly ranges from 75-100%.

REFERENCES

Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 73rd Annual International Meeting Society of Exploration Geophysicists, Expanded Abstracts, 2221-2224.

Bhattacharya, A. 1943, On a measure of divergence between two statistical populations defined by their probability distributions: Bulletin of the Calcutta Mathematical Soceity, **35**, 99-109

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, **10**, No. 1, 215-234.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, Development of the generative topographic mapping: Neurocomputing, **21**, No. 1, 203-224.

Bouma A. H., 2000, Fine Grained, mud-rich turbidite systems: model and comparison with coarse-grained, sand-rich systems, in A. H. Bouma, and C. G. Stone, 2000, Fine- grained turbidite systems: AAPG Memoir 72, and Society for Sedimentary Geology Special Publication No. 68, 9–20.

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.

Cruz-Helú P., V. R. Verdugo, and P. Bárcenas, 1977, Origin and Distribution of Tertiary conglomerates, Veracruz Basin, Mexico: AAPG Bulletin, **61**, 207–226.

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, B, **39**, no. 1:1 38.

Hernández-Martínez, R., 2009, Modelo Geológico del Campo, Internal report and presentation: PEMEX Exploración y Producción, unpublished.

Johnson, W. E., R. O. Louden, D. D. Lehman, and D. L. Edwards, 2001, Seismic has its many attributes: AAPG Explorer, http://www.aapg.org/explorer/geophysical_corner/2001/03gpc.cfm.

Michelena, R. J., Gonzales, E. and Capello de P., M., 1998, Similarity analysis: A new tool to summarize seismic attributes information: The Leading Edge, **17**, 545-548.

Poupon, M., Gil, J., D. Vannaxay, and B. Cortiula, Tracking Tertiary delta sands (Urdaneta West, Lake Maracaibo, Venezuela): An integrated seismic facies classification workflow, The Leading Edge, September 2004, 909-912.

Prost, G., and M. Aranda, 2001, Tectonics and hydrocarbon systems of the Veracruz Basin, Mexico, in C. Bartolini, R. T. Buffler, and Cantú-Chapa, eds., The western Gulf of Mexico Basin: Tectonics, sedimentary basins, and petroleum systems: AAPG Memoir, **75**, 271–291.

Romero, Pelaez, A. S., 2012, Prediction of reservoir quality with seismic attributes in Eocene submarine conglomerates, Mexico: MS Thesis, The University of Oklahoma.

The GTM Matlab Toolbox v1.0 beta, M. Svensen, 1996.

Slatt, M. R., and Y. Abousleiman, 2011, Merging sequence stratigraphy and geomechanics for unconventional gas shales. The Leading Edge, 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.

Svensen, M., 1998, GTM: The Generative Topographic Mapping: PhD Thesis, Aston University.

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

CHAPTER 7

CONCLUSIONS

In this dissertation I have implemented and evaluated two latent-space clustering algorithms: Kohonen Self-Organizing Maps (SOM) and Generative Topographic Mapping (GTM). Implemented in LINUX, these algorithms are able to handle multiple large 3D attribute volumes representative of modern seismic surveys. Most published work shows application of SOM to either fluvial or deep water systems. I have applied these algorithms and developed workflows not only to a deep water/mass transport complex/fan system, but to a gas shale reservoir, a diagenetically altered siliceous limestone reservoir, and to a heterogeneous carbonate wash reservoir.

SOM clustering is empirical, assigning each data vector to the "closest" prototype vector which is then displayed in color. Traditional implementations, including my own, do not provide a measure of "how close" a given data vector lies to the best matching prototype vector, or whether it can be estimated almost as well by other prototype vectors. In contrast, GTM is based on sound probabilistic and statistical theories. GTM predicts not only which cluster best represents the data, but how well it is predicted by all other clusters. Thus GTM interfaces neatly with the modern risk analysis workflows. By plotting the cluster results as 2D latent space components, I link the GTM clusters to interactive crossplotting tools, thereby providing additional a posteriori user control and hypothesis testing. While GTM is computationally more demanding than SOM cluster analysis, it is easily parallelized, while the “sequential” SOM implementation presented here does not. For these reasons, I prefer GTM clustering over SOM clustering.

Chapter 3 shows how SOM can be used for seismic stratigraphy, by applying it on the multi-attribute dataset from deep water Gulf of Mexico. After testing alternative input attribute volumes, I conclude that coherent energy, eigenstructure coherence, GLCM entropy and variance texture attributes and the magnitude of reflector convergence, confirm and extend a more traditional manual-implemented seismic stratigraphy interpretation to the full 3D data volume. Although I do not claim this suite of attribute volumes to be the “best”, they provide clear differentiations of MTCs, channel deposits, and basin floor fans. This new workflow shows how over-defined classes clump together after training and how 2D gradational HSV color bars aid in visual interpretation.

Chapter 4 applies SOM to a very different geologic environment – the diagenetically altered Mississippi Lime play of Oklahoma and Kansas. The Mississippi Lime has undergone significant wrench faulting and erosion, suggesting the use of “structural” attributes to differentiate the lithologies. The Mississippi Lime has also undergone significant diagenetic alteration, suggesting the use of “texture” attributes to differentiate the lithologies. In both cases, the input 256 prototype vectors clumped into four or five “natural” clusters. A posteriori analysis using image logs acquired in two horizontal wells shows the three major clusters correlate to the tripolitic chert, the interbedded lime-chert and the tight St. Joe's Limestone within the reservoir zone. Given these observations, I used the borehole image logs from two wells, to define average attribute vectors about these three facies and implemented a supervised workflow based on minimum Euclidean distance (MED) similarity. All three workflows gave consistent results with each other and with a P-impedance result generated by Dowdell et al. (2012).

Chapter 5 introduces GTM which is a more recently introduced development that addresses the shortcomings of the Kohonen SOM algorithm. Other than initial work by colleagues Wallet et al. (2010), it does not appear that GTM has been used for seismic data analysis. My first workflow predicts EUR and a suite of completion and reservoir properties made in a suite of horizontal well parameters in one of the recent shale plays. In what I believe to be an innovative implementation of GTM, the trained clusters are projected onto the 2D latent space and stored as two separate components. These two projection volumes form a cross-plot histogram, allowing the interpreter to interactively construct polygons about hypothesized clusters and visualize the resulting clusters as 3D geobodies. I apply this workflow to a Barnett Shale survey, map clusters, and then classify the clusters as brittle vs. ductile geobodies using microseismic events as control.

Chapter 6 applies GTM to analyze a laterally and vertically highly heterogeneous carbonate conglomerate oil field in the Veracruz Basin of southern Mexico. Unsupervised analysis of reservoir units EOC-10 and EOC-30 shows a heterogeneous probabilistic distribution of different facies that with *a posteriori* well control correspond to good and bad reservoir rocks. I then use these wells in a supervised workflow to quantify the likelihood of finding similar well production (facies) for locations away from the wells within the survey, providing a preliminary risk analysis for the two reservoir units.

In summary, both SOM and GTM provide clusters that are ordered on a 1D, 2D, or 3D latent space. Such ordering and mapping to 1D, 2D, or 3D color bars eliminates the need to know *a priori* how many clusters exist in the data. If the number of clusters are over-specified to be 256 or 4092, both algorithms and human color perception will result in clumping into a much smaller number of natural clusters. Echoing observations by Barnes and Laughlin (2002), the power of clustering is strongly dependent on the choice of input attributes. Attributes such as P and S impedances are direct measurements of lithology. In contrast attributes such as spectral components, reflector convergence vs. parallelism, edge detectors, and textures are direct measures of the depositional and diagenetic history, which are in turn indirectly related to lithology. Given the appropriate 3D attribute volumes, SOM and GTM workflows will not only accelerate, but with GTM, quantify the identification of different lithofacies, petrotypes or heterogeneity present in the reservoir zone. As more wells are drilled within the survey the confidence of the unsupervised seismic facies clustered volume will increase.

RECOMMENDATIONS

The major hurdle in classification lies in selecting the appropriate attribute volumes to bring out the geology of the reservoir correctly. Based on my research I postulate the “best” combination of attribute volumes for the different depositional environments studied in Table 2.

However more work needs to be done on sensitivity analysis and to find out the most and the least contributing attributes during multi-attribute clustering. Such sensitivity analysis is commonly done in nonlinear seismic inversion, whereby the (linearized) Jacobian at the last iteration shows which model parameter influences which measurements, while the eigenvectors of the Hessian show which linear combination of model parameters are well (or not well) resolved. Since GTM is based on probability, it is reasonable to think that a similar sensitivity analysis could be applied to the input attributes and output clusters.

There are also improvements to my algorithmic implementation. At present, the GTM code runs sequentially but could be programmed to run in parallel under MPI. I attempted an implementation of the “batch” vs. “sequential” implementation of SOM, with the idea that the batch process could be parallelized, while the sequential could not. Unfortunately, I found the convergence of the batch implementation for my seismic attribute volumes to be very slow.

Another software improvement would be to allow interactive definition of clusters for supervised classification directly in my “AASPI” implementation. At present, I identify the zones around the wells in Petrel, manually specify the inlines, the crosslines and the start and end times as input in AASPI, and then form vectors out of them, which is not only tedious, but allows for possible user errors.

My implementations were done on target zones, constrained by the start and end time for the input volumes or with flattened volumes. A further software improvement would be to modify the SOM and the GTM code to limit the multi-attribute volume analysis to lie within a given user-defined target formation constrained by the top and below surfaces, thereby removing redundant data affecting the model while training.

Table 2. Best Choice of Attributes

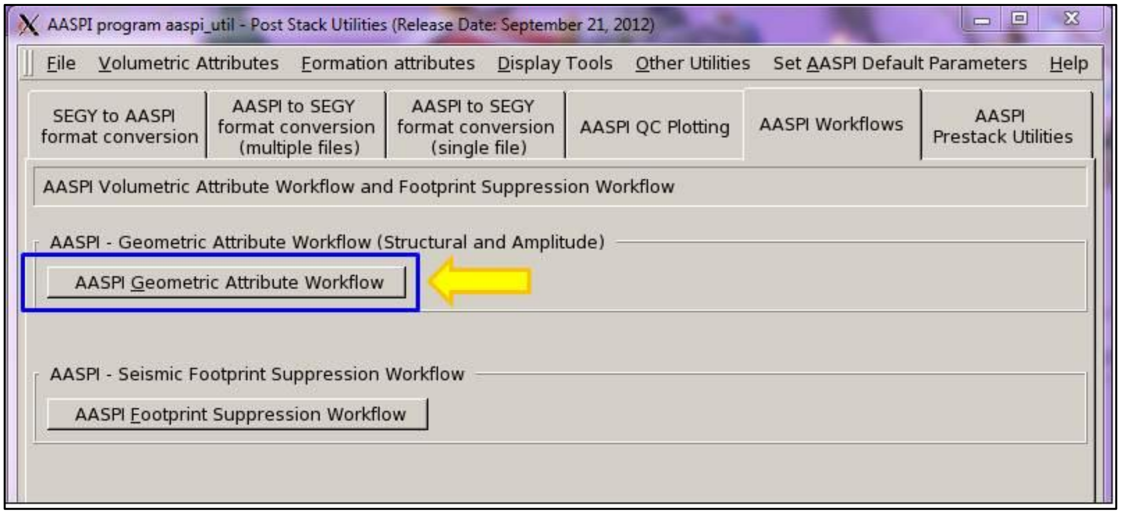
Reservoir Type	**"Best" Attribute Volumes**
Fluvial Depositional System/ Deep water depositional system	GLCM energy, GLCM entropy, GLCM homogeneity, dip magnitude, reflector convergence, coherence
Unconventional Shale plays	P-impedance, λρ, μρ and may be some anisotropy volumes
Mississippian Chert plays	GLCM texture attribute volumes, spectral bandwidth and the Impedance volume (if available)
Laterally and vertically heterogeneous conglomerate carbonate reservoirs	P-impedance, λρ, μρ, Vp/Vs, anisotropy volumes

APPENDICES

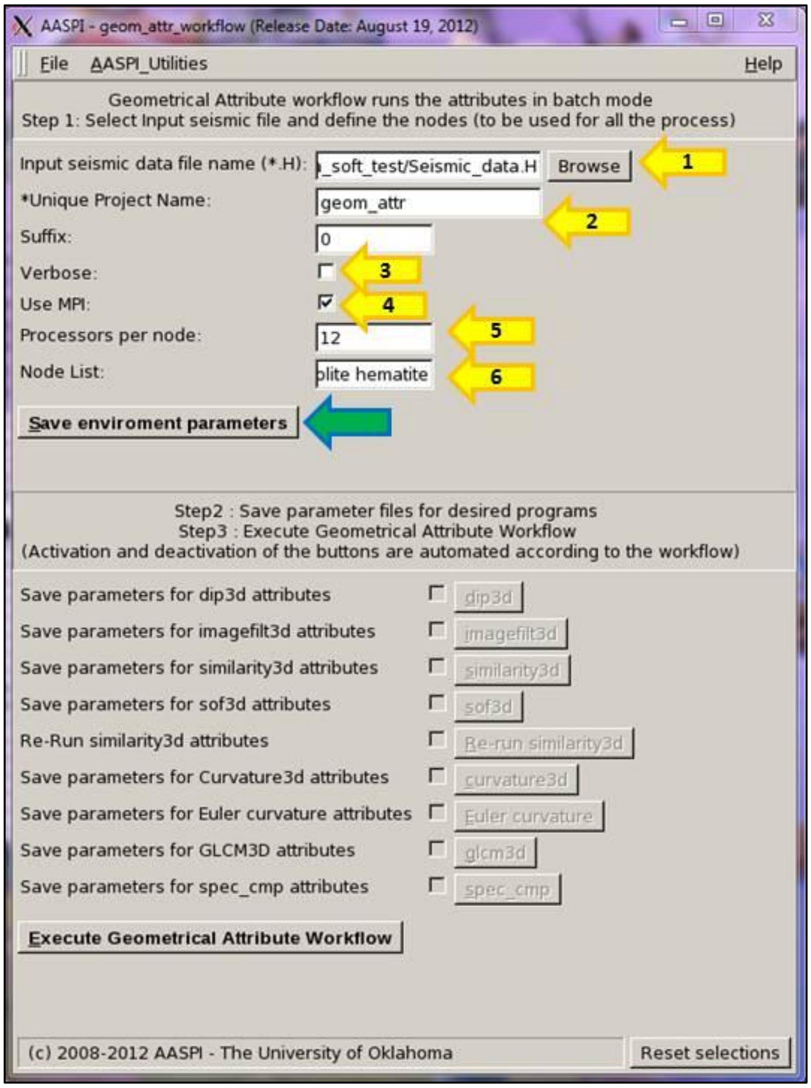
As a part of my research, I wrote all my programs and subroutines in FORTRAN90 using out AASPI I/O. In addition, I have generated interactive Graphic User Interfaces (GUIs), shell script and documentation so that OU students, staffs as well as AASPI sponsors find it easy to use my software. Below I discuss the GUIs, which are directly related to my research.

APPENDIX 1: Geometrical attribute Workflow GUI

Attribute computation of a very large data volumes can take considerable time. Experienced interpreters may already have familiarity with other data volumes from the same basin. Alternatively, they may need to have multi-attribute dataset to do some unsupervised of probabilistic seismic facies analysis. In this situation it may be useful to set up a workflow that will run a suite of attribute programs in the background, perhaps overnight.



The AASPI Geometrical Attribute Workflow GUI can be invoked from the main aaspi_util as shown above or by typing in aaspi_geom_attr_workflow separately in the terminal window. The following workflow GUI will then pop up.



AASPI - geom_attr_workflow (Release Date: August 19, 2012)

Geometrical Attribute workflow runs the attributes in batch mode
Step 1: Select Input seismic file and define the nodes (to be used for all the process)

Input seismic data file name (*.H): _soft_test/Seismic_data.H
*Unique Project Name: geom_attr
Suffix: 0
Verbose: ☐
Use MPI: ☑
Processors per node: 12
Node List: plite hematite

Step2 : Save parameter files for desired programs
Step3 : Execute Geometrical Attribute Workflow
(Activation and deactivation of the buttons are automated according to the workflow)

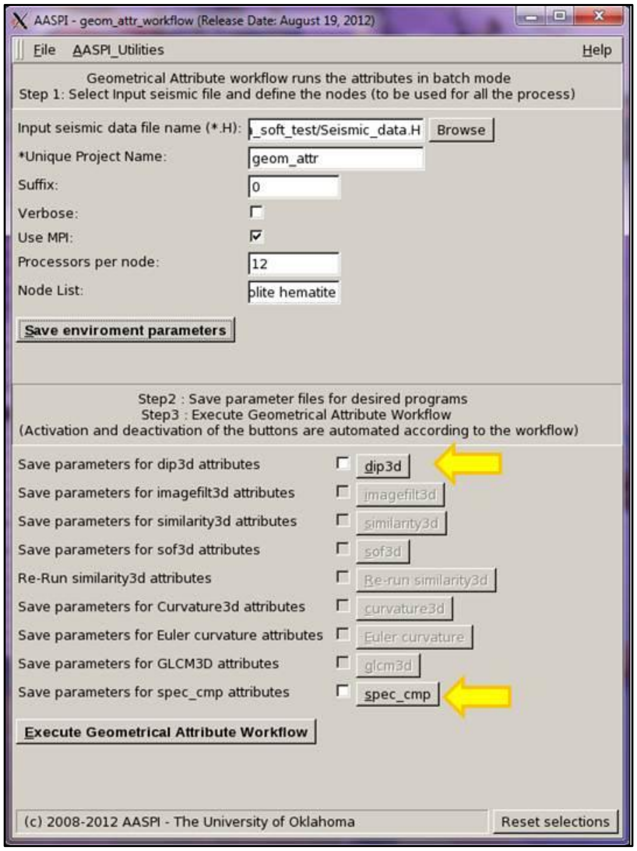
Save parameters for dip3d attributes ☐
Save parameters for imagefilt3d attributes ☐
Save parameters for similarity3d attributes ☐
Save parameters for sof3d attributes ☐
Re-Run similarity3d attributes ☐
Save parameters for Curvature3d attributes ☐
Save parameters for Euler curvature attributes ☐
Save parameters for GLCM3D attributes ☐
Save parameters for spec_cmp attributes ☐

(c) 2008-2012 AASPI - The University of Oklahoma

Step 1: Save the workflow environment parameters

In step 1 we need to input the seismic amplitude file and set up the project name and the MPI parameters which will be used for all the MPI processes. The seismic amplitude file is selected first (*Arrow 1*). Enter the project name and the suffix (*Arrow 2*). Verbose can be selected if required (*Arrow 3*). It is recommended to use MPI because except euler_curvature all the other processes run on MPIs (*Arrow 4*). Mention the processors per nodes and the node list. Each of our machines *tripolite.ou.edu* and *hematite.ou.edu* have 12 processors in it. Thus in the processors per node 12 is mentioned (*Arrow 5*) and in the node list *tripolite* and *hematite* is mentioned (*Arrow 6*).

After entering out all the parameters these parameters are saved (*Green Arrow*) which will be subsequently used for all the processes. Note that initially all the attribute buttons will be disabled. When the "Save Environment parameters" is clicked the *dip3d* and the *spec_cmp* buttons will be highlighted as shown. These two takes in only the seismic amplitude as inputs and are thus activated. The subsequent attribute buttons will be activated after their input file criterions are met.



Geometrical Attribute workflow runs the attributes in batch mode
Step 1: Select Input seismic file and define the nodes (to be used for all the process)

Input seismic data file name (*.H): _soft_test/Seismic_data.H
*Unique Project Name: geom_attr
Suffix: 0
Verbose: ☐
Use MPI: ☑
Processors per node: 12
Node List: lite hematite

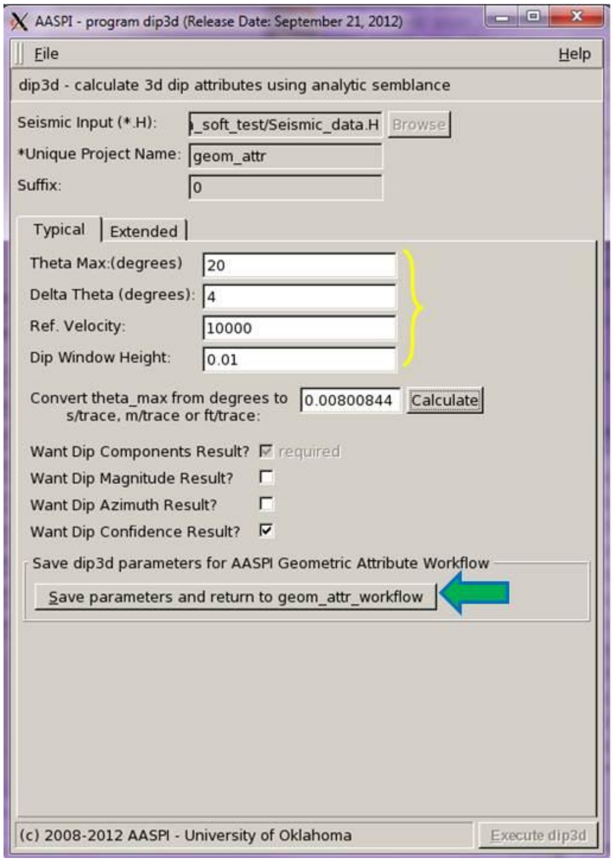
Step2 : Save parameter files for desired programs
Step3 : Execute Geometrical Attribute Workflow
(Activation and deactivation of the buttons are automated according to the workflow)

Save parameters for dip3d attributes ☐
Save parameters for imagefilt3d attributes ☐
Save parameters for similarity3d attributes ☐
Save parameters for sof3d attributes ☐
Re-Run similarity3d attributes ☐
Save parameters for Curvature3d attributes ☐
Save parameters for Euler curvature attributes ☐
Save parameters for GLCM3D attributes ☐
Save parameters for spec_cmp attributes ☐

(c) 2008-2012 AASPI - The University of Oklahoma

Step 2: Save the parameters for the volumetric attributes

In this step each of the attribute program is opened and their parameters are saved. The buttons are activated only when their input criterion are met. For example the imagefilt3d gets activated only after we open and save the dip3d parameters. The next figure shows the GUIs for dip3d and imagefilt3d. The parameters are mentioned and the Save and Exit button (*green arrow*) is pressed.



dip3d - calculate 3d dip attributes using analytic semblance

Seismic Input (*.H): _soft_test/Seismic_data.H
*Unique Project Name: geom_attr
Suffix: 0

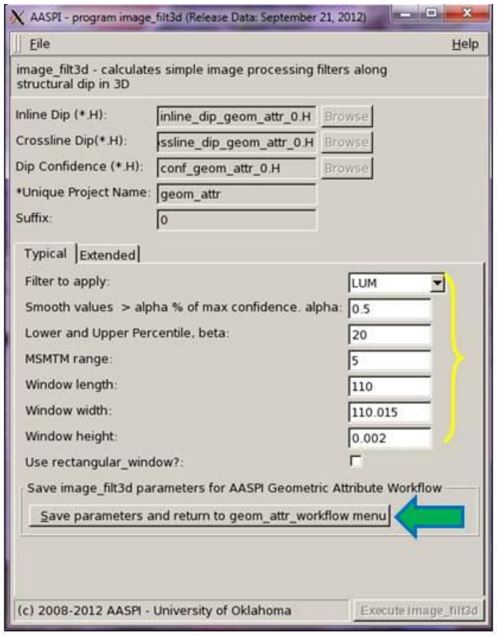
Theta Max:(degrees) 20
Delta Theta (degrees): 4
Ref. Velocity: 10000
Dip Window Height: 0.01

Convert theta_max from degrees to s/trace, m/trace or ft/trace: 0.00800844

Want Dip Components Result? ☑ required
Want Dip Magnitude Result? ☐
Want Dip Azimuth Result? ☐
Want Dip Confidence Result? ☑

Save dip3d parameters for AASPI Geometric Attribute Workflow
Save parameters and return to geom_attr_workflow

(c) 2008-2012 AASPI - University of Oklahoma



image_filt3d - calculates simple image processing filters along structural dip in 3D

Inline Dip (*.H): inline_dip_geom_attr_0.H
Crossline Dip(*.H): ssline_dip_geom_attr_0.H
Dip Confidence (*.H): conf_geom_attr_0.H
*Unique Project Name: geom_attr
Suffix: 0

Filter to apply: LUM
Smooth values > alpha % of max confidence. alpha: 0.5
Lower and Upper Percentile, beta: 20
MSMTM range: 5
Window length: 110
Window width: 110.015
Window height: 0.002
Use rectangular_window?: ☐

Save image_filt3d parameters for AASPI Geometric Attribute Workflow
Save parameters and return to geom_attr_workflow menu

(c) 2008-2012 AASPI - University of Oklahoma

Step 3: Execute the geometric attribute workflow

After saving all the *.parms (parameter) files they show up in the terminal window. We can do cat *aaspi_env.parms* to see the file contents. The other *.parms contents the saved parameters from all saved programs. After that, press the "Execute Geometrical Attribute Workflow" button. The reset selection button can be pressed if one wants to reset the program selections.

A typical workflow for *structural geometrical attributes* will be dip3d>imagefilt3d>similarity3d>sof3d>rerun-similarity3d>k_curvature3d>k_euler_curvature>

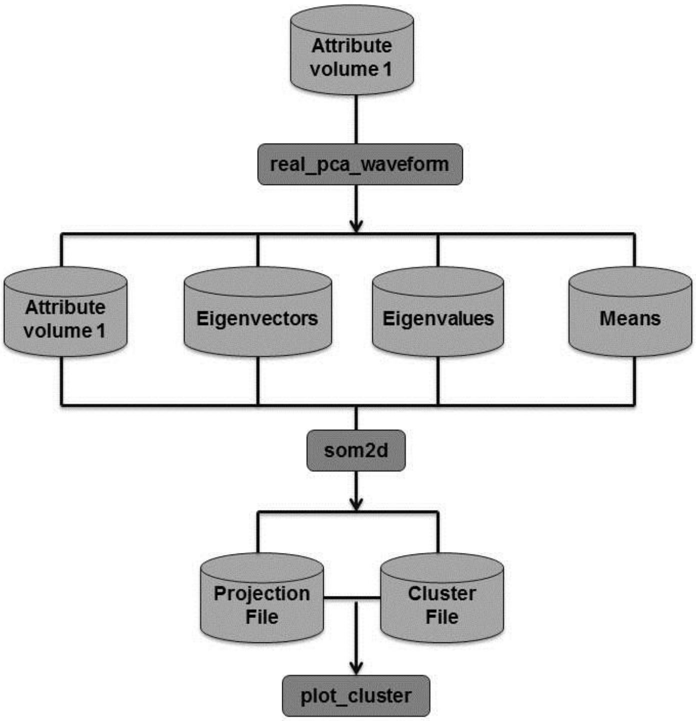
A typical workflow for *amplitude geometrical attributes* will be dip3d>imagefilt3d>similarity3d>sof3d>rerun-similarity3d>e_curvature3d>e_euler_curvature>glcm3d>spec_cmp

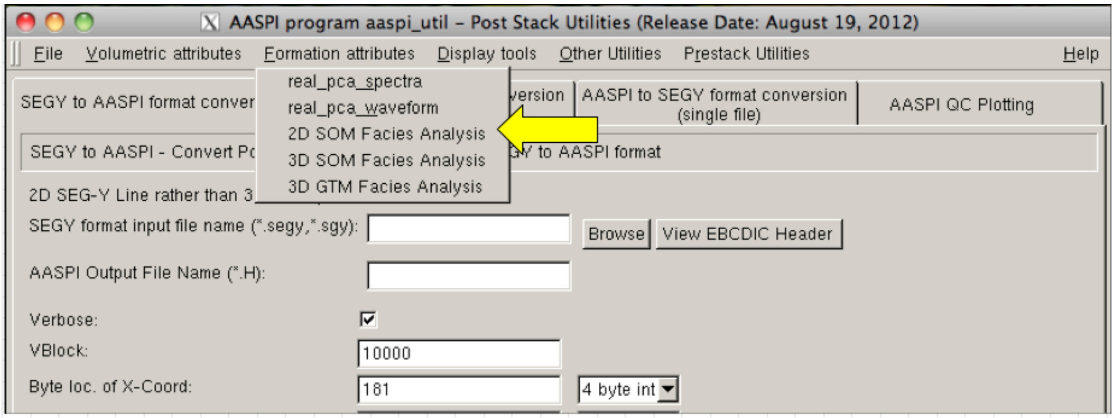
At any time, the terminal window will show the progress of the workflow. The text file *aaspi_geom_attr_workflow.out* can be checked to see the completion status of the workflow or whether there is any error in the execution of the program.

APPENDIX 2: 2D Kohonen SOM GUI

Computation flow chart

This 2D Facies classification analysis is comprised of three separate modules; **real_pca_waveform**, **som2d** and **plot_clusters**. The *real_pca_waveform* preconditions the different input attribute volume, which goes into as input to the som2d. It calculates the eigenvalues and eigenvectors from input dataset which will be used to project the input data vector into the latent space. The *som2d* program trains the data based on Kohonen Self-organizing Maps. The last module - plot_cluster assigns colors to the different trained facies into a 2D RGB gradational scale and plots the output seismic facies map (Matos et al., 2009, Roy et al., 2011). Below is the flowchart showing the workflow of 2D seismic facies analysis.





SEGY format input file name (*.segy,*.sgy):

AASPI Output File Name (*.H):

Verbose: ☑

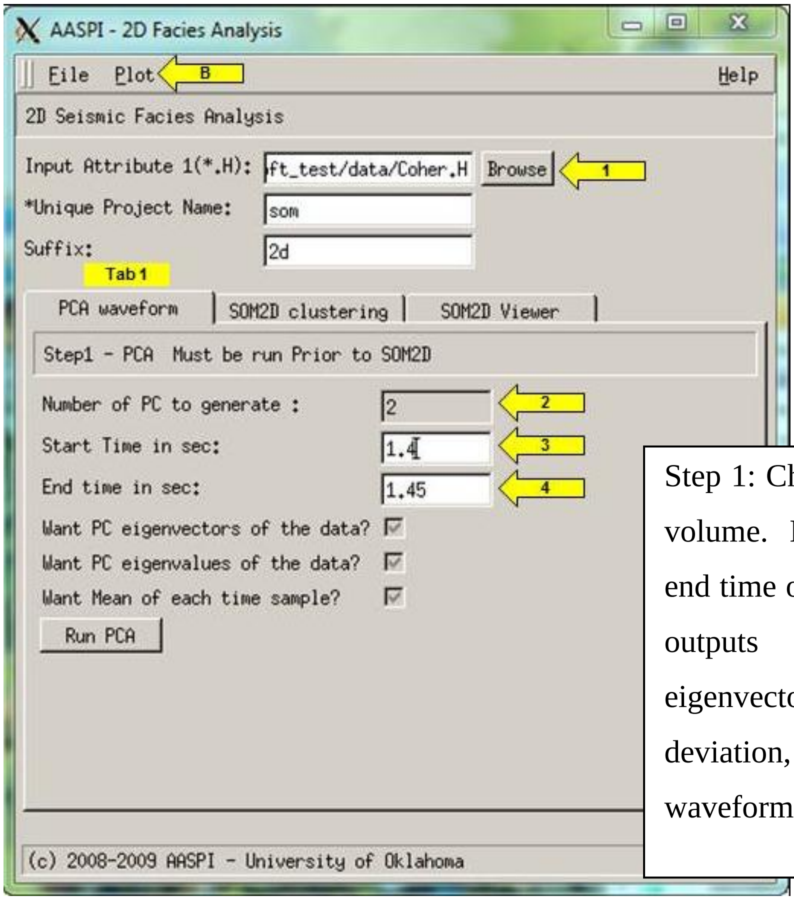
VBlock: 10000

Byte loc. of X-Coord: 181 4 byte int

This Program **2D Facies Analysis** is launched from the *Formation Attributes* in the main **aaspi_util** GUI.

Computing real_pca_waveform module

This is the first step of analysis. The input attribute volume is selected on which the waveform classification is to be done. Only the first two eigenvectors are used for som2d, thus it has been fixed at 2 (yellow arrow 2). The start time and the end time (arrows 3 and 4 respectively) are used to define the window of data used for the analysis.



2D Seismic Facies Analysis

Input Attribute 1(*.H): ft_test/data/Coher.H

*Unique Project Name: som

Suffix: 2d

Step1 - PCA Must be run Prior to SOM2D

Number of PC to generate : 2

Start Time in sec: 1.4

End time in sec: 1.45

Want PC eigenvectors of the data? ☑

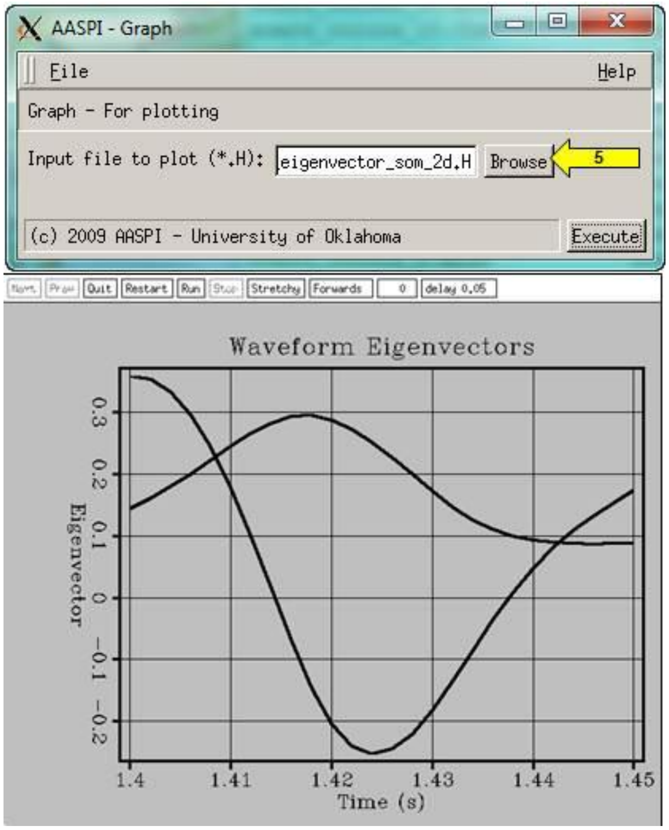
Want PC eigenvalues of the data? ☑

Want Mean of each time sample? ☑

(c) 2008-2009 AASPI - University of Oklahoma

Step 1: Choose the input attribute volume. Mention the start and end time of the data. This module outputs the eigenvalues, eigenvectors, the standard deviation, mean and the average waveform of the input data

To QC the outputs from the real_pca_waveform program we can plot the eigenvectors, the eigenvalues and the means in the simple graph utility as shown above. The eigenvalues and the eigenvectors, which form the initial set or a priori training vectors, are shown below. The horizontal axis represents the samples of the waveform used in the analysis. The plot of the eigenvectors is the plot of the first two eigenvectors.



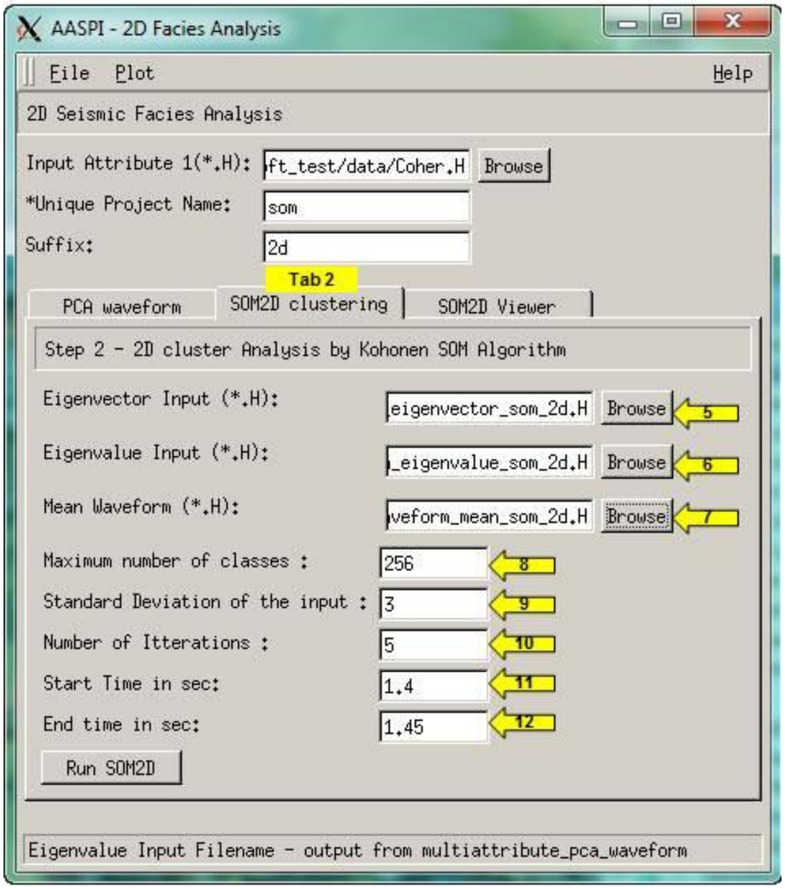
Graph - For plotting

Input file to plot (*.H): eigenvector_som_2d.H

(c) 2009 AASPI - University of Oklahoma

Computing som2d module

The som2D-clustering is the second step for analysis step and appears in the next tab (Tab 2). The input for the real_pca_waveform should remain the same for this som2D analysis. The output from the previous real_pca_waveform serves as input for som3D (shown in yellow arrows). The maximum *number of classes* should be <= 256, because most visualization software can only display 256 colors. The eigenvalues are used to scale the dimension of the latent space. The eigenvectors serve as the first approximation to the latent space forming the initial set of untrained vectors. The maximum *number of iterations* is required. The rate of convergence is printed in the output file and can be use guide subsequent jobs. The start time should be same as the *real_pca_waveform*.



2D Seismic Facies Analysis

Input Attribute 1(*.H): ft_test/data/Coher.H

*Unique Project Name: som

Suffix: 2d

Step 2 - 2D cluster Analysis by Kohonen SOM Algorithm

Eigenvector Input (*.H): eigenvector_som_2d.H

Eigenvalue Input (*.H): _eigenvalue_som_2d.H

Mean Waveform (*.H): veform_mean_som_2d.H

Maximum number of classes : 256

Standard Deviation of the input : 3

Number of Itterations : 5

Start Time in sec: 1.4

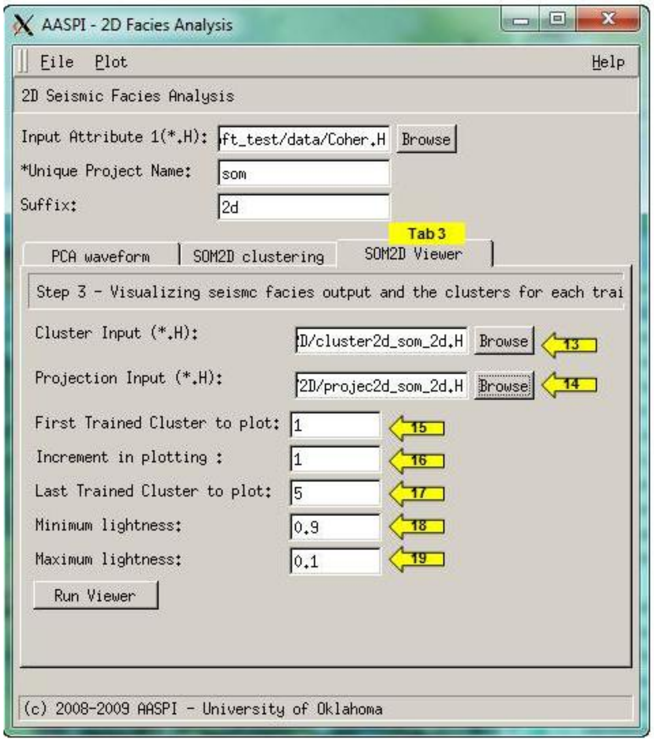
End time in sec: 1.45

Eigenvalue Input Filename - output from multiattribute_pca_waveform

Step 2: The input attribute volume should remain the same as real_pca_waveform. The eigenvectors, the eigenvalues and mean values are selected as input (marked with yellow arrows 5,6,7). The number of classes has a maximum limit of 256 because of the display limitations of the most visualization software which has a max of 256 colors that can be imported into it. Select the number of iterations (arrow10). Keep the start and end time same as before. This program outputs the 2D clustered data set for each iteration and the projected Trained Vectors (Prototype vectors). It also outputs the final trained 2D dataset, which can be exported as a SEGY file into other visualization software.

Plotting: plot_clusters module

The 3rd step is to color the trained 2D seismic dataset and highlight the variation in seismic facies. Go to Tab3 for the SOM 2D viewer. Different facies are represented by different colors. The outputs from the SOM2D serve as input to this module (shown in yellow arrows). This module helps in QC the facies volume after each training. This also generates a suite of color files which can be taken as input in visualization software like Petrel.



2D Seismic Facies Analysis

Input Attribute 1(*.H): ft_test/data/Coher.H

*Unique Project Name: som

Suffix: 2d

Step 3 - Visualizing seismc facies output and the clusters for each trai

Cluster Input (*.H): D/cluster2d_som_2d.H

Projection Input (*.H): 2D/projec2d_som_2d.H

First Trained Cluster to plot: 1

Increment in plotting : 1

Last Trained Cluster to plot: 5

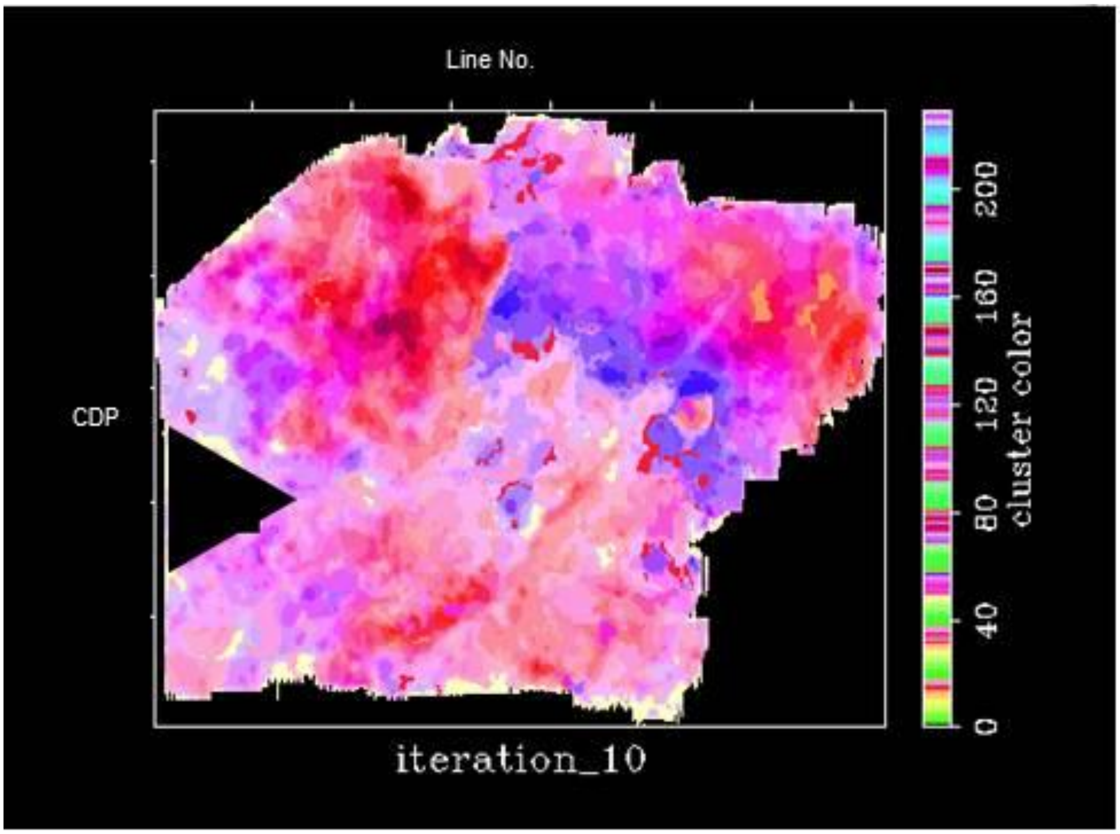
Minimum lightness: 0.9

Maximum lightness: 0.1

(c) 2008-2009 AASPI - University of Oklahoma

Step 3: This step creates the various color-files and colors the projected trained vectors and the 2D trained dataset. The cluster file (cluster2d*) and the projection file (project2d*) are taken as input (arrows 13 and 14). The Minimum lightness and the maximum lightness values help in changing the minimum and maximum saturation value of the 2D gradational colorbar (arrows 18 and 19). The *.alut file generated by this module can be imported into Petrel (shown in the later section) for seismic facies map visualization.

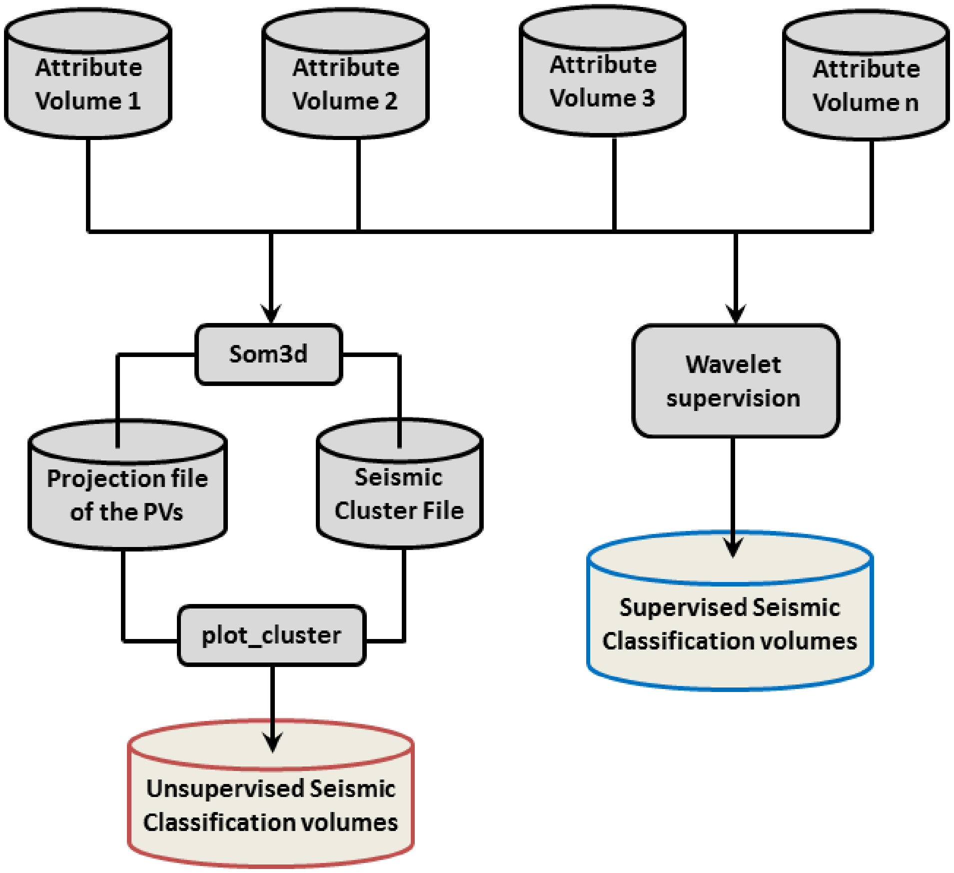
Below shows the output of a seismic facies map considering the coherent energy as the input volume after 10 iterations.

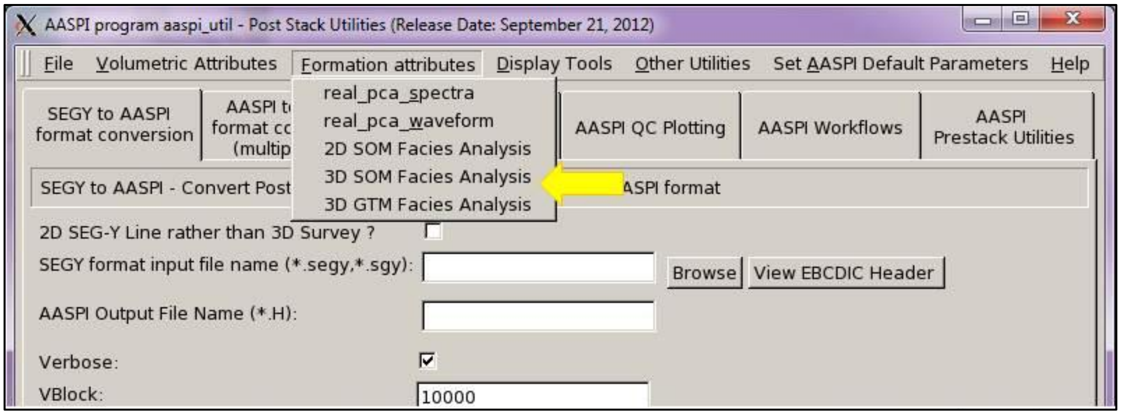


APPENDIX 3: 3D Multi-attribute Kohonen SOM GUI

Computation flow chart

This 3D Facies classification analysis is comprised of two separate modules; **som3d**, **plot_clusters**. Also there is a **wavelet supervision** program included within the som3d code which compares the dataset near the wells with the multi-attribute input volume. The last module - plot_cluster assigns colors to the different trained facies into a 2D RGB gradational scale and plots the output seismic facies volume. Below is the flowchart showing the workflow of 3D seismic facies analysis.





2D SEG-Y Line rather than 3D Survey ? ☐

SEGY format input file name (*.segy,*.sgy):

AASPI Output File Name (*.H):

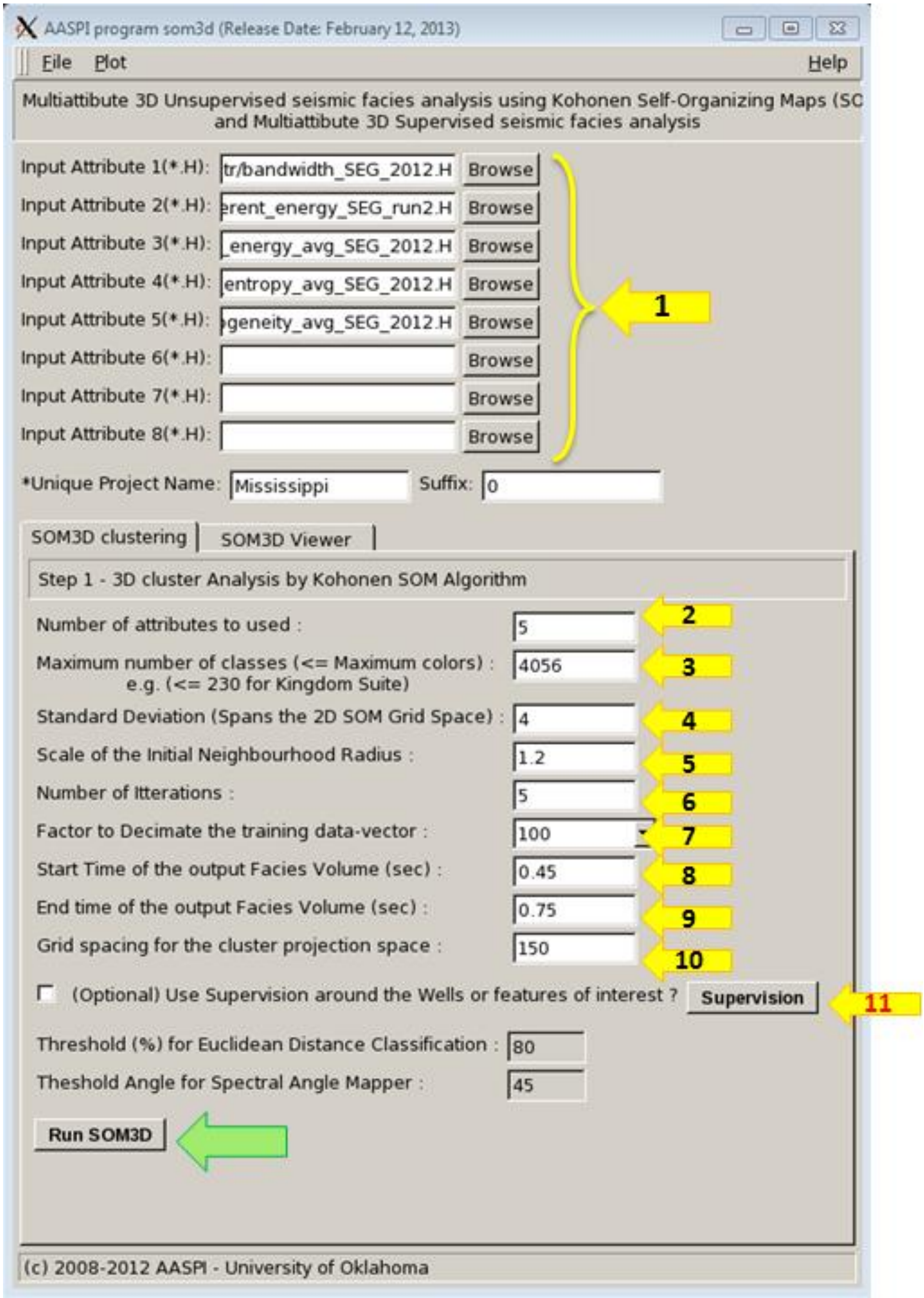
Verbose: ☑

VBlock: 10000

This Program **3D Facies Analysis** is launched from the Formation Attributes in the main **aaspi_util** GUI

Computing som3d module

The som3D-clustering is the first step of analysis. Use the browser on the first eight lines to choose the input seismic data file (Arrow 1). It is not mandatory to take in eight inputs. The number of inputs can vary from two – eight. Specify the number of input attributes in the field labeled “Number of attributes to use” (Arrow 2). The maximum number of classes can be any large number (Arrow 3). Most of the commercial visualization software can only display 256 colors thus generally is <= 256. However the more the uniform sampling of the latent space takes place we generally have more confidence in clustering. Thus, in this case we take a 4096 classes which later can be represented by a 64 – by – 64 colorscale. The initial neighborhood radius can be specified as the constant factor of the initial SOM 2D grid spacing of the PVs (Arrow 5). The number of iteration is mentioned next (Arrow 6). Select the decimation factor for using a subset of dataset used for training (Arrow 7). For example the value 100 means every 100th data-vector is used for training. The start time and the end time are automatically read in from the input files (Arrow 8 and 9).However it is better to mention them to focus the analysis only within the reservoir zone. The GUI for the SOM3D is shown in the next page.



Multiattibute 3D Unsupervised seismic facies analysis using Kohonen Self-Organizing Maps (SC
and Multiattibute 3D Supervised seismic facies analysis

Input Attribute 1(*.H): tr/bandwidth_SEG_2012.H

Input Attribute 2(*.H): erent_energy_SEG_run2.H

Input Attribute 3(*.H): _energy_avg_SEG_2012.H

Input Attribute 4(*.H): entropy_avg_SEG_2012.H

Input Attribute 5(*.H): geneity_avg_SEG_2012.H

Input Attribute 6(*.H):

Input Attribute 7(*.H):

Input Attribute 8(*.H):

*Unique Project Name: Mississippi Suffix: 0

Step 1 - 3D cluster Analysis by Kohonen SOM Algorithm

Number of attributes to used : 5

Maximum number of classes (<= Maximum colors) :
e.g. (<= 230 for Kingdom Suite) 4056

Standard Deviation (Spans the 2D SOM Grid Space) : 4

Scale of the Initial Neighbourhood Radius : 1.2

Number of Itterations : 5

Factor to Decimate the training data-vector : 100

Start Time of the output Facies Volume (sec) : 0.45

End time of the output Facies Volume (sec) : 0.75

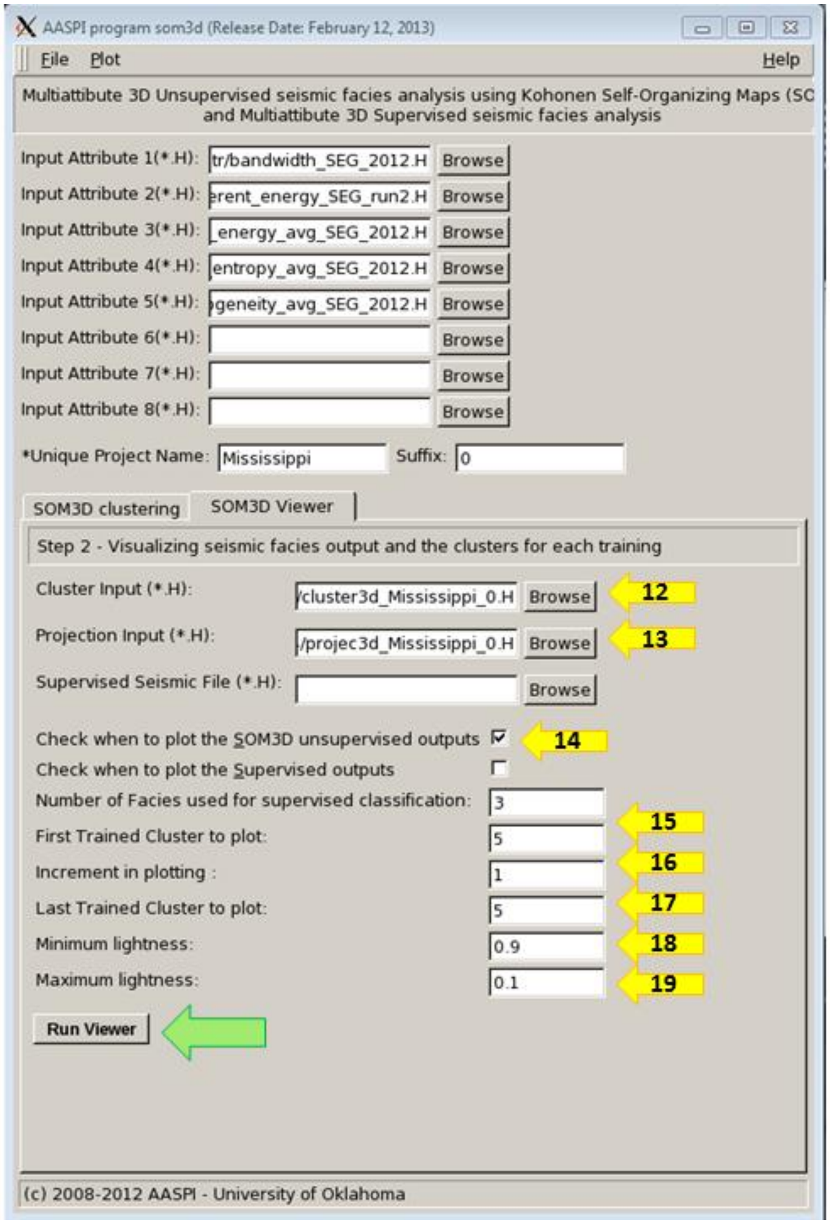
Grid spacing for the cluster projection space : 150

☐ (Optional) Use Supervision around the Wells or features of interest ?

Threshold (%) for Euclidean Distance Classification : 80

Theshold Angle for Spectral Angle Mapper : 45

(c) 2008-2012 AASPI - University of Oklahoma



AASPI program som3d (Release Date: February 12, 2013)

Multiattibute 3D Unsupervised seismic facies analysis using Kohonen Self-Organizing Maps (SO
and Multiattibute 3D Supervised seismic facies analysis

Input Attribute 1(*.H): tr/bandwidth_SEG_2012.H
Input Attribute 2(*.H): erent_energy_SEG_run2.H
Input Attribute 3(*.H): _energy_avg_SEG_2012.H
Input Attribute 4(*.H): entropy_avg_SEG_2012.H
Input Attribute 5(*.H): geneity_avg_SEG_2012.H
Input Attribute 6(*.H):
Input Attribute 7(*.H):
Input Attribute 8(*.H):

*Unique Project Name: Mississippi Suffix: 0

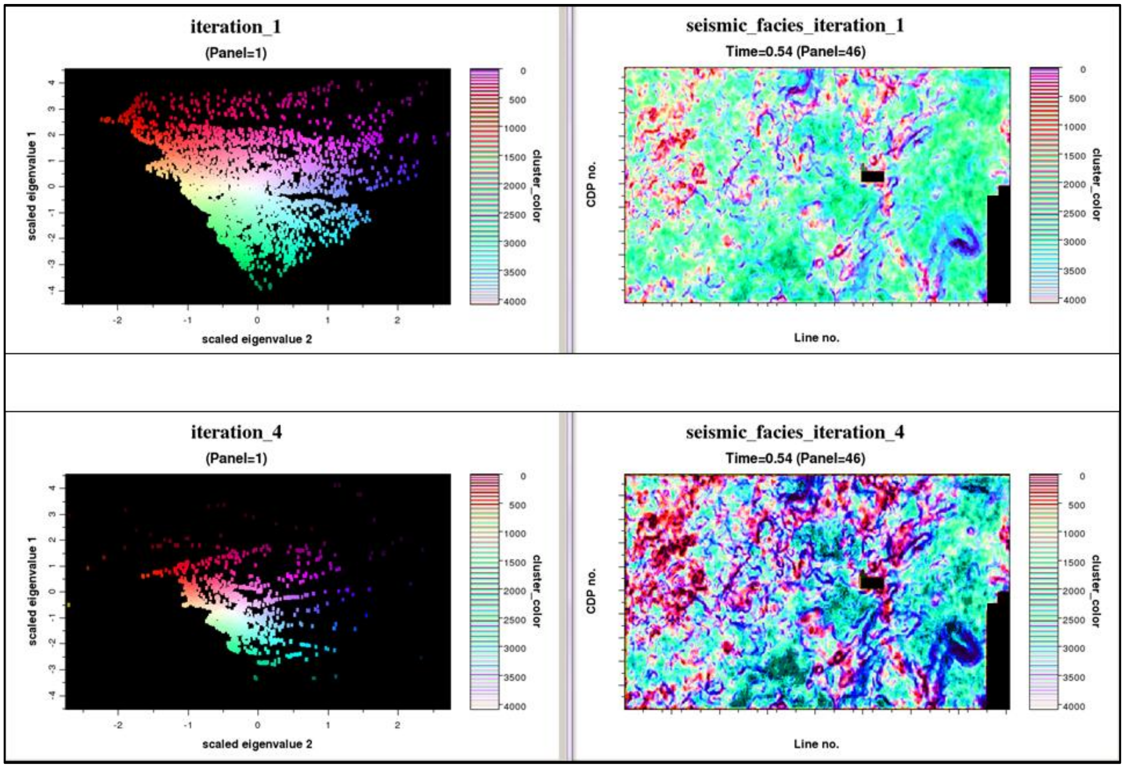
Step 2 - Visualizing seismic facies output and the clusters for each training

Cluster Input (*.H): /cluster3d_Mississippi_0.H
Projection Input (*.H): /projec3d_Mississippi_0.H
Supervised Seismic File (*.H):

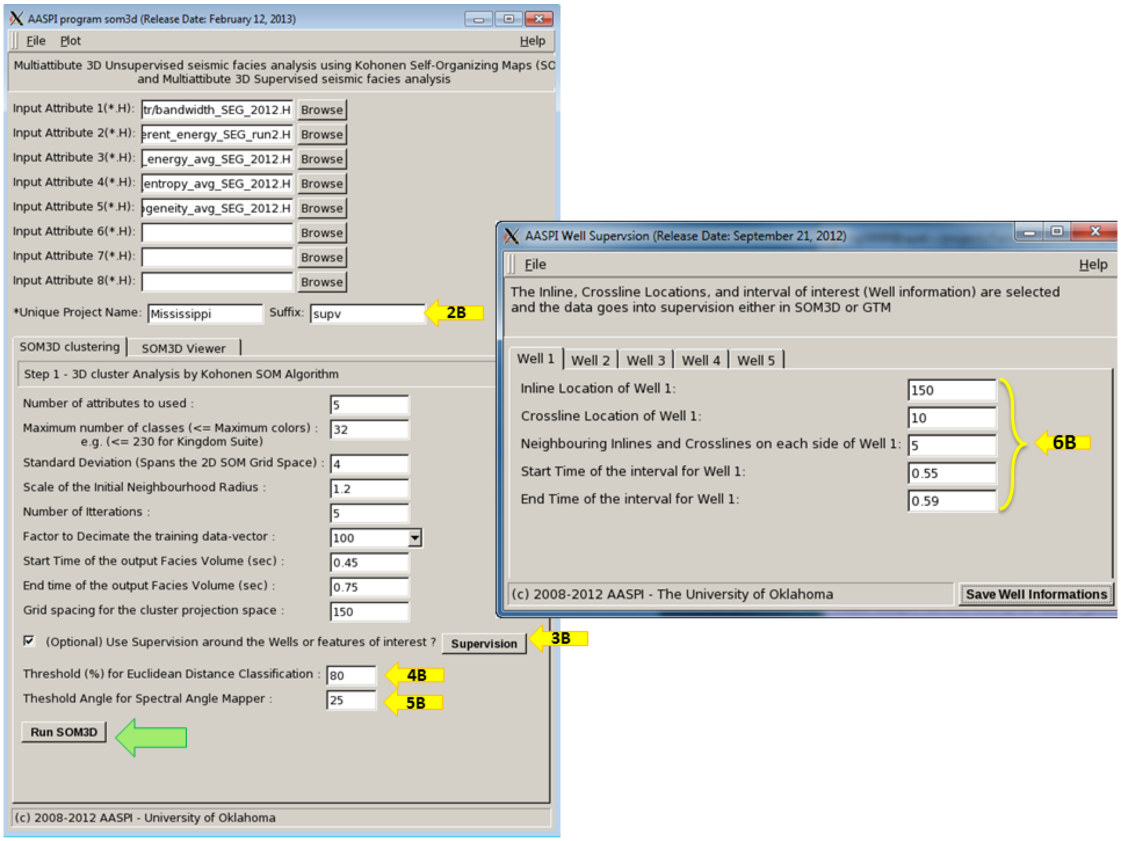
Check when to plot the SOM3D unsupervised outputs ☑
Check when to plot the Supervised outputs ☐
Number of Facies used for supervised classification: 3
First Trained Cluster to plot: 5
Increment in plotting : 1
Last Trained Cluster to plot: 5
Minimum lightness: 0.9
Maximum lightness: 0.1

(c) 2008-2012 AASPI - University of Oklahoma

The 2nd step is to color the trained 3D seismic dataset and highlight the variation in seismic facies. Different facies are represented by different colors. The SOM3D Viewer in the next tab. The outputs from the SOM3D serve as input to this module (Arrow 12 and 13). This module helps in QC the facies volume after each training. This also generates a suite of color files, which can be taken as input in visualization software like Petrel.



Output generated from the SOM3D facies clustering module with 4096 number of initial classes. Note that clusters starts forming for iteration 4 and the overdefined 4096 classes have actually represents 4-5 colors. In the visualization software like Petrel only a max of 256 colors can be imported. Thus we can use a lesser number of initial classes (256).



AASPI program som3d (Release Date: February 12, 2013)

Multiattibute 3D Unsupervised seismic facies analysis using Kohonen Self-Organizing Maps (SC
and Multiattibute 3D Supervised seismic facies analysis

Input Attribute 1(*.H): tr/bandwidth_SEG_2012.H
Input Attribute 2(*.H): erent_energy_SEG_run2.H
Input Attribute 3(*.H): _energy_avg_SEG_2012.H
Input Attribute 4(*.H): entropy_avg_SEG_2012.H
Input Attribute 5(*.H): geneity_avg_SEG_2012.H
Input Attribute 6(*.H):
Input Attribute 7(*.H):
Input Attribute 8(*.H):
*Unique Project Name: Mississippi
Suffix: supv

Step 1 - 3D cluster Analysis by Kohonen SOM Algorithm

Number of attributes to used : 5
Maximum number of classes (<= Maximum colors) : e.g. (<= 230 for Kingdom Suite) 32
Standard Deviation (Spans the 2D SOM Grid Space) : 4
Scale of the Initial Neighbourhood Radius : 1.2
Number of Itterations : 5
Factor to Decimate the training data-vector : 100
Start Time of the output Facies Volume (sec) : 0.45
End time of the output Facies Volume (sec) : 0.75
Grid spacing for the cluster projection space : 150
☑ (Optional) Use Supervision around the Wells or features of interest ?
Threshold (%) for Euclidean Distance Classification : 80
Theshold Angle for Spectral Angle Mapper : 25

(c) 2008-2012 AASPI - University of Oklahoma

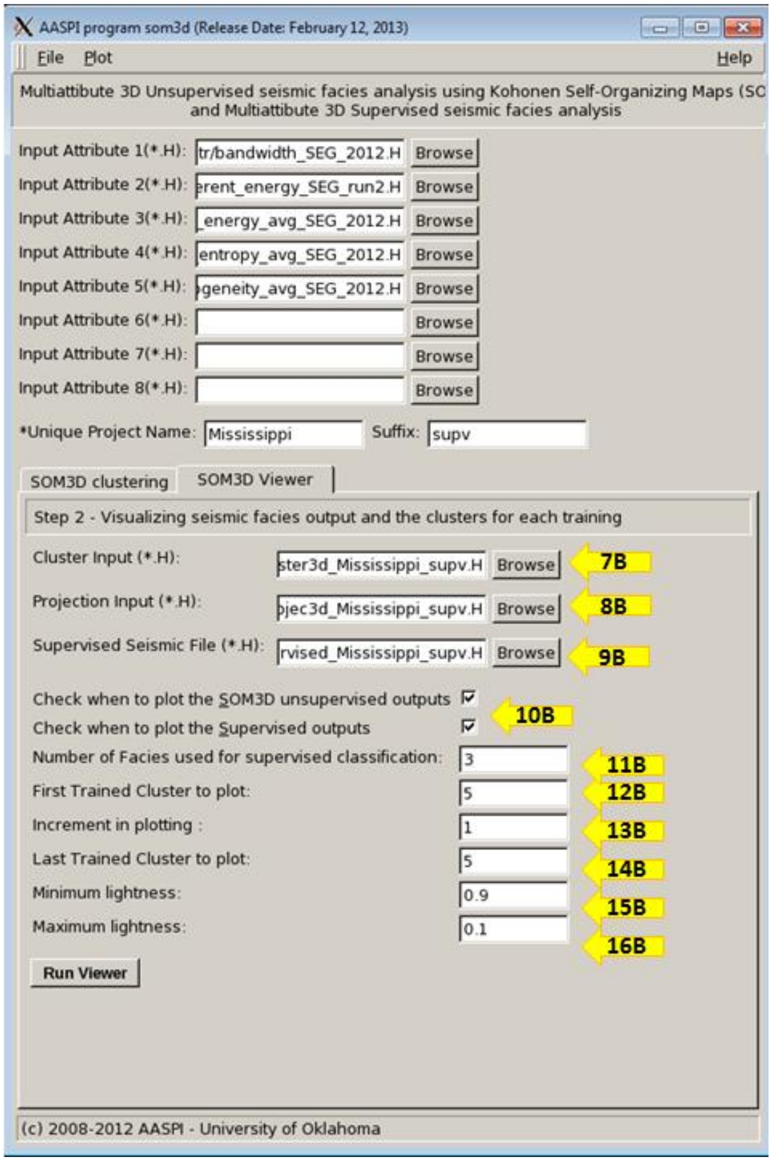
AASPI Well Supervsion (Release Date: September 21, 2012)

The Inline, Crossline Locations, and interval of interest (Well information) are selected
and the data goes into supervision either in SOM3D or GTM

Inline Location of Well 1: 150
Crossline Location of Well 1: 10
Neighbouring Inlines and Crosslines on each side of Well 1: 5
Start Time of the interval for Well 1: 0.55
End Time of the interval for Well 1: 0.59

(c) 2008-2012 AASPI - The University of Oklahoma

Supervised classification can be done based on either the Minimum Euclidean Distance (MED) or Spectral Angle Mapper (SAM). When Supervised (Arrow 3B) is pressed the well parameter window appears where we can define a cube specifying the inline, crossline, neighboring radius and start and end time. An average attribute data-vector is calculated from this cube, which becomes the training data-vector. A cutoff can be given for MED (Arrow 4B) and for SAM (Arrow 5B). The MED parameter is the similarity ratio and the SAM parameter is the cutoff angle.



Multiattibute 3D Unsupervised seismic facies analysis using Kohonen Self-Organizing Maps (SC
and Multiattibute 3D Supervised seismic facies analysis

Input Attribute 1(*.H): tr/bandwidth_SEG_2012.H
Input Attribute 2(*.H): erent_energy_SEG_run2.H
Input Attribute 3(*.H): _energy_avg_SEG_2012.H
Input Attribute 4(*.H): entropy_avg_SEG_2012.H
Input Attribute 5(*.H): geneity_avg_SEG_2012.H
Input Attribute 6(*.H):
Input Attribute 7(*.H):
Input Attribute 8(*.H):

*Unique Project Name: Mississippi
Suffix: supv

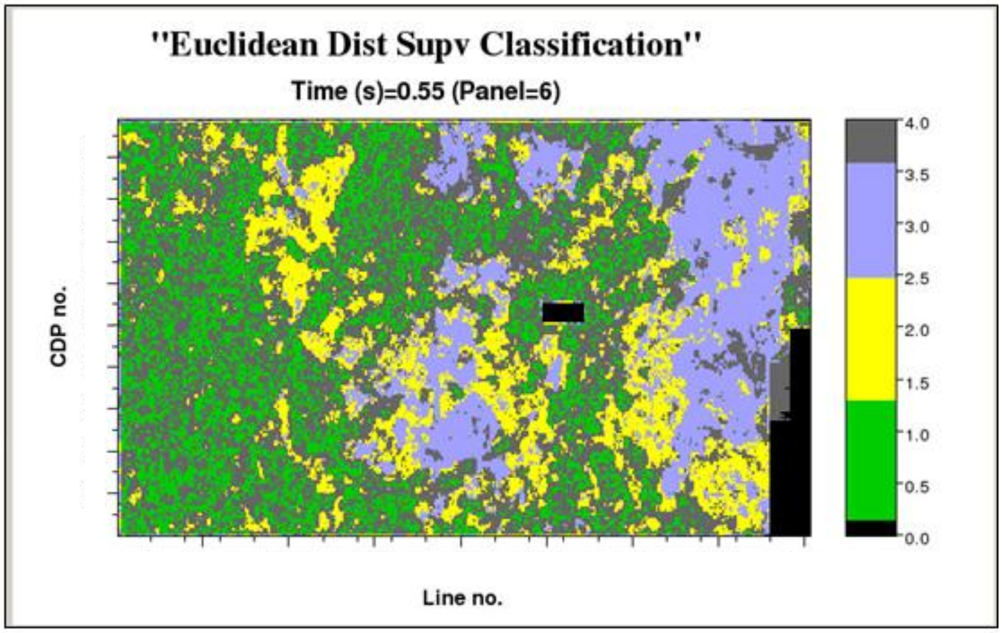
Step 2 - Visualizing seismic facies output and the clusters for each training

Cluster Input (*.H): ster3d_Mississippi_supv.H
Projection Input (*.H): pjec3d_Mississippi_supv.H
Supervised Seismic File (*.H): rvised_Mississippi_supv.H

Check when to plot the SOM3D unsupervised outputs ☑
Check when to plot the Supervised outputs ☑
Number of Facies used for supervised classification: 3
First Trained Cluster to plot: 5
Increment in plotting : 1
Last Trained Cluster to plot: 5
Minimum lightness: 0.9
Maximum lightness: 0.1

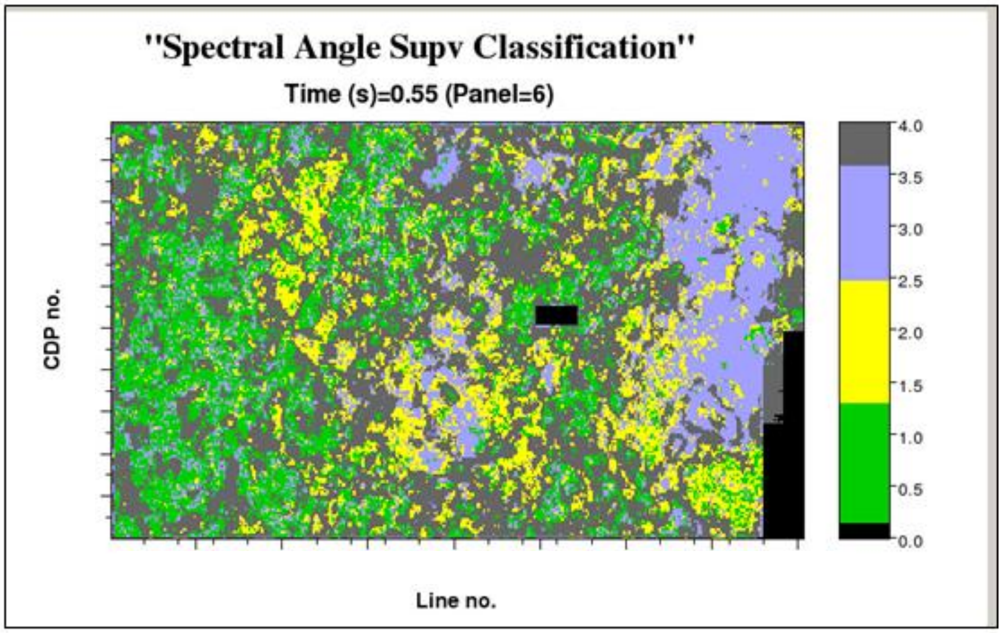
(c) 2008-2012 AASPI - University of Oklahoma

The results of the MED and the SAM classifier can be visualized by clicking on the checkbox (Arrow 10B). The number of facies used for the classification is equal to the number of wells used (Arrow 11B). All the other parameters are not required until one wants to visualize the SOM facies classification outputs too. The results of the MED and the SAM classifier are given in the next page.



"Euclidean Dist Supv Classification"

Time (s)=0.55 (Panel=6)

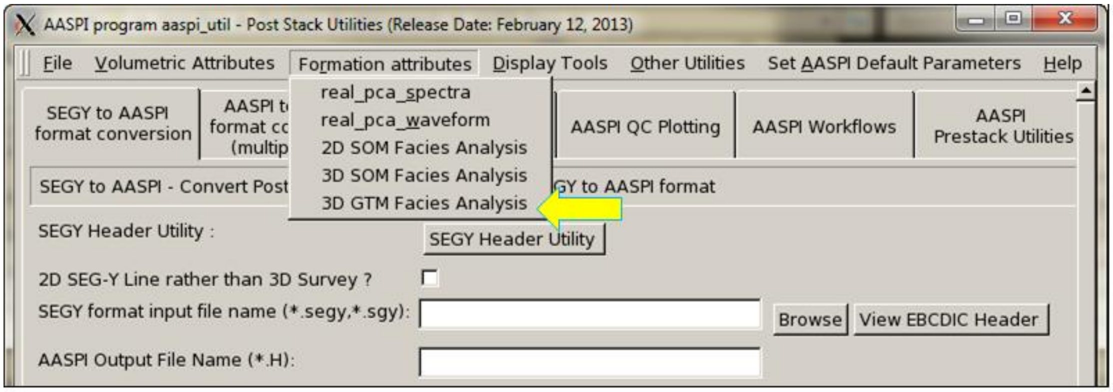


"Spectral Angle Supv Classification"

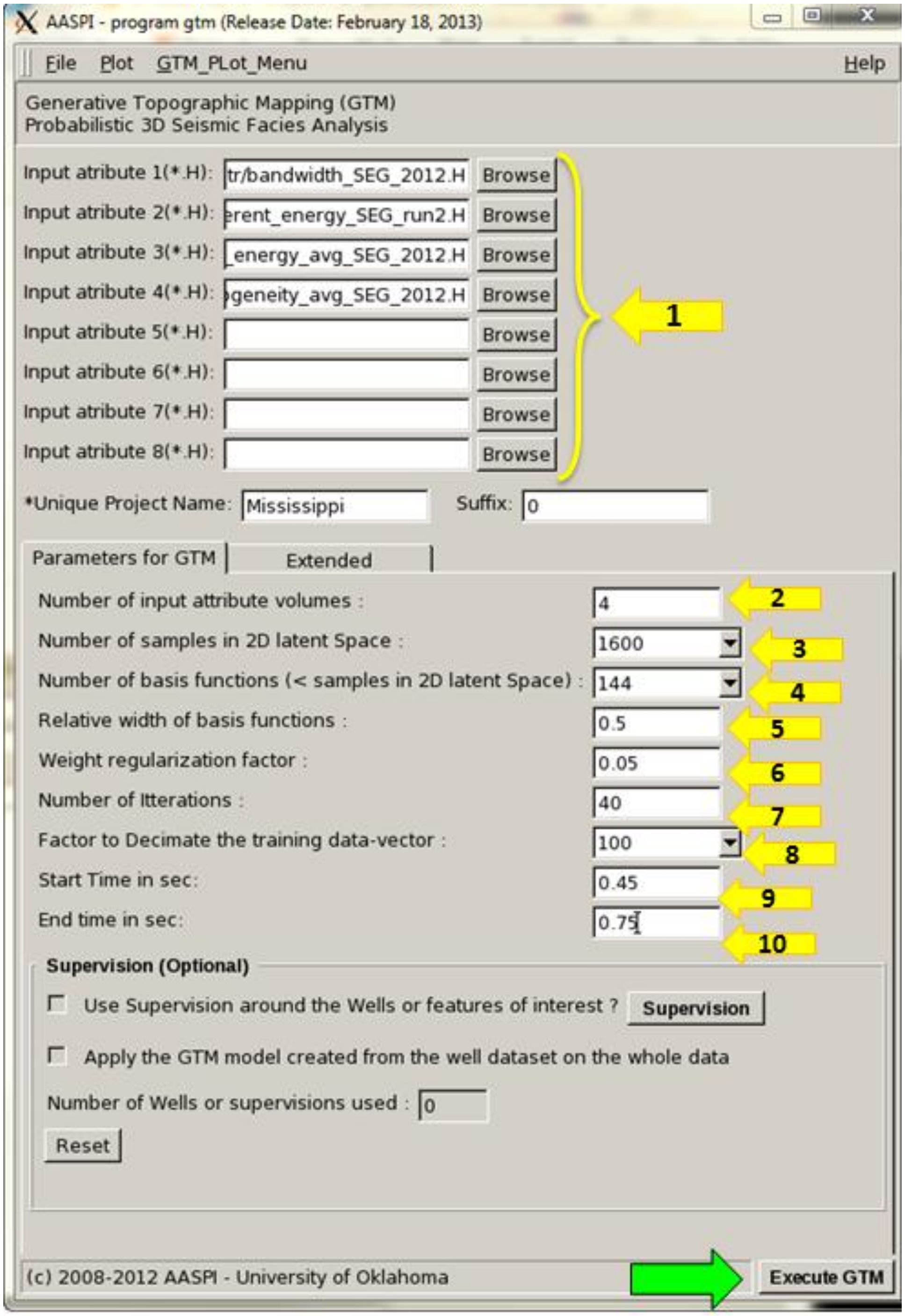
Time (s)=0.55 (Panel=6)

APPENDIX 4: Probabilistic 3D seismic facies classification GTM GUI

In the AASPI software, the GUI for GTM can be invoked by typing ***aaspi_gtm_pf*** or from the main ***aaspi_util*** window by selecting the Formation Attributes drop down menu (**3D GTM Facies Analysis**):



The following GUI will pop up:



AASPI - program gtm (Release Date: February 18, 2013)

Generative Topographic Mapping (GTM)
Probabilistic 3D Seismic Facies Analysis

Input atribute 1(*.H): tr/bandwidth_SEG_2012.H
Input atribute 2(*.H): erent_energy_SEG_run2.H
Input atribute 3(*.H): _energy_avg_SEG_2012.H
Input atribute 4(*.H): geneity_avg_SEG_2012.H
Input atribute 5(*.H):
Input atribute 6(*.H):
Input atribute 7(*.H):
Input atribute 8(*.H):

*Unique Project Name: Mississippi
Suffix: 0

Number of input attribute volumes : 4
Number of samples in 2D latent Space : 1600
Number of basis functions (< samples in 2D latent Space) : 144
Relative width of basis functions : 0.5
Weight regularization factor : 0.05
Number of Itterations : 40
Factor to Decimate the training data-vector : 100
Start Time in sec: 0.45
End time in sec: 0.75

Supervision (Optional)

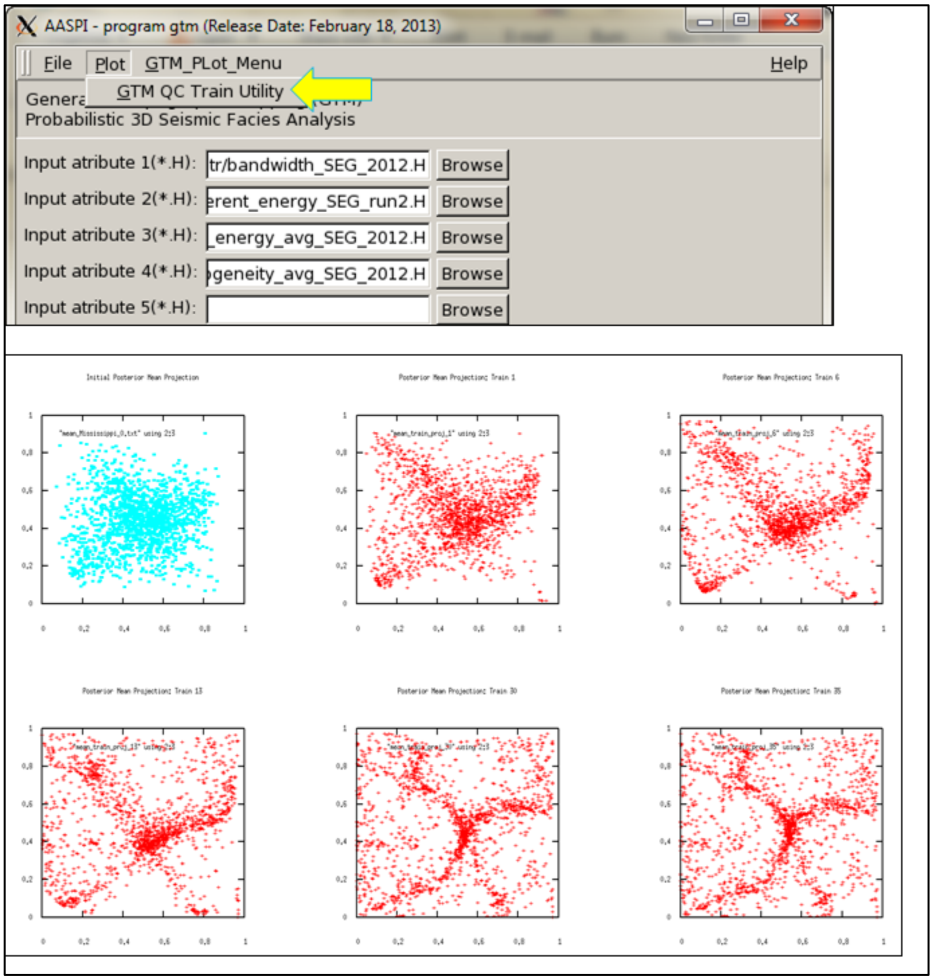
☐ Use Supervision around the Wells or features of interest ?

☐ Apply the GTM model created from the well dataset on the whole data

Number of Wells or supervisions used : 0

(c) 2008-2012 AASPI - University of Oklahoma

As with SOM, the input consists of a suite of **(1)** seismic attribute volumes that the interpreter has chosen to differentiate different seismic facies, rock types, lithologies, or other clusters. For example, a mass transport complex may be characterized by relatively low coherence, strongly converging reflectors, and high entropy (measured by the GLCM algorithm). Surrounding marine shales may be characterized by moderate coherence, low reflector convergence (i.e. parallel reflectors) and low GLCM entropy. Next, **(2)** enter the number of input volumes represents the dimensionality of the dataset (automatically updated). Then **(3)** select the number of grid points to span the 2D latent space, K, (GTM theory Chapter 5). These points are mapped to the data-space. Then **(4)** select the number of non-linear basis functions, J, that form a regular array of Gaussian functions. A linear combination of these basis functions is used to map the points in the latent space to the data space. Both the latent space samples and the basis functions should be squared integers value, (e.g. (256…400…625…900 etc.….). They can be automatically selected from some pre-defined values in dropdown menu. Care should be taken so that the number of basis functions (J) should be less than the number of grid points in the 2D latent space (K). Next, **(5)** Enter the width of the basis functions relative to the distance between two neighboring basis function centers. This width is used to define the standard deviation of the non-linear basis functions, which is constant for a GTM model. If $\varepsilon=2$ the basis functions will have widths (std. dev) equals to two times the distance between two neighboring basis function centers. Initially the code runs a multiattribute PCA to initialize the starting values of $\mathbf{W}$ and β (see GTM theory Chapter 5). Next, **(6)** enter the regularization factor,α used to stabilizing the linear equation for solving the new $\mathbf{W}$. This prevents any division by zero. Next, **(7)**, enter the number of iterations to run GTM. To minimize run times, only a fraction of the input dataset is used for training. Therefore **(8)** enter the decimation factor for using a subset of dataset used for training (yellow arrow 8). For example the value 100 means every 100th data-vector is used for training. Different training percentage of the data can be selected from the dropdown menu. It is recommended to run GTM only in the prospective zone because it is a computation intensive job. Finally, enter the **(9)** start time and the **(10)** end time of the data for GTM seismic facies classification. The "Save GTM Parameters" is an optional feature to create a parameter file, which can be used for a batch workflow (not included in this release). After providing all these parameters click "Execute GTM" program (green arrow).

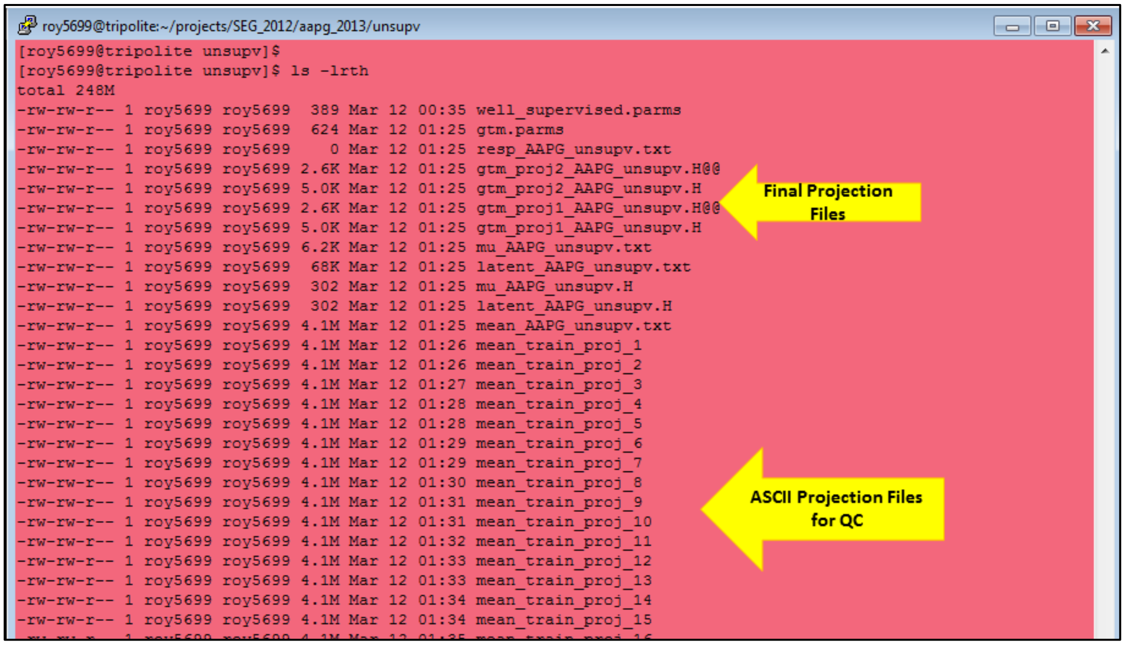


Genera
Probabilistic 3D Seismic Facies Analysis

Input atribute 1(*.H): tr/bandwidth_SEG_2012.H
Input atribute 2(*.H): erent_energy_SEG_run2.H
Input atribute 3(*.H): _energy_avg_SEG_2012.H
Input atribute 4(*.H): geneity_avg_SEG_2012.H
Input atribute 5(*.H):

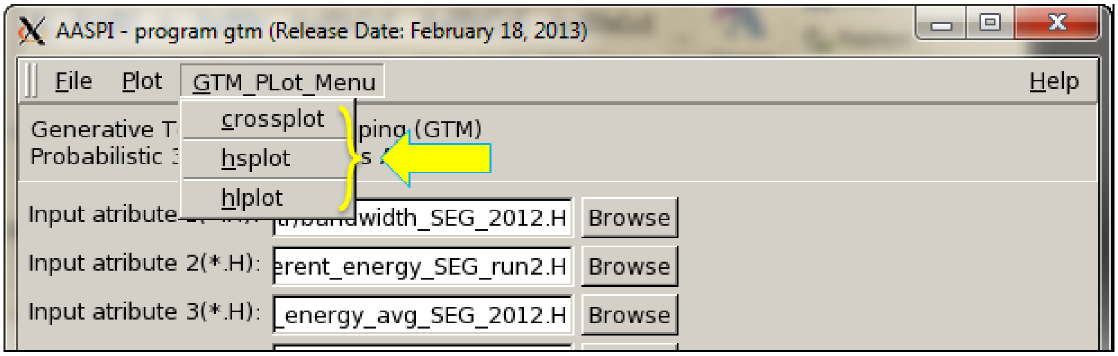
After the GTM model building is done, click on "GTM QC Train Utility" to QC the mean distribution of the posterior probability (responsibility) projections of the training data on the 2D latent space (presently this utility uses **gnuplot**). Otherwise the ASCII files *mean_train_proj_${train}* (shown below) generated can be viewed in any other graph utility manually.

The output files will be created from the GTM analysis is shown below. The gtm_proj1_${project_name}_${suffix}.H and the gtm_proj2_${project_name}_${suffix}.H are final projection files which stores the mean posterior probabilities along two axis. These two files should be imported into any visualization software for interactive picking the clusters (as discussed in the theory).



```
[roy5699@tripolite unsupv]$
[roy5699@tripolite unsupv]$ ls -lrth
total 248M
-rw-rw-r-- 1 roy5699 roy5699  389 Mar 12 00:35 well_supervised.parms
-rw-rw-r-- 1 roy5699 roy5699  624 Mar 12 01:25 gtm.parms
-rw-rw-r-- 1 roy5699 roy5699    0 Mar 12 01:25 resp_AAPG_unsupv.txt
-rw-rw-r-- 1 roy5699 roy5699 2.6K Mar 12 01:25 gtm_proj2_AAPG_unsupv.H@@
-rw-rw-r-- 1 roy5699 roy5699 5.0K Mar 12 01:25 gtm_proj2_AAPG_unsupv.H
-rw-rw-r-- 1 roy5699 roy5699 2.6K Mar 12 01:25 gtm_proj1_AAPG_unsupv.H@@
-rw-rw-r-- 1 roy5699 roy5699 5.0K Mar 12 01:25 gtm_proj1_AAPG_unsupv.H
-rw-rw-r-- 1 roy5699 roy5699 6.2K Mar 12 01:25 mu_AAPG_unsupv.txt
-rw-rw-r-- 1 roy5699 roy5699  68K Mar 12 01:25 latent_AAPG_unsupv.txt
-rw-rw-r-- 1 roy5699 roy5699  302 Mar 12 01:25 mu_AAPG_unsupv.H
-rw-rw-r-- 1 roy5699 roy5699  302 Mar 12 01:25 latent_AAPG_unsupv.H
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:25 mean_AAPG_unsupv.txt
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:26 mean_train_proj_1
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:26 mean_train_proj_2
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:27 mean_train_proj_3
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:28 mean_train_proj_4
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:28 mean_train_proj_5
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:29 mean_train_proj_6
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:29 mean_train_proj_7
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:30 mean_train_proj_8
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:31 mean_train_proj_9
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:31 mean_train_proj_10
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:32 mean_train_proj_11
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:33 mean_train_proj_12
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:33 mean_train_proj_13
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:34 mean_train_proj_14
-rw-rw-r-- 1 roy5699 roy5699 4.1M Mar 12 01:34 mean_train_proj_15
```

The crossplot, hsplot and the hlplot utility GUIs can be invoked by clicking on the "GTM Plot Menu" as shown in the GUI.

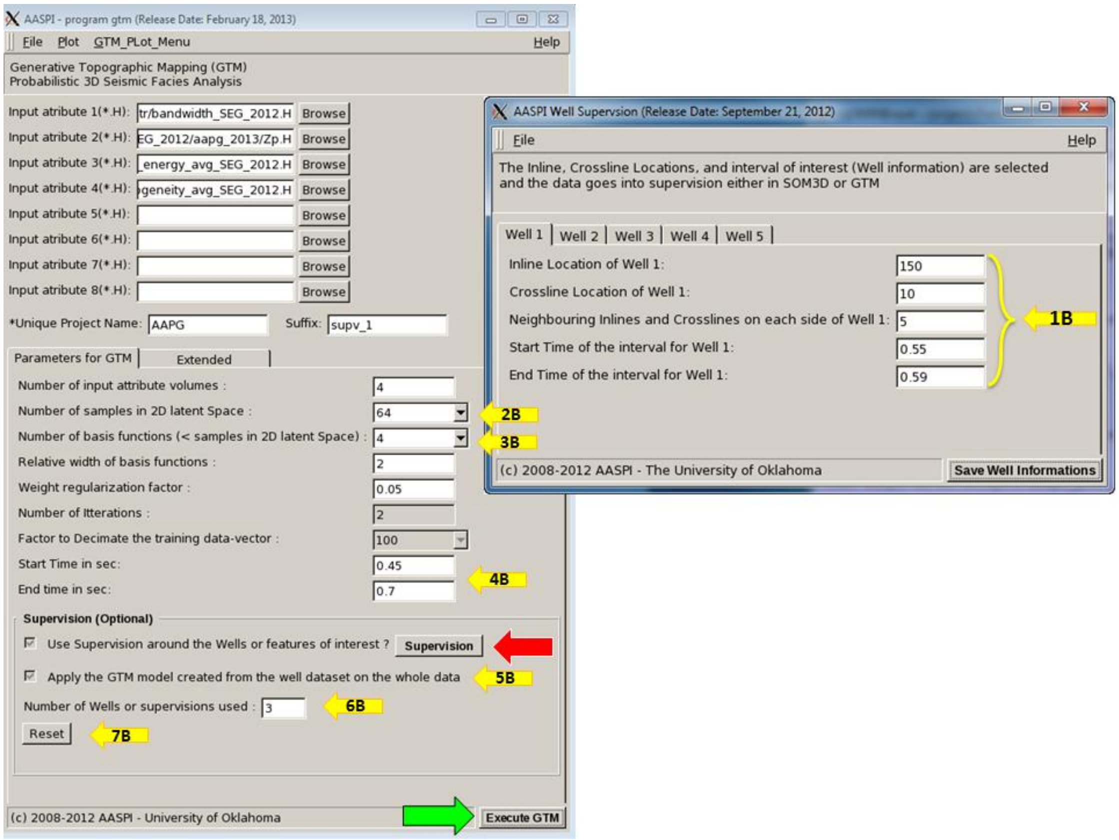


Input atribute 2(*.H): erent_energy_SEG_run2.H

Input atribute 3(*.H): energy_avg_SEG_2012.H

For a better analysis of the how likely one type of facies occurs we need to input some average wavelets around the wells within the survey or areas of interest. This is done by clicking the “supervision” button (red arrow). The well supervision GUI pops up and the inline, crossline, start-time, end-time and the neighborhood radius information are given (Arrow 1B).

With this supervised case the number of latent space variables (Arrow 2B) and the number of basis functions (Arrow 3B) should be reduced. The number of wells used for supervision is mentioned (Arrow 6B). Thus the new GTM model will be applied on the whole dataset to find the most likely facies corresponding to the wells. The reset button (Arrow 7B) can be used to start over unsupervised GTM analysis.


AASPI - program gtm (Release Date: February 18, 2013)

Generative Topographic Mapping (GTM)
Probabilistic 3D Seismic Facies Analysis

Input atribute 1(*.H): tr/bandwidth_SEG_2012.H
Input atribute 2(*.H): EG_2012/aapg_2013/Zp.H
Input atribute 3(*.H): _energy_avg_SEG_2012.H
Input atribute 4(*.H): ɔgeneity_avg_SEG_2012.H
Input atribute 5(*.H):
Input atribute 6(*.H):
Input atribute 7(*.H):
Input atribute 8(*.H):
*Unique Project Name: AAPG
Suffix: supv_1

Number of input attribute volumes : 4
Number of samples in 2D latent Space : 64
Number of basis functions (< samples in 2D latent Space) : 4
Relative width of basis functions : 2
Weight regularization factor : 0.05
Number of Itterations : 2
Factor to Decimate the training data-vector : 100
Start Time in sec: 0.45
End time in sec: 0.7

Supervision (Optional)
☑ Use Supervision around the Wells or features of interest ?
☑ Apply the GTM model created from the well dataset on the whole data
Number of Wells or supervisions used : 3

(c) 2008-2012 AASPI - University of Oklahoma

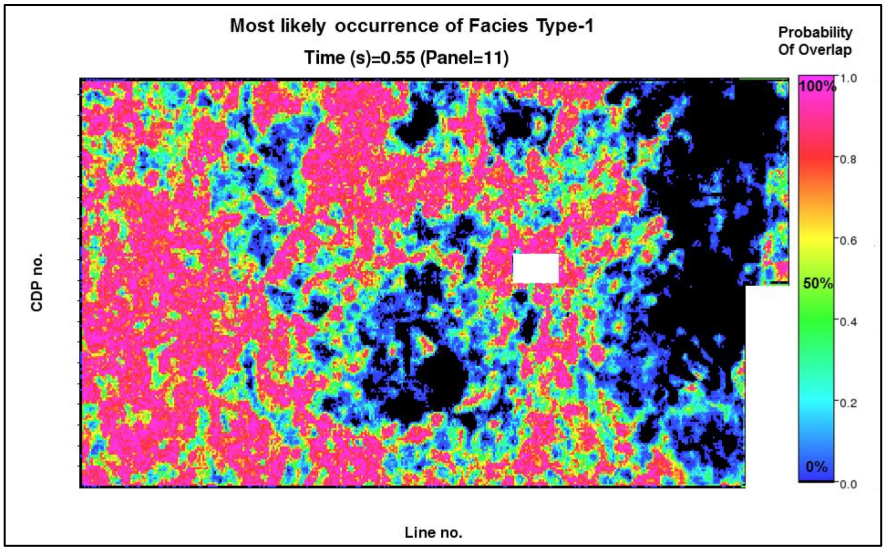
AASPI Well Supervsion (Release Date: September 21, 2012)

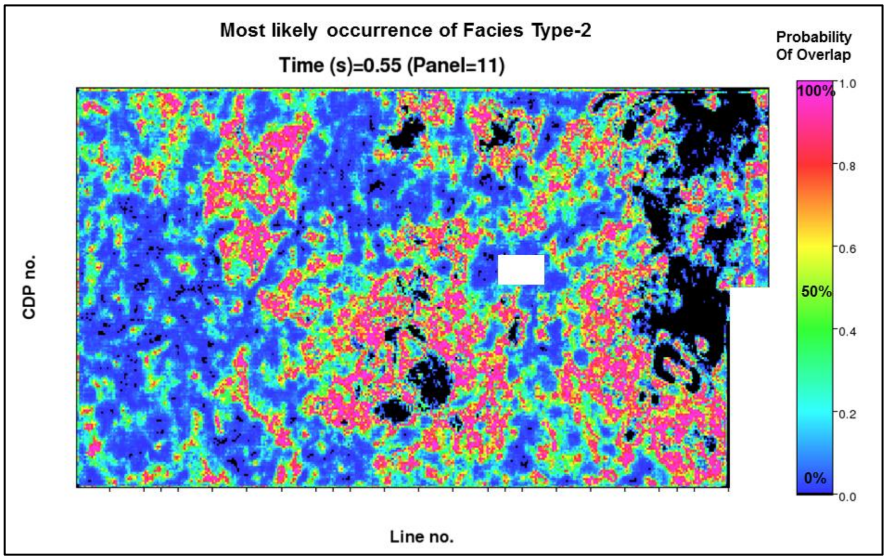
The Inline, Crossline Locations, and interval of interest (Well information) are selected
and the data goes into supervision either in SOM3D or GTM

Inline Location of Well 1: 150
Crossline Location of Well 1: 10
Neighbouring Inlines and Crosslines on each side of Well 1: 5
Start Time of the interval for Well 1: 0.55
End Time of the interval for Well 1: 0.59

(c) 2008-2012 AASPI - The University of Oklahoma

Three different facies types are given in the well information panels. Bhattacharya measure is used to calculate the overlap value between two PDFs (see chapter 6).





The above two figures show the most likely occurrence of facies Type 1 and Type 2. The magenta color highlights regions with the highest probability (90-100 %) of occurrence of the facies similar to the input well. The blue regions have very low likely that the facies is similar to the input well facies. And the black regions have no similarity to the facies type of the input well.