Cluster assisted 3D and 2D unsupervised seismic facies analysis, an example from the Barnett Shale Formation in the Fort Worth Basin, Texas.
Atish Roy * and Kurt J. Marfurt, The University of Oklahoma, Norman

Summary
The most popular seismic attributes fall into broad categories – those that are sensitive to lateral changes in waveform and structure, such as coherence and curvature, and those that are sensitive to lithology and fluid properties – such as AVO and impedance inversion. Unfortunately neither of these two attribute families works well in differentiating the depositional packages characterized by subtle changes in the stratigraphic column or lateral changes in texture. Automatic seismic facies analysis aims to classify similar seismic traces based on amplitude, phase, frequency and other seismic attributes. This paper reviews Kohonen Self Organizing Maps as one of the clustering algorithms that can generate 3D seismic facies volumes and maps using multiple attributes as input. The present area of study is the Mississippian Barnett Shale of the Fort Worth Basin in Texas. The aim of the study is to visualize the variation in shale and possible relationship between these rock types and their seismic expression and try to delineate by passed play after hydraulic fracturing.

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. This interpreter examines successive vertical sections through seismic volumes to determine the seismic character, which is then and 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 provide an understanding of the seismic facies volume and help qualitatively to understand the depositional environment and reservoir heterogeneity.

There are many different supervised and unsupervised clustering algorithms. K-means clustering works well where the data are noise free and with high relative dimensionality. Statistically, seismic data have high continuity, a low relative dimensionality and can be noisy and thus do not possess the pre-requisite features for successful analysis of the seismic trace shapes by K-means clustering (Coleou et al., 2003).

Another commonly used commercial product is based on Kohonen Self Organizing Maps. Coleou et al., (2003) and their colleagues were among the first to publish Kohonen SOMs based on seismic waveform classification using a 1D latent space. More recently Matos et al., (2009) and Roy et al., (2010), showed the advantage of extending the latent space to 2D and 3D, with corresponding 2D and 3D colorbars used to delineate the subsurface depositional environment.

Our focus in this study is the seismic facies analysis of the Barnett Shale in the Fort Worth Basin (FWB) in Texas. The Barnett Shale with an average thickness of ~350 feet is a laterally heterogeneous massive column of rock composed of three geomechanical facies and at least 14 gamma ray parasequences associated with the depositional environment of the shale. Properly mapping such heterogeneities will aid in understanding the non-uniform production and making engineering decisions.

Our first Kohonen SOM application will use multiple spectral component volumes as input and generate a single 3D seismic facies volume of the Barnett Shale formation. Each point is defined by its spectrum and the spectra classified this output seismic volume, which will be used to understand the heterogeneity present in the Barnett shale formation.

Our second Kohonen SOM application will use Lambda-Rho and Mu-Rho volumes, which are generated from pre-stack amplitude gathers. Here we will use not only the values of $\lambda_p$, $\mu_p$ at each sample, but all their vertical relationship. The output 2D seismic facies map can be compared to engineering data, thereby helping us to understand the geomechanical layering in the survey.

Kohonen SOM Clustering Analysis
SOM (Kohonen, 1995) clusters the data such that the statistical relationship between multidimensional data is converted into a much lower dimensionality latent space that preserves the geometric relationship between the data points. Mathematically each SOM unit (or Prototype Vector) within the latent space preserves the metric relationships and topologies of the multi-dimensional input data. The Prototype vectors (PVs) that are the training set of vectors have the same dimension as the input data. These PVs are arranged in a regular low-dimensional grid or map, thereby topologically connecting a PV to its neighbors.

To avoid guessing at the number of clusters necessary to represent the data, we have over-defined the number of initial clusters (PVs), which after subsequent iterations
3D and 2D unsupervised seismic facies analysis

tends to converge to a lesser number of clusters. To create an initial 2D map we have projected the eigenvectors from Principal Component Analysis (PCA) of the input datasets to generate a “N-dimensional” initial set of 256 Prototype vectors (PVs), where 256 is the maximum number of colors supported by our visualization software. For the 3D SOM algorithm N is the number of different input attributes and while the 2D SOM algorithm N is the number attributes at each sample times the number of stratal slices used to form the vertical patterns. This N represents the dimensionality of the input dataset. The SOM can be understood as a 2D sheet with the interconnected PVs imbedded on it. We then train the PVs in accordance to the input dataset by the Kohonen SOM neighborhood-training rule. As part of SOM training process, an input vector is initialized and is compared with all N dimensional PVs on the 2D grid latent space. The PV with the best match (the winner PV) will be updated. By the neighborhood-training rule a certain number of PVs around the “winner” PV are also updated. With successive iterations, this neighborhood radius decreases. The number of iterations continues until the convergence criterion (which depends on the lowest neighborhood radius and the minimum distance between all the projections of the clusters) is reached. Unless specific clusters are fixed such as at well locations, or to represent a desired behavior, this training and subsequent clustering is not biased by any input, giving rise to an unsupervised clustering.

After the training is complete the modified PVs are then color-coded by using a 2D gradational colorscale (Matos et al., 2009). The distance of the PVs in the latent space from their center of mass is plotted against lightness and the azimuths of the PVs are plotted against hue. These colored PVs are compared with the input data set. Those traces with similar seismic nature are assigned the same colors. Thus the resulting 3D seismic facies volume or the 2D seismic facies map is not forced into the process but comes from the clustering of the PVs in an unsupervised manner.

The output is either a seismic facies volume or a map where different changing facies are represented by a continuous gradational (Hue Saturation Lightness) HLS colorbar. The spatial distribution of color change is due to the change in the seismic response of the reflectors. The change in seismic response may be due to the change in depositional environment, lithology change, the fluid present, the nature of the rock and various other inherent natures of the rock.

3D seismic attribute generation

The 3D seismic consists of a 200 square miles survey in the NE Fort Worth Basin. The data are sampled at 110ft by 110ft by 2ms. For attribute generations the seismic data between the Marble faults horizon and the Viola limestone is considered.

We flattened the seismic above Viola Limestone and used a matching pursuit spectral decomposition algorithm to compute spectral components ranging between 10Hz to 70Hz in steps of 5Hz. These 13 different spectral component volumes are input to the 3D SOM algorithm. Thus the input dataset is 13 dimensional with each inline, crossline and sample of the input data having 13 different attributes assigned to it. These volumes are normalized using a Z-score algorithm that is survey consistent so that there is no bias of the input datasets. The 13 dimensional prototype vectors are projected onto 2D map as shown in Figure 1a. The initial regularly-spaced PVs will maintain their interconnectivity but with each iteration move away or towards each other to best represent the 13-dimensional data.

While the vertical pattern of seismic amplitudes can be used to cluster seismic waveforms (e.g. Poupon et al., 2004), such waveforms depend on both the reflectivity and the seismic wavelet. Stephens et al. (2011) used a 1D SOM algorithm to classify stacking patterns of inverted values of Poisson's ratio, delineating areas in the Eagleford Shale of south Texas more prone to hydraulic fracturing. Given the correlation of "fracability" to values of Young's modulus vs. Poisson's ratio, or alternatively of lambda-rho vs. mu-rho, we generalize our algorithm to simultaneous cluster the stacking pattern of the lambda-rho, mu-rho pairs. We generated 11 stratal slices for each of the two volumes between the Lower Barnett top and Viola Limestone top. The two stratal slice volumes of Lambda-Rho and Mu-Rho are concatenated to form an N = 2 x 11 dimensional vector at each trace location. The 2D SOM analysis is focused within the Lower Barnett Shale only and considers the rock-physics properties for better analysis.

3D SOM analysis

The 3D SOM algorithm trained the PVs in the 2D latent space (Figure 1a), with their positions updated after each iteration, resulting in the new updated position of the PVs is shown in Figure 1b. There was very little updating of the PVs in their latent space after the 10th iteration, allowing to stop the training process. The map shows how the PVs have moved close to each other in their latent space and also to the data points. The HLS colors are assigned to the PVs according to their distance from their center of mass and their azimuth. These trained PVs are compared back with the dataset to give the seismic facies volume. Thus two dissimilar neighboring traces in the seismic volume will be far apart in the latent space and thus have different colors. Conversely two similar traces in the seismic volume will have nearly same color.
3D and 2D unsupervised seismic facies analysis

Figure 1: (a) The Prototype vectors plotted in a latent space. The x and the y-axis extends in the $V_1$ and $V_2$ directions $\pm 3\lambda_1$ by $\pm 3\lambda_2$, where $(\lambda_1, V_1)$ and $(\lambda_2, V_2)$ are the first and second eigenvalue-eigenvector pairs. Each point (white squares) is a projected Prototype vector having the same dimensionality, $N$, as the input data. (b) Projection of the PVs after 10 iterations with the 3D SOM application. The inset shows the HLS colorbar.

Figure 2: Output Seismic facies volume from 3D SOM analysis of spectral component inputs. (a) Timeslice at 1.37s in the Upper Barnett Shale (b) Timeslice at 1.43s in the Lower Barnett Shale. The green colored facies more in the central part and the cyan/light pink colored facies is more abundant in the southern part of the survey (c) Along an Inline AB (VE>50). The volume is flattened at Viola horizon. Other horizons are marked along the inline.

Figure 2a shows a time slice at 1.37sec of the seismic facies within the Upper Barnett. The heterogeneous nature of the Upper Barnett is shown by different colors. From the colored PVs (Figure 1b) we interpret the cyan and the violet facies to be close to each other since they are near each other in the latent space. In contrast the green and the blue facies are far apart in the latent space should represent considerably different seismic facies.

Figure 2b shows a time slice at 1.43sec in the lower Barnett Shale. Similar heterogeneity is present in the Lower Barnett shale also. The facies colored in green is more abundant in here. The heterogeneity pattern in the Lower Barnett shale should be calibrated with some well control. Figure 2c is a vertical slice through the timeslices showing the variation of seismic facies in Barnett Shale. Note that there is lateral as well as vertical change in seismic facies within the Barnett shale.

2D SOM analysis

To better estimate geomechanical properties, the 2D SOM algorithm was applied on Lamé parameter Lambda-Rho and Mu-Rho volumes generated using a model based prestack inversion workflow. Figure 3 shows the output 2D facies map of the lower Barnett shale using the volumes. The $k_1$ principal curvature extracted on the top of Viola Limestone is co-rendered on the seismic facies map to understand the compartmentalization in the Lower Barnett shale.

Figure 3: Output seismic facies map based on 11 stratal slices through $\lambda\rho$, $\mu\rho$ volumes in the Lower Barnett shale. The $k_1$ principal curvature extracted on the top of Viola limestone is co-rendered on the seismic facies map ($k_1$ in gray-black). Note the compartmentalization of the Lower Barnett shale based on similar seismic facies and less complexity in structure. Two such prominent zones - green and violet are highlighted with arrows. It is noted that there is a correlation between the $k_1$ principal curvature and the seismic facies boundaries.
3D and 2D unsupervised seismic facies analysis

We compare these results with the average P-wave Impedance (Figure 4a) and the average density (Figure 4b) within the Lower Barnett Shale. Overall the average impedance results shows that they are having higher values in the East and in the Southwest. The central part of the survey shows generally lower values of impedance for Lower Barnett. Similarly, the average density of Lower Barnett shows a higher density in the northern half of the survey compared to the southern half.

![Figure 4: (a) Average impedance of Lower Barnett shale. (b) Average density of the lower Barnett shale.](image)

Correlating the facies map of Lower Barnett Shale shows a general lithology change that follows combined effect of the impedance and density variations. The East and the southeast have better facies continuity and structurally less complex. These results should however be correlated to well logs for better understanding them.

Discussions

From the 3D and the 2D SOM analysis it is evident that the characteristics of the Barnett formation changes from Northeast to Southwest. There is a fairly uniform facies characteristic in the East shown in green color and in the southeast shown in cyan (Figures 2b and 3). As we go west of the survey the lithological complexity of the shale formation increases. In other words the probability of encountering similar lithology is more expected in this eastern and the southeastern part of the survey.

The output facies from the SOM analysis of the λρ, μρ volumes differentiates the Lower Barnett shale into different lithological boundaries. The extracted principal curvatures on the top of the Viola horizon, k1 co-rendered on the seismic facies map helps in delineating different lithological boundaries within the Lower Barnett shale. Before making any conclusive decisions more studies have to done considering the information from the wells in this region.

Towards supervised classification

Ideally each seismic facies in the volume should be interpreted and calibrated to well information and rock properties. Such a posteriori classification requires a large number of wells encompassing all lithological variation in a region, which in most of the times is insufficient. Thus there is a need of some supervision in the algorithm.

If we have well information in certain areas of the survey we can assign their corresponding attribute patterns to Prototype vectors that will be fixed for all iterations. The other PVs are modified using the data and their proximity to all the other (including the fixed) prototype vectors, thereby introducing some supervision in the application.

Conclusions

We have presented a workflow that provided a volumetric estimation of seismic facies. These facies can either be sample by sample or according to their vertical stacking pattern. Barnes and Laughlin (2002) found that the power of classification is heavily dependent on the choice of attributes. For the Barnett Shale we feel attributes such as λρ, μρ, VTI, HTI intensities will be most closely coupled to geomechanical behavior.

Acknowledgements

Thanks to Devon Energy for permission to use and publish their seismic from Northeast Fort Worth Basin. We thank the sponsors of the OU Attribute-Assisted Processing and Interpretation Consortium. We also thank our colleague Roderick Perez for providing the inversion volumes.