Unsupervised clustering of gamma-ray spectrometry data using Gaussian Mixture Models

Jessica Derkacz Weihermann [a], Matheus Pinheiro Ferreira [b], Luís Gustavo de Castro [a], Francisco José Fonseca Ferreira [a], Adalene Moreira Silva [c].
[a] Federal University of Paraná, [b] Military Institute of Engineering, [c] University of Brasília

This paper was prepared for presentation during the 16th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 19-22 August 2019.

Abstract

Airborne gamma-ray spectrometry (AGRS) is a useful tool for geological mapping, mineral exploration, and geomorphological studies. The variations observed in AGRS data are typically used to differentiate geological units. Geological mapping with AGRS data is usually performed by visual interpretation, which requires expert knowledge and time. In this paper, we apply unsupervised clustering on principal component analysis images derived from primary variables (potassium, uranium and thorium) to automatically map geological units in Mara Rosa – Goiás, Brazil. Gaussian Mixture Models (GMMs) is an unsupervised clustering method that automatically estimates the number of clusters in a dataset and was poorly explored in the context of geological mapping. The GMM classification of principal component analysis images showed correspondence with geology, facilitating the interpretation.

Introduction

Gamma-ray spectrometry is widely used for surface geological mapping and mineral exploration (e.g. Graham & Bonham-Carter 1993; Charbonneau et al. 1997, Shives et al. 2000, Wilford 1997).

The airborne gamma-ray spectrometry (AGRS) method measures the relative abundance or concentration of potassium (K), uranium (eU) and thorium (eTh) in rocks and weathered materials up to 30–45 cm deep by detecting the gamma radiation emitted by the natural radioactive decay of these elements (Dickson & Scott, 1997; Minty, 1997; Wilford et al., 1997; Wilford, 2002; IAEA, 2003). The main sources of gamma radiation are derived from the disintegration of potassium 40 (^{40}K), uranium 238 (^{238}U) and thorium 232 (^{232}Th) series.

The standard AGRS processing consists in basic K, eTh, eU maps and ternary (K-eTh-eU) images to evaluate lithological information, and secondary maps to analyze features of interest (mineral exploration, environmental issues, among others). Secondary maps are based in operations between the radioelements. Usually, the geophysical - geological interpretation is based on visual pattern recognition. In order to evaluate this data with a quantitative analysis, we combined unsupervised clustering and principal component analysis of AGRS data.

Study area

We tested the proposed methodology in an area located in the State of Goiás (Fig.1), inserted the Aerogeophysical Project Mara Rosa Magmatic Arc (CPRM, 2004) illustrated in figure 1. The Mara Rosa Magmatic Arc (Fig.2) is inserted in the geological framework of the Tocantins Province, which corresponds to an expressive neoproterozoic (Brasiliano / Pan-African) orogenic zone. The Province eastern portion is occupied by the Brasília Belt, which comprises mainly a thick metasedimentary sequence and a large area where juvenile Neoproterozoic arc rocks are exposed (Goiás Magmatic Arc) (Silva et al., 2007).

The Goiás Magmatic Arc is a Neoproterozoic crustal accretion, subdivided in Arenópolis Arc and Mara Rosa Arc, southern and northern portions, respectively (Pimentel 2016).

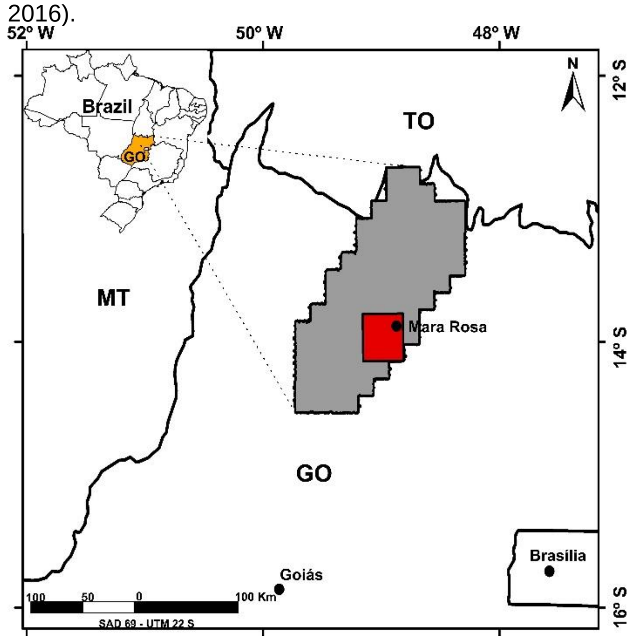

***Figure 1 –** Location of the study area in red, indicating the geophysical project in grey.*

The Mara Rosa Magmatic Arc consists of elongated volcano-sedimentary belts with NNE direction, individualized by orthogneisses with predominantly tonalitic (Arantes et al. 1991, Pimentel et al. 1997). The supracrustal sequences include expressive volume of mafic metavolcanic rocks, in addition to intermediate to felsic representatives, intercalated to metasedimentary rocks of detrital and chemical nature. The mineral paragenesis with granada, staurolite and cyanite indicate conditions of metamorphism in the greenschist to amphibolite facies (Della Giustina 2007).

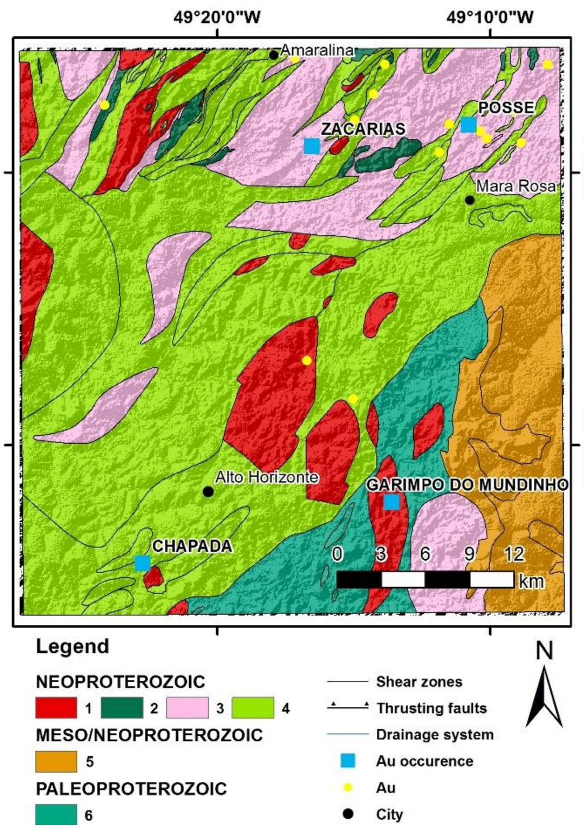

Figure 2 – *Geological map of the study area. 1 - Sin-Pos tectonic granites; 2- Gabbro-diorite; 3 - Orthogneiss-Biotite gneiss; 4 - Mara Rosa Volcano-Sedimentary Sequence; 5- Serra da Mesa group; 6 - Campinorte Volcano-Sedimentary Sequence.*

Method

Geological-geophysical maps based on AGRS data are usually produced by visual interpretation of a ternary image, which is a color composite image of K, eTh and eU in the red, green and blue channels respectively. Ternary images are used by geologists to manually map and delineate geological features of a given area. Geological mapping based on visual interpretation highly depends on the experience and knowledge of the interpreter and can be costly and operationally prohibitive to perform over large spatial extents. Thus, alternative techniques are needed to optimize the process and ensure reproducibility of results.

Clustering analysis is a pattern recognition technique that aims to separate data into a set of clusters, such that each cluster contains similar data. Unsupervised clustering is applied when there is no available information on the classes present in a given study area. The *k*-means algorithm is one of the first unsupervised clustering method developed (Jain, 2010) and has been applied on clustering of AGRS data (e.g., Lima & Marfurt, 2018). The *k*-means algorithm performs iterative clustering, which aims to separate the data into a set of *k* classes (clusters). It uses the distance of each sample to the cluster centroid as a classification function and aims to minimize the sum of the distances among the samples and the closest cluster centroid. The main disadvantage of *k*-means is that the user should provide beforehand the number of clusters to partitioning the data, which usually leads to erroneous results. Gaussian Mixture Models (GMMs) is an unsupervised clustering method that automatically estimates the number of clusters in a dataset and was poorly explored in the context of geological mapping.

The GMM algorithm aims to estimate the parameters of the Gaussian distributions associated with each cluster (mean vector and covariance matrix) and the respective mixture components. Thus, each cluster is modeled not only by its mean vector, which is the case of *k*-means, but also by the covariance matrix, giving the model more flexibility to fit the data. To use the GMMs, it is necessary to estimate the proportion of each class and the respective mean vectors and covariance matrices. The *expectation-maximization* (EM) algorithm (Friedman et al., 2001) is usually employed to this task, allowing the estimation of the Gaussian components in an iterative and automatic fashion. In a quantitative approach to estimate the number of classes using GMMs, the *Bayesian information criterion* (BIC) (Schwarz, 1978) is commonly used. The number of classes that provides the lowest BIC is considered the optimal value to cluster the data (Fig.3).

In this study, we performed unsupervised clustering of AGRS data using GMMs on principal component analysis of primary variables (K, eTh and eU). Principal Component Analysis (PCA) is a multivariate statistical method that was successfully used to classify gamma-ray spectrometry data (Dragović & Onjia, 2006). PCA performs an orthogonal mathematical projection operation to convert the original data features into new uncorrelated features called principal components (PCs).

Firstly, we have processed the data of the Mara Rosa Project (CPRM, 2004) resulting in basic maps - potassium, thorium and uranium. The ternary image (Fig. 4a) was obtained from these basic maps (K, eTh, eU). We transformed the ternary image (Fig. 4a) into a principal component image (Fig. 4b) and applied to it the GMM clustering method. Before applying the PCA transformation, we normalized the measured K, eTh and eU to the [0-1] range with the z-score method. This procedure was necessary because the primary variables varied in different ranges (K in percentage, eU and eTh in ppm). We then applied the GMM algorithm with classes ranging from one to ten, computing the BIC criteria at each realization. Finally, the PCA image was partitioned with the number of classes that showed the lowest BIC. The classified image was confronted against the geological map produced by visual interpretation to evaluate the potential of GMMs to automatically distinguish geological contacts in the study area.

Results and Discussion

The lowest BIC was obtained with seven classes (Fig. 3), thus we applied the GMMs with seven classes.

Ternary maps show the variation of K, eTh and eU at each location. The image is created by combining potassium (red), thorium (green) and uranium (blue) to form a single image. The intensity of the three different colors reflects the importance of the K, eTh and eU individually. Combination of colors indicate a combination of relative elements. White and black colors indicate high and low values in all three radionuclides.

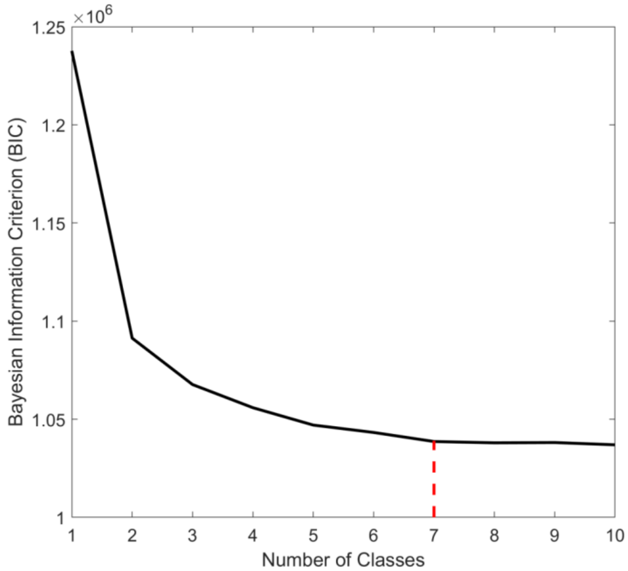

Figure 3 – *Bayesian information criterion (BIC) obtained using one to ten classes to cluster the PCA image (Fig. 4b) of the ternary (K-eTh-eU) (Fig. 4a).*

Figure 4 indicates the ternary image and its correlation with the color composite of PCA image. The PCA analysis allows a quantitative interpretation. It is possible to note the similarity between the two images, except for the colors.

Figure 5 indicates the final PCA classification with seven classes, and it shows a good correlation with geology. The granitic bodies (number 1 of Fig.5b) are well delineated and are mainly correlated with classes 1, 2, 5 and 7. The Gabbro-diorites (number 2 of Fig. 5b) are mainly related to class 1.

The orthogneisses (number 3 of Fig. 5b) are correlated mainly with classes 1 and 2, there is also a well-delineated body associated with class 3 in the northeast. Related to the Mara Rosa Sequence (number 4 of Fig. 5b) are observed classes the 1, 2 and 5.

The Serra da Mesa Group (number 5 of Fig.5b), has a particular signature that differs from all geological units in the area. It is mainly related to classes 4, 5 and 6. However, class 7 highlights one of its geological units. These classes may be related to the carbonate schist of the meta-psamo-pelitic unit.

The Campinorte Sequence (number 6 of Fig. 5b) consists of quartzites and a wide variety of mica schists. Within this larger unit are observed three smaller units (meta-rhyodacite, meta-rhyolite and schists), that apparently are related to the radioelements highest values. In these units, the predominant classes are 3, 5 and 7.

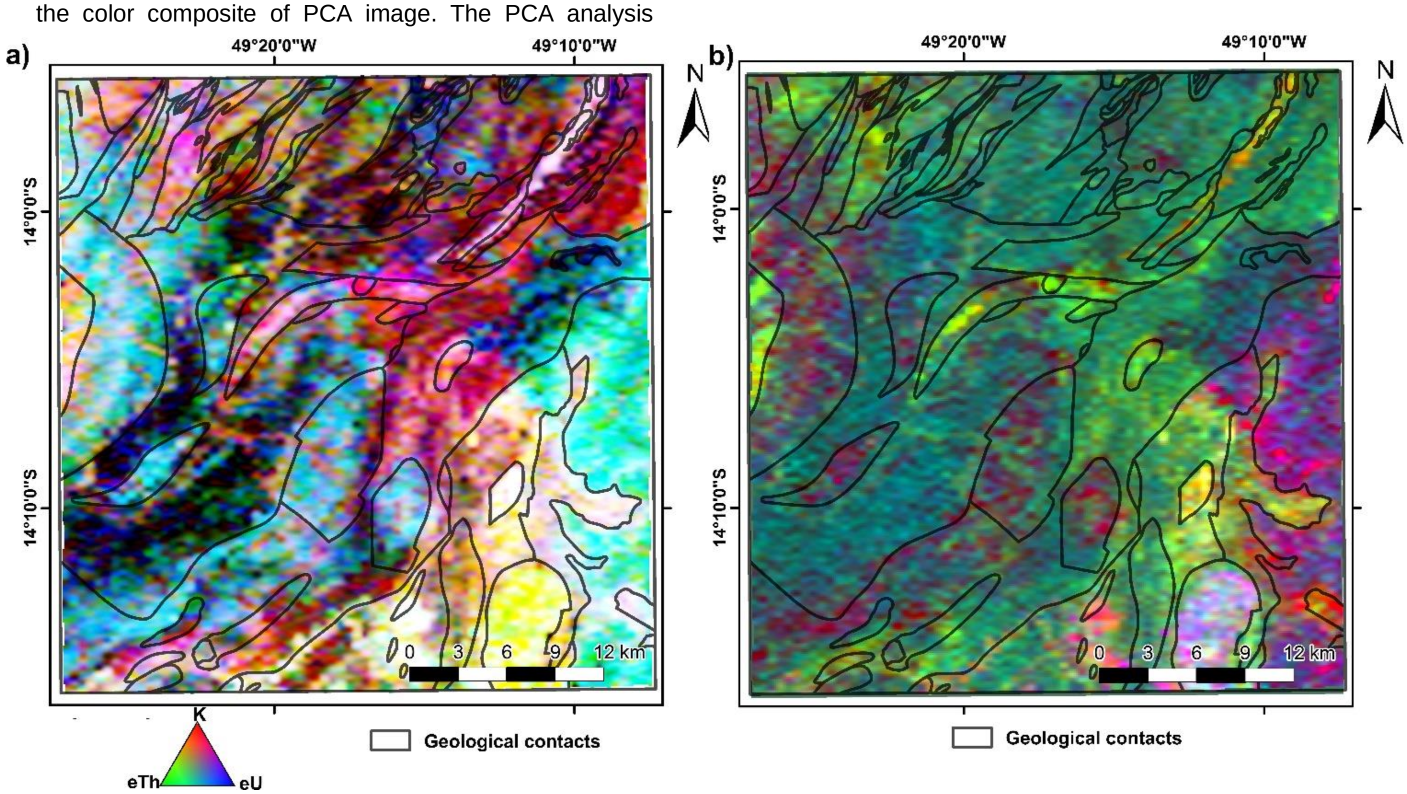

Figure 4 – *a) Ternary (K-eTh-eU) map, b) Color composite of PCA images (R=PC1, G=PC2, B=PC3).*

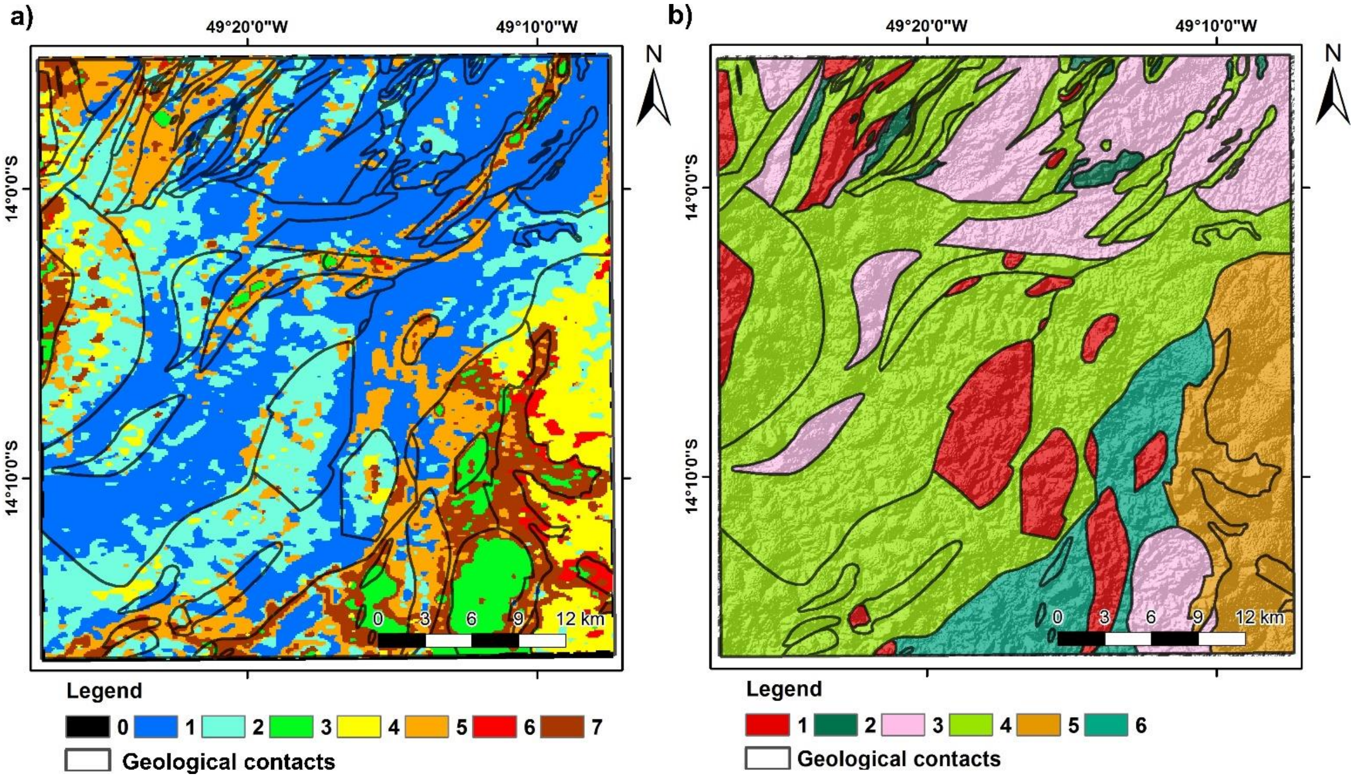

Figure 5 – *a) Classification map obtained with Gaussian Mixture Models and principal component analysis, b) Geological map.*

Some geological features are highlighted more than others, this could possibly be related to its mineralogical compositions and should be further investigated.

Conclusions

Gamma-ray spectrometry maps are correlated with the geology of the study area. Usually, the geophysical - geological interpretation is based on visual pattern recognition. The combination of unsupervised clustering and PCA to airborne gamma-ray spectrometry data allows a better quantitative analysis of the geological contacts of the study area. The GMM classification of PCA images has a direct correspondence with geology, and some geological features are well delineated by the obtained classes.

Acknowledgments

J.D. Weihermann and L.G. Castro were supported in this research by fellowships from Fundação Capes, Ministério da Educação. F.J.F Ferreira. and A.M. Silva were supported in this research by National Council for Scientific and Technological Development (CNPq, Brazil) (contracts 303826/2018-5, and 307177/2014-9 respectively). We thank S. P. Oliveira (DMAT-UFPR) for his suggestions throughout this work.

References

ARANTES, D., BUCK, P.S., OSBOURNE, G.A., PORTO, C.G. A Sequência Vulcanosedimentar de Mara Rosa e Mineralizações Auríferas Associadas. Boletim Informativo da SBG, Núcleo Centro-Oeste, 27-40, 1991.

CHARBONNEAU, B.W. Geochemical evolution and radio-active mineralogy of the Fort Smith radioactive belt, Northwest Territories, Canada. Primary Radioactive Minerals (The Textural Patterns of Radioactive Mineral Paragenetic Associations). Theophrastus Publications, Athens, Greece, pp. 21–48, 1991.

CPRM. Levantamento aerogeofísico do Estado de Goiás – Arco Magmático de Mara Rosa – Relatório Final, Volume I – Texto Técnico, 2004.

DELLA GIUSTINA, M. E. S. Geologia e geocronologia da sequência Vulcano-sedimentar Campinorte e do Complexo Uruaçu, província Tocantins. Dissertação de Mestrado - Universidade de Brasília, Brasília. 106pp, 2007.

DRAGOVIĆ, S.; ONJIA, A. Classification of soil samples according to their geographic origin using gamma-ray spectrometry and principal component analysis. Journal of environmental radioactivity, v. 89, n. 2, p. 150-158, 2006.

DICKSON, B.L., SCOTT, K.M. Interpretation of aerial gamma-ray surveys — adding the geochemical factors. AGSO J. Aust. Geol. Geophys. 17 (2), 187–200, 1997.

FRIEDMAN, J, HASTIE, T, TIBSHIRANI, R. The elements of statistical learning, vol. 1. Springer series in statistics New York, NY, USA, 2001.

GRAHAM, D.F., BONHAM-CARTER, G.F. Airborne radio-metric data: a tool for reconnaissance geological mapping using a GIS. Photogram. Eng. Remote. Sens. 58 (8), 1243–1249, 1993.

IAEA. Guidelines for Radioelement Mapping Using Gamma Ray Spectrometry Data.Technical Report Serie, n° 1363. International Atomic Energy Agency, Vienna. 179 pp., 2003.

JAIN, A.K. Data clustering: 50 years beyond K-means.Pattern recognition letter**s**, v. 31, n. 8, p. 651-666, 2010.

DE LIMA, R.P.; MARFURT, K.J. Principal component analysis and K-means analysis of airborne gamma-ray spectrometry surveys. In:SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists, p. 2277-2281, 2018.

MINTY, B.R.S. Fundamentals of airborne gamma-ray spectrometry. AGSO J. Aust. Geol. Geophysics. 17 (2), 39–50, 1997.

SILVA, A.M., OLIVEIRA, C.G., MARQUES, G.C., PIRES, A.C.B. Geophysical responses of hydrothermal rocks associated with copper-gold mineralization in the Neoproterozoic Mara Rosa Magmatic Arc, central Brazil. In:Proceedings of the fifth decennial international conference on mineral exploration, 2,1179-1183, 2007.

SCHWARZ G. Estimating the Dimension of a Model. Annals of Statistics 6: 461–464, 1978.

SHIVES, R.B.K., CHARNONNEAU, B.W., Ford, K.L. 2000. The detection of potassic alteration by gamma-ray spectrometry – Recognition of alteration related to mineralization. Geophysics, 65:2001, 2011.

PIMENTEL, M. M. The tectonic evolution of the Neoproterozoic Brasília Belt, central Brazil: a geochronological and isotopic approach. Brazilian Journal of Geology, 46, 67-82, 2016

PIMENTEL, M.M., WHITEHOUSE, M.J., VIANA, M.G., FUCK, R.A., MACHADO, N. The Mara Rosa Arc in the Tocantins Province:further evidence for Neoproterozoic crustal acretion in central Brazil. Precambrian Research, 81. p.299-310, 1997.

WILFORD, J. Airborne gamma-ray spectrometry. In: Papp (Ed.), Geophysical and Remote Sensing Methods for Regolith Exploration. CRCLEME Open File Report 144, pp. 46–52, 2002.

WILFORD, J.R., BIERWIRTH, P.N., CRAIG, M.A. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO J. Aust. Geol. Geophysics. 17 (2), 201–216, 1997.