

# Agrupamento e visualização de dados sísmicos através de quantização vetorial

Ernesto Marchioni Fleck, Carlos Eduardo Pedreira e Rogério Santos

**Abstract**—This work suggests the use of a new method of seismic data clustering that can aid in the visualization of seismic maps. Seismic data have asymmetric distributions. They are traditionally classified by methods that lead the proposed groups' references to their mean values. The mean value is, however, sensitive to noise and outliers and the classification methods that make use of this estimator are, consequently, subjected to generating distorted results. Although other works have suggested the use of the median in cases where the distributions are asymmetric – due to the fact that the estimator is robust with respect to noise and outliers – none have proposed a method that would lead the groups' references to the median while treating seismic data.

**Palavras chaves**—Agrupamento de dados, mapas sísmicos, mediana, quantização vetorial, tratamento de distribuições assimétricas.

## I. INTRODUÇÃO

A determinação da dimensão de reservatórios de petróleo é um importante desafio para a indústria de hidrocarbonetos. Em geral, a estrutura destes reservatórios é geologicamente complexa devido, entre outros fatores, às descontinuidades das estruturas que unem as áreas de grande concentração de petróleo. Os maiores desafios acontecem quando ocorrem estruturas de pequenas espessuras [1]. Tradicionalmente, uma ferramenta essencial para o dimensionamento dos reservatórios são os mapas sísmicos que realçam as formas geológicas encontradas. Com o aumento da potencialidade dos recursos computacionais, a evolução dos mapas de dados sísmicos adquiriu novas possibilidades e pode ser resumida da seguinte forma: (a) Mapas que evidenciam, em cores, contrastes de faixas de valores dos dados sísmicos atribuídas pelo geofísico. Nestes casos, os mapas dependem do conhecimento do intérprete que determina faixas associadas a cores para a definição de estruturas. O sucesso da pesquisa depende de considerável tempo de dedicação de um profissional altamente qualificado; (b) Mapas que mostram

formas geológicas resultantes de processos de agrupamento de dados sísmicos, realizados através de técnicas de inteligência computacional.

A opção b tem mostrado melhores resultados [2]. No entanto, os grupos formados por vezes não atendem satisfatoriamente às necessidades de visualização das formas geológicas mais complexas desejadas pelos geofísicos. O entendimento da não satisfação está relacionado às características dos dados sísmicos pouco consideradas nas suas classificações pelos métodos existentes [3] [4] [5]. Em um método de agrupamento de dados sísmicos, espera-se que um tratamento do sinal determine, pelo menos, formas estruturais (falhas, bordas, dunas, diques...) e formas estratigráficas (canais, lobos, leques, barras,...).

A média e a mediana são dois estimadores de posição possíveis. A escolha de um destes estimadores pode influenciar na determinação dos grupos estudados. Normalmente, os dados sísmicos distribuem-se assimetricamente e uma consequência desta assimetria é a desigualdade de valores entre a média e a mediana da distribuição dos dados. A mediana é o centro de simetria de uma distribuição [6] motivo pelo qual, neste artigo, é proposto um método que induz os centros de referências dos grupos à mediana. A aplicação deste método mostrou-se eficaz no agrupamento de dados sísmicos objetivando a visualização em mapas. O método utiliza os conceitos de aprendizagem não supervisionada, particularmente da Quantização Vetorial (QV).

## II. DADO SÍSMICO

No que concerne à aquisição, sinais sísmicos são oriundos de ondas refletidas resultantes, por exemplo, de uma fonte impulsiva, como os explosivos (tipo dinamite) ou, no caso marítimo, disparos de cápsulas de ar comprimido por canhões especiais. Em cada tipo diferente de terreno que a onda incidente encontra, uma outra onda é refletida, captada por um conjunto de receptores e registrada (gravada em meio magnético). O valor do sinal sísmico captado é chamado de amplitude. Os receptores que captam os sinais são distribuídos sobre a superfície do terreno ou sobre a superfície do mar em coordenadas  $x$  (“*in line*”) e  $y$  (“*cross line*”). Uma terceira coordenada  $z$ , que demarca a profundidade, é determinada pela captação dos sinais no tempo. Então, com um período constante, as amplitudes são registradas nas coordenadas  $x$  e  $y$ . O conjunto  $x$ ,  $y$ ,  $z$  e amplitude forma um volume sísmico ou dados sísmicos.

E. M. Fleck Depto. Informática - Tecgraf – PUC-Rio (e-mail: [fleck@tecgraf.puc-rio.br](mailto:fleck@tecgraf.puc-rio.br))

C. E. Pedreira Depto. Engenharia Elétrica – PUC-Rio (e-mail: [carlosp@centroin.com.br](mailto:carlosp@centroin.com.br))

R. Santos Petrobras – E&P (e-mail: [r.santos@petrobras.com.br](mailto:r.santos@petrobras.com.br))

O tratamento dos dados sísmicos engloba itens tais como a adequação dos tempos dos sinais para “perto” das superfícies. Este processo é conhecido como “correção estática” ou simplesmente “estática”. Esta correção determina a adequação dos dados captados às irregularidades das superfícies da crosta terrestre que são analisadas [7] [25].

Os dados sísmicos, aqui tratados, são monovariáveis. A única variável medida na pesquisa sísmica é o sinal captado ou amplitude.

As amplitudes do volume sísmico possuem grande quantidade de ruído com valores próximos às amplitudes naturais e, portanto, de difícil exclusão pelos métodos existente na etapa de tratamento destes dados [8].

Após a coleta e o tratamento dos dados sísmicos, uma etapa de interpretação destes dados é realizada. Uma ferramenta da etapa de interpretação são os mapas. Este artigo, portanto, está inserido na etapa de interpretação dos dados sísmicos.

A base de dados utilizada neste artigo é formada de 772800 registros de amplitude sísmica. Estes dados correspondem a uma pequena parte da bacia de Campos e foram fornecidos pela Petrobras. Os dados possuem dois tipos de variáveis. O primeiro tipo é a amplitude do sinal. O segundo tipo é a posição da amplitude e compreende quatro variáveis: as coordenadas x, y e z e uma variável que indica a camada a que pertence à amplitude. Os dados foram fornecidos em 8 camadas. A variável z determina a distorção das camadas. A figura 1 mostra a distribuição dos dados sísmicos fornecidos pela Petrobrás.

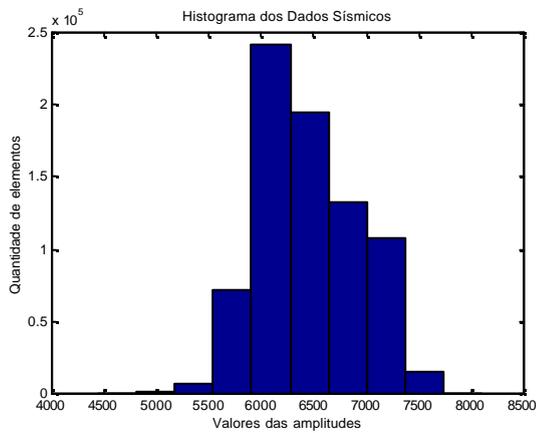


Fig. 1 Distribuição dos dados sísmicos testados

O conjunto dos dados tem as seguintes características:

Quant	Média	Mediana	Desvio
772800	6446.7	6373	465.94

### III. AGRUPAMENTO DE DADOS

O conceito de Quantização Vetorial (QV) [9] [10], vem sendo aplicado com sucesso em diversas áreas [11] [12]. A idéia básica da QV é representar os grupos através de um número finito de vetores, que aqui vamos denominar como protótipos.

Do ponto de vista de classificação, atribuí-se a cada classe um ou mais protótipos utilizando o princípio dos vizinhos mais próximos [13] [14].

A idéia central é formar  $k$  grupos de tal forma que os elementos dos dados de entrada  $x = \{x_1, x_2, x_3, \dots, x_m\}$  sejam classificados nos grupos  $w = \{w_1, w_2, w_3, \dots, w_k\}$ . Então, cada elemento do conjunto  $x_j$  é alocado em um grupo cuja distância entre o elemento do dado de entrada e o protótipo seja mínimo:

$$w_c = \min_{i=1, \dots, k} (d(x_j, w_i)) \quad (1)$$

onde  $c \in [1, k]$  indica o grupo escolhido

A busca de uma melhor solução para a determinação dos grupos é definida pela localização das referências que é feita através de um processo iterativo que inclui a atualização dos valores das referências  $w$  e o cálculo de um erro. Busca-se, a cada iteração, a minimização do erro em cada grupo “ $j$ ” através, por exemplo, da seguinte expressão:

$$e_j = \left( \sum_{i=1}^{n_j} |x_i - w_j|^2 \right)^{\frac{1}{2}} \quad (2)$$

Os valores de  $w$  são determinados por uma função de custo. Considere a seguinte função:

$$F_{x_j} : \mathfrak{R} \rightarrow \mathfrak{R} \quad w \rightarrow \sum_{j=1}^m \frac{(x_j - w)^2}{2} \quad (3)$$

$$F_x = \sum_{j=1}^m \frac{(x_j - w)^2}{2} \quad (4)$$

Cujas, primeira e segunda derivadas, são dadas por:

$$F'_x(w) = \sum_{j=1}^m (w - x_j) \quad (5)$$

e

$$F''_x(w) = m \quad (6)$$

Por (6) podemos dizer que  $F$  é convexa (segunda derivada positiva:  $m > 0$ ) e, portanto, tem um ponto de mínimo local no ponto onde a primeira derivada é zero. O posicionamento de  $w$  no ponto mínimo de  $F$  é encontrado para a primeira derivada nula. Neste caso tem-se então que:

$$F'_x(w) = \sum_{j=1}^m (w - x_j) = 0 \Rightarrow$$

$$mw - \sum_{j=1}^m x_j = 0 \Rightarrow w = \frac{1}{m} \sum_{j=1}^m x_j$$

Então, o valor para  $w$  que coincide com a média do grupo proposto é o ponto que, para uma distribuição de dados, implica na obtenção de um erro mínimo e de um ponto de mínimo local da função  $F$ .

#### A. Algoritmos de indução à média

Os algoritmos que utilizam a função de custo  $F$  são ditos algoritmos que induzem  $w$  à média. O Gradiente Descendente é uma técnica freqüentemente utilizada para a determinação de  $w$  utilizando a função  $F$  através da equação:

$$w_c^{t+1} = w_c^t + \mathbf{h}(X^t - w_c^t) \quad (7)$$

onde  $t$  indica a iteração do processo e

$$X \equiv \underbrace{\{x_1, x_2, \dots, x_m\}}_{\text{época 1}}, \underbrace{\{x_1, x_2, \dots, x_m\}}_{\text{época 2}}, \dots, \underbrace{\{x_1, x_2, \dots, x_m\}}_{\text{época final}}$$

Os algoritmos que induzem  $w$  à média determinam bons resultados para dados com distribuições simétricas. As equações (1), (2) e (7) são à base de vários algoritmos que utilizam funções de custo para agrupamento orientados à média, dentre eles o SOM de Kohonen [15] [16] [17]. Este método é freqüentemente utilizado para agrupamento de dados sísmicos [18] [19] [20]. No entanto, os métodos de agrupamento que tratam distribuições assimétricas (caso dos dados sísmicos) são evidentemente menos eficientes se utilizam médias como estimador do centro dos grupos formados [21] [22].

#### B. Algoritmos de indução à mediana

Na literatura, são encontrados algoritmos que atualizam as referências  $w$  através do cálculo dos valores da mediana do grupo proposto. Por exemplo, Pitas et al. [21] propõem um modelo QV com uma atualização de  $w$  baseada no cálculo da mediana. O processo pode ser resumido da seguinte maneira: (1) Seja  $Z_i^t$  um vetor com os elementos já selecionados do grupo  $i$  na iteração  $t$ ; (2) Seja  $x$  um elemento do conjunto de origem  $X$  que está sendo classificado no grupo  $Z_i^{t+1}$  na iteração  $t+1$ ; (3) Então, a atualização de  $w$  é realizada obedecendo o seguinte critério:  $w^{t+1} = \text{mediana}\{x \cup Z_i^t\}$ . Kohonen e Somervuo [23] e Georgakis et al. [24] elaboraram propostas a partir da idéia de Pitas et al. [21] sugerindo diferentes formas de atualizar os conjuntos vizinhos ao  $w$  vencedor.

#### C. Algoritmo proposto

O algoritmo proposto não calcula a mediana; ele induz as referências dos grupos à mediana. Duas vantagens são verificadas na utilização do algoritmo proposto comparativamente aos existentes na literatura. A primeira concerne o menor tempo de processo. A segunda vantagem está relacionada a melhores classificações avaliadas através dos erros quadráticos.

Considere a inclusão de uma função  $\mathbf{b}$  em (7):

$$w^{t+1}(x) = w^t(x) + \mathbf{h}\mathbf{b}^t(X^t - w^t) \quad (8)$$

$$\mathbf{b} \equiv 1 - \frac{c(2\mathbf{s}^2 - (X - w)^2)}{\mathbf{s}^4} \exp\left(\frac{-(X - w)^2}{2\mathbf{s}^2}\right) \quad (9)$$

onde,  $c \in \mathfrak{R}$ ,  $c \geq 0$  é uma constante que determina o grau de não linearidade de  $\mathbf{b}$ ,  $w \in \mathfrak{R}$  é a referência de um grupo de dados de entrada  $X$  cujo desvio padrão é  $\mathbf{s} \in \mathfrak{R}$ ,  $\mathbf{s} > 1$ .

Limitando-se os valores de  $1 \leq \mathbf{b} \leq 0$  e valendo-se das propriedades: (a)  $\mathbf{b}$  é uma função simétrica; (b) o valor  $\min \mathbf{b}$  é sempre coincidente com  $w$ ; (c) o valor  $\min \mathbf{b}$  é inversamente proporcional ao valor de  $c$ ; (d) para o valor

$\min \mathbf{b} = 0$  o valor de  $\max c = \frac{\mathbf{s}^2}{2}$ ; é possível mostrar que a equação (8) induz  $w$  à média e à mediana dependendo do valor da  $c$ . Na proposta da equação (8), a indução à média é realizada para  $c = 0$ . Para valores de  $c$  acima de zero e menores do que  $c = \frac{\mathbf{s}^2}{2}$ , dependendo da assimetria da distribuição dos dados,  $w$  é induzido à mediana.

## IV. RESULTADOS

Os resultados aqui apresentados são comparativos. Eles mostram a classificação de um algoritmo usualmente utilizado para a classificação de dados sísmicos (SOM de Kohonen) e o algoritmo proposto no que concerne ao atributo adquirido. Busca-se observar as diferenças entre a classificação orientada à média (Kohonen) e a classificação orientada à mediana (algoritmo proposto).

As classificações foram realizadas em nove grupos, que é uma quantidade que permite a visualização de formas estruturais, estratigráficas e o isolamento de ruídos. O algoritmo de Kohonen utilizou 5000 épocas e o algoritmo proposto 300 épocas.

Para cada processo, junto com os mapas, são calculados os seguintes dados estatísticos dos grupos: valores mínimos e máximos, média, desvio padrão e quantidade de elementos classificadas em cada grupo.

Os mapas foram produzidos em 8 camadas. No entanto, para a comparação dos resultados entre os dois algoritmos

somente uma camada é escolhida em cada mapa. A tabela 1 mostra a classificação dos dados através da aplicação do algoritmo de Kohonen, a tabela 2 refere-se ao algoritmo proposto. A figura 2 mostra a camada 3 do mapa com os grupos propostos pelo algoritmo de Kohonen. A figura 3 mostra a mesma camada 3 classificada pelo algoritmo proposto. As figuras 4 e 5 mostram os mapas da camada 1 com uma área assinalada evidenciada nas figuras 6 e 7.

TABELA I  
CLASSIFICAÇÃO DO ALGORITMO DE KOHONEN (MÉDIA)

	Quant	Média	Desvio	Valor mínimo	Valor máximo
G 1	127670	5815	176.16	4445	6002
G 2	69484	6057	30.42	6002	6107
G 3	68987	6318	31.07	6266	6374
G 4	120973	6185	45.31	6107	6266
G 5	21931	6392	11.00	6374	6412
G 6	83750	6713	60.17	6612	6821
G 7	97157	6509	58.18	6412	6612
G 8	43286	6886	38.24	6821	6952
G 9	139562	7115	146.48	6952	8097

A visualização no mapa mostra que a área ocupada pelo grupo 1 (Kohonen) é mais detalhada no algoritmo proposto.

TABELA II  
CLASSIFICAÇÃO DO ALGORITMO PROPOSTO (MEDIANA)

	Quant	Média	Desvio	Valor mínimo	Valor máximo
G 1	5610	5255	185.22	4445	5460
G 2	122796	6487	71.47	6370	6615
G 3	84342	6718	60.84	6615	6827
G 4	152241	5984	83.69	5809	6111
G 5	184806	6234	73.76	6111	6370
G 6	69129	6932	61.31	6827	7038
G 7	42281	5695	86.87	5460	5809
G 8	92034	7174	81.64	7038	7340
G 9	19561	7436	80.29	7340	8097

O agrupamento resultante da aplicação do algoritmo de Kohonen apresenta dois grandes grupos (grupos 1 e 9) com grandes quantidades de elementos e com grandes desvios padrões. A busca da minimização dos desvios nos grupos determina o aparecimento destes grandes grupos com grandes desvios. Tal comportamento pode ter sido ocasionado pela existência de ruídos nos dados. Com efeito, observa-se que os grupos 1 e 2 do algoritmo de Kohonen (orientação à média) têm a mesma extensão de valores de amplitudes sísmicas dos grupos 1, 4 e 7 do algoritmo de orientação à mediana. A quantidade de elementos é similar: 197154 nos grupos 1 e 2 contra 200132 dos grupos 1, 4 e 7. A classificação orientada à mediana mostra o grupo 1 com pequeno número de elementos e grande desvio, sugerindo um isolamento de ruídos de baixa amplitude sísmica. E ainda, os outros grupos possuem desvios que podem ser considerados com valores homogêneos.

A figura 2 mostra um pacote de informações sísmicas, aqui denominada camada 3, onde também este detalhamento pode ser verificado. Outra observação que reforça a idéia do isolamento de ruídos no grupo 1 do algoritmo proposto é a nitidez das estruturas geológicas nos mapas.

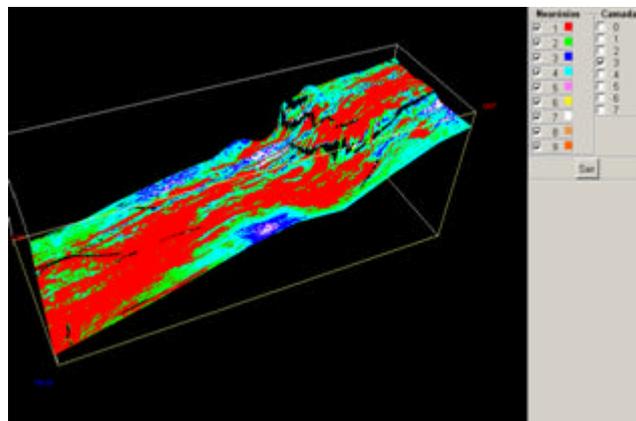


Fig. 2 Mapa resultante do algoritmo de Kohonen - camada 3

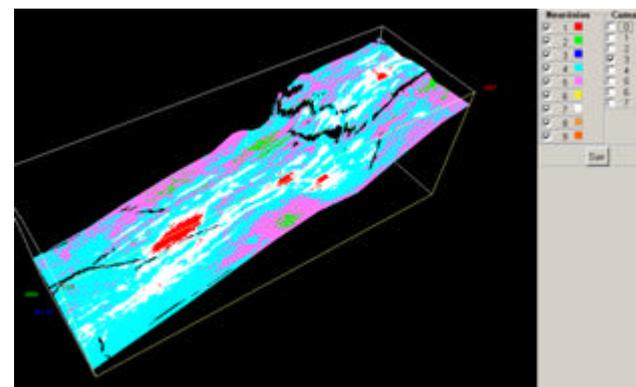


Fig. 3 Mapa resultante do algoritmo proposto - camada 3

As figuras 4 e 5 mostram a camada 1 das classificações realizadas.

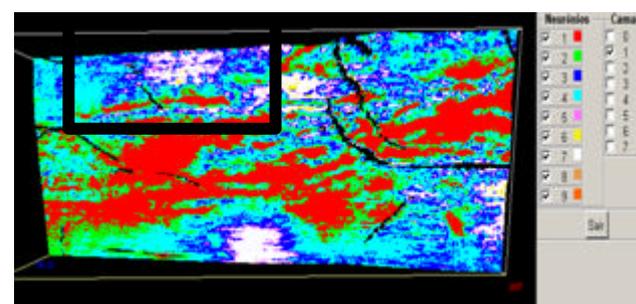


Fig. 4 Mapa resultante do algoritmo de Kohonen - camada 1

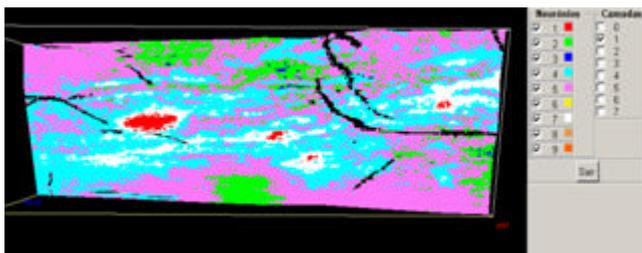


Fig. 5 Mapa resultante do algoritmo proposto -camada 1

Além do detalhamento verificado na camada 3, uma área da camada 1 é selecionada (notar um retângulo na parte superior esquerda do mapa) e observada separadamente através de uma ampliação nas figuras 6 e 7. As formas observadas na figura 7 são mais nítidas e bem formadas do que as formas observadas na figura 6 originada do algoritmo de Kohonen.

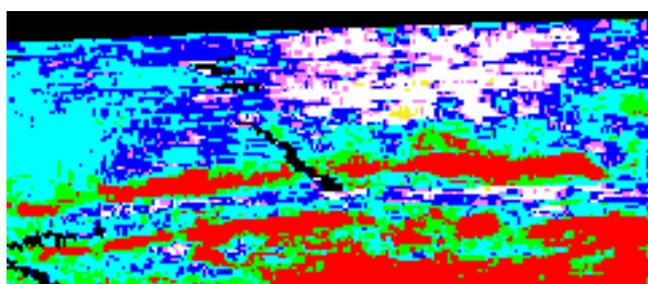


Fig. 6 Área realçada no mapa do algoritmo de Kohonen



Fig. 7 Área realçada no mapa do algoritmo proposto

## V. CONCLUSÃO

O método proposto gera soluções sísmico-geológicas (formas geológicas a partir dos dados sísmicos) onde são possíveis a visualização e o dimensionamento de reservatórios de petróleo.

Duas são as principais contribuições deste artigo. A primeira é uma nova visão de agrupamento para dados sísmicos, que tem distribuição de dados assimétrica. As distribuições assimétricas são melhores agrupadas por processos que induzam as referências dos grupos à mediana.

A segunda contribuição encerra uma proposta de algoritmo para o agrupamento de distribuições assimétricas. Para tal, o algoritmo faz com que as referências dos grupos sejam induzidas à mediana.

Dentre os desafios para trabalhos futuros está o de adequar a função  $b$  para  $\mathfrak{R}^n$ . Esta tarefa é complexa dada a existência de múltiplas definições para mediana no  $\mathfrak{R}^n$ .

A determinação da dimensão de reservatórios de petróleo é um importante desafio para a indústria de hidrocarbonetos e a utilização deste método tem mostrado importantes resultados em testes com dados reais.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Kubichek, R., Quincy, E., Statistical Modeling and Feature Selection for Seismic Pattern Recognition, Pattern Recognition, vol 18, no. 6 pp 441-448, 1985.
- [2] Aminzadeh, F., Barhen, J., Glover, C. W., Toomarian, N. B., Reservoir parameter estimation using a hybrid neural network, Computers & Geosciences 26, 2000, 869-875.
- [3] Muñoz, A., Muruzábal, J., Self-organizing maps for outliers detection, Neurocomputing 18 (1998) 33-60.
- [4] Wong, P., Gedeon, T., Taggart, J., An Improved Technique in Porosity Prediction: A Neural Network Approach, IEEE transaction On Geoscience and Remote Sensing, vol. 33, no. 4, July (1995) 971-980.
- [5] Strelitz, R., Keshet, Y., Integral Transform, Data Compression, and Automatic Analysis of Seismic Sections, IEEE transaction on Geoscience and Remote Sensing, vol. 28, no.6 November 1990, 982-991.
- [6] Small, C. G., A Survey of Multidimensional Medians, International Statistical Review (1990), 58, 3, 263-277.
- [7] Yilmaz, Ö., "Seismic Data Processing", Society of Exploration Geophysicist, Tulsa, 1987.
- [8] Houlding, S. W., "3-D Geoscience Modeling Computer Techniques for Geological Characterization", Springer-Verlag, Berlin, 1994.
- [9] Linde, Y., Buzo, A., Gray, R.M., An Algorithm for Vector Quantizer Design, IEEE Transaction on Communications, Vol com-28, no. 1, January 1980.
- [10] Gersho, "Asymptotically Optimal Block Quantization", IEEE trans. on information theory IT-25 pp 373-380, 1979.
- [11] Saito, K., Nakano, R., Extracting regression rules from neural networks, Neural Network, vol. 15, I 10, December 2002, 1279-1288.
- [12] Rizzo, F., Storer, J. A., Carpentieri, B., Overlap and channel errors in Adaptive Vector Quantization for image coding, Information Sciences, vol. 171, I 1-3, March 2005, 125-143.
- [13] Bishop, C. M., Neural Networks for Pattern Recognition, Oxford University Press, New York, 1995.
- [14] Cover, T. "Estimation by the Nearest Neighbor Rule" IEEE Transaction on Information Theory, vol. IT-14, no. 1, January 1968.
- [15] Chakraborty, B., Kaustubha, R., Hegde, A., Pereira, A., Acoustic Seafloor Sediment Classification Using Self-Organizing Maps, IEEE Transaction on Geoscience and Remote Sensing, vol. 39, no. 12, December 2001, 2722-2725.
- [16] Kangas, J., Kohonen, T., Laaksonen, Variants of Self-Organizing Maps, IEEE transaction on Neural Networks, vol. 1 no. 1 March 1990, 93-99.
- [17] Kohonen, T., "Self-Organizing Maps", Springer, Third Edition, 2001.
- [18] Essenreiter, R., Karrenbach, M., Treitel, S., Identification and classification of multiple reflections with self-organizing maps, Geophysical Prospecting, 49 (2001) 341-352.
- [19] Strecker, U., Uden, R., Data mining of 3D poststack seismic attribute volumes using Kohonen self-organizing maps, The Leading Edge, October, (2002), 1032-1037.
- [20] Tarvainen, M., Recognizing explosion sites with a self-organizing network for unsupervised learning, Physics of the Earth and Planetary Interiors 113 (1999) 143-154.

- [21] Pitas, I., Kotropoulos, C., Nikolaidis, N., Yang, R., Gobbouj, M., Order Statistics Learning Vector Quantizer, IEEE Transaction on Image Processing, vol. 5, no. 6, June 1996.
- [22] Jacobs, D., Weinshall, D., Gdalyahu, Y., Classification with Nonmetric Distances: Image Retrieval and Class Representation, IEEE trans. on Pattern Analysis and Machine Intelligence, vol 22, no. 6, June 2000, 583-600.
- [23] Kohonen, T., Somervuo, P., How to make self-organizing maps for nonvectorial data, Neural Network, 15 (2002) 945-952.
- [24] Georgakis, A., Kotropoulos, C., Xafopoulos, Pitas, I., Marginal median SOM for document organization and retrieval, Neural Networks, 17(2004) 365-377.
- [25] Gadallah, M. R., "Reservoir Seismology – Geophysics in Nontechnical Language", PennWell Books, Tulsa, 1994.