Revista Brasileira de Geofísica
RBGf
Brazilian Journal of Geophysics

Revista Brasileira de Geofísica (2010) 28(4): 631-642
© 2010 Sociedade Brasileira de Geofísica
ISSN 0102-261X
www.scielo.br/rbg

SEISMIC INTERPRETATION OF SELF-ORGANIZING MAPS USING 2D COLOR DISPLAYS

Marcílio Castro de Matos[1,2], Kurt J. Marfurt[2] and Paulo R.S. Johann[3]

Recebido em 29 julho, 2009 / Aceito em 23 agosto, 2010
Received on July 29, 2009 / Accepted on August 23, 2010

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, can be applied to enhance seismic interpretation analysis associated with one and two-dimensional colormaps.

Keywords: self-organizing maps, Kohonen, classification, seismic facies.

RESUMO. Classificação não supervisionada de padrões em grupos é formalmente chamada de agrupamento. Dependendo da área de aplicação estes padrões são chamados de listas, observações ou vetores. Na exploração geofísica, padrões são associados a atributos sísmicos, formas de onda sísmicas ou fácies sísmicas. O principal objetivo deste artigo é mostrar como um dos mais populares algoritmos de agrupamento – mapas auto-organizáveis de Kohonen, associado a mapas de cores em uma e duas dimensões, podem ser aplicados a interpretação sísmica.

Palavras-chave: mapas auto-organizáveis, Kohonen, classificação, fácies sísmicas.

[1]Independent consultant <http://www.matos.eng.br> – E-mail: marcilio@matos.eng.br

[2]The University of Oklahoma, ConocoPhillips School of Geology and Geophysics, Room 872 SEC, 100 E. Boyd St., Norman, OK, USA, 73019-1009. Fax: 1-405-325-3140 – E-mail: kmarfurt@ou.edu

[3]Petrobras, E&P Production Engineering, Reservoir Geophysics Management, Av. República do Chile, 330, 9th floor, Rio de Janeiro, RJ, Brazil. Fax: +55 21 3224-6737 – E-mail: johann@petrobras.com.br

INTRODUCTION

One of the most important goals of seismic stratigraphy is to recognize and analyze seismic facies with regard to the geologic environment (Dumay & Fournier, 1988). According to Sheriff (2002), seismic facies analysis is done by examining seismic traces within an analysis window to characterize the amplitude, abundance, continuity and configuration of reflections in order to predict the stratigraphy and depositional environment.

The human brain excels at recognizing 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 to interpreters is provided by seismic attributes, which represent complex multisample waveforms by a reduced number of more relevant measurements designed to delineate geologic 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 (Poupon et al., 1999; Barnes & Laughlin, 2002) associated with 1D and 2D colormaps to help seismic interpretation. Notwithstanding SOM can also be used to estimate the number of clusters (Matos et al., 2007), in this paper, we show how to associate SOM to 1D and 2D colormaps to help interpreters visually identify the clustering structure of the input seismic attributes. Then, we apply the proposed SOM visualization technique to a seismic dataset acquired in the Campos Basin, offshore Brazil.

KOHONEN SELF-ORGANIZING MAPS (SOM)

The SOM (Kohonen, 2001) clustering is one of the most commonly used tools for non-supervised seismic facies analysis, with SOM providing ordered clusters that can be mapped to a gradational colorbar (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 $\Re^n$, $a_j = [a_{j1}, a_{j2}, \dots, a_{jN}]$, $j = 1, 2, \dots, J$; where N is the number of seismic attributes and J is the number of seismic traces when SOM is applied to surface attributes or is the number of voxels (Matos et al., 2005) when SOM is applied to volumetric attributes. The objective of the algorithm is to organize the dataset of input seismic attributes, into a geometric structure called the SOM.

If we assume that the self-organizing map has P units, defined as prototype vectors, then, there will exist P N-dimensional prototype vectors m_i, $m_i = [m_{i1}, \dots, m_{iN}]$, $i = 1, 2, \dots, P$; connected to its neighbors by a grid of lower dimension than P. Usually, this grid has dimension one or two and is related to SOM dimensionality. 2D SOM is most commonly represented by hexagonal or rectangular structural grids. After initializing the SOM prototype vectors to reasonably span the data space, the 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 (Kohonen, 2001). This iterative process stops either when the SOM converges or the training process reaches a predetermined number of iterations.

SOM places the prototype vectors on a regular low-dimension grid in an ordered fashion (Kohonen, 2001) and after training, the prototype vectors form a good representation of the input dataset of seismic attributes. Next, we label each input seismic attribute vector by the index of the closest SOM prototype vector, i.e., the SOM index with highest cross-correlation to the input data vector. This labeling process is called classification (Kohonen, 2001). SOM can be considered an unsupervised classification algorithm because no previous information is used to generate the prototype vectors. While SOM can easily be supervised (Kohonen, 2001), we will not do so in this paper.

The number of prototype vectors in the map determines both 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 reduced version of the input n-dimensional data ruled by a lower dimensional grid that attempts to preserve the original topological structure and since seismic data measures the changes in geology, SOM approximates 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 visualization difficult. For this reason, we exploit the topological relation among the prototype vectors as a visualization tool to display 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

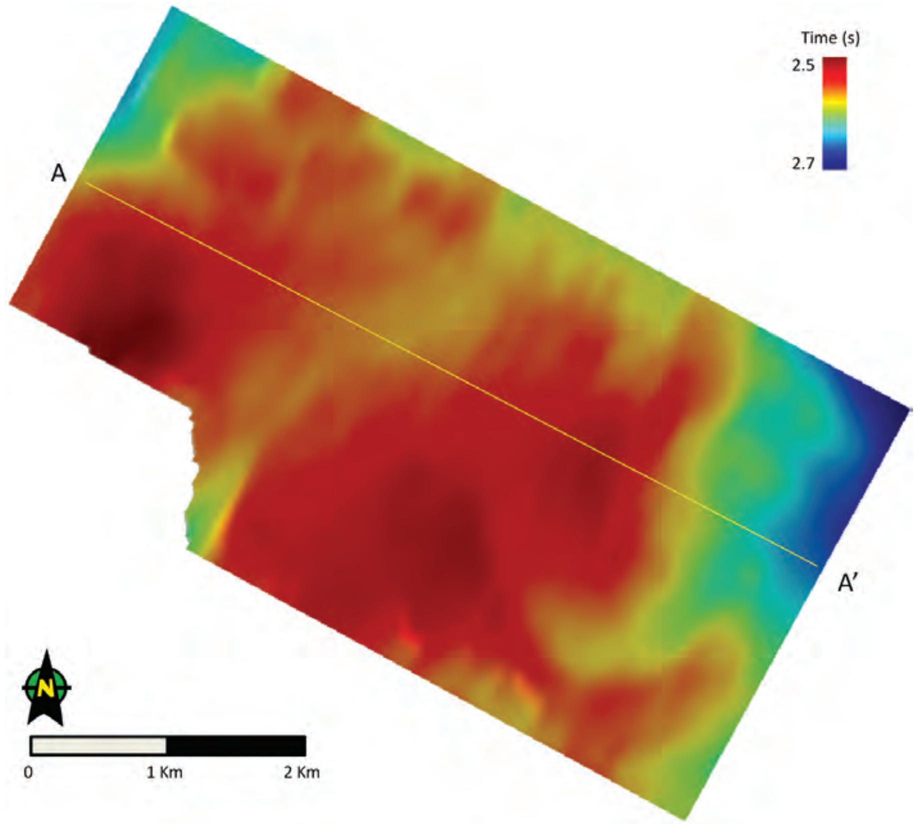


Figure 1 – Time-structure map of the base of the reservoir.

1D, 2D or 3D colorbars to the SOM topology to represent the location of each prototype vector.

SOM can be applied to volumetric or surface attributes. In this paper, we applied SOM to a suite of stratal slices through the seismic amplitude volume resulting in a cluster index map that can be displayed in the same manner as other horizon-based attributes (Chopra & Marfurt, 2007).

Geologic objective

Before we present the SOM methodologies, we introduce the seismic problem addressed in this paper. The main goal was to delineate a channel in the basal stratigraphic unit of a turbidite reservoir from the Campos Basin, offshore Brazil. Figure 1 shows the two-way time-structure map of the base of the reservoir. Figure 2a shows a seismic inline and Figure 2b shows a zoomed version with the proportional horizon slices generated between the base of the reservoir and an intermediate stratigraphic horizon, while Figure 3 shows an amplitude horizon slice at the base of the reservoir.

1D SOM plotted against 1D colorbar

The main objective here is to classify the waveforms represented by the amplitudes illustrated in Figure 2b by using the 1D and 2D SOM displayed against 1D and 2D colorbars. Therefore, the input seismic attributes are the instantaneous amplitude horizon maps of each proportional slice. In general, attributes other than amplitude can be used – for example, Angelo et al. (2009) applied 2D SOM to seismic textures computed using a gray-level co-occurrence matrix.

First, a one-dimensional SOM was trained. Then each prototype vector was assigned a color using an HSV color model (Guo et al., 2008) with hue ranging between H=0° (red) and H=270° (blue) and fixed values of saturation, S=1.0 and value V=1.0. 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 assigning it to the color of the closest prototype vector. In general, classification can be done on any suite of attributes through the use of the Mahalanobis distance. On our Campos Basin example shown in Figure 1, our attributes are simply seis-

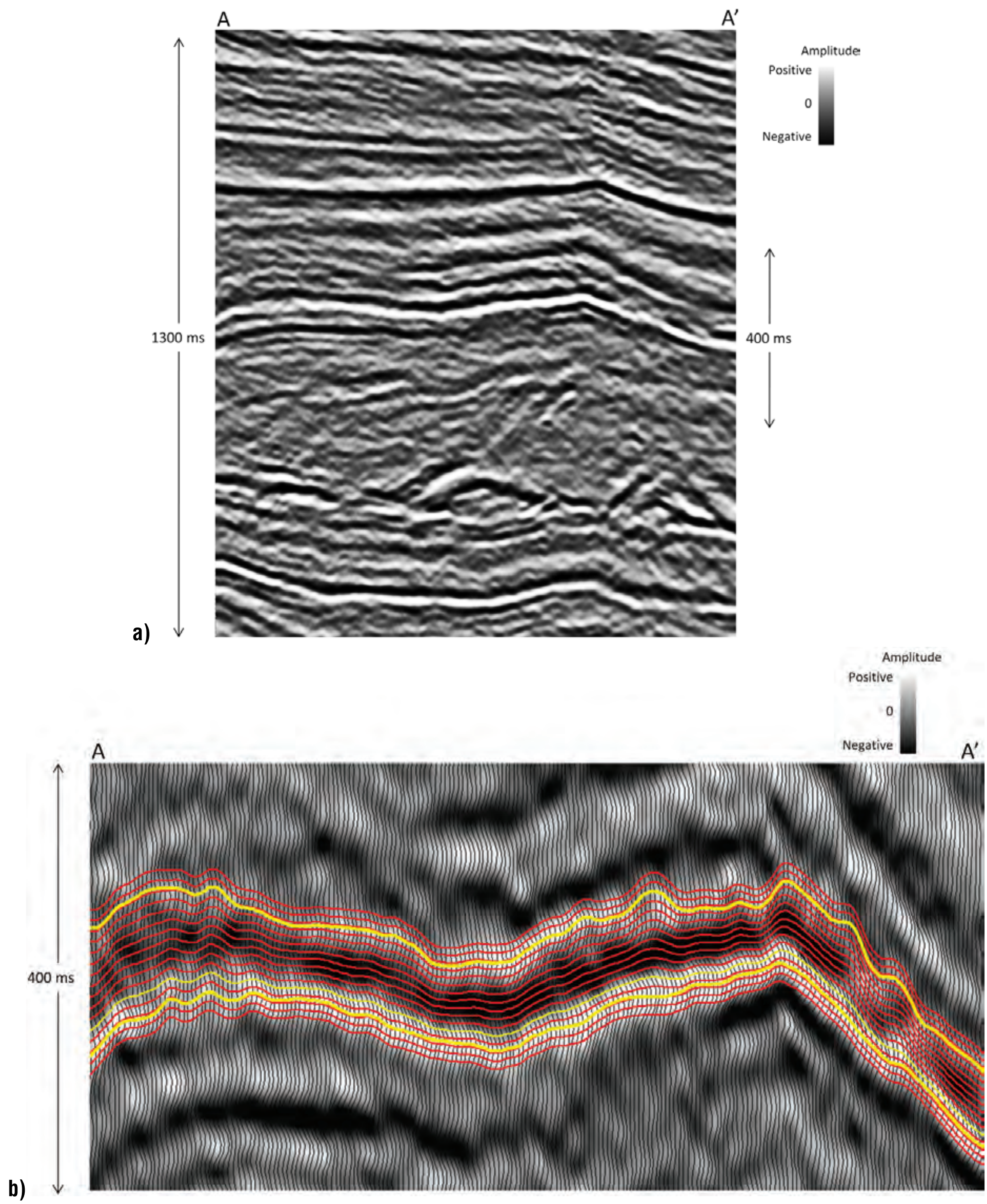


Figure 2 – a) Seismic line; b) Proportional horizon slices between the base of the reservoir and an overlying intermediate stratigraphic horizon.

mic amplitudes on subsequent stratal slices, such that the Mahalanobis distance is replaced by the simpler Pythagorean distance. When viewed vertically, each prototype vector takes on the appearance of a waveform shape, giving rise to what is called “waveform shape classification” (e.g. Coléou et al., 2003). Figure 4a shows the result using 19 classes labeled by 19 colors uniformly

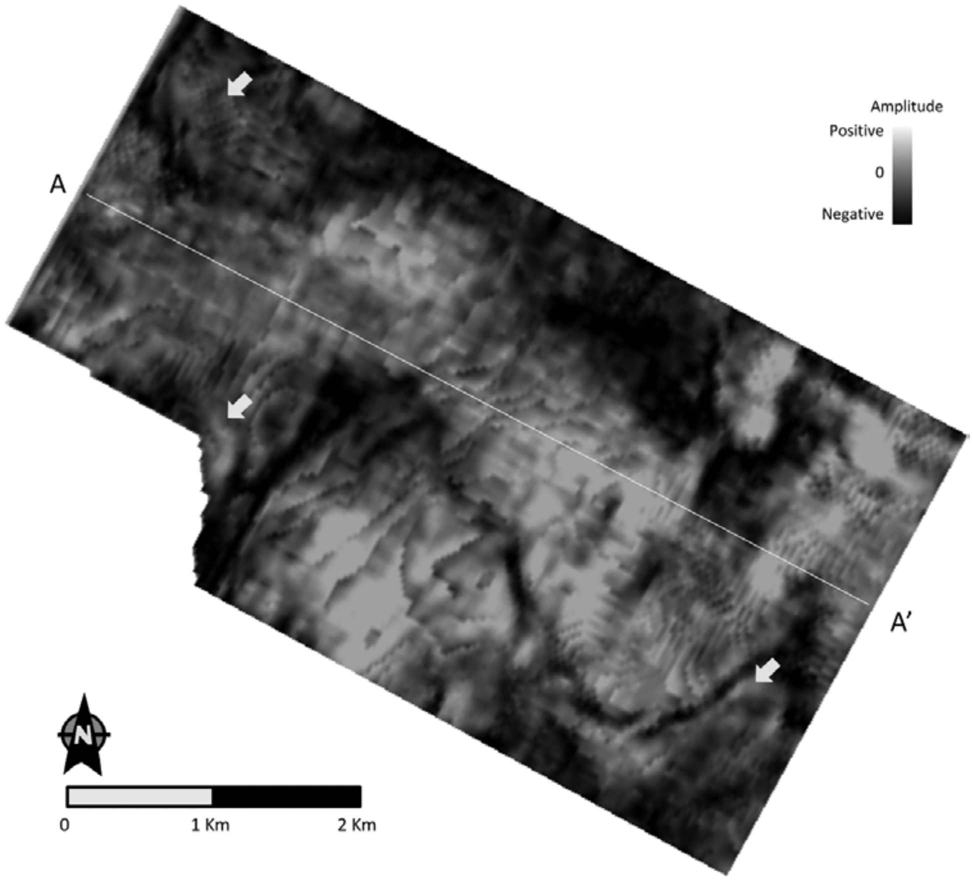


Figure 3 – Amplitude horizon slice at the base of the reservoir. Block arrows indicate where high amplitudes are not associated with the channel waveform shape.

distributed along the hue (azimuth) defined by Eq. (1):

$$H(i) = \frac{2\pi i}{N-1} \times \frac{3}{4} = \frac{270° i}{N-1}, \; i = 0, \ldots, N-1 \qquad (1)$$

where $N = 19$ is the number of colors.

This representation neither takes into account the distances between the prototype vectors nor shows the clustering structure. Figure 4b shows the same 1D SOM colored by using the distances between neighboring prototype vectors defined by Eq. (2):

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

Using Eq. (2) we note that waveforms that have a similar shape (i.e. the classes are near each other in n-space) have similar colors, which facilitates the visual identification of the seismic facies (Fig. 4b). Note that the numerator of Eq. (2) is the location of cluster i in latent space, while the denominator is the length of the total 1D latent space.

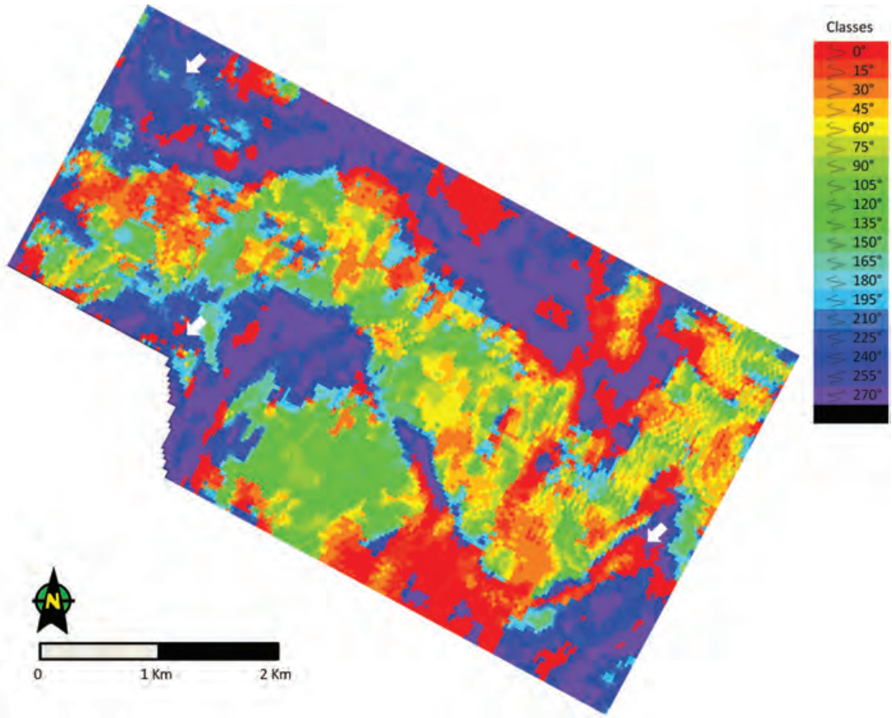
By increasing the number of prototype vectors, clusters, and colors to 256 (Fig. 5), we generate intermediate clusters which further delineate subtle features for the human interpreter.

Specifically, we clearly see that some regions in Figure 3 with high amplitudes indicated by block arrows are not associated with the channel waveform shape as shown in Figures 4 and 5.

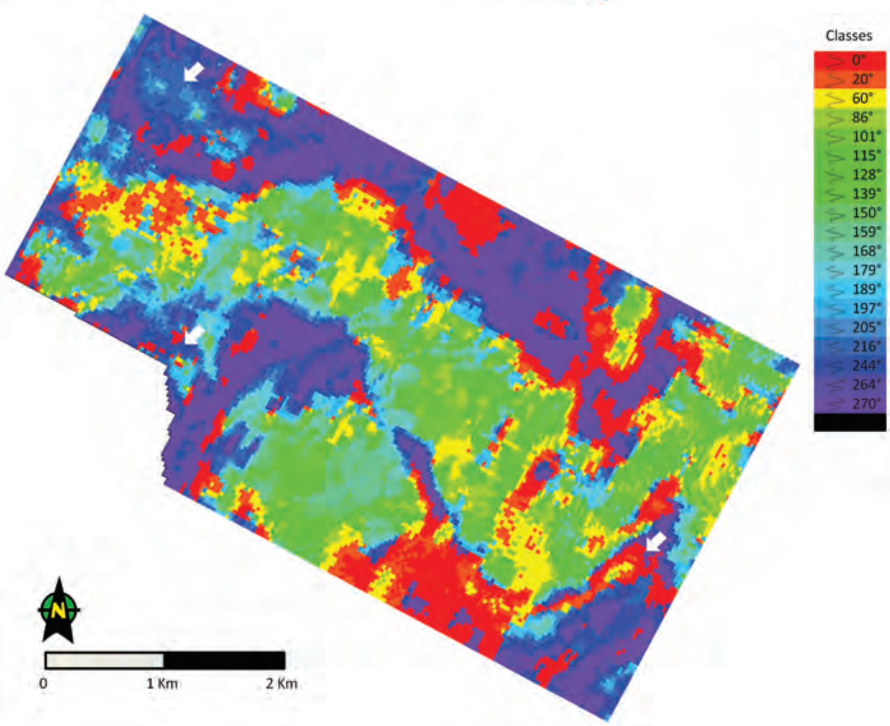
SOM plotted against 2D colormaps

Although 1D SOM provides very good visualization results it is less effective in identifying the number of clusters in the data (Matos et al., 2007).

Measuring the distances between SOM prototype vectors is one way to identify the number of clusters in the data. Figure 6 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



a)



b)

Figure 4 – 1D SOM with 19 classes; a) Colors proportional to prototype vector index; b) Colors proportional to the prototype vector location on a 1D reference space.

extracted from a complex sandstone turbidite system. The choice of the seismic attributes for the classification of seismic patterns is fundamental to obtain geologically relevant results.

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

Figure 7 shows the classification results using a 2D color-

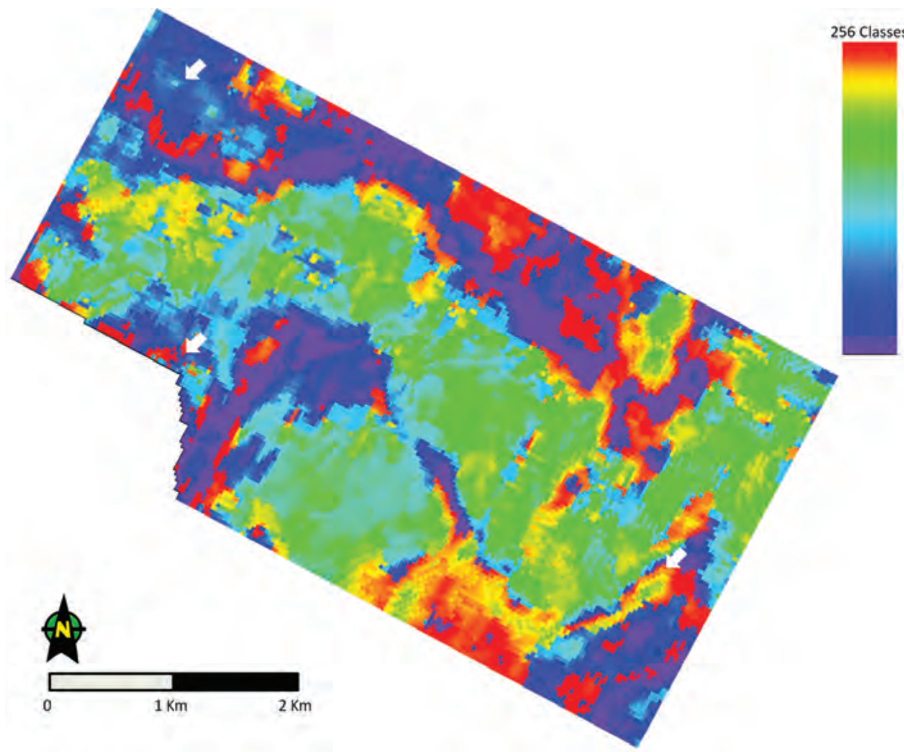


Figure 5 – 1D SOM with 256 classes. Colors are proportional to the prototype vector location on a 1D reference space.

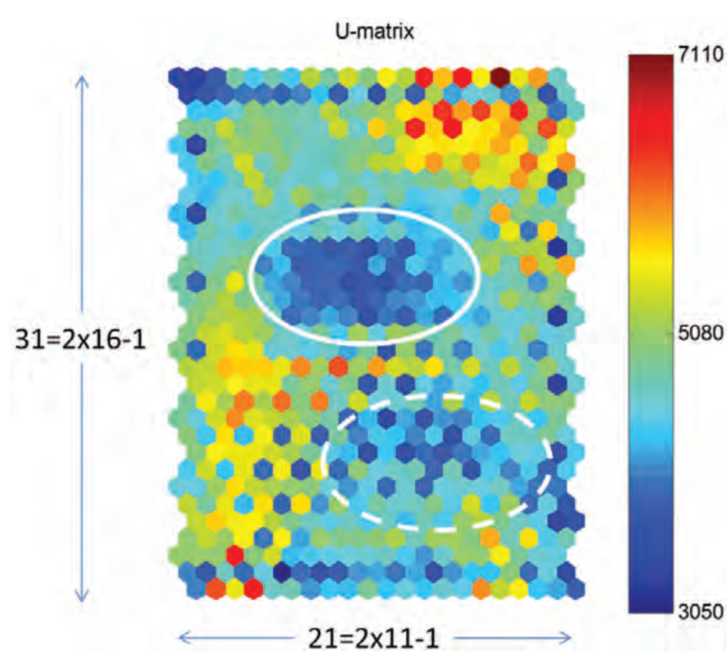


Figure 6 – U-matrix, where colors correspond to the distances between prototype vectors. U-matrix size is equal to twice 2D SOM size minus one (see Matos et al., 2007). Solid white ellipse indicates one good cluster. Dotted white ellipse indicates a more diffuse cluster. In general the data do not cluster well, but can still be ordered.

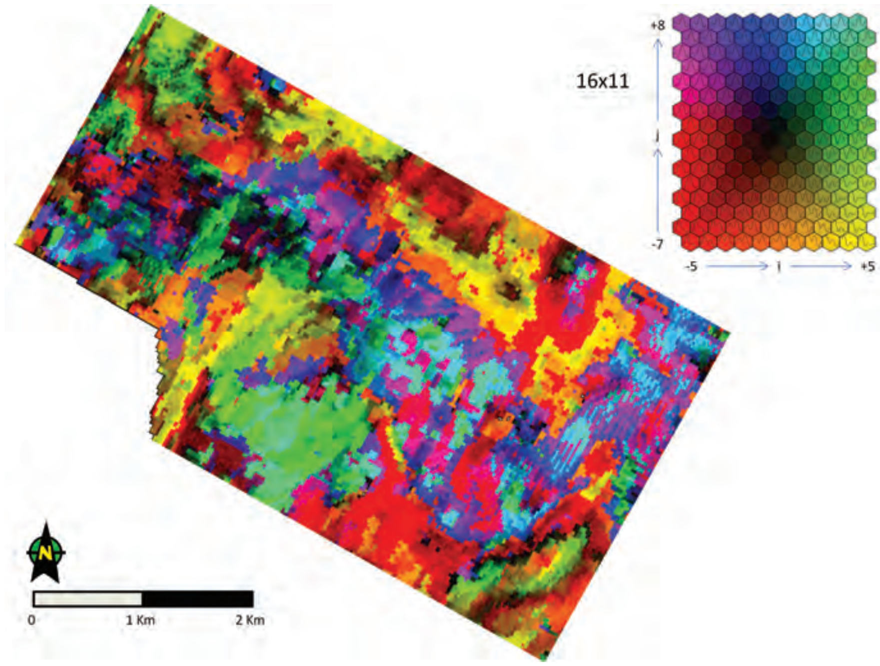


Figure 7 – 2D SOM with 16×11 classes. Colors are a simple function of the prototype vector indices x_{ij}, where $-5 \leq i \leq +5$ and $-8 < j \leq +8$. The hue is proportional to $\tan^{-1}(j/i)$ while the saturation is proportional to $(i^2 + j^2)^{1/2}$.

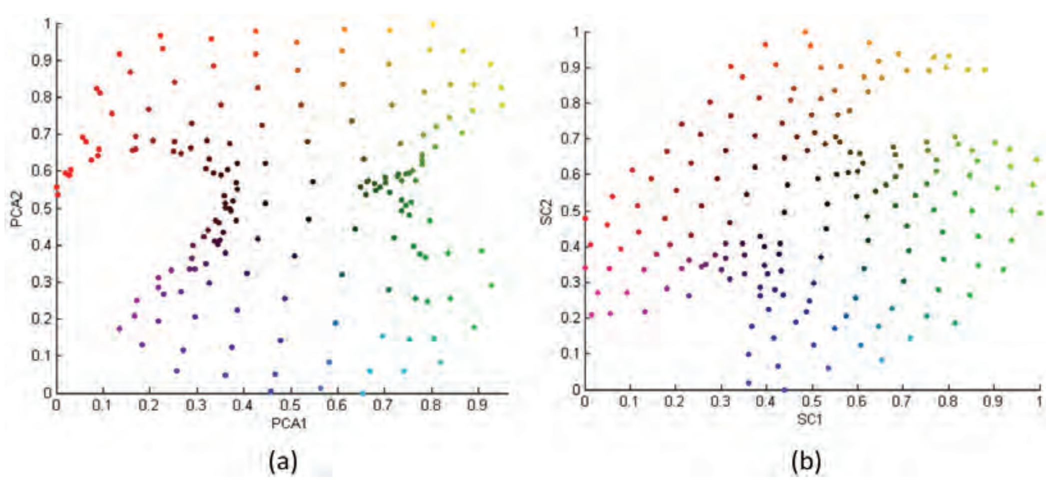


Figure 8 – SOM using a) a PCA projection and b) a Sammon projection. The hue is proportional to $\tan^{-1}\left(\frac{y-1/2}{x-1/2}\right)$ while the saturation is proportional to $\left[(x-1/2)^2 + (y-1/2)^2\right]^2$, such that they map to the 2D (PCA or SC) reference space.

bar (without taking into account the distances among the SOM prototype vectors). Although accounting for the distances is not as direct as with 1D SOM, Himberg (1998) suggests several alternative measures. In this paper we project the SOM prototype vectors using Principal Component Analysis and Sammon mapping onto a two-dimensional plane. We then apply the HSV color model to the 2D projections and color the SOM units. Figure 8a shows the 2D PCA projection of the SOM prototype vectors while Figure 8b shows the Sammon projection. Figure 9a shows the SOM classification results using PCA and Figure 9b shows the results using Sammon mapping. We could also use the first three PCA and Sammon components of the SOM prototype vectors projections to create a similar 3D HSV or RGB color model (Wallet et al., 2009).

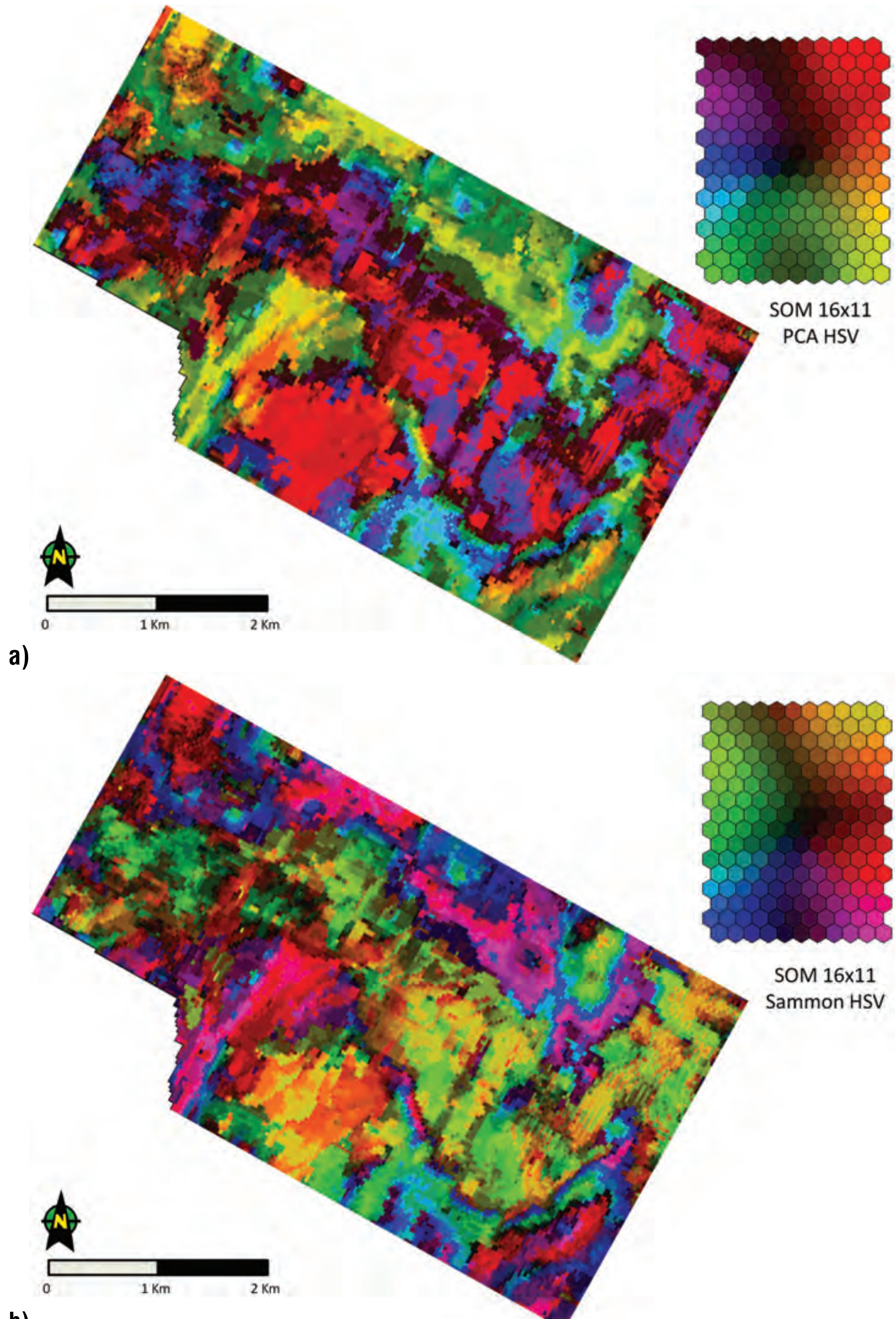


Figure 9 – 2D SOM with 16×11 classes. Colors are proportional to the location using a) the PCA and b) the Sammon mapping projections displayed in Figure 8. PCA and SC have a similar color pattern, but it seems PCA and SC 2D projections of the SOM prototype vectors are rotated between each other.

Another way to obtain a 2D projection is by contracting the topological coordinates of the pre-specified 2D grid using the distances among the prototype vectors. New grid coordinates can be estimated as the weighted average of the prototype vector locations (Himberg, 2000). The weights are defined as the similarities between prototype vector pairs, s_{ij}. We then construct a function, f, that sets the similarity, $s_{ij} = 1$ for a distance $d_{ij} = 0$ between prototype vectors i and j for identical vectors, and

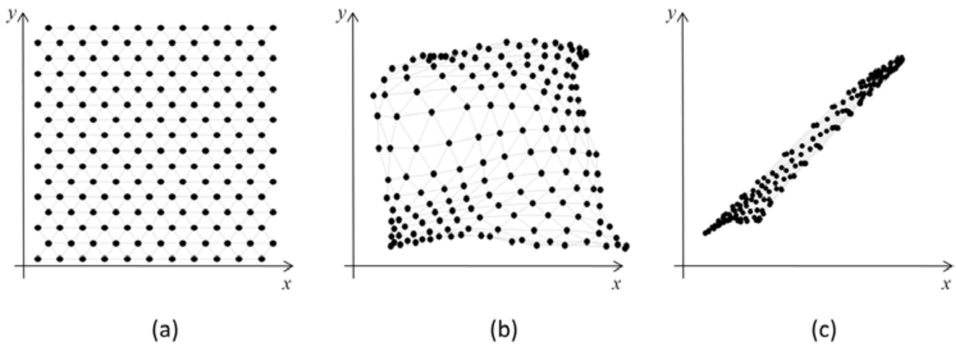


Figure 10 – Grid coordinates location during contraction. The (x, y) SOM grid coordinates are updated by Eq. (4).

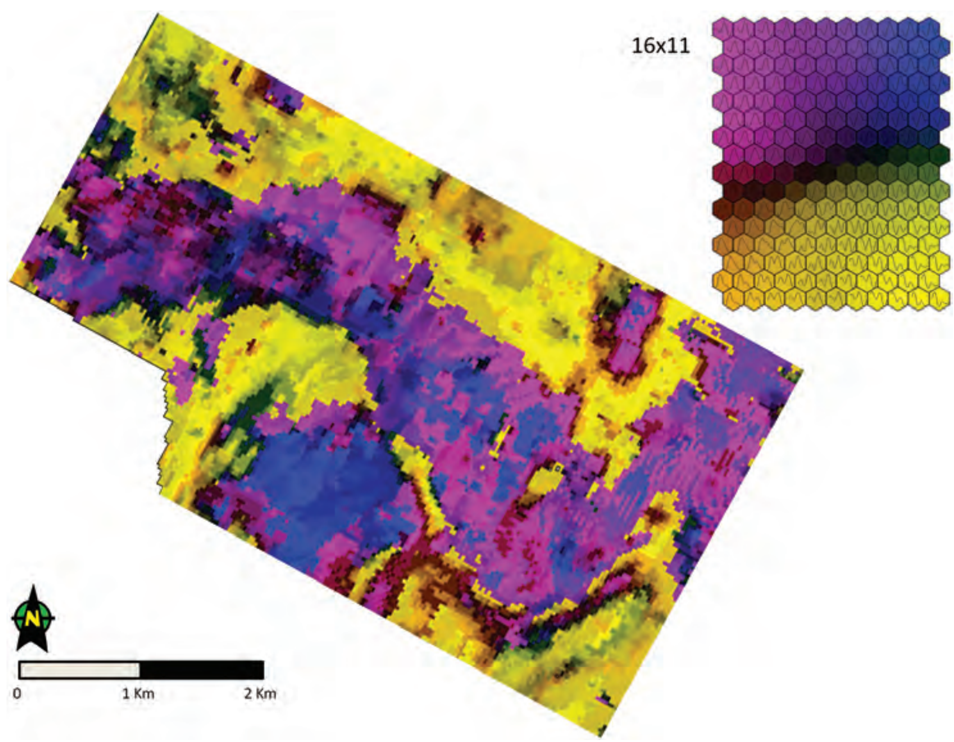


Figure 11 – 2D SOM with 16×11 classes. Colors are proportional to the location of the prototype vectors after grid contraction.

$s_{ij} = 0$ for very large distances for highly dissimilar vectors:

$$s_{ij} = f\left(d_{ij}\right) = \exp\left(d_{ij}^2/\sigma^2\right), \tag{3}$$

where σ^2 is a parameter that controls the width of f.

After each row of the matrix $\mathbf{S}$ is normalized, the grid coordinates are updated by using:

$$x_{i+1} = \mathbf{S}x_i,\ i = 0, 1, \ldots, r, \tag{4}$$

where the x_i vectors are the coordinates in the original grid, $\mathbf{S}$ is the similarity matrix and r is number of iterations. Figure 10 shows the contraction progress of the grid coordinates.

After the contraction, the grid coordinates are used to create a 2D colorbar as before. Figure 11 shows the classification result.

We can see from Figures 9 and 11 that the channel is clearly delineated and the relationship among waveforms in the 2D SOM colorbar helps to interpret the geology.

CONCLUSIONS

Most seismic clustering workflows attempt to estimate the number of clusters prior to labeling the data. In this paper, we avoid this difficulty by oversampling the latent space with a very large number of clusters and plotting them against a continuous color model. In our examples using 8-bit color display software, we limited ourselves to 256 colors. Using this methodology, the ordered data are then 'clustered' in the mind of the interpreter. By this way we showed that color-coding the SOM is a powerful means to visualize the relationship among n different attributes in one, two and three-dimensional color space. Since this is an unsupervised technique the major user intervention before interpretation is the actual choice of which attributes to use in the classification.

ACKNOWLEDGEMENTS

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

REFERENCES

ANGELO SM, MATOS M & MARFURT KJ. 2009. Integrated seismic texture segmentation and clustering analysis to improved delineation of reservoir geometry. In: 79th Annual International Meeting of the SEG. Expanded Abstracts, 1107–1111.

BARNES AE & LAUGHLIN K. 2002. Investigation of methods for unsupervised classification of seismic data. In: 72th Annual International Meeting, SEG. Expanded Abstracts, 2221–2224.

CHOPRA S & MARFURT KJ. 2007. Seismic Attributes for Prospect Identification and Reservoir Characterization. Society of Exploration Geophysicists. p. 7–8.

COLÉOU T, POUPON M & AZBEL K. 2003. Interpreter's Corner – Unsupervised seismic facies classification. A review and comparison of techniques and implementation. The Leading Edge, 22: 942–953.

DUMAY J & FOURNIER F. 1988. Multivariate statistical analyses applied to seismic facies recognition. Geophysics, 53: 1151–1159.

GUO H, LEWIS S & MARFURT KJ. 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. p. 446–466.

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

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

KOHONEN T. 2001. Self-organizing maps. 3rd ed. Springer-Verlag. p. 105–114.

MATOS MC, OSÓRIO PLM & JOHANN PRS. 2005. Vertical Seismic Facies Detection through Unsupervised 3D Voxel based Seismic Facies Classification Applied to a Turbidite Field in Campos Basin, Brazil. In: 9th International Congress of the Brazilian Geophysical Society. Expanded Abstracts. CD-ROM.

MATOS MC, OSÓRIO PLM & JOHANN PRS. 2007. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. Geophysics, 72: P9–P21.

POUPON M, AZBEL K & INGRAM JE. 1999. Integrating seismic facies and petro-acoustic modeling. World Oil Magazine, June, 75–80.

SHERIFF RE. 2002. Encyclopedic dictionary of applied geophysics. 4th ed. SEG.

ULTSCH A. 1993. Knowledge extraction from self-organizing neural networks. In: OPITZ O, LAUSEN B & KLAR R (Eds.). Information and Classification: Concepts, Methods and Applications. Proceedings of the 16th Annual Conference of the "Gesellschaft für Klassifikation e.V.", Universität Dortmund, 1.-3.4.1992, Springer-Verlag, Berlin. Springer-Verlag, 301–306.

WALLET BC, MATOS MC & KWIATKOWSKI T. 2009. Latent space modeling of seismic data. An overview. The Leading Edge, 28: 1454–1459.

NOTES ABOUT THE AUTHORS

Marcílio Castro de Matos received the B.S. and M.Sc. degree in electrical engineering from Instituto Militar de Engenharia (IME) in 1988 and 1994, respectively, and the doctor degree from Pontifícia Universidade Católica do Rio de Janeiro in 2004. He served from 1989 to 1999 as military engineer at the Brazilian Army Test Center and at IME as military professor from 1999 to 2010. He was visiting scholar at The University of Oklahoma (OU) from January 2008 to January 2010 and is currently co-investigator of the Attribute-Assisted Seismic Processing & Interpretation research consortium at OU. As an independent consultant his main research interests include applied seismic analysis, digital signal processing, spectral decomposition, and seismic pattern recognition.

Kurt J. Marfurt began his geophysical career as an assistant professor teaching mining geophysics at Columbia University's School of Mines in New York. After five years, he joined Amoco Tulsa Research Center where he won five patents. Joined the University of Houston in 1999 as a professor in the Department of Geosciences and as director of the Center for Applied Geosciences and Energy. He joined the University of Oklahoma in 2007. His primary research interest is in the development and calibration of new seismic attributes to aid in seismic processing, seismic interpretation, and reservoir characterization.

Paulo Roberto Schroeder Johann joined Petrobras in 1981. He has been Reservoir Geophysics manager since 2007 and has been involved with internal Petrobras training program for new geophysicists and geologists since 1985. He was PRAVAP (Research IOR Program) coordinator in geophysical technology from 2001 to 2007. He has vast experience in the petroleum industry and his career has encompassed geophysical acquisition, geophysical interpretation, and reservoir geophysics. He received the B.S. degree in geology from UNISINOS University, Brazil in 1980, the D.E.A. and the Ph.D. degrees in reservoir geophysics from Paris VI University, France in 1994 and 1997, respectively. He received MBA degrees in Project Management and in Advanced Management from FGV and COPPEAD, respectively. He was the first SEG Latin America Honorary Lecturer (2008) and was vice-president of SBGf (2003-2005) and of SEG (2008-2009).