Self-Organizing Maps in Airborne Geophysical Data applied to Geological Mapping in Amazonian Region

Cleyton de Carvalho Carneiro, University of Campinas (UNICAMP)
Stephen James Fraser, Australian Commonwealth Scientific and Industrial Research Organization (CSIRO)
Alvaro Penteado Crósta, University of Campinas (UNICAMP)
Adalene Moreira Silva, University of Brasília (UNB)
Carlos Eduardo de Mesquita Barros, Federal University of Paraná (UFPR)

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

Abstract

Airborne geophysical methods are commonly applied as tools to provide a basis or improve geological mapping and to assist in assessing mineral prospectivity. This paper details a multi-variate analysis of airborne geophysical data, using a Self-Organizing Map approach. The SOM analysis indicated that the data could be grouped into eleven clusters. These clusters could be analyzed individually, according to the contribution of each variable. These results allow for an enhanced interpretation and insight into several geological processes.

Introduction

The Self-Organizing Map (SOM) is a tool for the analysis and visualization of high-dimensional data, based on principles of vector quantization (Kohonen 2001). Fraser & Dickson (2007) state that most of SOM procedures can be considered exploratory, and the method can be used to perform broad categories of operations, such as, prediction or estimation, clustering, classification, pattern recognition and/or noise reduction. SOM is an unsupervised method; therefore, no prior knowledge is required as to the nature or number of groupings within the dataset. SOM can be applied to both categorical and continuous variables. All of these advantages make the SOM technique ideal for the analysis of complex and disparate geoscientific data.

Research has demonstrated many examples of SOM analyses on geological, mining and exploration data sets (Fraser et al. 2006, Bierlein et al. 2008). However, little work has been published on the use of SOM to analyze airborne geophysical datasets.

The Anapu-Tuerê area is located in the central east region of Pará state, in the Amazonian region, in the north of Brazil. In this region, occurrences of gold, chromium, nickel, copper, titanium, vanadium and platinum were regionally registered. In 2004, the Brazilian Geological Survey (CPRM) concluded the “Anapu-Tuerê Airborne Geophysical Project”. This project collected high-resolution magnetic and gamma ray spectrometry data. Therefore, it became possible to review the geological mapping of the region based on recent techniques of spatial analysis.

Method

Airborne Geophysical Data

Acquisition of the airborne geophysical data as part of the Anapu-Tuerê Project (Table 1) took place between August and October, 2004. The project covered an area of 24,735 km^2, consisting of 53,331 km of line profiles. These profiles were placed in 310 lines of production and 30 control lines.

Table 1: *Characteristics of Anapu-Tuerê Aerogeophysical Project.*

Direction of production lines	N-S
Spacing between production lines	0,5 km
Direction of control lines	E-W
Spacing between control lines	10,0 km
Interval between consecutive geophysical measurements	0,1 s (magnetometer); 1,0 s (espectrometer)
Average flight height	100 m
Average flight speed	260 km/h

The magnetic data were processed as the anomalous magnetic field, or as the total measured field corrected for the diurnal variation, the main International/Definitive Geomagnetic Reference Field (IGRF/DGRF), and leveling errors. The gamma spectrometry data were processed into energy channels with reference to the total energy, or total count channel (TC), which was expressed in mR/hr. The potassium (K) channel data were expressed in percentage; while the uranium (eU) and thorium (eTh) channel data were expressed as micro-equivalents. The data base was corrected, according to Minty (1997), for dead-time, energy variations, or spectral stabilization of the respective background levels of radiation, altitude variations relative to the nominal value for the project and for scattering due to the Compton Effect. The following ratios, eU/eTh, eU/K and eTh/K were calculated after the corrections.

Eight geologically significant, geophysical parameters were then selected for further analysis as input to the SOM procedure: K, eTh, eU, TC, eU/eTh, eU/K, eTh/K and AS.

Self-Organizing Maps Analysis

In a SOM analysis, each sample is treated as an n-dimensional (nD) vector in a data space defined by its variables (Fraser and Dickson, 2007). Next, a number of “artificial seed-vectors” are modified to represent the distribution and structure of the original input data using measures of vector similarity. The number of seed-vectors equates to the size of the self-organized “map” chosen for the analysis. Once trained these seed-vectors are known as “Best Matching Units” (BMUs); and it is these best matching units or vectors that are projected onto the enveloping hyper-surface and transformed to produce the self-organized map representation of the data. Once computed, the “map” can be visualized in many ways. This study used three “map” visualizations: (i) the “Unified distance Matrix” (U-Matrix); (ii) “component plots”; and (iii) a K-means clustering of the BMU vector-values.

For the SOM analysis of the Anapu-Tuere area, eight variables were used as input: gamma ray spectrometry channels of potassium (K), thorium (eTh), uranium (eU), total-count (TC), the ratios eU/eTh, eU/K and eTh/K, in addition to the analytical signal (AS) of the anomalous magnetic field. These parameters together represent a formidable set of exploration tools, capable of mapping a range of different geological characteristics and processes.

Firstly, there was a need to ensure that the above data sets were co-registered onto a grid with a common geographic origin and cell size. Secondly, the eight input variable values for each cell were exported as a centroid value for that cell. These located values were used as input to the SOM analysis.

The following SOM setup and processing parameters were chosen for this study. The data space was randomly initiated or seeded; a hexagonal lattice was chosen for display; and the surface of a toroidal hypervolume was used for the BMU projection. A map size of 40 rows x 35 columns was chosen as appropriate for this exploratory study.

To visualize the spatial distribution of the SOM output data, the cluster output values were gridded using “nearest neighbor” for discrete data.

Results

A Self-Organized Map analysis, using a 40x35 sized map, was calculated on the eight input geophysical parameters. The U-Matrix resulting from this analysis is shown on Figure 1(A). The U-Matrix shows three areas with high dissimilarity, which coincide with elevated contributions of eU, eU/eTh ratio and AS. High similarity in U-matrix is related with mediums contributions of K, eTh, eU, TC and eU/eTh.

In a next step, a Davies-Bouldin analysis (Davies and Bouldin 1979) was run and 11 clusters were chosen. Using *K*-means the 11 clusters were then produced (Figure 1 (B)), based on the similarity between the node vectors (hexagons). Figure 1 (C) shows the spatial distribution of input samples assigned their SOM-derived cluster number.

Then, there were used the reference values derived from the total data set, to classify each SOM-derived cluster’s component contribution, as shown in Table 2.

From the individual cluster characteristics and the spatial distribution of samples belonging to each cluster, it is possible to relate the SOM results to geological responses and processes and understand that:

(i) Cluster 10 is related to rocks of Três Palmeiras Sequence. Possibly though geological effects, like as metamorphism and hydrothermal alteration, this sequence was also expressed in others clusters (for example, Clusters 1 and 2);

(ii) The occurrences of gold mineralization in the study area are related to Clusters 1, 2 and 10. Significantly, Cluster 2 was found to be related to rocks from hydrothermal alteration zones. Sulfide-bearing diorites were found to be related to Cluster 11, which also suggests some evidence for gold mineralization;

(iii) All the lateritic outcrops were related to Cluster 8. This relationship suggests that the signature “low K, medium eTh, eU/eTh and AS, high TC, eU, eU/K and eTh/K” are a good signature for mapping these rocks in the Amazonian region.

(iv) Different types of granites (*sensu latu*) were related to Clusters 3, 5, 6, 7 and 9. Those rocks appear to belong to the domains of João Jorge Intrusive Suites and Bacajai Complex. However, further petrographic and geochemical analyses would be needed to define more clearly the characteristics between the SOM-derived clusters and the collected rocks.

(v) The signature “medium values of TC, K, eTh and all ratios, low eU and high magnetic gradient AS” present in Cluster 4, shows high correspondence with gabbroic composition dykes. Such dykes do not appear in the existing geological maps, but should be added to subsequent updates of these maps.

Conclusions

The integrated analysis and spatial display of multi-variate geoscience data sets can contribute substantially towards identifying and understanding geological effects and processes from such data. However, an appreciation of the effects and influence of each input variable’s characteristics is necessary to understand the significance of the resultant SOM classes. The SOM analysis of the Anapu-Tuerê airborne geophysical datasets produced 11 clusters, which were related to different rocks, mineral prospecttivity and petrogenetic history.

In this study, SOM has proven to be an effective tool to map various populations with similar, characteristics from airborne geophysical data sets. Both the magnetic and gamma ray spectrometric data that we have analyzed contain subtle relationships, which allow us to detect and extract geological information. However, the magnetic field needs to be express adequately. Because of the low latitudes of the Amazon region, the best product to show the magnetic gradient is the analytic signal. Correlations between the different geophysical data sets could be observed on the SOM results.

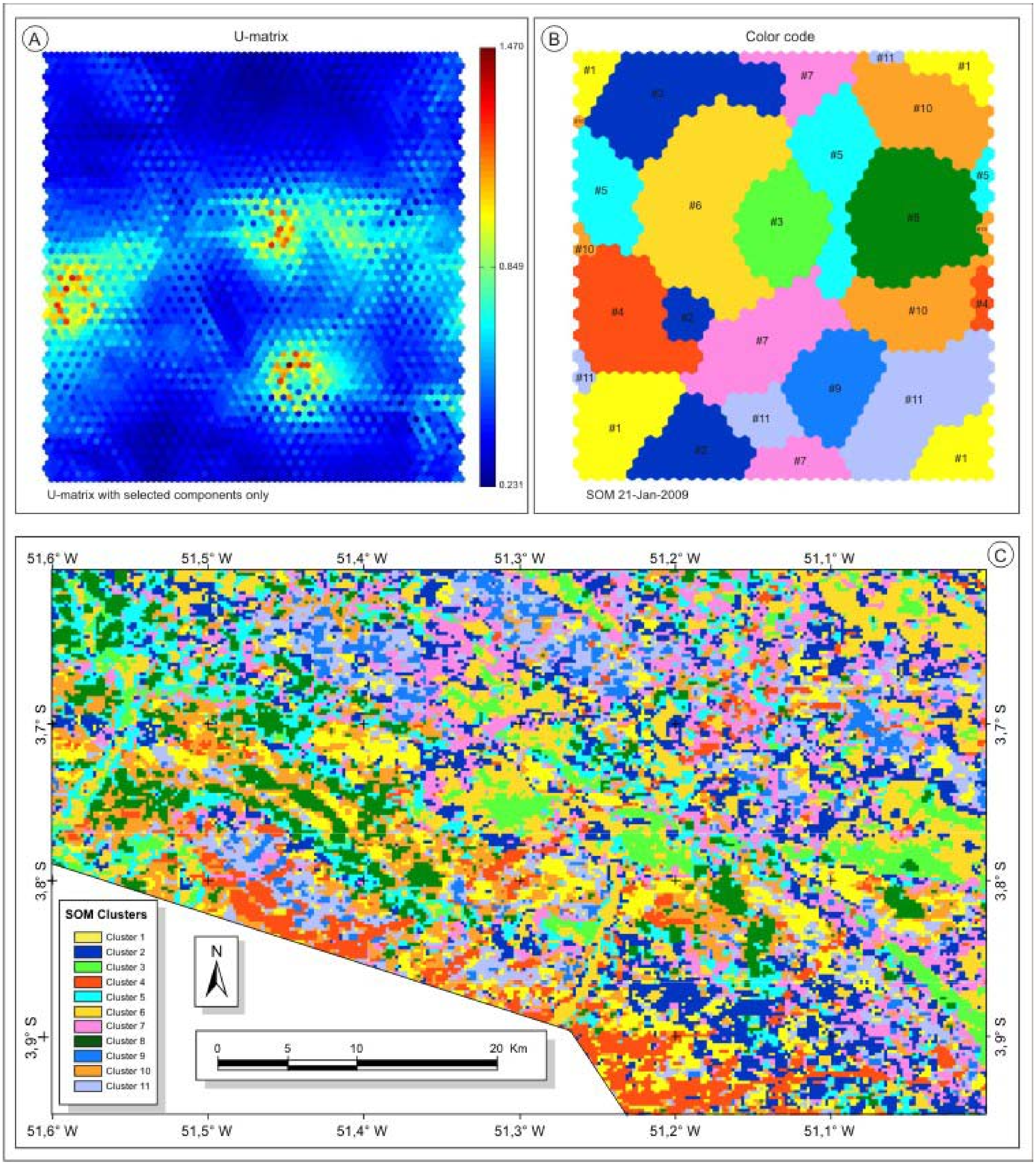

Figure 1: (A) U-Matrix, with the nodes colured to represent their similarity; (B) K-means cluster classification of the BMU vector values on the self-organized map; (C) Spatial map showing each sample coded by it's cluster colour derived from the SOM analysis.

Table 2: *Influence of variables in each cluster from SOM analysis.*

		K	*eTh*	*eU*	*TC*	*eU/eTh*	*eU/K*	*eTh/K*	*AS*
Cluster 1	Average (Classification)	0,6688 (Medium)	6,3390 (Low)	1,7040 (Low)	2,3207 (Low)	0,0888 (Low)	5,9327 (Medium)	89,7042 (Medium)	0,1426 (High)
Cluster 2	Average (Classification)	1,0945 (High)	9,4319 (Medium)	2,1912 (Low)	4,1724 (Medium)	0,1142 (Low)	5,3032 (Low)	75,6534 (Low)	0,1141 (Medium)
Cluster 3	Average (Classification)	1,4637 (High)	18,7528 (High)	4,1892 (High)	8,7251 (High)	0,1772 (High)	7,3437 (High)	82,9028 (Medium)	0,1230 (Medium)
Cluster 4	Average (Classification)	0,7151 (Medium)	7,8708 (Medium)	2,1962 (Low)	3,1410 (Medium)	0,1375 (Medium)	7,9336 (Medium)	94,6270 (Medium)	0,6246 (High)
Cluster 5	Average (Classification)	0,7625 (Medium)	13,7326 (High)	2,9508 (High)	5,4064 (High)	0,1382 (Medium)	9,6212 (High)	109,2608 (High)	0,1329 (High)
Cluster 6	Average (Classification)	1,6067 (High)	15,2030 (High)	2,9205 (High)	7,0752 (High)	0,1246 (Medium)	5,3700 (Low)	74,3382 (Low)	0,1695 (High)
Cluster 7	Average (Classification)	0,9591 (Medium)	8,0937 (Medium)	2,7859 (High)	3,9593 (Medium)	0,2116 (High)	7,0255 (Medium)	77,1660 (Medium)	0,1255 (Medium)
Cluster 8	Average (Classification)	0,4872 (Low)	14,7413 (Medium)	2,8300 (High)	5,1918 (High)	0,1227 (Medium)	20,1791 (High)	219,9992 (High)	0,1260 (Medium)
Cluster 9	Average (Classification)	0,7102 (Medium)	4,0848 (Low)	2,4586 (Medium)	2,1918 (Low)	0,3614 (High)	8,9230 (High)	77,4451 (Medium)	0,1013 (Medium)
Cluster 10	Average (Classification)	0,5170 (Low)	8,4121 (Medium)	2,4040 (Medium)	3,1317 (Medium)	0,1550 (Medium)	13,2180 (High)	132,2628 (High)	0,1438 (High)
Cluster 11	Average (Classification)	0,7032 (Medium)	4,7218 (Low)	2,0437 (Low)	2,1189 (Low)	0,1926 (High)	7,0447 (Medium)	79,6574 (Medium)	0,1263 (Medium)

Acknowledgments

The authors would like to thank: the “State University of Campinas” (UNICAMP/Brazil) where the initial project was planned; the “State of São Paulo Research Foundation” (FAPESP/Brazil), for granting a PhD scholarship (Process 05/59368-0) to the first author; the “National Counsel of Technological and Scientific Development” (CNPq/Brazil), for concession of the project (Process 478792/2007-7), which paid for the field-work; the “Geological Survey of Brazil” (CPRM/Brazil), for provision of the airborne geophysical dataset and some field-data; Australia's Commonwealth Scientific and Industrial Research Organization’s Division of Exploration and Mining (CSIRO/EM), located at the “Queensland Centre for Advanced Technologies” (QCAT), in Brisbane, for the facilities to execute the SOM analysis; the “University of Brasilia” (UNB/Brazil) for support regarding the geophysical dataset; the “Federal University of Paraná” (UFPR/Brazil) for petrographic analyses; and to Helen Field and Jane Hodgkinson for discussions and contributions.

Refences

Bierlein, F. P., Fraser, S. J., Brown, W. M. and Lees, T. 2008. Advanced methodologies for the analysis of databases of mineral deposits and major faults. Australian Journal of Earth Sciences, 55:1, 79 – 99.

Davies, D.L., Bouldin, D.W. 1979. A cluster separation measure. *IEEE Transcations on Pattern Analysis and Machine Intelligence*, 1(2):224–227.

Fraser, S. J., Mikula, P.A., Lee, M. F., Dickson, B.L. and Kinnersly, E., 2006, Data Mining Mining Data – Ordered Vector Quantisation and Examples of its Application to Mine Geotechnical Data Sets. In, Dominy S. (Editor): Sixth International Mining Geology Conference, “Rising to the Challenge”, August 21-23, 2006 Darwin. AusIMM. Publication Series No 6/2006, pp 259-268.

Fraser, S.J., Dickson, B.L. 2007. A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps. *In:* "Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration" edited by B. Milkereit, 2007, p. 907-910.

Kohonen, T. 2001. Self-Organizing Maps, third extended edition, Springer Series in Information Sciences, Vol 30 (Springer: Berlin, Heidelberg, New York).

Minty, B.R.S. 1991. Simple Micro-Levelling for Aeromagnetic Data. *Exploration Geophysics*. 22:591-592.