Seismic Color Self-Organizing Maps

Marcílio Castro de Matos, Kurt J. Marfurt, The University of Oklahoma
Paulo Roberto Schroeder Johann, Petrobras

This paper was prepared for presentation during the 11th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Abstract

Classification without supervision of patterns into groups is formally called clustering. Depending on the application area these patterns are called data lists, observations or vectors. For exploration geophysicists, these patterns are usually associated with seismic attributes, seismic waveforms or seismic facies.

The main objective of this paper is to show how one of the most popular clustering algorithms: Kohonen Self-organizing Maps, should be applied to enhance seismic interpretation analysis associated with one and two-dimensional color maps.

Introduction

One of the most important goals of seismic stratigraphy is to recognize and analyze seismic facies with regard to the geologic environment (Dumay and Fournier, 1988). According to Sheriff (2002), seismic facies analysis is done by examining seismic traces to identify the characteristics of a group of reflections involving amplitudes, abundance, continuity and configuration of reflections in order to predict the stratigraphy and depositional environment.

The human brain excels at recognizing image patterns. Indeed, successful interpreters have developed during their careers a mental library of seismic facies based on their work history. They then compare new facies they encounter against their catalogue. Given the ever increasing size of 3D seismic data volumes the human brain can use some help. Considerable help is provided by seismic attributes, which represent complex multisample waveforms by a reduced number of more relevant measurements designed to delineate geolocic features of interest. The goal of clustering is to organize these seismic attributes in a way that further enhances otherwise hidden geologic features.

Kohonen self organizing maps (SOM) is one of the most effective seismic clustering tools (Barnes and Laughlin, 2002) because it can be associated with 1D and 2D colormaps to help seismic interpretation. Specifically, we applied the visualization technique to a real seismic data set from Campos Basin, offshore Brasil.

Kohonen self organizing maps (SOM)

The SOM (Kohonen, 2001) and K-means clustering are the two most commonly used tools for non-supervised seismic facies analysis with SOM providing ordered clusters that can be mapped to a gradational color bar (Coléou et al, 2003).

SOM is closely related to vector quantization methods (Haykin, 1999). We begin by assuming that the input variables, i.e., the seismic attributes, can be represented by vectors in the space $\Re^n$, $\boldsymbol{x=[x_1,x_2,...,x_n]}$. The objective of the algorithm is to organize the dataset of input seismic attributes, delineated by a geometric structure called the SOM. Each SOM unit, defined as a "vector prototype", is connected to its neighbors, which in 2-D usually forms hexagonal or rectangular structural maps.

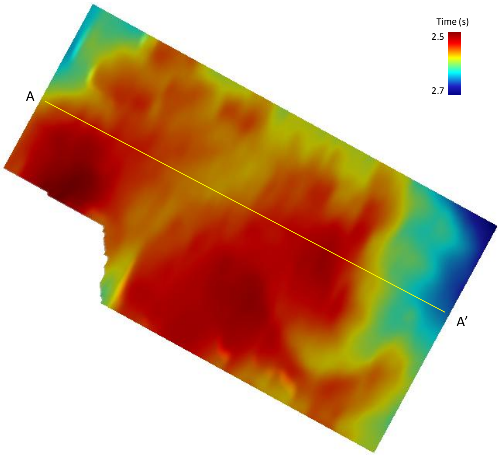
We assume that the map has P elements, then, there will exist P n-dimensional prototype vectors $\boldsymbol{m_i}$, $\boldsymbol{m_i=[m_{i1},...,m_{in}]}$, $i=1, 2,..., P,$ where n is the number of input seismic attributes. After SOM training, the prototype vectors are a good representation of the input dataset of seismic attributes.

The number of prototype vectors in the map determines its effectiveness and generalization capacity. During the training, the SOM forms an elastic net that adapts to the "cloud" formed by the input seismic attribute data. Data that are close to each other in the input space will also be close to each other in the output map. Since the SOM can be interpreted as a mapping of the input n-dimensional space onto a two-dimensional grid that preserves the original topological structure, and since seismic data measures the changes in geology, SOM preserves the topological relation of the underlying geology.

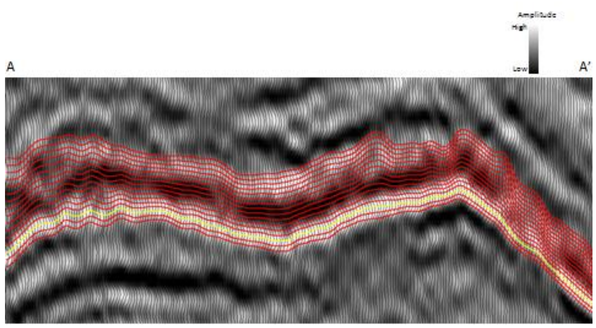
Although the prototype vectors represent the input data very well they have the same dimension of the input data making vizualization difficult. However, topological relation among the prototype vectors can be used as a visualization tool showing the different data characteristics and structuring. One way to visualize cluster formation of the SOM prototype vectors is by computing the distance among the vectors thereby generating a U-matrix (Ultsch, 1993). Another way is by mapping continuous 1D, 2D or 3D colorbars to the SOM topology to represent the location of each prototype vector.

1D SOM plotted against 1D colorbars

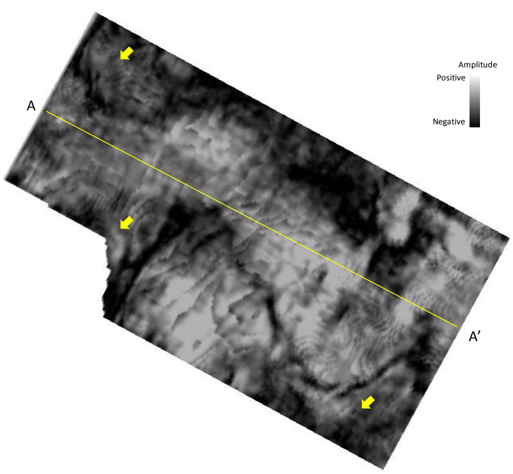
Before we present the SOM methodologies, we introduce the real seismic problem addressed in this paper. The main goal was to delineate a channel in the basal stratigraphic unit of a turbidite reservoir from Campos Basin, offshore Brasil. Figure 1a shows the two way time structure map of the base of the reservoir. Figure 1b shows a seismic inline with the proportional horizon slices generated between the base of the reservoir and an intermediate stratigraphic horizon, while Figure 1c shows an amplitude horizon slice at the base of the reservoir.



(a)



(b)



(c)

Figure 1: a) Time-structure map of the base of the reservoir; b) Proportional horizon slices between the base of the reservoir and an intermediate stratigraphic horizon; c) Amplitude horizon slice at the base of the reservoir.

The main objective here is to classify the waveforms represented by the amplitude illustrated in Figure 1b by using the SOM 1D, SOM 2D and colors.

First, a one dimensional SOM was trained; each prototype vector was assigned a color taken from the hue circle, i.e., considering saturation and value equal to one in the HSV color model (Guo et al., 2008). Since the SOM prototype vectors represent the complete input seismic data in the analysis window, classification is achieved by comparing each input trace with the SOM prototype vectors and plotted using the corresponding labeled color of the closest one. In general, classification can be done on any suite of attributes through the use of the Mahalanobis distance. On our Campos Basin shown in Figure 1, our attributes are simply seismic amplitudes on subsequent strata slices, such that the Mahalanobis distance is replaced by the simpler Pythagorean distance. When viewed vertically, each prototype takes on the appearance of a waveform shape, giving rise to what is called "waveform shape classification" (e.g. Couleou et al., 2003). Figure 2a shows the result using 12 colors. The classes were labeled by using 12 colors uniformly distributed along the hue circle as defined by:

$$hue(i) = \frac{2\pi i}{N}, i = 0, \ldots, N-1 \qquad (1)$$

where N=12 is the number of colors.

However, this representation does not take into account the distances between the prototype vectors and it does not show the clustering structure. Figure 2b shows the same 1D SOM colored by using the distances between neighboring prototype vectors, as defined by:

$$\begin{cases} hue(0) = 0, \\ hue(i) = 2\pi \dfrac{\sum_{j=1}^{i} \|m_{j+1} - m_j\|}{\sum_{j=1}^{N-1} \|m_{j+1} - m_j\|}, i = 1, \ldots, N-1. \end{cases} \qquad (2)$$

We can clearly see that classes near to each other have similar colors in Figure 2b, which facilitates the visual identification of the seismic facies.

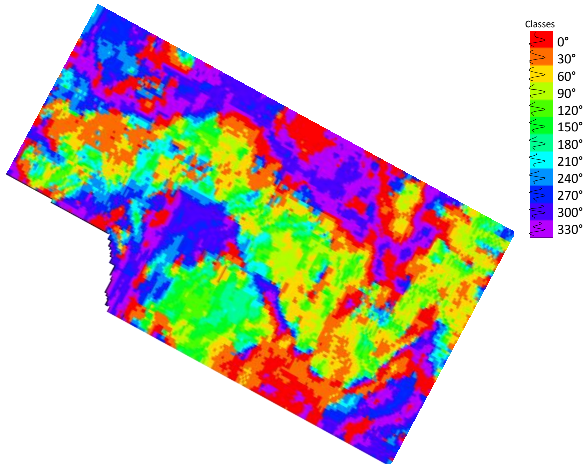
By increasing the number of prototype vectors, clusters and colors to 256 (Figure 3), we generate intermediate clusters which further delineate subtle features for the human interpreter.

Specifically, we clearly see that some regions in Figure 1 with high amplitudes, as the north-west and south-east, indicated by block arrows, are not associated with the channel waveform shape as shown in Figures 2 and 3.

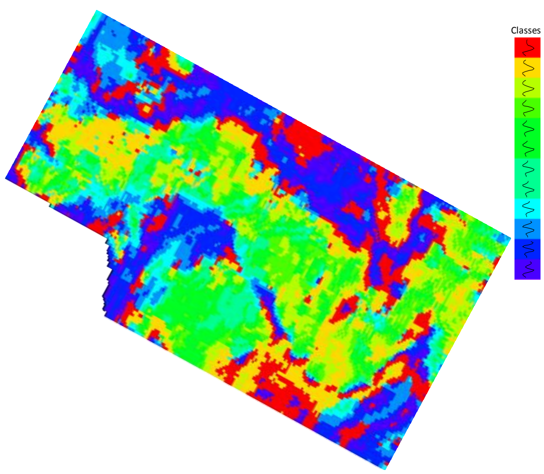
2D SOM plotted against 2D colormaps

Although 1D SOM provides very good visualization results it is not recommended to identify the number of clusters in the data (Matos et al., 2007).

Measuring the distances between SOM prototype vectors is one way to identify clusters in the data. Figure 4 shows the 2D SOM U-matrix obtained from the same seismic waveforms classified using the 1D SOM. We note that there is no obvious number of seismic facies. In this case, the choice of seismic trace amplitudes was inappropriate for seismic facies identification. Geologically, we expect a wide range of waveform variations in the area of interest because the seismic data were extracted from a complex sandstone turbidite system. The choice of the seismic attributes for the classification of seismic patterns is fundamental to obtain coherent results.



(a)



(b)

Figure 2: 1D SOM with 12 classes; a) Coloring without taking into account the distances among the prototype vectors; b) Coloring taking into account the distances. The actual prototype vectors are plotted as the shapes in the colorbar.

Although we cannot identify a discrete the number of seismic facies from the SOM when using the attributes chosen in this paper we can use gradational colors to visualize the more continuous relation among the waveforms.

Figure 5 shows the classification results using the value color code gamut. In this case by a 2D color bar.

In Figure 5, we did not take into account the distances among the SOM prototype vectors to create the 2D color map. Although the problem is not as direct as with 1D SOM, there are different ways to do it (Himberg, 1998). In this paper we project the SOM prototype vectors using Principal Component Analysis and Sammon mapping onto a two dimensional plane and then apply the HSV color to the 2D projections and color the SOM units. Figure 6a shows the 2D PCA projection of the SOM prototype vectors while Figure 6b shows the Sammon projection. Figure 7a shows the SOM classification results using PCA and Figure 7b shows the results using Sammon mapping.

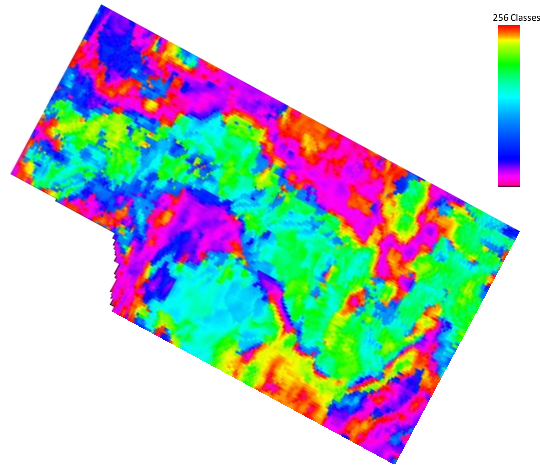


Figure 3: 1D SOM with 256 classes coloring taking into account the distances among the prototype vectors.

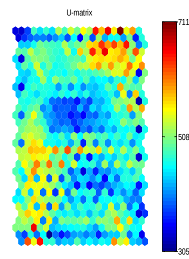


Figure 4: U-matrix, where colors correspond to the distances between prototype vectors (see Matos et al., 2007).

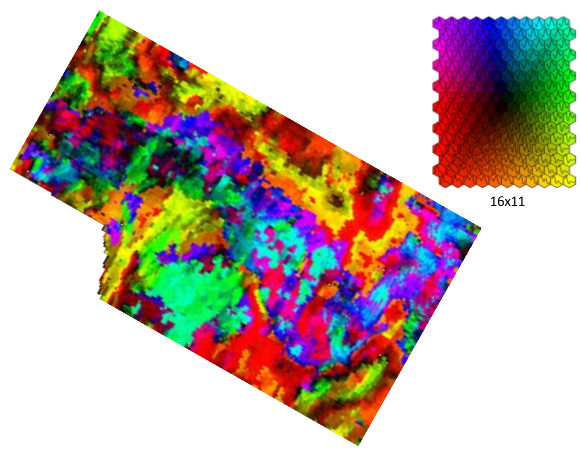
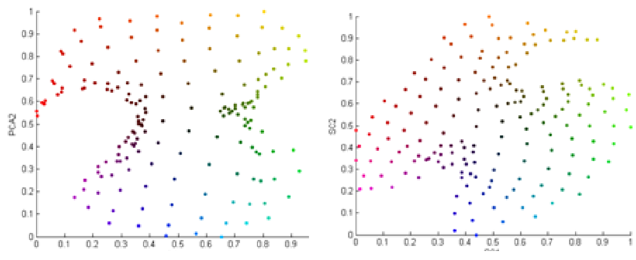


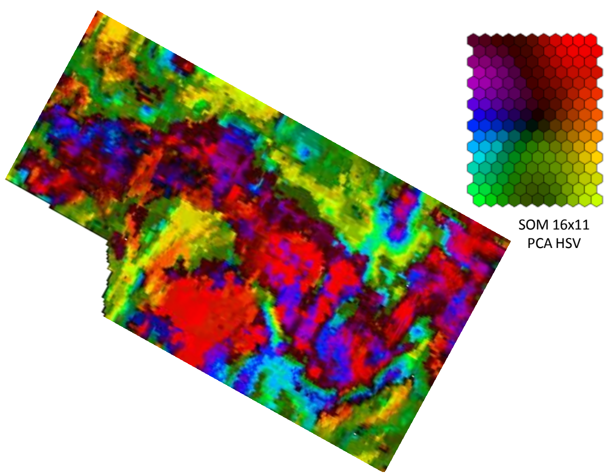
Figure 5: 2D SOM with 16x11 classes; Coloring without taking into account the distances between the prototype vectors. The colors were designated by using polar coordinates: values vary from 0 to 1, hues are function of the angle of the prototype vectors in the SOM topology plane and the saturation is equal to one.



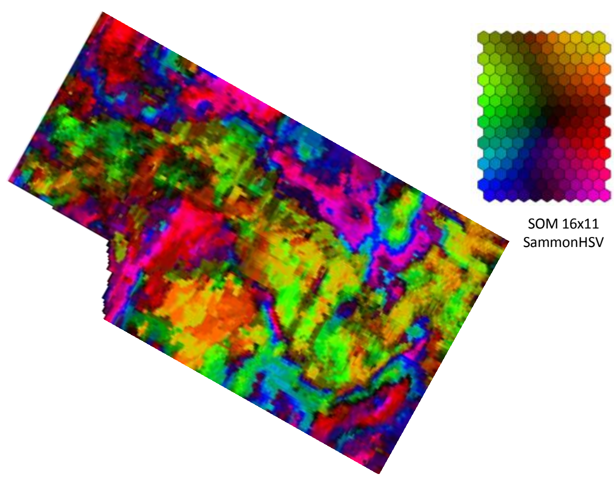
(a) (b)

Figure 6: SOM a) PCA projection; b) Sammon projection.

Once again, we can see from Figure 7 that the channel is clearly delineated and the relationship among waveforms in the 2D SOM color bar helps to interpret the geology.



(a)



(b)

Figure 7: 2D SOM with 16x11 classes. Coloring taking into account the distances between the prototype vectors using a) PCA projection and b) Sammon mapping projection.

Conclusions

Coding the SOM is a very good tool to visualize the relationship among different attributes. This technique is unsupervised and can be directly applied by the user. However, it should be emphasized that the choice of the seismic attributes for the classification of seismic patterns is fundamental to obtain coherent results.

Acknowledgments

The authors would like to thank Petrobras for their cooperation in providing the data and the authorization to publish this work. The first two authors would also like to thank the support from the University of Oklahoma Attribute-Assisted Seismic Processing and Interpretation Consortium.

References

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

Coléou,T.,M. Poupon, and K. Azbel, 2003, Interpreter's Corner—Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, 22, 942–953.

Dumay, J., and F. Fournier, 1988, Multivariate statistical analyses applied to seismic facies recognition: Geophysics, 53, 1151–1159.

Guo H., Lewis S. and K. J. Marfurt, 2008, Mapping multiple attributes to three- and four-component color models — A tutorial, GEOPHYSICS, 73, W7–W19.

Haykin, S., 1999, Neural networks: A comprehensive foundation, 2nd ed.: Prentice Hall.

Himberg, J., 1998, Enhancing the SOM-based Visualization by Linking Different Data Projections, First International Symposium on Intelligent Data Engineering and Learning (IDEAL'98), 427-434.

Kohonen, T., 2001, Self-organizingmaps, 3rd ed.: Springer-Verlag.

Matos, M.C., Osório, P.L.M and P.R.S. Johann, 2007, Unsupervised seismic facies analysis using wavelet transform and self-organizing maps, Geophysics, 72, P9–P21.

Sheriff, R. E., 2002, Encyclopedic dictionary of applied geophysics, 4th ed.: SEG.

Ultsch, A., 1993, Knowledge extraction from self-organizing neural networks, in Opitz et al., eds., Information and classification: Springer-Verlag.