

AGRUPAMENTO DE MAPAS AUTO ORGANIZÁVEIS DE KOHONEN APLICADO A ANÁLISE DE FÁCIES SÍSMICAS

MARCÍLIO CASTRO DE MATOS*[†], PAULO LÉO M. OSÓRIO[†], PAULO ROBERTO SCHROEDER JOHANN[§]

**Departamento de Engenharia Elétrica, Instituto Militar de Engenharia
Pr. Gen. Tibúrcio, 80, Rio de Janeiro, BRASIL, 22290-270*

*†Laboratório de Processamento de Sinais, Departamento de Engenharia Elétrica, PUC-RIO
Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ, BRASIL, 22453-900*

*§Departamento de Exploração e Produção, Petrobras S.A.
Av. República do Chile, 65/1704B, Rio de Janeiro, RJ, BRASIL, 20035-900*

E-mails: marcilio@ime.eb.br, osorio@ele.puc-rio.br, johann@petrobras.com.br

Abstract— The estimation of the geological parameters directly from the seismic signal is called seismic facies analysis. The seismic facies analysis is obtained using pattern recognition techniques, where through the combination of the seismic attributes appropriately it could be possible to estimate reservoir properties. This work proposes to use the clustering of the self-organizing maps as a tool to seismic facies analysis.

Keywords— Seismic Facies Analysis, Pattern Recognition, Self-Organizing Maps, Kohonen Maps.

Resumo— A estimação de parâmetros extraídos das reflexões sísmicas é chamada de análise de fácies sísmicas. A análise de fácies sísmicas é realizada através do emprego de técnicas de reconhecimento de padrões, onde através da combinação adequada de atributos sísmicos, busca-se a identificação de características geológicas dos reservatórios. Este trabalho propõe a utilização do agrupamento dos mapas auto organizáveis de Kohonen como ferramenta para análise de fácies sísmicas.

Palavras-chave— Mapas auto organizáveis de Kohonen, Análise de fácies sísmicas, Reconhecimento de padrões.

1 Análise de Fácies Sísmicas

Os modelos geológicos dos reservatórios de petróleo são formados a partir de estimativas de características específicas das rochas ou sedimentos. Estimativas das propriedades das rochas, tais como porosidade, permeabilidade, tipo do fluido, litologia e outras propriedades dos reservatórios são fundamentais para a exploração, desenvolvimento e produção de campos de petróleo. Assim, a busca por modelos mais exatos e confiáveis é uma questão de grande impacto para as companhias que exploram petróleo, pois pequenos erros quantitativos no dimensionamento dos reservatórios podem representar grandes perdas, ou ganhos, econômicos.

Toda a sorte de informação é válida para a construção dos modelos de reservatório. Uma das mais importantes é a proveniente de poços perfurados. Informações como perfis, amostras estratigráficas e dados de produção são bastante precisas. Entretanto, as informações de poços são pontuais e não refletem o comportamento do reservatório como um todo, além de não estarem disponíveis na fase inicial de exploração. Modelos construídos utilizando somente dados de poço são extrapolações de propriedades pontuais para a área do reservatório.

Ao contrario dos dados de poços, os dados sísmicos 3D abrangem grandes áreas e embora não tenham a mesma resolução, assumindo que mudanças

na litologia, nas propriedades das rochas e/ou dos fluidos afetam os traços sísmicos com respeito a amplitude, forma e coerência lateral, pode-se considerar que a análise das variações dos atributos sísmicos na região do reservatório analisado, possa ser utilizada como ferramenta para a construção de modelos de reservatório. A descrição e interpretação de parâmetros extraídos das reflexões sísmicas, incluindo geometria, continuidade, amplitude, frequência e velocidade intervalar é chamada de análise de fácies sísmicas (Mitchum, 1977).

A análise de fácies sísmicas é realizada através do emprego de técnicas de reconhecimento de padrões, onde através da combinação adequada de atributos sísmicos, busca-se a identificação de características geológicas dos reservatórios. Dentre as técnicas usadas, destacam-se as redes neurais.

2 Métodos Utilizados para Análise de Fácies Sísmicas

A análise de fácies sísmicas não é uma tarefa simples, tampouco determinística. Pois, não existe, ainda, consenso sobre uma forma padrão para análise, levando os interpretes a diversificar a forma de análise.

A análise de fácies sísmicas pode ser não-supervisionada ou supervisionada.

Diz-se que a análise de fácies sísmicas é não-supervisionada, sempre que as informações sobre a geologia da região forem incompletas ou inexistentes, ou seja, no período de exploração ou quando a área analisada ainda não tiver sido suficientemente explorada através da perfuração de poços.

A análise de fácies sísmicas será supervisionada quando existirem dados suficientes da região para se depurar a análise.

Independente da análise de fácies sísmica ser supervisionada ou não, geralmente, a sua realização se dá por meio da: segmentação espacial e temporal dos dados sísmicos 3D com orientação geológica; seleção dos atributos sísmicos que serão utilizados; escolha do número de classes, fácies ou padrões, e o número de iterações que serão utilizados pelo algoritmo; treinamento e classificação dos atributos selecionados utilizando alguma ferramenta estatística ou redes neurais; construção e interpretação dos mapas de fácies sísmicas.

A escolha do algoritmo usado para a classificação dos atributos sísmicos é função do conhecimento geológico da região em análise. Quando a análise for não supervisionada, utilizam-se algoritmos que busquem a clusterização ou agrupamentos dos dados que tenham características em comum. Os algoritmos mais utilizados pelos programas comerciais são o K-means (Duda et al., 2001) e o de Aprendizado Competitivo (Gurney, 1997).

Quando a análise é supervisionada, utilizam-se algoritmos estatísticos tradicionais ou redes neurais como o algoritmo de retro-propagação (Haykin, 1999).

O resultado da análise de fácies sísmicas é chamado de mapa de fácies sísmicas. Através deste mapa, pode-se avaliar características em comum existentes na região analisada. Associando informações de poços é possível calibrar o mapa para a característica geológica desejada.

Pode-se verificar facilmente, que apesar de sua comprovada eficácia (Johann, 1999), todo o processo descrito deve ser realizado com bastante cuidado, pois pequenos erros em qualquer etapa podem levar a resultados totalmente diferentes da realidade.

Atualmente, os Mapas Auto Organizáveis (SOM) ou mapas de Kohonen (Kohonen, 2001) vêm se tornando uma das ferramentas mais comuns para a análise não-supervisionada (Morrice et al, 1996). Entretanto, a escolha do número de classes e a melhor combinação de atributos sísmicos, normalmente, são realizadas de forma empírica pelos sistemas existentes atualmente no mercado (Matos et al., 2003).

3 Mapas Auto Organizáveis de Kohonen

Os mapas auto-organizáveis de Kohonen, SOM, além de produzirem uma espécie de similaridade entre os dados de entrada, são uma ferramenta efetiva para a visualização e clusterização de dados multidimensionais, convertendo relações estatísticas en-

tre dados multidimensionais em relações geométricas simples dos pontos correspondentes no espaço de dimensão inferior, tipicamente bidimensional. Logo, o SOM pode ser interpretado como uma forma de se preservar as relações métricas e topológicas do espaço de entrada multidimensional em uma rede bidimensional, que pode ser usada como uma ferramenta de visualização, mostrando diferentes características dos dados, como sua possível estruturação em grupos.

O mapeamento realizado pelo SOM é implementado de forma similar à quantização vetorial. Assumindo que o conjunto das variáveis de entrada pode ser representado por vetores no espaço \mathcal{R}^n , $x=[x_1, x_2, \dots, x_n]$, o objetivo do algoritmo é mapear o conjunto de dados de entrada em um espaço de dimensão inferior, delimitado por uma estrutura geométrica, chamada de mapa auto organizável. As unidades do SOM são conectadas às suas vizinhas, formando mapas com estruturas, normalmente, hexagonais ou retangulares.

O mapeamento é realizado através de vetores protótipos que mantêm uma relação ordenada entre si. Especificamente, para cada elemento do SOM existe um vetor protótipo, que por sua vez está associado aos seus vetores protótipos vizinhos. Supondo que o mapa tenha P elementos, relacionados a estes existirão P vetores protótipos m_i , $m_i=[m_{i1}, \dots, m_{in}]$, i variando de 1 à P , onde n é a dimensão do vetor de entrada. Após, o treinamento do SOM, os vetores protótipos passam a representar o conjunto de dados de entrada.

O número de unidades no mapa determina a sua eficácia e capacidade de generalização. Durante o treinamento, o SOM forma uma rede elástica que se adapta à “nuvem” formada pelos dados de entrada. Dados localizados próximos entre si no espaço de entrada, também serão próximos entre si no mapa formado. Desta forma, o SOM pode ser interpretado como um mapeamento do espaço de entrada em um arranjo ou reticulado bidimensional que preserva a estrutura topológica original. Tal característica o torna bastante atraente para aplicações geológicas onde as grandezas físicas são distribuídas de forma quase contínua pela natureza.

Normalmente, o SOM é treinado de forma iterativa. Inicialmente, os vetores protótipos podem ser inicializados aleatoriamente ou ter os seus valores arbitrados usando, por exemplo, as projeções nos dois maiores autovetores dos dados de entrada. Então, a cada passo do treinamento, um vetor de entrada x é aleatoriamente escolhido do conjunto de dados de entrada. A distância entre x e todos os vetores protótipos é calculada. A unidade do mapa com menor distância ao vetor de entrada x (BMU – “best matching unit”) é chamada de b :

$$\|x - m_b\| = \min_i \{\|x - m_i\|\} \quad (1)$$

Em seguida, o vetor protótipo correspondente ao BMU e seus vizinhos são atualizados, sendo, portanto, movidos em direção ao vetor de entrada ven-

cedor no espaço de entrada. A regra de atualização para a unidade i é dada por:

$$m_i(t+1) = m_i(t) + \alpha(t) h_{bi}(t) [x - m_i(t)] \quad (2)$$

onde t denota o tempo, $\alpha(t)$ é a taxa de aprendizado e $h_{bi}(t)$ é o tamanho da vizinhança centrada na unidade vencedora que decresce ao longo do aprendizado e é dada por:

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}} \quad (3)$$

onde r_b e r_i são as posições dos neurônios b e i no reticulado do SOM.

Após o treinamento, os vetores protótipos representam uma forma de mapear os dados de entrada no SOM, bem como representam uma boa aproximação dos vetores de entrada.

Como o principal objetivo deste trabalho é a identificação de agrupamentos de dados, serão utilizados, para visualização do SOM, técnicas que enfatizem as correlações entre os pontos do mapa.

Visualizar as distâncias entre os vetores protótipos vizinhos é uma forma bastante interessante de se identificar similaridades entre os elementos do SOM. Especificamente, a matriz unificada de distância (Matriz-U) (Utsch, 1993) é uma forma de representar tais distâncias.

Para a formação da matriz-U, após o treinamento, para cada elemento do SOM é calculada a distância entre os vetores protótipos vizinhos além da média entre eles. A matriz-U além de mostrar a distância média entre cada elemento, também ilustra o gradiente entre os mesmos. Normalmente, a matriz-U é visualizada como uma imagem e indica possíveis agrupamentos formados pelos atributos representados pelo SOM, sugerindo, inclusive, o número de grupos existentes no dado.

4 Agrupamento do SOM

Em processos de mineração de dados, independente da área do conhecimento, o volume de dados costuma ser muito grande. Entretanto, as informações a respeito destes dados são, geralmente, precárias. Analogamente, em reconhecimento de padrões sísmicos, algumas informações não são conhecidas previamente. Por exemplo, a estimativa do número de fâcies sísmicas existente nos dados, em análises não-supervisionadas, é realizada de forma empírica.

Arbitrando-se o número de vetores protótipos do SOM muito maior que o número de grupos esperados no dado, a partir da visualização do SOM, pode-se estimar a quantidade de agrupamentos existentes no dado. Portanto, a visualização da formação de agrupamentos de dados com características em comum através do SOM é uma ferramenta bastante poderosa para análise de sinais. Entretanto, através da visualização, somente informações qualitativas podem ser obtidas.

Para se obter uma descrição quantitativa das propriedades dos dados, os grupos formados devem

ser selecionados no SOM. Logo, baseado na visualização, os grupos poderiam ser escolhidos manualmente. Entretanto, como tal procedimento seria tedioso e geraria incertezas, propõe-se o agrupamento semi-automático dos dados através do agrupamento do SOM utilizando métodos tradicionais como o K-means.

A abordagem utilizada (Vesanto e Alhoniemi, 2000) neste trabalho consiste em se agrupar os vetores protótipos em lugar dos dados originais. Desta forma, grandes conjunto de dados, representado pelos vetores protótipos do SOM podem ser indiretamente agrupados. O método proposto, além de dar um melhor entendimento sobre a formação dos grupos, tem como vantagem a redução drástica do tempo computacional (Vesanto, 2002). Outro benefício desta metodologia é a redução de ruído, visto que os vetores protótipos representam médias locais dos dados.

Cabe ressaltar que o objetivo deste trabalho não é a obtenção de uma forma ótima para clusterização de dados, mas sim de identificar e separar diferentes características em grandes conjuntos de dados, especificamente, dados sísmicos. Obviamente, esta abordagem somente é válida quando os grupos encontrados utilizando o SOM são similares àqueles do dado original. Outro fator a se cuidar é que sem um bom entendimento e uma preparação cuidadosa dos dados o resultado obtido pode ficar ininteligível, ou seja, para o caso específico de dados sísmicos, a escolha equivocada dos atributos sísmicos pode levar a conclusões errôneas.

O agrupamento do SOM pode ser realizado utilizando qualquer método tradicional, aglomerativo ou partitivo. Outras alternativas podem ser também utilizadas, entre elas, a de segmentar a matriz-U utilizando algoritmos de processamento de imagem (Costa, 1999). Especificamente, neste trabalho será utilizado um algoritmo de clusterização partitivo: o K-means.

Através da visualização do SOM pode-se arbitrar e verificar a escolha do número de grupos existente no dado. Entretanto, com o objetivo de auxiliar e automatizar o processo de classificação optou-se por utilizar o índice de Davies-Bouldin (Davies e Bouldin, 1979) como forma de avaliar o particionamento do K-means. De acordo com índice de validação de Davies-Bouldin, o melhor agrupamento minimiza a eq.(4):

$$\frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{cc}(Q_k, Q_l)} \right\} \quad (4)$$

onde C é o número de grupos, S_c é a distância entre as amostras dos grupos Q_k e Q_l e d_{cc} a distância entre os centróides dos respectivos grupos. Portanto, valores do índice de Davies e Bouldin menores que um representam agrupamentos separados, enquanto valores maiores que um representam grupos possivelmente cruzados.

5 Agrupamento do SOM aplicado à análise de fácies sísmicas

Baseado na clusterização do SOM, propõe-se para análise não-supervisionada de padrões sísmicos o seguinte algoritmo:

a) Segmentação espacial e temporal com orientação geológica;

b) Seleção dos atributos sísmicos que serão utilizados como dados de entrada da análise;

c) Formação e treinamento do SOM com número de vetores protótipos muito maior que o número esperado de fácies sísmicas;

d) Utilizando a visualização da U-matrix do SOM em comparação com o índice de Davies-Bouldin para diferentes números de agrupamentos é estimado o número de fácies sísmicas;

e) Clusterização e rotulação dos vetores protótipos do SOM utilizando o algoritmo partitivo K-means;

f) Após os elementos do SOM serem rotulados com o número de fácies sísmicas estimada, os atributos sísmicos para cada ponto do reticulado de entrada são comparados com os vetores protótipos do SOM e então classificados de acordo com a fácies do agrupamento que pertencer o vetor protótipo mais próximo.

g) Construção e Interpretação dos mapas de fácies sísmicas.

Uma grande contribuição desta metodologia está no auxílio na escolha do número de fácies sísmicas, normalmente, arbitrada de forma empírica (Johann, 1999).

5.1 Análise de fácies sísmicas de um dado sintético.

Para ilustrar o procedimento proposto um dado sísmico 3D sintético convolucional, composto por apenas duas linhas sísmicas foi gerado a partir do modelo da figura 1.

Pode-se verificar que, na região onde ficaria localizado o suposto reservatório de petróleo, o referido modelo possui três diferentes fácies sísmicas, caracterizadas pelas velocidades de propagação de 3100 m/s, 3200 m/s e 3300 m/s (Santos, 1997).

Para esta análise foram utilizados como atributos sísmicos as amplitudes em torno do reservatório em questão (Coulèou et al, 2003), i.e., a região marcada em preto na figura 2a. A utilização das amplitudes instantâneas como atributos de entrada equivale a classificação das formas de onda na região de interesse. Tal procedimento utilizando outros algoritmos de classificação de fácies é diretamente aplicado por alguns softwares comerciais (Stratimagic, 2001). A análise com o algoritmo proposto, encontra-se na figura 2. Verifica-se para este caso em especial, que

os três grupos ou fácies são facilmente identificados pela matriz-U e o resultado da classificação foi excelente. Entretanto, o índice mínimo de Davies-Bouldin 5 não correspondeu ao número de fácies existente. Portanto, a escolha do número de fácies deve sempre que possível ser feita de forma semi-automática, ou seja, deve-se confirmar a estimativa do número de fácies após a visualização da matriz-U.

A mesma análise foi realizada acrescentando-se ruído gaussiano ao modelo sísmico sintético. O resultado ilustrado na figura 3 confirma novamente o bom resultado obtido utilizando-se a metodologia proposta e a importância da visualização da matriz-U para a estimativa do número de fácies sísmicas.

Entretanto, ao simular uma interpretação ruidosa dos horizontes que delimitam a região de interesse, a análise de fácies sísmicas utilizando a metodologia proposta, obteve um resultado muito ruim, conforme ilustrado na figura 4. Tal fato se deve ao atributo sísmico utilizado, amplitude instantânea, ser sensível a deslocamentos no tempo (Rankel&Mitchel, 2003). Logo, a escolha dos atributos sísmicos para a classificação de padrões sísmicos é fundamental para a obtenção de resultados coerentes.

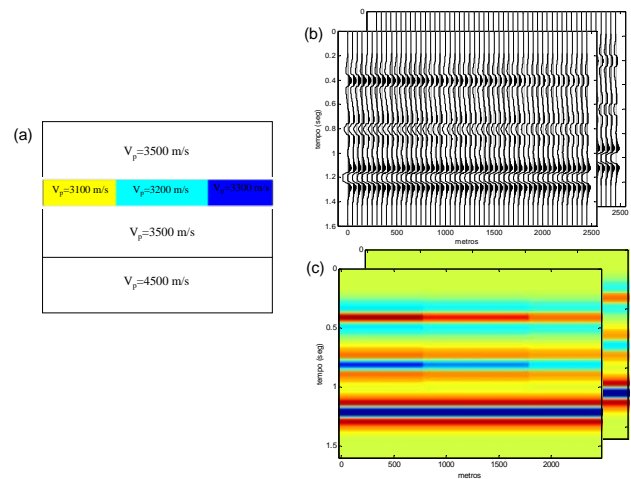


Figura 1: (a) Modelo esquemático com as velocidades de propagação de cada camada; (b) Sinal sísmico sintético gerado a partir do modelo de velocidades; (c) O mesmo sinal sísmico representado como uma imagem.

5.2 Análise de fácies sísmicas de um dado real.

A metodologia proposta para análise de fácies sísmicas foi aplicada em um dado sísmico 3D real da Bacia de Campos. Nove amostras da unidade estratigráfica superior do reservatório (Johann, 1999) foram utilizadas como atributo de entrada. Como a variação de espessura do reservatório é bastante grande, para evitar a utilização de amostras fora da área de interesse, foram geradas nove amostras interpoladas dentro da referida unidade estratigráfica.

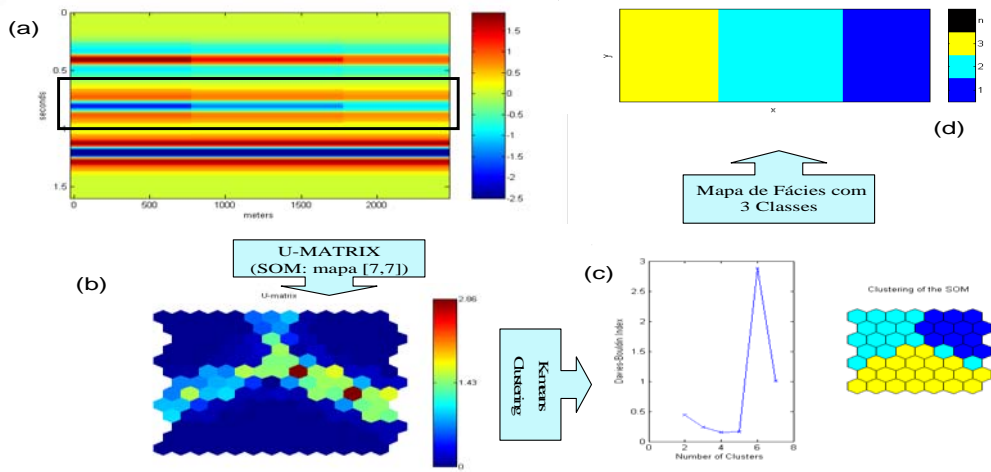


Figura 2: (a) Segmentação temporal do sinal sísmico; (b) Matriz-U formada a partir do SOM com dimensões [7,7]; (c) Análise dos possíveis grupos formados pelo SOM via índice de Davies-Bouldin; (d) Mapa de fácies sísmicas gerado.

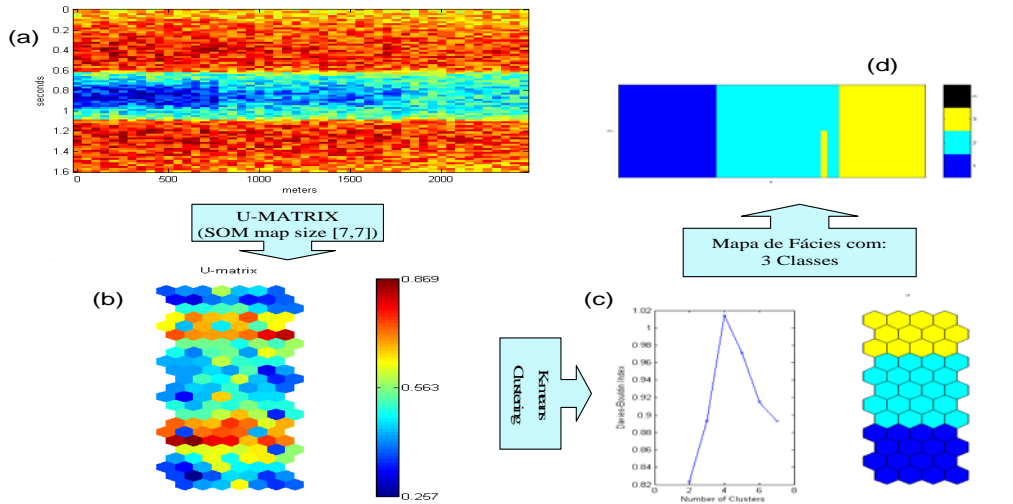


Figura 3: (a) Segmentação temporal do sinal sísmico com ruído aditivo; (b) Matriz-U formada a partir do SOM com dimensões [7,7]; (c) Análise dos possíveis grupos formados pelo SOM via índice de Davies-Bouldin; (d) Mapa de fácies sísmicas gerado.

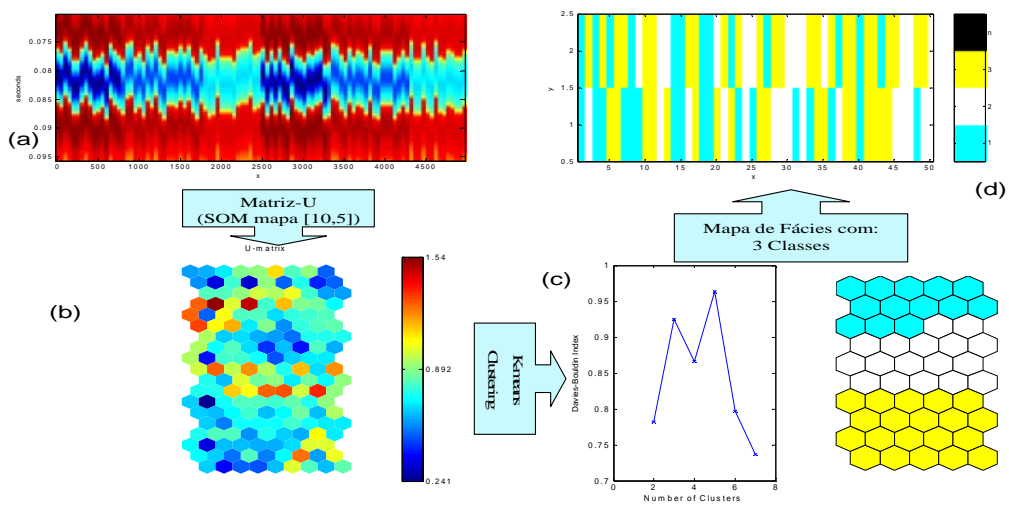


Figura 4: (a) Segmentação temporal do sinal sísmico com ruído de interpretação; (b) Matriz-U formada a partir do SOM com dimensões [10,5]; (c) Análise dos possíveis grupos formados pelo SOM via índice de Davies-Bouldin; (d) Mapa de fácies sísmicas gerado.

A análise de fácies sísmicas está ilustrada na figura 5. Verifica-se que não se pode confirmar a existência de grupos bem definidos através da matriz-U, mas sim da existência de um único grupo com possíveis pequenas divisões, o que significa que não existe uma divisão clara entre as formas de onda preponderantes no dado real analisado. Estimando-se o número de fácies através do índice mínimo de Davies-Bouldin, observa-se que o resultado do agrupamento obtido com seis fácies sísmicas é aceitável, mas não totalmente justificável. Sabe-se da análise petrofísica (Johann, 1999) do dado em questão, que o número real de fácies é quatro. O respectivo mapa com este número de fácies também é mostrado na figura 5. Neste caso, em especial, a escolha do atributo amplitude instantânea não se mostrou como a mais adequada para a identificação das fácies sísmicas. Provavelmente, por causa da grande variação das formas de onda na região de interesse, visto se tratar de um dado originário de um sistema turbidítico bastante complexo.

6 Conclusão

A metodologia proposta para análise de fácies sísmicas se mostrou bastante eficaz. Principalmente, como ferramenta para estimação do número de fácies sísmicas presente no dado. Entretanto, a preparação e seleção dos atributos sísmicos devem ser realizadas de forma bastante criteriosa, refletindo corretamente a natureza da característica que se deseja analisar e identificar (Matos et al., 2003) e (Matos et al., 2004).

Agradecimentos

Os autores agradecem a ANP pela permissão do uso dos dados sísmicos e à Petrobras pelo suporte financeiro através do PRAVAP 19.

Referências Bibliográficas

- Coléou T., Poupon M., and Azbel K. (2003). Interpreter's Corner—Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge*, p. 942, Vol 22, Issue 10.
- Costa, J.A.F. (1999). Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis. Tese de Doutorado. Universidade de Campinas.
- Duda, R.O.; Hart P.O.; Stork D. G. (2001). *Pattern Classification* 2nd ed. John Wiley & Sons, Inc.
- Gurney, K. (1997). *An Introduction to Neural Networks*, UCL Press, 1 Gunpowder Square, London EC4A 3DE, UK.
- Haykin, S. (1999). *Neural Networks a Comprehensive Foundation* 2nd ed. Prentice Hall.
- Johann, P.R.S. (1999). Inversion sismo-stratigraphique et simulations stochastiques: reservoir turbiditique, offshore du Brésil. *Intégration géologique, géophysique et géostatistique en 3-D*, Ph. D. Thesis, Université de Paris VI.
- Matos, M.C., Osório, P.L.M., Johann, P.S. (2003). Using Wavelet Transform and Self Organizing Maps for Seismic Reservoir Characterization of a Deep-Water Field, Campos Basin, Brazil. 65th Mtg: EAGE.
- Matos, M.C., Osório, P.L.M., Johann, P.S. (2004). Unsupervised Seismic Facies Classification Using Matching Pursuit and Self Organizing Maps; 66th EAGE (European Association of Geoscientists & Engineers) Conference & Exhibition, Paris-FR;
- Mitchum, R.M. (1977). Glossary of Terms Used in Seismic Stratigraphy. *AAPG Memoir* 26. P. 205-212.
- Morice, M., Keskes, N., Jean, F. (1996). Manual and automatic seismic facies on SISMAGE workstation: Un example of image analysis technique applied to the exploration of Perdana-Nakhoda turbidic fan complex (Pliocene offshore Brunei), IMA3G research project, Elf aquitaine production. SEG Meeting abstracts.
- Rankey E. C., Mitchell J. C. (2003). Interpreter's Corner—That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis. *The Leading Edge*, p. 820; Vol 22, Issue 9.
- Santos, M.S. (1997). Caracterização de Reservatórios via Redes Neurais. Dissertação de Mestrado, UFBA.
- STRATIMAGIC Basic Training Course. (2001). *Paradigm Geophysical*.
- Ultsch, A. (1993). Knowledge Extraction from Self-Organizing Neural Networks. In: Opitz et al. Ed. *Information and Classification*. Springer-Verlag, Berlin.
- Vesanto, J., Alhoniemi, E. (2000). Clustering of the Self-Organizing Map: *IEEE Transactions on Neural Networks*, Volume 11, Number 3, pp. 586-600.
- Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. (1999). Self-Organizing Map in Matlab: The SOM Toolbox. *Proceedings of the Matlab DSP Conference*. Espoo, Finland.

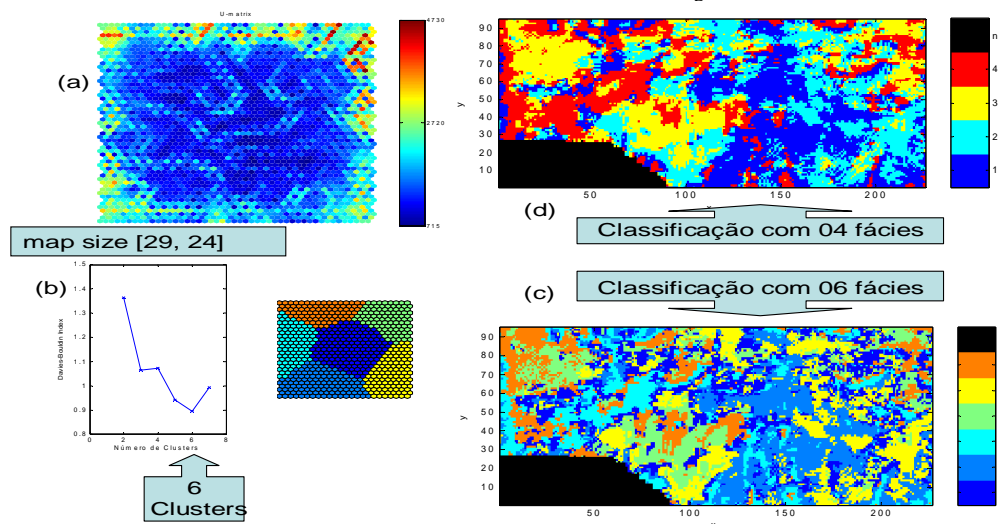


Figura 5: Análise de fácies sísmicas utilizando dados reais