

MAPAS AUTO-ORGANIZÁVEIS APLICADOS AO MAPEAMENTO DO POTENCIAL MINERAL NA REGIÃO DE SERRA LESTE, PROVÍNCIA MINERAL DE CARAJÁS, PARÁ

Emilson Pereira Leite e Carlos Roberto de Souza Filho

Recebido em 4 junho, 2009 / Aceito em 26 abril, 2010
Received on June 4, 2009 / Accepted on April 26, 2010

ABSTRACT. A Self-Organizing Map (SOM) was designed with the aim of integrating and searching for patterns in airborne geological and geophysical gamma-spectrometric and magnetic data of the Serra Leste region, Carajás Mineral Province. SOM is an unsupervised Artificial Neural Network method that performs a non-linear mapping from a high-dimensional data space to a 2-dimensional grid, whereas preserving the topological relations in the original data. The SOM grid can be efficiently used in an integrated visualization and understanding of the internal relationships in the data. The K-means algorithm is applied to the SOM grid to reduce the number of mapped patterns so as to facilitate interpretation. Unfolding of the clustered SOM grid associates each mapped pattern with the spatial position of each data point. The SOM reclassified map was compared with a classified map obtained with the Fuzzy C-means method for the same input data and with the same number of classes. The results show the potentiality of SOM in producing higher quality integrated maps to support mineral exploration.

Keywords: Carajás Mineral Province, airborne geophysics, self-organizing maps, fuzzy clustering, mineral potential mapping.

RESUMO. Um Mapa Auto-Organizável (SOM – *Self-Organizing Map*) foi elaborado com o objetivo de integrar e procurar por padrões em dados geológicos e geofísicos (gamaespectrométricos e magnéticos) da região de Serra Leste, Província Mineral de Carajás. O SOM é um método de Rede Neural Artificial não-supervisionada que realiza um mapeamento não-linear de um espaço de dados multidimensionais para um arranjo bidimensional, preservando as relações topológicas dos dados originais durante o processo. O algoritmo *K-means* foi aplicado ao mapa produzido pelo SOM para reduzir o número de padrões mapeados de forma a facilitar a interpretação. O mapa reclassificado foi associado à posição geográfica de cada ponto na área de estudo e comparada a outro mapa classificado pelo método *Fuzzy C-means*, obtido com os mesmos dados de entrada e com o mesmo número de classes. Os resultados mostram a potencialidade do SOM para a produção de mapas integrados de maior qualidade a fim de dar suporte à exploração mineral.

Palavras-chave: Província Mineral de Carajás, aerogeofísica, mapas auto-organizáveis, agrupamento *Fuzzy*, mapeamento do potencial mineral.

INTRODUÇÃO

A formação e localização de depósitos minerais são o resultado da interação entre diversos processos físicos e químicos que deixam assinaturas na forma de feições geológicas associadas aos depósitos. Tais feições são caracterizadas por respostas típicas em um ou mais conjuntos de dados (e.g. geofísicos, multiespectrais, geológicos). A relação entre esses diferentes tipos de dados e depósitos minerais conhecidos pode ser extremamente complexa para ser adequadamente modelada por aproximações lineares. Além disso, os dados podem ser esparsos, descontínuos e formarem distribuições não-normais, tornando ineficiente a aplicação de métodos de estatística multivariada tradicionais.

Os Mapas Auto-organizáveis (SOM – *Self-Organizing Maps*; Kohonen, 2001) e outros métodos de Redes Neurais Artificiais (e.g. Leite & Souza Filho, 2009a,b) são mais apropriados nestes casos, sendo que SOM tem a vantagem de não requerer dados supervisionados (e.g. localização de depósitos e ocorrências minerais conhecidas) na fase de treinamento da rede. Este método tem sido utilizado para análise e detecção de padrões em dados multidimensionais, tais como geoquímicos (e.g. Penn, 2004); multi-espectrais (e.g. Saveliev & Dobrinin, 1999); hiperespectrais (e.g. Penn, 2002); geofísicos (e.g. Klose, 2006), entre outros.

O objetivo deste trabalho é demonstrar a potencialidade da utilização do SOM na interpretação conjunta de dados aerogeofísicos e geológicos, de modo a fornecer elementos adicionais para identificação de padrões associados às mineralizações presentes na região de Serra Leste, Província Mineral de Carajás (PMC), norte do Brasil (Fig. 1). Por se tratar de um método pouco utilizado para este tipo de aplicação, a comparação com métodos mais populares de reconhecimento de padrões aplicados a dados geocientíficos torna-se necessária. Neste trabalho, os resultados de SOM foram comparados com aqueles obtidos por um método de classificação de dados baseado em lógica nebulosa e agrupamento nebuloso; o *Fuzzy C-means* (FCM). Os métodos de agrupamento nebuloso também se enquadram entre os métodos de classificação não-supervisionados e são frequentemente utilizados como ferramentas para o mapeamento do potencial mineral, entre outras aplicações em Geociências (e.g. Demicco & Klir, 2003; Mansour et al., 2009).

A área de estudo está localizada entre os paralelos 634.000 E e 666.000 E e os meridianos 9.325.000 N e 9.347.400 N, Zona 22S, e representa uma área chave na PMC devido à diversidade de depósitos minerais ali presentes (Figs. 1 e 2). Compreende diversos tipos de depósitos relevantes, incluindo: (i) o depósito de Serra Pelada (Au-Pd-Pt), hospedado nos metassedimentos do Grupo Rio Fresco, Formação Águas Claras (~ 2.68 Ga; Cabral et

al., 2002a,b); (ii) depósitos de Cu-Mo-Au (e.g., Serra Verde; Villas & Santos, 2001) hospedado em rochas meta-vulcânicas do Grupo Rio Novo, próximas ao contato com o Granito Estrela de 2.76 Ga (Marschik et al., 2005); (iii) depósitos de ferro em rochas arqueanas de 2.76 Ga do Supergrupo Itacaiúnas (e.g., Serra Leste); (iv) depósitos de manganês (e.g., Sereno; Coelho & Rodrigues, 1986) hospedado no Grupo Rio Fresco; e depósitos de cromo-EGP (Elementos do Grupo da Platina; e.g., Luanga) associados a complexos máfico-ultramáficos (Suito, 1988).

Embora métodos indiretos tenham sido tema de estudos recentes na PMC para investigação de depósitos de Elementos do Grupo da Platina (EGP) (Souza Filho et al., 2007; Leite & Souza Filho, 2009a,b), este trabalho pretende avançar o conhecimento sobre métodos mais robustos para extração de informações de interesse exploratório aplicáveis a essa região.

DADOS

Os dados aero-gamaespectrométricos foram coletados em 1993 pela empresa GEOMAG, utilizando detectores de cristais de NaI ativados por Tálcio, a 60 m de altitude, ao longo de linhas de voo orientadas na direção E-W, perpendicular ao *trend* de rochas do Grupo Rio Novo. Ao longo dessas mesmas linhas de voo, foram coletados os dados aeromagnéticos pela Geoterrex-Dighem, em 1999, a 120 m de altitude.

Os dados geofísicos foram pré-processados para remoção de dados espúrios e interpolados utilizando o algoritmo de curvatura mínima (Briggs, 1974), formando *grids* de concentração de Th, K, U e do campo magnético total, com espaçamento de 62,5 m (1/4 do espaçamento entre as linhas de voo). O IGRF-1995 (*International Geomagnetic Reference Field*) (IAGA, 1996) foi removido do campo magnético total e os dados foram micro-nivelados para remoção de ruídos remanescentes ao longo das linhas de voo. Posteriormente, foram calculadas as derivadas direcionais do campo magnético anômalo e amplitude do sinal analítico (ASA) (Roest et al., 1992). Os mapas das derivadas direcionais (em x e y) serviram para traçar eixos magnéticos. Um mapa de *buffers* de 200 m foi gerado ao redor desses eixos, até um máximo de 2000 m.

O mapa de litotipos foi utilizado como dado categórico, no qual a sigla respectiva a cada litotipo foi associada a uma célula do *grid*. A partir deste *grid* categórico foi efetuada uma conversão para dados numéricos através da atribuição de graus de pertinência para cada unidade litológica (Tab. 1), visando destacar aquelas mais favoráveis à ocorrência de mineralizações de Cr-EGP (Leite & Souza Filho, 2009a). Com base nos dados da literatura sobre a associação desse tipo de depósito às uni-

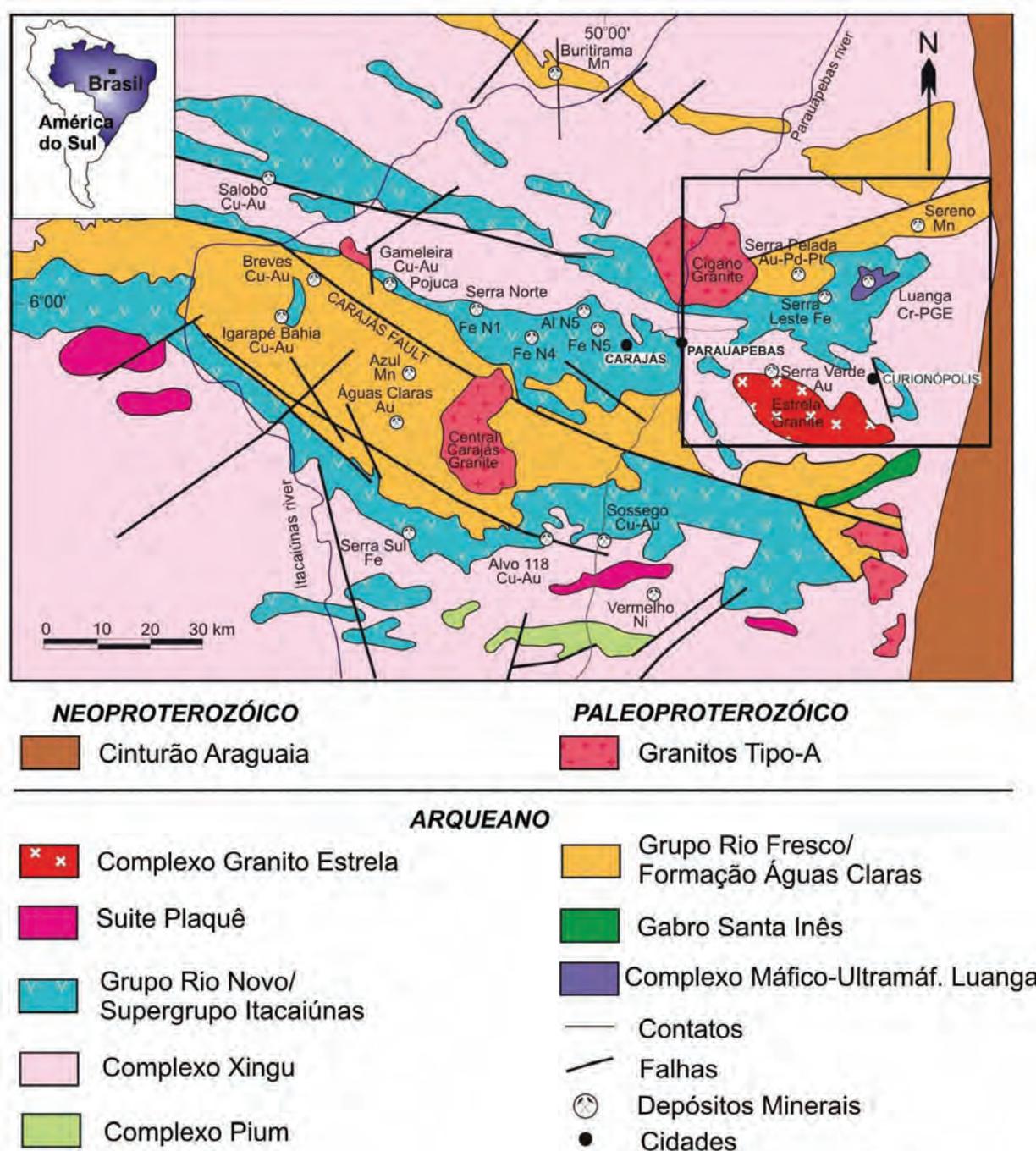


Figura 1 – Mapa geológico da Província Mineral de Carajás (PMC), incluindo a localização dos principais depósitos minerais (adaptado de Docege, 1988). A região de Serra Leste está destacada dentro do retângulo preto na porção leste da PMC.

dades observadas em Serra Leste: (i) as rochas máficas e ultramáficas do Complexo Luanga são consideradas como as mais favoráveis (pertinência 9), pois contém mineralizações conhecidas; (ii) seguidas das rochas meta-vulcano-sedimentares do Grupo Rio Novo (pertinência 75) e meta-sedimentares do Grupo

Rio Fresco (pertinência 50). As coberturas quaternárias foram também assinaladas com pertinência 50 visto que recobrem rochas potencialmente relacionadas à mineralização alvo do estudo. As rochas consideradas menos favoráveis (grau de pertinência 5) estão compreendidas no embasamento granítico Xingu.

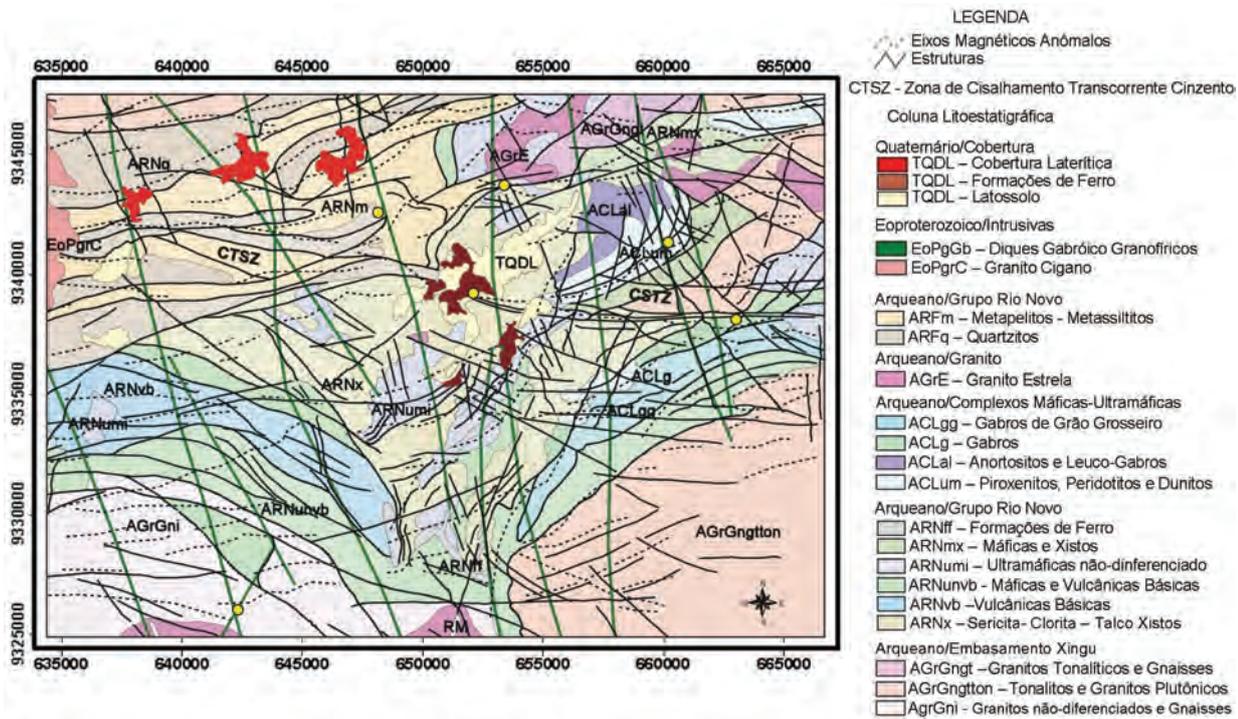


Figura 2 – Mapa geológico da região de Serra Leste (modificado de Veneziani & Okida, 2001). Os círculos em amarelo representam os depósitos minerais.

Tabela 1 – Grau de pertinência fuzzy da litologia para mineralizações do tipo Cr-EGP.

Litologias	Grau de pertinência
ACLal, ACLg, ACLum	95
ARNif, ARNmx, ARNumi, ARNunvb, ARNx	75
ARFm, ARFq, TQDL	50
EoPgrC	25
AGrE, AGrGngt, AGrGngtt, AGrGni	5

Todos os *grids* utilizados são compostos por 511 × 360 células. Cada *grid* contém certa quantidade de células sem dados. A simultaneidade espacial de células contendo dados entre os *grids* foi utilizada como condição *sine qua non* para a realização das classificações. Essa seleção resultou em 167.512 células processadas para cada tipo de dado. Os *grids* são os mesmos daqueles utilizados apresentados em Leite & Souza Filho (2009a) e os eixos magnéticos podem ser visualizados na Figura 2.

MÉTODOS

Mapas Auto-Organizáveis (SOM – Self-Organizing Maps)

O método SOM, também conhecido como Rede de Kohonen (Kohonen, 2001), é uma Rede Neural Artificial (RNA) estruturada em duas camadas. A primeira camada representa os dados de

entrada, definidos na forma de uma matriz

$$X = [x_{k=1}, \dots, x_{k=M}]$$

de tamanho $P \times M$, com $x_k = [x_1 \dots x_P]$. No caso de dados georreferenciados, M é a quantidade de pontos definidos no espaço geográfico e P é a quantidade variáveis observadas em cada ponto. A segunda camada é um arranjo de neurônios, normalmente 2-D, com uma topologia específica (e.g. hexagonal, quadrada, retangular, aleatória) e completamente conectado à primeira camada. Cada neurônio possui um peso sináptico m_i associado, também conhecido como vetor protótipo, de mesma dimensionalidade em relação aos vetores x_k , onde $i = 1, 2, \dots, nu$, sendo nu o número de neurônios da segunda camada do SOM. O principal objetivo de um Mapa Auto-Organizado é o mapeamento não-linear do conjunto de dados de entrada, preservando-se sua estrutura topológica, de forma que padrões adjacentes nos dados sejam associados a neurônios adjacentes no arranjo do SOM.

A imagem de um vetor x_k de dados de entrada na matriz SOM é o vetor m_b que mais se aproxima de x_k . Este vetor é chamado de unidade de melhor ajuste (*Best Matching Unit – BMU*) ou unidade vencedora. Usando uma medida de dissimilaridade $d(x_k, m_i)$, como por exemplo a distância Euclidiana, esta uni-

dade é representada pelo índice

$$b_k = \arg \min_i \{d(\mathbf{x}_k, \mathbf{m}_i)\}. \quad (1)$$

No algoritmo tradicional, empregado neste trabalho, cada vetor protótipo é atualizado iterativamente segundo a expressão

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t) \cdot h_{b_k}^{(i)}(t) \cdot [\mathbf{x}_k - \mathbf{m}_i(t)], \quad (2)$$

onde t é a iteração; onde $\alpha(t)$ é a razão de aprendizado e $h_{b_k}^{(i)}(t)$ representa a função núcleo de vizinhança em torno da BMU. Neste trabalho, $h_{b_k}^{(i)}(t)$ foi definida por uma função gaussiana

$$h_{b_k}^{(i)}(t) = \exp\left(-\frac{(d_{b_k}^{(i)})^2}{2\sigma^2}\right), \quad (3)$$

na qual σ é o raio da vizinhança e $d_{b_k}^{(i)} = d(\mathbf{m}_{b_k}, \mathbf{m}_i)$ é a distância entre as unidades b e i do SOM. A função $h_{bi}(t)$ define a região de influência que a amostra de entrada tem sobre o SOM. $\alpha(t)$ decresce durante o processo de treinamento para que ocorra estabilização das BMUs, segundo a equação

$$\alpha(t) = \alpha_0 / \left(1 + 100 \frac{t}{T}\right), \quad (4)$$

onde T é o total de iterações e α_0 é a razão de aprendizado inicial. Assim, como vetores de entrada contendo valores similares são mapeados em posições vizinhas no mapa de saída, SOM tem como característica principal preservar a relação topológica existente nos dados.

Foram analisadas duas medidas quantitativas da qualidade do mapeamento pelo SOM: (1) erro de quantização (qe), que indica a resolução do SOM; e (2) erro topográfico (te), que indica o quanto o SOM preserva a topologia dos dados de entrada. Elas são respectivamente expressas por:

$$qe = \frac{1}{M} \sum_{k=1}^M \|\mathbf{x}_k - \mathbf{m}_{b_k}\| \quad \text{e} \quad te = \frac{1}{M} \sum_{k=1}^M f(\mathbf{x}_k), \quad (5)$$

onde M é a quantidade de vetores de dados \mathbf{x} . A função $f(\mathbf{x}_k)$ é igual a '1' se a primeira e a segunda BMUs de \mathbf{x}_k são adjacentes, e '0' (zero) caso contrário (Kohonen, 2001).

O exato algoritmo SOM utilizado neste trabalho pode ser resumido nos seguintes passos:

- 1) Faça $t = 0$. Para $i = 1, 2, \dots, nu$, atribua à $\mathbf{m}_i(t)$ valores aleatórios situados no intervalo (0,1) e extraídos de uma distribuição uniforme padrão.

- 2) Normalize todos os vetores \mathbf{x}_k para que seus valores se situem no intervalo (0,1).
- 3) Selecione um vetor \mathbf{x}_k qualquer e determine b_k pela Equação (1).
- 4) Calcule $\mathbf{m}_i(t+1)$ utilizando a Equação (2).
- 5) Calcule qe e te através das Equações (5).
- 6) Enquanto $t \leq T$ ou $qe \leq \varepsilon$, sendo ε um número suficientemente pequeno, incremente t e retorne ao passo 3.

Para melhor visualização dos agrupamentos dos nós do SOM após o treinamento, foi utilizada a matriz unificada de distâncias (matriz-U; Ultsch, 1992, 1993). A matriz-U é construída sobre o arranjo de neurônios do SOM. Cada elemento nesta matriz contém o valor médio dos pesos dos neurônios adjacentes e entre eles são inseridos elementos que contêm a distância entre esses pesos. Portanto, para um arranjo SOM de $nu = nx \times ny$ elementos, a matriz-U terá $3 \times nx \times ny - nx - ny$ elementos. As dimensões nx e ny foram determinadas a partir da relação heurística $nu = 5 \times \sqrt{M}$, definida em Vesanto et al. (2000). A razão entre nx e ny foi calculada com base na razão entre os dois maiores valores singulares da matriz de covariância dos dados e nx e ny foram estimados de forma que seu produto fosse o mais próximo possível de nu , sem perda significativa da proporção entre eles. Por este método, dado que $M = 167512$, o tamanho escolhido foi de 60×34 . Desta forma, $i = 1, \dots, 2040$ na Equação 2. A topologia escolhida para as análises foi a hexagonal.

A quantidade de vetores \mathbf{x} associados a cada BMU não foi aqui listada, porém os tamanhos dos hexágonos que compõem a matriz-U são proporcionais a esta quantidade. A dimensão de \mathbf{x} é dada pela quantidade de variáveis geofísicas e geológicas utilizadas. Neste trabalho, como foram utilizadas cinco variáveis geofísicas e uma variável geológica, cada vetor de entrada tem dimensão 6×1 . O algoritmo foi aplicado em duas etapas, sendo $T = 1000$ e $\varepsilon = 10^{-3}$ nas duas. Na primeira, foram atribuídos os valores $\alpha_0 = 0,9$ e $\sigma = \max(nx, ny)/4 = 15$; na segunda, para refinamento do SOM, foram utilizados $\alpha_0 = 0,1$ e $\sigma = 1$.

Com o objetivo de formar agrupamentos a partir das unidades do SOM, foi utilizado o algoritmo *K-means* em sua forma tradicional (e.g. Cios et al., 1998; Seber, 2004), onde o número de agrupamentos K foi definido com auxílio do índice Davies-Bouldin (Davies & Bouldin, 1979).

O processamento dos dados foi efetuado utilizando o *software* MatLab[®] v. 7.0(R14) e a biblioteca SOM Toolbox 2.0[©] (Vesanto et al., 2000). Esta biblioteca está disponível gratuitamente¹

¹<http://www.cis.hut.fi/somtoolbox/about.shtml>

e contém funções para montagem, treinamento do SOM, cálculo da matriz-U, classificação por *K-means*, além de ferramentas diversas para visualização de resultados.

Fuzzy C-means (FCM)

O algoritmo FCM é utilizado para determinar centros de agrupamentos de vetores p-dimensionais de dados e os graus de pertinência de cada vetor em relação a esses agrupamentos (classes). Quanto mais próximo do centro de um agrupamento um determinado dado estiver, maior será seu grau de pertinência a esse agrupamento (Bezdek, 1981). Neste trabalho, esta proximidade foi calculada pela norma Euclidiana, porém outras normas podem ser utilizadas. O algoritmo tem como base a minimização da seguinte função-objetivo:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty, \quad (6)$$

onde m é qualquer número real maior do que 1, u_{ij} é o grau de pertinência de x_i à classe j , x_i é o i -th vetor p-dimensional de dados de entrada, c_j é o centro p-dimensional da classe j , C é a quantidade de classes, N (neste caso) é a quantidade total de vetores \mathbf{x} , e $\|\cdot\|$ é qualquer norma vetorial que expressa a similaridade entre o dado de entrada e o centro da classe. O particionamento nebuloso é conduzido por uma otimização iterativa da função-objetivo, onde as pertinências e os centros das classes são atualizados, respectivamente, pelas equações:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad (7)$$

e

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}. \quad (8)$$

Cada iteração termina quando $\max_{ij} \left\{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \right\} < \varepsilon$, onde ε é um valor mínimo situado entre 0 e 1 e definido pelo intérprete, e k é o contador de iterações. Este procedimento converge para um mínimo local de J_m (Bezdek, 1981). Passo a passo, o exato algoritmo FCM aplicado foi:

- 1) Faça $k = 0$. Inicialize a matriz $U^{(k)} = [u_{ij}]$, atribuindo a cada u_{ij} valores aleatórios situados no intervalo (0,1) e extraídos de uma distribuição uniforme padrão.
- 2) No passo k , calcule os vetores $C^{(k)} = [c_j]$ pela Equação 7, utilizando $U^{(k)}$.

- 3) Calcule $U^{(k+1)}$ aplicando a Equação 8.
- 4) Se $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ ou $k > K$, sendo K o número máximo de iterações, finalize o processo. Caso contrário, incremente k e retorne ao passo 2.

Os parâmetros fixos do FCM utilizados neste trabalho foram: $m = 2$; $k_{\max} = 200$; e $\varepsilon = 10^{-5}$. Assim como no caso do SOM, todo o processamento dos dados foi conduzido no *software* MatLab[®] v. 7.0(R14). Neste caso, basicamente foi utilizada a função *fcm*, a qual contém o algoritmo com os passos descritos aqui, além de outras funções auxiliares do próprio MatLab².

Avaliação da classificação dos dados: índice Davies-Bouldin

O índice Davies-Bouldin (Davies & Bouldin, 1979) é uma medida da similaridade entre as classes de dados. Este índice pode ser utilizado para validação da classificação dos dados e para comparação entre diferentes distribuições de dados entre as classes. O índice Davies-Bouldin independe do número de classes e do método de classificação, o que o torna aplicável aos dois métodos utilizados neste trabalho. Este índice é dado por:

$$\frac{1}{c} \sum_{k=1}^C \max_{C \neq l} \left\{ \frac{S_C(Q_k) + S_C(Q_l)}{d_{Ce}(Q_k, Q_l)} \right\} \quad (9)$$

onde C é a quantidade de classes, $S_C(Q_k)$ representa a distância intra-classe (dispersão interna da classe Q_k), $d_{Ce}(Q_k, Q_l)$ representa a distância entre as classes Q_k e Q_l . Os valores mais baixos do índice Davies-Bouldin indicam as melhores classificações dos dados. Convém ressaltar que este índice deve servir apenas como guia e não como um indicador absoluto de melhor classificação de dados (Ultsch, 1993).

RESULTADOS

As contribuições individuais de cada variável de entrada para o SOM são apresentadas nos planos de componentes (Kohonen, 2001; Fig. 3). A Figura 4 apresenta a matriz-U obtida a partir do arranjo SOM de 60×34 unidades. Os menores valores nesta matriz indicam as menores dissimilaridades entre os vetores protótipo e os vetores de entrada e são representados por cores frias. Por outro lado, os maiores valores indicam as maiores dissimilaridades e são representados por cores quentes. Portanto, os agrupamentos são tipicamente visualizados como regiões homogêneas com tonalidades próximas ao azul. As separações entre os agrupamentos são marcadas pelas regiões mais lineares de cores mais próximas ao vermelho.

²<http://mathworks.com/access/helpdesk/help/toolbox/fuzzy/fcm.html>

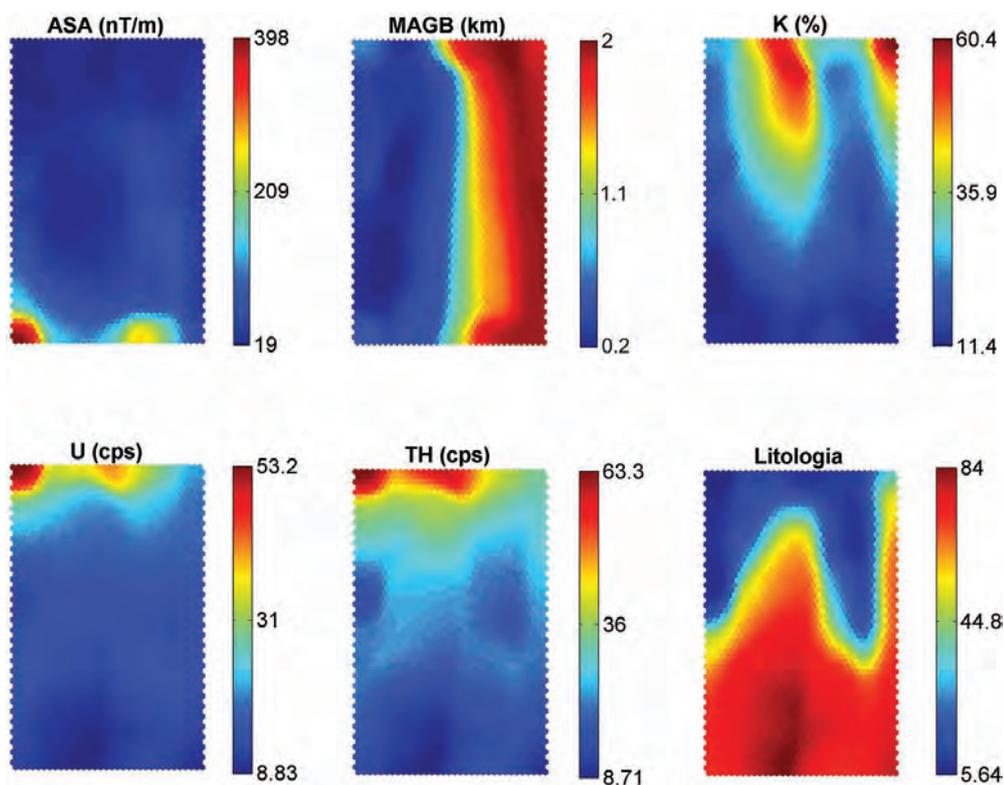


Figura 3 – Planos de componentes de cada variável de entrada. ASA = Amplitude do Sinal Analítico, MAGB = *Buffers* em torno de eixos magnéticos, K = concentração de Potássio, U = concentração de Urânio, TH = concentração de Tório. As litologias foram codificadas numericamente de acordo com a maior ou menor favorabilidade para hospedagem de mineralização do tipo Cr-EGP (Luanga) (Tab. 1).

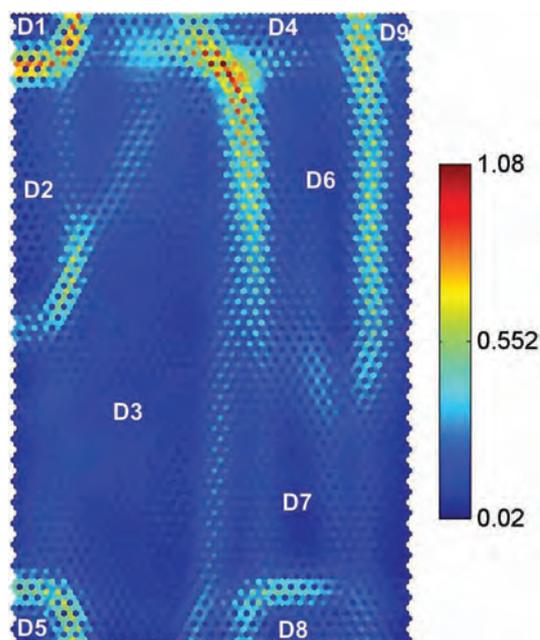


Figura 4 – Matriz-U calculada sobre o SOM. As depressões indicadas no mapa (baixos valores) apontam para os centros das classes e as regiões com altos valores apontam para os limites entre as classes (Ullsch, 1992, 1993).

A Figura 5a mostra os índices Davies-Bouldin calculados para até 10 agrupamentos. O algoritmo de *K-means* (e.g. Cios et al., 1998; Seber, 2004) foi utilizado para re-classificar as BMUs definidas pelo SOM apenas para efeito de visualização dos agrupamentos observados na matriz-U. Este resultado é apresentado na Figura 5b. Nesta figura, cada hexágono da matriz-U foi re-escalado de acordo com a quantidade de vetores de entrada utilizados para gerar a BMU correspondente. Quanto maior o hexágono, maior a quantidade de vetores de entrada utilizada. É notável que as menores quantidades de dados de entrada estão associadas às BMUs próximas das bordas dos agrupamentos. Comparando-se com a matriz-U da Figura 3, isso mostra que os dados de entrada mais dispersos (i.e., as regiões onde existem as maiores heterogeneidades na distribuição dos dados de entrada), servem como delimitadores de agrupamentos de dados mais homogêneos, pois são as BMUs para as quais os valores da matriz-U são mais elevados. Os diagramas de dispersão da Figura 6 mostram a variabilidade dos dados geofísicos em relação aos diferentes grupos de litologias, definidos de acordo com a Tabela 1. A distribuição espacial do mapa SOM reclassificado

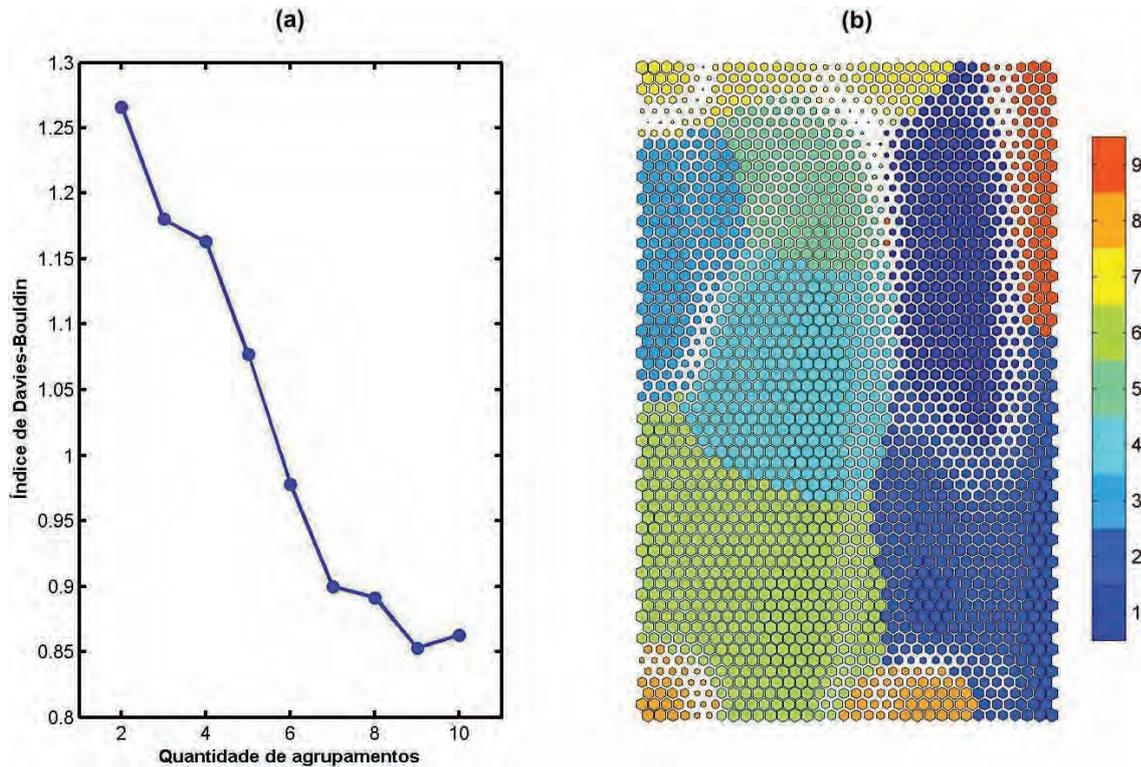


Figura 5 – Agrupamento das unidades da matriz-U obtidos com o *K-means*. O menor índice de Davies-Bouldin foi obtido para 9 agrupamentos (a). O tamanho dos hexágonos da matriz-U é proporcional à quantidade de vetores de entrada associados à BMU correspondente (b).

pelo *K-means* e o mapa classificado por FCM pode ser vista na Figura 7. A Figura 8 apresenta os erros de quantização obtidos com os dois métodos. A Tabela 2 contém as medidas estatísticas associadas às distribuições dos erros de quantização (*qe*) obtidos com os métodos SOM e FCM.

Tabela 2 – Medidas estatísticas das distribuições dos erros de quantização (*qe*) obtidos com os métodos SOM e FCM.

	<i>qe</i> (SOM)	<i>qe</i> (FCM)
Valor mínimo	0.027	0.115
Valor máximo	14.033	16.616
Média	0.747	1.251
Desvio padrão	0.656	0.897
% > 1,000	17.359	49.013
% > 2,000	3.9239	12.416

DISCUSSÃO

Analisando os planos de componentes (Fig. 3) em conjunto com a matriz-U (Fig. 4), nota-se que a região D1 é delimitada por altos valores das concentrações de eU e eTh e forma o agrupamento 7 definido pelo SOM/*K-means* (Fig. 5b). A região D2 mostra baixos valores de todas as variáveis geofísicas e também baixas pertinências *fuzzy* da litologia, indicando que esta região

provavelmente tem baixo potencial para conter mineralizações de Cr-EGP. Em contraste, a região D3, em sua maior parte, representa altos valores de pertinência *fuzzy* da litologia. Nota-se também que esta região é a que engloba a maior parte dos pontos e foi subdividida pelo SOM/*K-means* em 3 agrupamentos menores (4, 5 e 6; Fig. 5b). O agrupamento 6 contém os pontos com maiores graus de pertinência *fuzzy* litológica e os menores valores das concentrações de eU, eTh e K; o agrupamento 5 representa altos valores de K e baixa pertinência *fuzzy* litológica; e o agrupamento 4 difere do 6 apenas por conter um pouco mais elevados da concentração de K.

A região D7 difere de D3 somente no fato da primeira representar pontos com maiores distâncias aos eixos magnéticos. O SOM/*K-means* separou D7 no agrupamento 2. Da mesma forma, a região D6 é separada da D3 pela diferença nestes mesmos valores, além de D6 representar pontos com menores pertinências *fuzzy* da litologia. A região D4 é delimitada por altas distâncias (>1000 m) em relação aos eixos magnéticos, baixos valores da ASA e de pertinência *fuzzy* da litologia, além de valores intermediários das concentrações radiométricas. Esta região é englobada pelo agrupamento 7 do SOM/*K-means*. As regiões D5 e D8 representam um pequeno grupo de pontos com alta pertinência

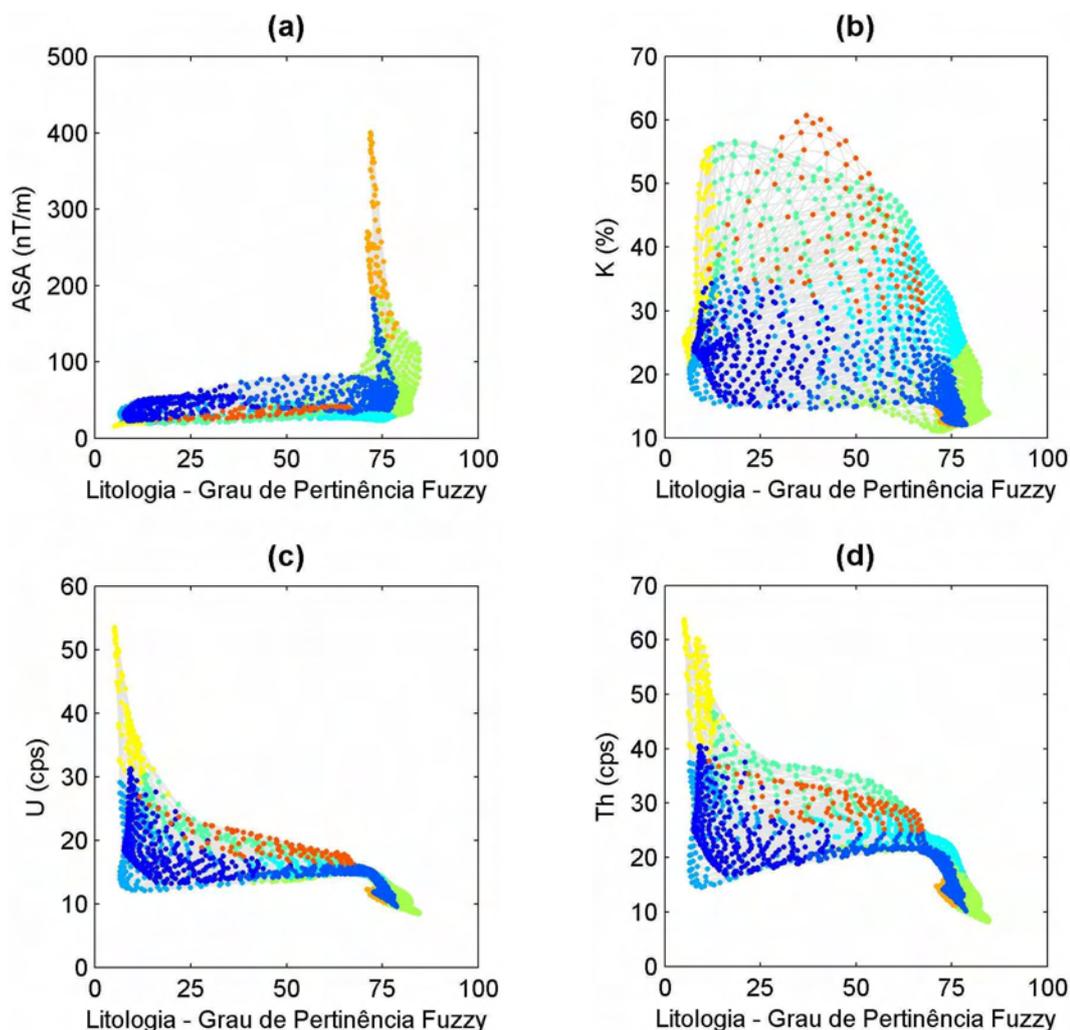


Figura 6 – Diagramas de dispersão dos valores das variáveis geofísicas vs. graus de pertinência *fuzzy* da litologia. Os graus de pertinência estão apresentados na Tabela 1. Apenas as BMUs foram incluídas nos diagramas. As cores representam as nove classes obtidas via *K-means* e seguem o mesmo padrão da Figura 5b.

fuzzy da litologia (agrupamento 8). Entretanto a região D5 apresenta, relativamente à D8, valores mais elevados da ASA e menores distâncias aos eixos magnéticos. Altas concentrações de K, altos valores de pertinência *fuzzy* da litologia e grandes distâncias aos eixos magnéticos definem a região D9, separada no agrupamento 9 do SOM/*K-means*.

Os diagramas de dispersão apresentados na Figura 6 mostram o comportamento das variáveis geofísicas em relação aos litotipos (cf. Tab. 1). Aparentemente, as regiões pertencentes ao agrupamento 6 devem ser prioritárias para investigações mais detalhadas sobre locais que permitam concluir sobre a presença ou ausência de depósitos do tipo Cr-PGE. Isso pois esse agrupamento representa os maiores valores de pertinência *fuzzy* da

litologia, além de intervalos relativamente pequenos para os valores das variáveis geofísicas, incluindo as menores distâncias aos eixos magnéticos. Como pode ser observado no mapa da Figura 7a, este agrupamento engloba o depósito conhecido de Cr-PGE do Complexo Luanga, situado na porção nordeste da área de estudo. O agrupamento 8 também representa altos valores de pertinência *fuzzy* da litologia, além de altos valores da ASA e baixos valores dos elementos radiométricos, os quais mostram-se agrupados em pequenas regiões nos diagramas de dispersão. O agrupamento 7 destaca-se por concentrar os menores valores da ASA e baixas pertinências *fuzzy* da litologia associadas às quatro variáveis geofísicas.

O mapa da Figura 7b mostra a distribuição espacial dos 9

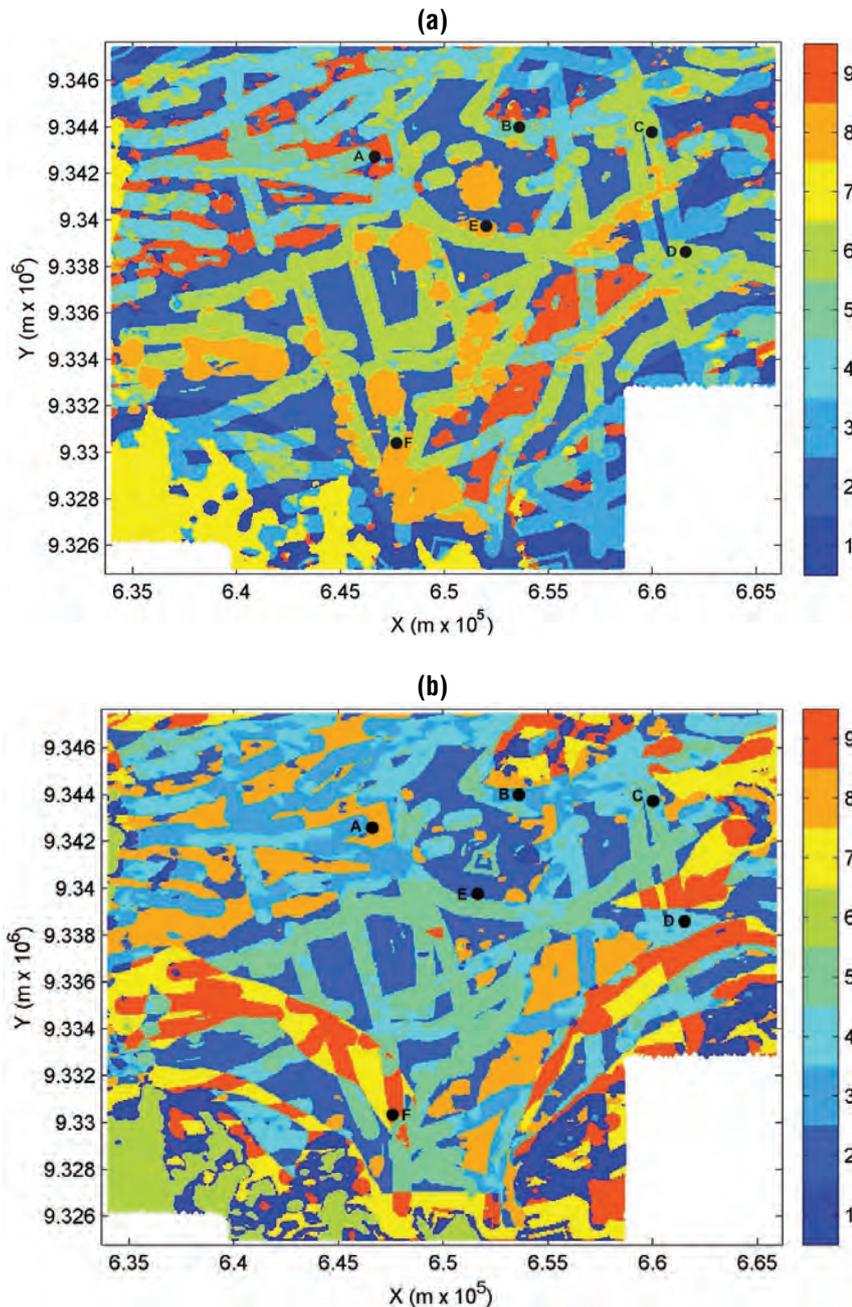


Figura 7 – Mapas com 9 agrupamentos para a região de Serra Leste, Província Mineral de Carajás, obtidos a partir dos métodos: (a) SOM/*K-means*; (b) FCM. Os círculos pretos representam as mineralizações de: (A) Serra Pelada (Au, Pt, Pd); (B) Formiga (Au); (C) Luanga (Cr, Pt, Pd); (D) Cutia (Au); (E) Serra Verde (Cu, Au); e (F) Serra Leste (Fe).

agrupamentos obtidos com o FCM, para comparação com os resultados do SOM/*K-means*. Neste mapa, o agrupamento 5 engloba o depósito de Cr-EGP de Luanga e é o mais similar ao agrupamento 6 do mapa classificado pelo SOM/*K-means*. É notável que o agrupamento 8 do SOM/*K-means* engloba as regiões D5 e D8 do SOM, ignorando suas significativas diferenças

em relação aos dados de entrada. Da mesma forma, não existe um agrupamento único correspondente no mapa do FCM, indicando que o SOM/*K-means* falhou na separação desses dois grupos de dados.

De maneira geral, percebe-se um aspecto mais ruidoso do mapa obtido com o FCM e maior suavidade no mapa obtido com

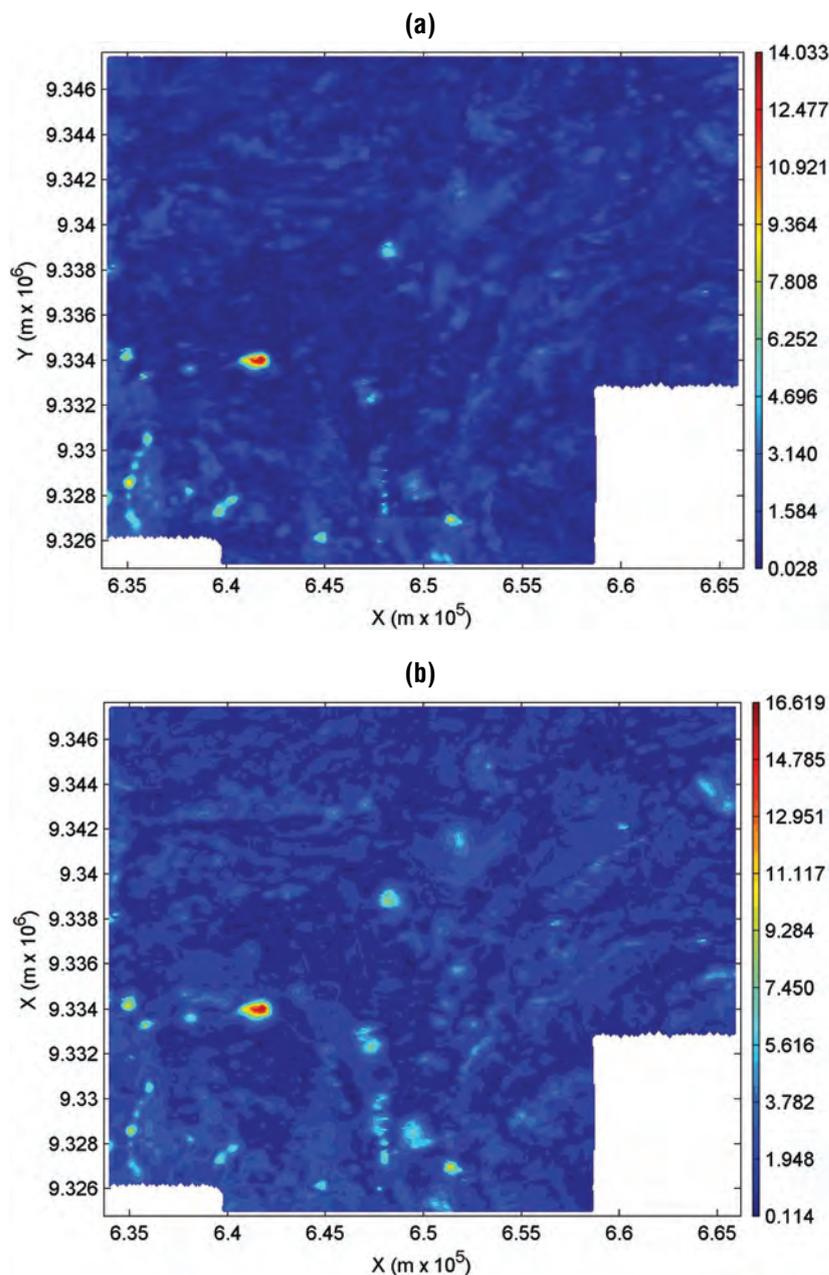


Figura 8 – Distribuição geográfica do erro de quantização do SOM (a) e do FCM (b).

o SOM/*K-means*. Os mapas de erros de quantização da Figura 8 apresentam distribuição espacial semelhante, sendo que os maiores valores ocorrem devido aos altos valores da ASA nessas regiões, principalmente na região sudoeste dos dois mapas. Isto reflete a distribuição dos dados do campo magnético, a qual possui um desvio padrão maior do que aqueles associados aos dados radiométricos. Exatamente nesses pontos a diferença entre os valores dos dados de entrada e os centros dos agrupamentos (ou BMUs no caso do SOM) é maior nos dois casos. Entretanto,

os valores máximo e mínimo, a média e o desvio padrão são menores no caso do SOM. Nota-se também que somente 17,36% dos valores de $qe(\text{SOM})$ são maiores do que 1,00, enquanto que quase a metade dos valores de $qe(\text{FCM})$ são maiores do que 1,00.

CONCLUSÕES

Neste trabalho, o método de Mapas Auto-Organizáveis (SOM – *Self-Organizing Maps*) foi utilizado para classificação de dados geofísicos e geológicos na região de Serra Leste, Província Mine-

ral de Carajás, Pará. Um resultado de classificação obtido com o método de *Fuzzy C-means* (FCM) foi utilizado para comparação com os resultados do SOM. O enfoque da classificação foi o mapeamento de potenciais alvos para futuras investigações locais sobre a presença de mineralizações de Cr-EGP. A interpretação da matriz-U, em conjunto com os planos de componentes das variáveis geofísicas e das pertinências *fuzzy* associadas às unidades litológicas, permitiu a identificação da origem dos agrupamentos de dados. A matriz-U apresentou alguns agrupamentos que puderam ser visualmente identificados. Adicionalmente, o algoritmo clássico *K-means*, quando aplicado sobre o SOM previamente calculado (SOM/*K-means*), foi capaz de separar esses e outros agrupamentos, definindo seus limites. Isto foi importante para a interpretação posterior da distribuição geográfica dos agrupamentos. Entretanto, duas regiões com características distintas no SOM foram erroneamente agrupadas pelo SOM/*K-means*. A quantidade de agrupamentos foi escolhida com base no menor índice Davies-Bouldin, o que resultou na escolha de um mapa com nove agrupamentos para interpretação.

O mapa obtido pelo método FCM, também para nove agrupamentos, mostrou algumas similaridades e diferenças em relação ao mapa classificado pelo SOM. Porém, em geral, ambos apresentam a mesma consistência em relação aos dados de entrada. Particularmente, o SOM separou uma classe de dados que representa os valores mais altos do campo magnético, possivelmente porque tais valores encontram-se muito distantes do valor médio. Além disso, o agrupamento onde está inserido o depósito de Cr-EGP de Luanga é consistente com a assinatura geofísica mapeada no trabalho de Souza Filho et al. (2007).

A distribuição dos erros de quantização do SOM apresentou um valor médio 40% menor e um desvio padrão 27% menor em relação à distribuição de erros obtida com o FCM. A análise da distribuição espacial das variáveis de entrada em conjunto com o a distribuição geográfica dos erros de quantização indicou que os maiores erros de quantização estão associados aos maiores gradientes do campo magnético da região. Este fato reflete apenas uma particularidade dos dados utilizados neste trabalho, onde a amplitude do sinal analítico do campo magnético anômalo (ASA) apresenta desvio padrão muito mais elevado do que os desvios-padrão dos demais dados que compuseram os vetores de entrada do SOM.

Desta forma, os resultados deste trabalho mostram a potencialidade do SOM na integração e interpretação conjunta de dados geofísicos e geológicos, aplicados ao mapeamento do potencial mineral em áreas de densa cobertura vegetal, como a Província Mineral de Carajás.

AGRADECIMENTOS

E.P. Leite e C.R. Souza Filho agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa de pós-doutorado (Proc. N. 05/04453-3) e pelo suporte financeiro (Proc. N. 03-09916-6), respectivamente.

REFERÊNCIAS

- BEZDEK JC. 1981. Pattern Recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell, 256 pp.
- BRIGGS IC. 1974. Machine contouring using minimum curvature. *Geophysics*, 39: 39–48.
- CABRAL AR, LEHMANN B, KWITKO R & CRAVO CHC. 2002a. The Serra Pelada Au-Pd-Pt deposit, Carajás Mineral Province, northern Brazil; reconnaissance mineralogy and chemistry of very high grade palladian gold mineralization. *Econ. Geol.*, 97: 1127–1138.
- CABRAL AR, LEHMANN B, KWITKO R & CRAVO CHC. 2002b. Palladium and platinum minerals from the Serra Pelada Au-Pd-Pt deposit, Carajás Mineral Province, Northern Brazil. *Can. Mineral*, 40: 1451–1463.
- CIOS KJ, PEDRYCZ W & SWINIARSK RM. 1998. Data Mining Methods for Knowledge Discovery. Kluwer Academic Publishers, Norwell, 495 pp.
- COELHO CES & RODRIGUES OB. 1986. Jazida de manganês do azul, Serra dos Carajás, Pará. In: SCHOBENHAUS C & COELHO CES (Eds.). Principais depósitos minerais do Brasil. DNP/M/CVRD, Vol. 2, 145–152.
- DAVIES D & BOULDIN D. 1979. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(2): 224–227.
- DEMICO RV & KLIR GJ. 2003. Fuzzy Logic in Geology. Elsevier Science Academic Press, San Diego, 347 pp.
- DOCEGEO. Rio Doce Geologia e Mineração S.A. 1988. Revisão litoestratigráfica da Província Mineral de Carajás – litoestratigrafia e principais depósitos minerais. In: Congresso Brasileiro de Geologia, 35., Belém: SBG, Anexo aos Anais, pp. 11–54.
- IAGA DIVISION 5, WORKING GROUP 8. 1996. International geomagnetic reference field, 1995, revision. *Geophys. J. Int.*, 125: 318–321.
- KLOSE CD. 2006. Self-Organizing Maps for Geoscientific Data Analysis: Geological Interpretation of Multidimensional Geophysical Data. *Comput. Geosci.*, 10: 265–277.
- KOHONEN T. 2001. Self-Organizing Map. 3 ed., Springer-Verlag, Berlin, pp. 113.
- LEITE EP & SOUZA FILHO CR. 2009a. Probabilistic neural networks applied to mineral potential mapping for platinum-group elements in the Serra Leste region, Carajás Mineral Province, Brazil. *Comput. Geosci. (UK)*, 35: 675–689.

- LEITE EP & SOUZA FILHO CR. 2009b. Artificial neural networks applied to mineral potential mapping for copper-gold mineralizations in the Carajás Mineral Province. *Geophys. Prospect.*, 57(6): 1049–1065.
- MANSOUR Z, ALI P & MAHDI Z. 2009. A computational optimized extended model for mineral potential mapping based on WofE method. *Am. J. Applied Sci.*, 6(2): 200–203.
- MARSHIK R, MATHUR R, RUIZ J, LEVEILLE RA & ALMEIDA AJ DE. 2005. Late Archean Cu-Au-Mo mineralization at Gameleira and Serra Verde, Carajás Mineral Province, Brazil: constraints from Re-Os molybdenite ages. *Miner. Deposita*, 39: 983–991.
- PENN SB. 2002. Using self-organizing maps, histograms, and standard deviation to detect anomalies in hyperspectral imagery data. In: *Proceedings of the Fifth International Airborne Remote Sensing Conference*, Miami, Florida, 22–24.
- PENN SB. 2004. Using self-organizing maps to visualize high-dimensional data. *Comput. Geosci. (UK)*, 31(5): 531–544.
- ROEST WR, VERHOEF J & PILKINGTON M. 1992. Magnetic interpretation using the 3-D analytical signal. *Geophysics*, 57: 116–125.
- SAVELIEV AA & DOBRININ DV. 1999. Hierarchical multispectral image classification based on self organized maps. In: *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99*, Piscataway, NJ, 5: 2510–2512.
- SEBER GAF. 2004. *Multivariate observations*. 2 ed., John Wiley & Sons, Inc., Hoboken, NJ, 686 pp.
- SOUZA FILHO CR, NUNES AR, LEITE EP, MONTEIRO LVS & XAVIER RP. 2007. Spatial analysis of airborne geophysical data applied to geological mapping and mineral prospecting in the Serra Leste region, Carajás Mineral Province, Brazil. *Surv. Geophys.*, 28: 377–405.
- SUITA MTF. 1988. *Geologia da área Luanga com ênfase na petrologia do complexo básico-ultrabásico Luanga e depósitos de cromita associados*, Pará. Dissertação de Mestrado, Brasília, UnB, 83 pp.
- ULTSCH A. 1992. Knowledge acquisition with self-organizing neural networks. In: ALEKSANDER I & TAYLOR J (Eds.). *Artificial Neural Networks*. Elsevier Science Publishers, B.V., 735–740.
- ULTSCH A. 1993. U*-matrix: a tool to visualize clusters in high dimensional data. University of Marburg, Department of Computer Science, Technical Report, 36, 12 pp.
- VENEZIANI P & OKIDA R. 2001. Mapeamento geológico-estrutural da região do Projeto Platina (Serra Pelada) baseado em dados integrados RADARSAT-TM e aerogeofísica. Relatório Técnico, 52 pp.
- VESANTO J, HIMBERG J, ALHONIEMI E & PARHANKANGAS J. 2000. *SOM Toolbox for Matlab 5*. Helsinki University of Technology, Technical Report, A57, 59 pp.
- VILLAS RN & SANTOS MD. 2001. Gold deposits of the Carajás Mineral Province: deposit types and metallogenesis. *Miner. Deposita*, 36: 300–331.

NOTAS SOBRE OS AUTORES

Emilson Pereira Leite. Professor Doutor do Departamento de Geologia e Recursos Naturais do Instituto de Geociências da UNICAMP. Obteve seu doutorado em Geofísica pelo IAG/USP com estágio no *Department of Earth and Space Science and Engineering* da Universidade de York em Toronto, Canadá. Suas principais áreas de interesse são métodos potenciais em Geofísica, inversão de dados geofísicos, sensoriamento remoto e integração de dados geofísicos e geológicos, caracterização e previsão de alvos exploratórios.

Carlos Roberto de Souza Filho. Engenheiro Geólogo (Universidade Federal de Ouro Preto, 1988), Mestre em Metalogênese (UNICAMP, 1991), PhD (Open University, Inglaterra, 1995), Jovem-Pesquisador (FAPESP-UNICAMP, 1995-1997), Professor-Doutor (UNICAMP, 1997-2002), Livre-Docente (UNICAMP, 2002-2009). Desde 2009, é Professor Titular do DGRN do IG-UNICAMP. Atualmente é Coordenador-Geral dos Programas de Pós-graduação do IG, líder do Grupo de Geotecnologias e de Ciências Planetárias, pesquisador Nível 1A do CNPq, pesquisador da NASA nos programas ASTER e *Mars-Earth Analogs*, e Editor-Associado da Revista *Computers & Geosciences*. Sua área principal de interesse é em Mapeamento Geológico e Exploração Mineral e Petrolífera com base em dados e métodos de Sensoriamento Remoto, Aerogeofísica, Sistemas de Informações Georreferenciadas e Sistemas Especialistas.