Geological Characterization by Applying Automatic Clustering to Multiple Geophysical Inversions

Melo, A.T.[1, 2], Li, Y. [2]

1. Departamento de Geologia, Universidade Federal de Minas Gerais, Brazil
2. Department of Geophysics, Colorado School of Mines, United States

ABSTRACT

Geophysical inversions in mineral exploration have become fundamental to understanding physical property distribution in the 3D subsurface and as a means of quantitative geophysical interpretation. The interpretation process aiming to identify geologic units from geophysical inversions, however, is not straightforward due to limitations in representing the geology with physical properties. While progress has been made on incorporating geological knowledge into inversions, greenfield exploration still faces the challenges from the lack of a priori *knowledge. To reduce the impact of ambiguity of interpretation and reduce the drilling risk in the initial stages of exploration, methods are needed to connect geophysical inversions to subsurface geology. We present the results of applying automatic clustering to identify ore from multiple inverted physical property models in an iron-oxide copper gold (IOCG) deposit. Unconstrained independent inversions of magnetic, gravity gradient, and DC resistivity data were performed to construct 3D models of susceptibility, density, and conductivity, respectively. The results of the inversions show that the geophysical response of our target is obscured by the stronger response of an adjacent iron formation. We apply k-means clustering to automatically identify clusters in the inverted physical property values to aid in our interpretation. First, we propose the use of the L-curve criterion from inverse theory to plot the k-means objective function versus number of clusters to select the optimum value for the latter. This optimum number of clusters is then used as input in the correlation-based clustering, which has a more robust measure of similarity. The clustering result so derived shows a good spatial correspondence with the known geology from drilling information, and the method is able to identify the ore zone. The proposed method is entirely data-driven and has proven to work in a complex geological setting such as the Cristalino copper deposit.*

INTRODUCTION

The increasing use of geophysical inversions in mineral exploration has resulted in great successes in selecting and understanding prospects for drilling. These prospects commonly become associated with new ore discoveries, and the chance of success increases with the amount of geological knowledge available. In brownfield exploration, *a priori* geological information improves geophysical models. On the other hand, such *a priori* information is rarely available in greenfield exploration, which often only has regional geologic maps or nearly nothing if the area is covered by overburden. Consequently, the geophysical models are more influenced by non-uniqueness, and therefore face greater drilling risk.

Geology characterization, the process of identifying associations between geophysical bodies and different geologic units, becomes more difficult in areas with little auxiliary geological knowledge. For this reason, in this paper, we use multiple geophysical methods in a greenfield exploration targeting aiming to decrease the uncertainty in interpretation. However, appropriate techniques are necessary to produce integrated interpretations from multiple data. Therefore, we propose the use of automatic clustering to extract consistent information with geological meaning from multiple geophysical inversions.

In our present work, we first give a brief overview of some geology characterization methods. We then, present the method we have proposed, which consists of two parts: the application of k-means clustering to define the optimum number of clusters and the application of correlation-based clustering to refine the results. We demonstrate the success of this method on Cristalino iron-oxide copper gold (IOCG) deposit in northern Brazil. In Cristalino deposit, we use the correlation between three physical properties recovered from deterministic geophysical inversions to identify copper ore. We first build 3D models of susceptibility, density and conductivity from minimally constrained 3D inversions of magnetic, gravity gradient and DC resistivity data, respectively. We next identify the optimal number of clusters of recovered physical properties by applying k-means clustering and exploring the relationship between its objective function and the number of clusters. Although the use of k-means clustering provided an output that correlates well with the known geology, this method favors the identification of spherical clusters and the result was highly influenced by irrelevant attributes. To improve clustering of the physical properties, we used the defined optimum number of clusters to apply correlation-based clustering, which is a robust clustering technique capable of identifying clusters arbitrarily oriented in the parameter space of physical properties. We identified the copper ore cluster without the input of *a priori* information and, thereby, simulate a greenfield exploration process.

GEOLOGY CHARACTERIZATION METHODS

Extracting geological meaning from geophysical inversions is challenging, especially in areas with little auxiliary geological

In "Proceedings of Exploration 17: Sixth Decennial International Conference on Mineral Exploration" edited by V. Tschirhart and M.D. Thomas, 2017, p. 739–743

knowledge. Methods for geological characterization from geophysical models have been shown to benefit interpretation. Williams et al. (2004) use empirical relations to identify alteration zones from geophysical inversions on regional scale. Further exploring this method, Williams and Dipple (2007) apply a mineralogy unmixing technique through linear programming to define alteration zones. These methods require petrophysical relationships suitable for the type of target being investigated.

In the absence of prior geologic information at a site, Kowalczyk et al. (2010) use the cross plot of density versus susceptibility to define classes of different lithologies in regional scale. Martinez et al. (2011) and Martinez and Li (2015) take this approach further by applying specific ranges of density and susceptibility to assign specific lithology types. Melo et al. (2015) use a similar approach to identify copper mineralization guided by the trends in the cross plot of susceptibility and conductivity. The application of this method is feasible when two physical properties are being used.

When more physical properties are added to the interpretation, the dimensionality makes manual interpretation difficult. To overcome this difficulty, Paasche et al. (2006) apply fuzzy c-means clustering to automatically identify clusters in recovered physical property models and well log data to construct an integrated model. Fraser et al. (2012) apply self-organizing maps (SOM) to physical property models from multiple geophysical inversions. In their particular example, SOM identifies eight clusters, but not all of them have a direct spatial correlation with the mapped geological units; thus, this number of clusters seems to be over-fitting the data. Different clustering techniques use specific measures of similarity, and most of them require the number of clusters as a user input. Interpreters using automatic clustering techniques need to adequately choose the optimal number of clusters without overfitting the data. The correct choice of cluster number ensures the identification of all important groups, avoiding the formation of meaningless clusters.

COPPER DEPOSIT EXAMPLE

Cristalino (482 Mt @ 0.65% Cu and 0.06 g/t Au (NCL Brasil, 2005)) is a world class IOCG deposit located in the Carajás Mineral Province, a highly mineralized metallogenic region in Brazil. The copper deposit is hosted by a splay of the Carajás Fault, which is a major crustal fault. This splay fault cuts through a volcano-sedimentary sequence formed by iron formation interlayered with mafic and felsic volcanic rocks (Figure 1). This sequence is dipping approximately 60º to southwest, parallel to the fault plane that acted as a conduit for hydrothermal fluids. Furthermore, the mineralogy of each zone is highly variable depending on the host rock type (Huhn et al., 1999).

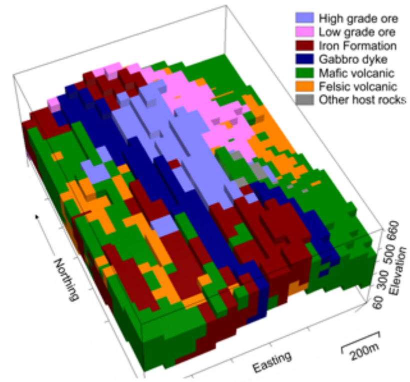
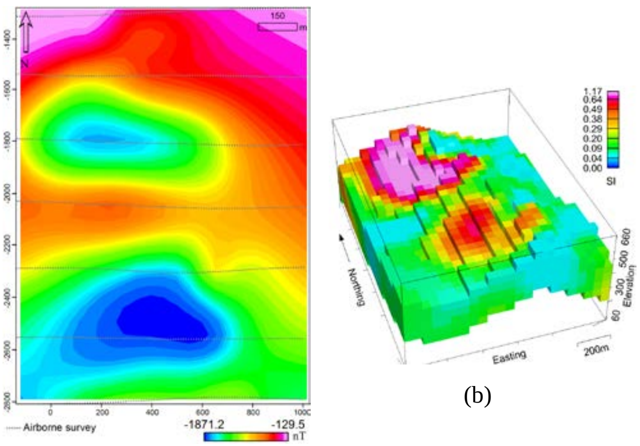


Figure 1: 3D geological model of Cristalino copper deposit constructed from 303 drillholes (adapted from Vale S.A., 2004).

The geological characterization scheme presented in this study used magnetic, gravity gradient and DC resistivity data over the deposit. The data corresponding to each geophysical method were independently inverted to build susceptibility, density and conductivity models. We used the 3D potential field inversion algorithm developed by Li and Oldenburg (1998) to invert the magnetic data, Li (2001) for the gravity gradient data and Li and Oldenburg (2000) for the DC resistivity data. The inverse solutions were obtained using Tikhonov regularization. The data of each geophysical method were inverted using the same mesh of cubic cells of 50x50x50 m to ensure the spatial compatibility. We performed a regional-residual separation using an inversion-based method (Li and Oldenburg, 1998) on the aeromagnetic data (Figure 2a). The recovered susceptibility model (Figure 2b) shows two magnetic bodies associated with the iron formation. The inversion of the gravity gradient data used all components. The recovered density contrast model (Figure 3) shows two anomalies of high density associated with the iron formation and between them a body associated with ore. The DC resistivity data were acquired along parallel east-west lines, using a dipole-dipole array with a dipole separation of 60 m and 6 n-spacings. The main anomaly in the recovered conductivity model (Figure 4) is located in the central part of the mesh, which is spatially coincident with the copper ore.

Defining the Number of Clusters

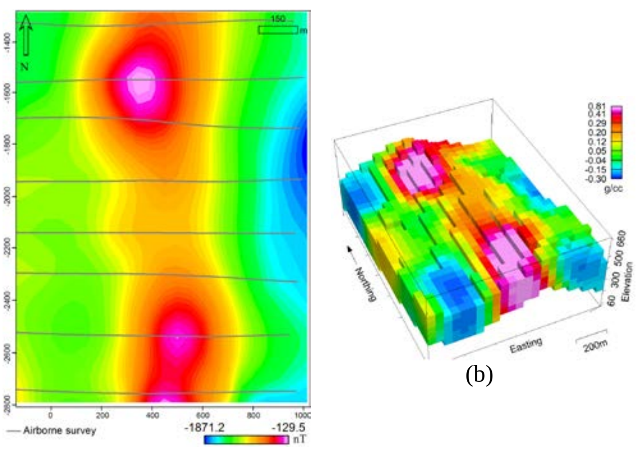
Most clustering techniques require the interpreter to provide the number of clusters. To address this issue, our first step is to define the number of main groups the combination of the three physical properties shows. We choose to use k-means clustering to test its consistency in clustering with different random initializations of cluster centres. K-means clustering minimizes the objective function of the squared distance between the points and cluster centres. Our goal is to find the number of clusters k that minimizes this objective function while keeping meaningful clusters. A procedure similar to the L-curve criterion in inverse theory is applied, where we pick as the optimum number of clusters the point of maximum curvature of the curve of k versus objective function plot (Figure 5). In this case $k = 4$ is the optimum value, and the curve shows that adding more than 4 clusters only decreases the k-means objective function slightly. The spatial distribution of the k-means clustering with $k = 4$ (Figure 6) shows that the clusters are associated with: i) iron formation, ii) high-grade ore, iii) host rocks, and iv) noise. The low-grade ore appears as host rock. The clusters associated with the iron formation and high-grade ore are robust and practically do not change if we increase to k = 5. The main difference between the two clustering results is in the host rock region, which becomes further subdivided.



(a)

(b)

Figure 2: (a) Residual magnetic anomaly obtained from inversion-based regional-residual separation, and (b) inverted 3D susceptibility model. The inducing field has an inclination of -3.5º, declination of -19º, and a strength of 25,500 nT.



(a)

(b)

Figure 3: (a) Tzz component of the gravity gradient data terrain corrected with density of 2.45 g/cc, and (b) inverted 3D density model.

Correlation-Based Clustering

After defining the optimum number of clusters, we propose the application of a clustering method less susceptible to irrelevant attributes. K-means clustering is more appropriate for spherical clusters, but the clusters in the recovered physical property models in this study have linear trends (Figure 6). For this reason, a different measure of similarity was applied. A robust method for detecting arbitrarily oriented clusters is the correlation-based clustering technique.

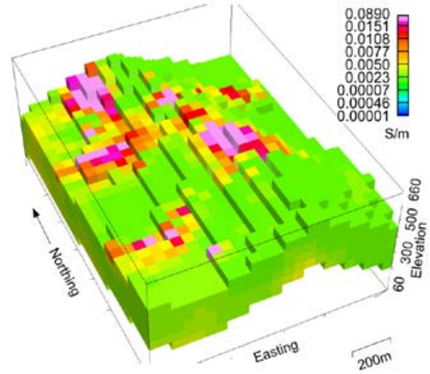


Figure 4: 3D conductivity model.

Figure 5: Objective function versus the number of clusters showing the optimal number of clusters k = 4 in the red circle.

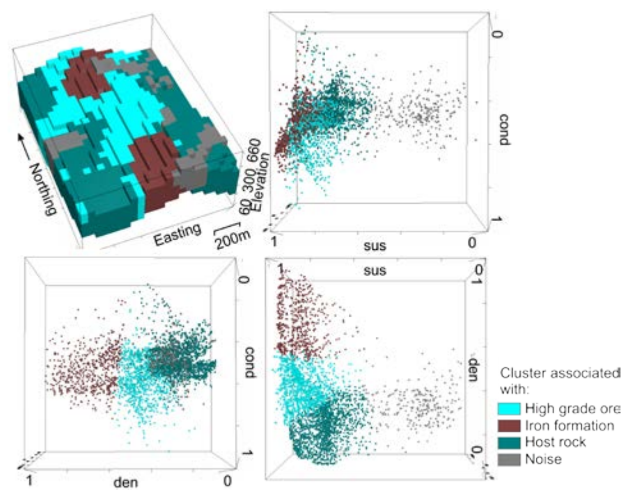


Figure 6: 3D plot of the result of k-means clustering with k = 4, and 3D scatter plots of the physical properties recovered from inversion color coded by the cluster type, showing the physical properties of each cluster.

This method was originally designed for clustering high dimensional data, but its concept of similarity applies to our problem because it searches for common correlations between points. We used the algorithm ORCLUS (arbitrarily ORiented projected CLUSter generation, Aggarwal and Yu (2000)), which does a principal component analysis of the data. The algorithm initializes with a large number of initial points or seeds sampled from the data, and applies k-means clustering to form the initial set of clusters. In the next step, it computes the covariance matrix of the points within each cluster, applies eigenvalue decomposition, and selects only the orthonormal eigenvectors with the least spread. Then it evaluates pairs of clusters and decides if two clusters fit into the same pattern of behavior, if so they are merged into a single cluster. Clusters that fit in the same pattern have a small projected energy (sum of the distances, projected on a plane, between points). The projected energy acts as an implicit objective function. The algorithm iteratively merges similar clusters until the user input number of clusters is reached.

The spatial distribution of the correlation-based clustering with k = 4 (Figure 7) shows an improved result compared with k-means clustering, forming spatially more compact clusters that are less influenced by irrelevant attributes from smoothness in the models. For example, in the southwestern area of the model, k-means clustering (Figure 6) classify some cells as belonging to the cluster associated with high-grade ore while correlation-based clustering does not. These cells lack any association with the corresponding geophysical signature of high-grade ore (high conductivity, intermediate density, and low susceptibility). The result is also less influenced by isolated anomalies, when only one physical property shows anomaly. The main improvement was in the higher spatial correlation with the high-grade ore (Figure 8).

Although the proposed method for selecting the optimum number of clusters proved to yield an estimation of k that is highly correlated with the known geology, this number should be taken as an initial estimation, which is important to avoid under- or over-fitting the data. In addition, one should explore the results of applying correlation-based clustering using k values close to the defined optimum number of clusters. This procedure might be able to identify clusters that are correlated to geological features, but were not identified by k-means clustering due to the difference in the measure of similarity.

CONCLUSIONS

In this work, we used deterministic inversions of magnetic, gravity gradient and DC resistivity data to identify the geophysical response of the geological units in the Cristalino deposit. We showed that the iron formation has the most anomalous values of susceptibility and density, and intermediate values of conductivity, making it an easy unit to identify. On the other hand, the high-grade ore anomaly has a mixed signature, with susceptibility varying over a wide range, density varying from intermediate to high values, and high conductivity. Therefore, DC data were necessary to identify this unit, which is our main target. Otherwise, it would not stand out from the anomalies associated with the iron formation. Although the use of this third physical property was fundamental for identifying our target, using more than two physical properties makes manual interpretation difficult. Consequently, we applied automatic clustering to our inversion results.

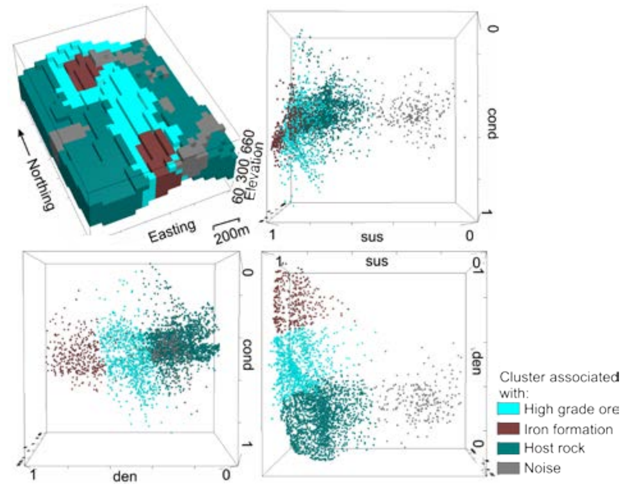
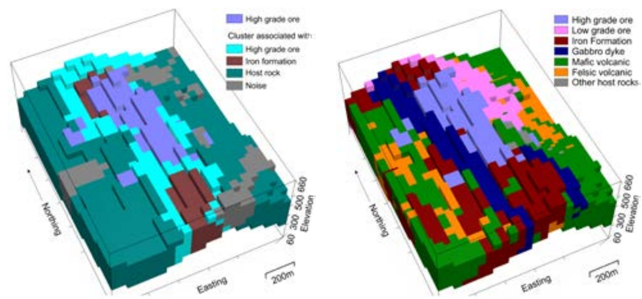


Figure 7: 3D plot of the result of correlation-based clustering with k = 4, and 3D scatter plots of the physical properties recovered from inversion color coded by the cluster type, showing that the result is less influenced by irrelevant attributes.



(a) (b)

Figure 8: Comparison between (a) the result of correlation-based clustering with k = 4 superimposed by the high-grade ore unit, and (b) the 3D geological model, showing the high spatial correspondence between clustering and known geology.

We demonstrated that applying an L-curve criterion to the relationship between the objective function and number of clusters helps define the optimal number of clusters objectively. However, as k-means clustering favors spherical clusters, the results are noisy and heavily influenced by irrelevant attributes from the smoothness in the models. Therefore, we applied a clustering method with a more robust measure of similarity. We used the defined optimum number of clusters to apply correlation-based clustering, which looks for maximum correlations between points and is capable of finding clusters with arbitrary orientations. The result of correlation-based clustering was less influenced by irrelevant attributes and improved the spatial association with the known geology. One of the known geological units, the low-grade ore, could not be identified in this process because the physical property contrast is small and the inverted models were not able to reproduce it. This method is entirely driven by the data and proved to work in a complex geological setting such as Cristalino. The work flow presented here can be applied to many types of target in greenfield exploration to increase the understanding and confidence in the drilling planning stages, decreasing the decision risk.

AKNOWLEDGEMENTS

The first author acknowledges the CNPq (Brazilian National Council for Scientific and Technological Development) for the sponsorship. We thank Vale S.A. for providing the data used in this study and for permission to publish results obtained from these data. We also thank all CGEM members, especially Jiajia Sun for discussions. This work is supported in part by the Gravity and Magnetics Research Consortium (GMRC).

REFERENCES

Aggarwal, C. C. and P. S. Yu, 2000, Finding generalized projected clusters in high dimensional spaces: SIGMOD Rec., 29, 70–81.

Fraser, S. J., G. A. Wilson, L. H. Cox, M. Cuma, M. S. Zhdanov, and M. A. Vallée, 2012, Self-organizing maps for pseudo-lithological classification of 3D airborne electromagnetic, gravity gradiometry and magnetic inversions: Australian Society of Exploration Geophysicists, Extended Abstracts 2012, 4 p.

Huhn, S., C. D. J. Souza, M. D. Albuquerque, E. Leal, and V. Brustolin, 1999, Descoberta do depósito Cu (Au) Cristalino: geologia e mineralização associada-região da Serra do Rabo - Carajás-PA: SBG/Núcleo Norte, Simpósio de Geologia da Amazônia, 6, 140–143.

Kowalczyk, P., D. Oldenburg, N. Phillips, T. Nguyen, and V. Thomson, 2010, Acquisition and analysis of the 2007-2009 Geoscience BC airborne data: Expanded Abstracts of Australian Society of Exploration Geophysicists – Petroleum Exploration Society of Australia Airborne Gravity Workshop. ASEG-PESA Airborne Gravity Workshop, Canberra, Geoscience Australia, Canberra, 115–124.

Li, Y., 2001, 3-D inversion of gravity gradiometer data: 71st Annual International Meeting, SEG, Expanded Abstracts, 1470–1473.

Li, Y. and D. W. Oldenburg, 1998, Separation of regional and residual magnetic field data: Geophysics, 63, 431–439.

Li, Y. and D. W. Oldenburg, 2000, 3-D inversion of induced polarization data: Geophysics, 65, 1931–1945.

Martinez, C. and Y. Li, 2015, Lithologic characterization using airborne gravity gradient and aeromagnetic data for mineral exploration: A case study in the Quadrilátero Ferrífero, Brazil: Interpretation, 3, SL1–SL13.

Martinez, C., Y. Li, R. Krahenbuhl, and M. Braga, 2011, Lithologic characterization using magnetic and gravity gradient data over an iron ore formation in the Quadrilátero Ferrífero, Minas Gerais, Brazil: Presented at the 12th International Congress of the Brazilian Geophysical Society.

Melo, A. T., J. Sun, and Y. Li, 2015, Geophysical inversions applied to geological differentiation and deposit characterization: A case study at an IOCG deposit in Carajás Mineral Province, Brazil: 85th Annual International Meeting, SEG, Expanded Abstracts, 2012-2016.

NCL Brasil, 2005, Revision de la estimación de recursos del proyecto Cristalino: Technical report, Vale S.A. Internal Report.

Paasche, H., J. Tronicke, K. Holliger, A. G. Green, and H. Maurer, 2006, Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses: Geophysics, 71, H33–H44.

Vale S.A., 2004, Relatório de reavaliação de reservas estudo de pré-viabilidade projeto Cristalino: Technical report, Vale S.A. Internal Report.

Williams, N. and G. Dipple, 2007, Mapping subsurface alteration using gravity and magnetic inversion models, in B. Milkereit, ed., Proceedings of Exploration 07, 461–472.

Williams, N. C., R. Lane, and P. Lyons, 2004, Towards 3D maps of alteration under cover: Regional constrained 3D inversion of potential field data from the Olympic Cu-Au province, South Australia: ASEG Extended Abstracts 2004, 4 p.