Applying Self Organizing Maps of Multiple Attributes, an Example from the Red-Fork Formation, Anadarko Basin

Atish Roy *, Kurt J.Marfurt, The University of Oklahoma, and Marcilio Castro de Matos, Instituto Militar de Engenharia, Brasil.

Summary

The Self Organizing Map (SOM) is one of the most effective pattern recognition techniques, and is commonly used tool for non-supervised seismic facies analysis. Early SOM implementations required estimating the number of clusters. Current implementations avoid this choice by over-defining the number of clusters and mapping them against continuous 1D, 2D and 3D colorbars, which the interpreter then visually clusters. We generate SOM clusters based on the wavelet shape, GLCM attributes as well as on the spectra of the Red-Fork formation and correlate the results to other seismic attributes and the knowledge of geology from extensive well control in the area.

Introduction

SOM (Kohonen, 2001) clusters data such that the statistical relationship between multidimensional data is converted into a much lower dimensional latent space that preserves the geometrical relationship among the data points. Mathematically, each SOM unit preserves the metric relationships and topologies of the multidimensional input data. SOM prototype vectors or neurons have the same dimension as the input data, and are arranged in a regular low-dimensional grid or map, thereby topologically connecting it to its neighbors.

Coleou et al. (2003) and their colleagues were among the first to publish SOM-based seismic waveform classification using a 1D latent space. More recently, Matos et al. (2009) showed the advantage of extending the latent space to 2D and 3D (with corresponding 2D and 3D colorbars) indelineating channels of a turbidite reservoir from the Campos Basin, offshore Brazil.

In this paper, we apply this workflow to seismic amplitude and spectral component volumes for a seismic survey acquired over the Anadarko Basin, Oklahoma, USA. We interpret these results using extensive well control and geological information in this area (Suarez et al., 2008). We also show how SOM can be used to extract features from a suite of spectral components and from multiple Grey Level Co-occurrence Matrix attribute volumes.

Kohonen Self Organizing Maps (SOMs)

The Kohonen SOM (Kohonen, 2001) is not only an effective way of visualizing multidimensional data but also preserves the original topological structure, making it amenable for seismic facies analysis. Initially we assume the input seismic attributes are represented by J vectors in the space $\mathcal{R}^n$, $x_j=[x_{j1}, x_{j2}, x_{j3} \dots x_{jN}]$ where N is the number of input seismic attributes and $j=1,2,\dots,J$ is the number of seismic traces analyzed. These vectors are in turn represented by P prototype vectors m_i, $m_i=[m_{i1,} m_{i2} \dots m_{iN}]$, where $i=1,2,\dots,P$. Prototype vectors are organized on a grid of lower dimension than P. After initializing the SOM prototype vectors to reasonably span the data space, the first next, or training, step in SOM is to choose a representative subset of the J input vectors. Each training vector is associated with the nearest prototype vector. After each iteration of the training, the mean and standard deviation of the input vectors associated with each prototype vector is accumulated, after which the prototype vectors are updated using a function of the distance between it and its neighbors. This iterative process stops either when the SOM converges or the training process reaches a predetermined number of iterations. One way to evaluate the SOM clustering results is to plot the distance between the neighboring prototype vectors thereby generating the Unified Matrix distance (U-Matrix) map. While the U-matrix can be used to estimate the number of clusters or data classes in the data, in this paper, we project the SOM to a 2D or 3D colorbar using Principal Component Analysis (PCA) after which the results are visually clustered in the interpreters brain.

Geology of Red-Fork Formation

The Red-Fork Formation from the Anadarko Basin in Oklahoma represents Pennsylvanian stratigraphic features which involved multiple stages of incised valleys fill. The depositional history of the Red-Fork is mainly divided into three coarsening upwards marine parasequences (Peyton et al., 1998). Figure 1 shows the four major depositional sequences in different colors and log interpretation done by Peyton et al., (1998) across the line AA'. The Red Fork formation is around 60 m thick.

Atish Roy. 100 East Boyd Street, Suite 710, Norman, Oklahoma 73019. E-MAIL: atish.roy@ou.edu

Comparison of SOM with Geology and other Seismic attributes

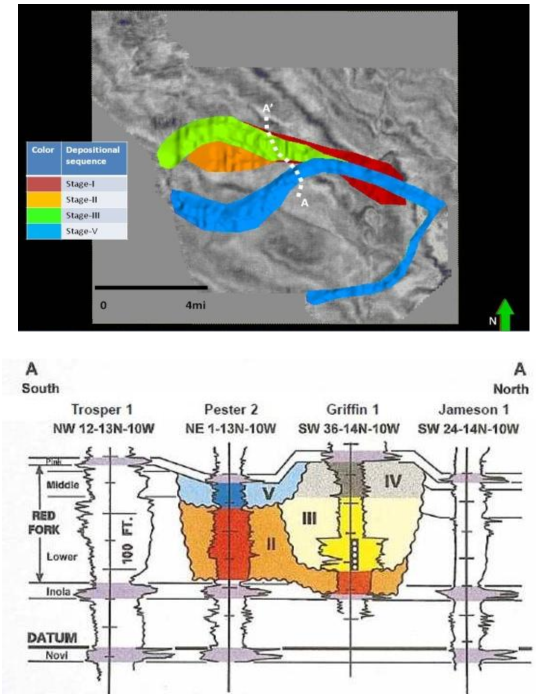


Figure 1: (a) Four major depositional stages in the Red Fork Formation. (b) Stratigraphic well-log cross section of AA' (Peyton et al., 1998).

We flattened the seismic amplitude volume on the Pink Lime horizon. The interpretation (Figure 2) of the different depositional stages of the Red-Fork Formation was done using the amplitude and coherency volumes coupled with well control.

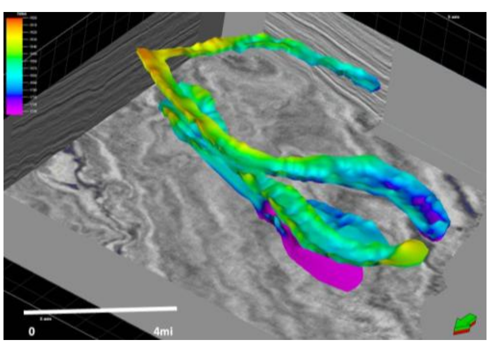


Figure 2: Interpretation of the different depositional stages of the Red-Fork Formation using amplitude, coherence and well control.

Method

Figure 3 shows the workflow we implemented using SOM. Since the time migrated dataset from the survey has a strong North - South and East - West acquisition footprint, we applied a footprint removal algorithm (Falconer and Marfurt, 2008) using the short wavelength most-positive principal curvature attribute (k_1) to calculate the footprints in the k_x- k_y domain. We then calculated the Grey Level Co-occurrence Matrix (GLCM) and also evaluated spectral decomposition of the filtered seismic data using frequencies ranging from 10 Hz to 90 Hz with a step of 10 Hz. We then selected the horizon slice 60ms below the Pink Lime horizon for SOM analysis.

Figure 3: Flowchart showing the workflow for SOM

Comparison of SOM with Geology and other Seismic attributes

2D SOM colormaps

We calculated the U- Matrix from the seismic data and also obtain a set of trained prototype vectors. These vectors are then projected in 2D *x-y* latent space either projecting on the two largest eigenvectors, or alternatively, by using a Sammon projection. Then these vectors are colored using an HSV color model with the hue proportional to $\boldsymbol{tan^{-1}\frac{y-0.5}{x-0.5}}$ and the saturation proportional to $\sqrt{(\mathbf{x}-\mathbf{0.5})\mathbf{2}+(\mathbf{y}-\mathbf{0.5})\mathbf{2}}$ as shown in Figure 5.

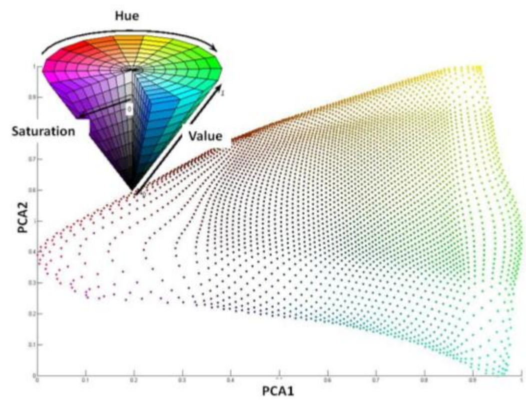


Figure 4: Projection of the GLMC attribute classification prototype vectors onto the two largest eigenvectors using the 2D color bar shown in Figure 6b.

Results of SOM on Red-Fork Formation

We applied our SOM workflow to three different attribute datasets. Each suite of attributes was extracted about a phantom horizon slice 60 ms below the Pink Lime. Our first example used seismic amplitude as input, resulting in a 'waveform' classification. The cropped flattened seismic volume consisted of 16 samples around the Red Fork formation. The trained prototype vectors (waveform classification) were projected using a Sammon Projection. The classification map is shown in Figure 5a using the 2D color map of the prototype vectors (Figure 5c). These prototype vectors correspond to the positions in the U-Matrix (Figure 5b). The U-Matrix (Figures 5b) shows the clustering of the prototype vectors. The blue in the color scale indicates that the distances between the vectors are small and thus that the neighboring prototype vectors are similar.

Our second example used spectral components ranging between 10-80 Hz at 10 Hz increments. as input, resulting in a 'spectral' classification. The trained prototype vectors are projected by PCA Projection. The 2D SOM classification (Figure 6a) distinguishes between the different channel fills and the floodplain deposits.

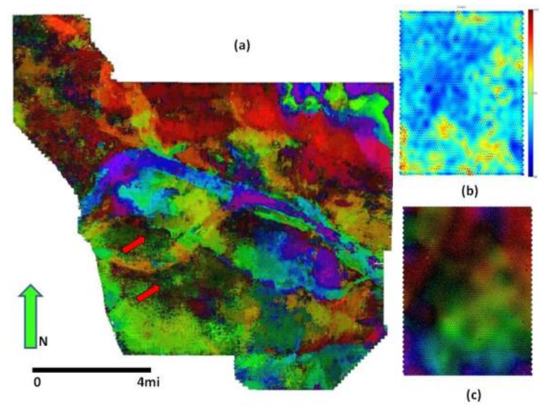
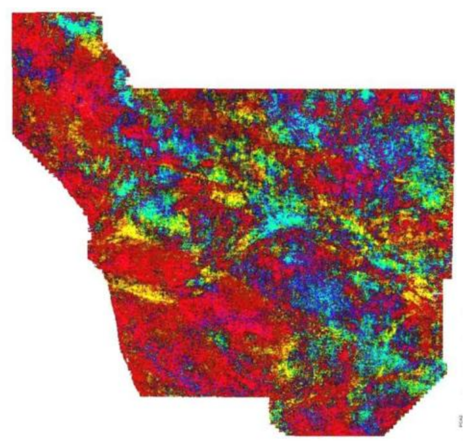


Figure 5: SOM waveform classification of 16 seismic amplitude samples as the input attribute. (a) 2D SOM output (b) U-Matrix of the prototype vectors. Blue area shows better clusturing of the prototype vectors. (c) 2D color of the prototype vectors. Red arrows indicate flood plain deposits.



Our third example used a suite of seven GLCM attribute - energy, mean, variance, homogeneity, dissimilarity, entropy and contrast, as input. These attributes were scaled to range between0-1 before they were merged.

Results from Seismic attributes

The Spectral Decomposition and the Grey Level Co-occurrence Matrix (GLCM) calculated are shown in Figures 7a to 8c. The GLCM attributes such as GLCM Energy (Figure 8a) measures the textural uniformity of the data, GLCM Entropy (Figure 8b) represents the degree of disorder of a system and GLCM Homogeneity (Figure 8c) measures the overall smoothness of the data. Thus we can infer that all the three GLCM attributes represents the same depositional history present in the data.

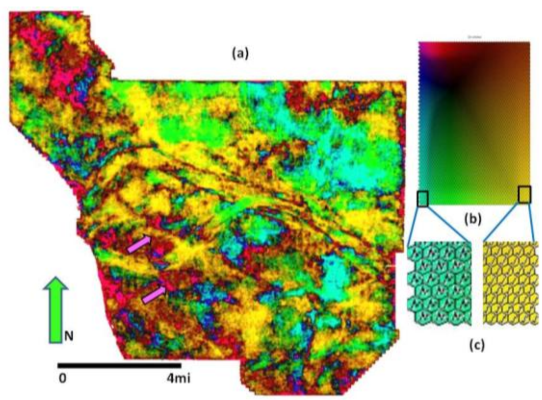


Figure 7: (a) 2D SOM of the seven GLCM texture attributes, (b) color coding of the prototype vectors, and (c) representative attribute patterns that make up each prototype vector.

Comparison of 2D SOMs with GLCM attributes

We observed that the floodplain deposits indicated by the red arrows in Figures 6-8 exhibits less textural uniformity or energy, higher entropy and low homogeneity. These areas can also be identified from the 2D SOM of Case 2 (Figure7). The channel fills and the floodplain deposits can also be clearly distinguished in the 2D SOM results.

Conclusion

The 2D SOM results show a clear demarcation between the different depositional stages (between Stage III and Stage V). The older stage II is also shown as different color in the SOM thus verifying the different sequences of Red Fork formation. The SOM results with 35Hz input show much better correlation with the GLCM and actual geology of Red-Fork formation. One limitation of mapping is a folding of the prototype vectors while doing PCA projection and thus two or more different prototype vectors can have the same color. However from this case study it can be inferred that SOM gives a decent knowledge of the subsurface depositional environment and can be used while exploring new areas where good geological information is not present.

Acknowledgements

Thanks to Chesapeake for permission to use and publish their seismic and well data from the Anadarko Basin. We also thank the sponsors of the OU Attribute-Assisted Processing and Interpretation consortium.