

Geology differentiation of geophysical inversions using machine learning

Aline Tavares Melo^{*1,2} and Yaoguo Li²

¹Departamento de Geologia, Universidade Federal de Minas Gerais, Brazil

²Department of Geophysics, Colorado School of Mines, United States

Summary

Multiple geophysical methods are often employed to improve subsurface understanding, especially in areas with little a priori geological information. Therefore, quantitative methods for integrating multiple physical property models are fundamental to taking the interpretation further into geology differentiation of distinct units. Hence, applications of machine learning are growing in geosciences due to its potential to integrate various sources of information. We evaluate the performance of density-, distribution-, centroid-, and correlation-based clustering methods in the identification of the three geologic units in density, susceptibility and conductivity models derived from a synthetic model, and show that correlation-based clustering gives the best results for geology differentiation. We apply the method to physical property models recovered from field data over a copper deposit and the results show a good spatial correspondence with the known geology from drilling information, allowing the construction of a quasi-geology model.

Introduction

Different geophysical methods probe subsurface geology through different physical phenomena and they collectively provide improved understanding of subsurface. Therefore, effective quantitative methods for integrating multiple inverted physical property models are necessary to extract the maximum amount of information and advance the interpretation further into differentiation of geologic units. Geology differentiation is the process of identifying associations between geophysical units and different geological units to improve interpretation. To accomplish this, machine learning (ML) provides an effective and important means due to its potential to improve interpretation of information from multiple sources of data. However, different from other fields where machine learning is already being used as a tool to make interpretations more accurate and consistent with all data, such approaches are still in the initial stages in mineral exploration. In brownfield exploration, supervised machine learning has been applied to train algorithms for the identification of new targets. Unfortunately, the lack of training data for greenfield exploration primarily allows the application of unsupervised machine learning at present, which has the ability of exploring hidden structures in the data. Therefore, considering a greenfield exploration scenario, where specific a priori geologic information is unavailable, we investigate the application of unsupervised machine learning in geology

differentiation based on independent minimally constrained inversions of multiple geophysical data sets.

We first examine the interpretation of magnetic, gravity gradient, and DC resistivity data over a synthetic geologic model, which is inspired by the Cristalino iron oxide copper gold deposit, in northern Brazil, and has three main units: the copper ore, iron formation, and mafic volcanic host rock. The inverted susceptibility, density, and conductivity models are used for geology differentiation by applying unsupervised machine learning, more specifically, clustering algorithms. We evaluate the performance of density-, distribution-, centroid-, and correlation-based clustering methods in the identification of the three geologic units. We show that correlation-based clustering yields the best results for the geology differentiation, and the resulting integrated model can be interpreted using references of ore deposit models to form a quasi-geology model. We then apply correlation-based clustering to the geophysical models derived from inversion of field data at Cristalino deposit and show the strong correspondence between the quasi-geology model constructed through the geology differentiation method using machine learning presented here and the geology model constructed from drilling.

Methodology

The synthetic geology model constructed for this study (Figure 1) is based on Cristalino iron oxide copper gold (IOCG) deposit, in northern Brazil. The copper deposit is hosted by iron formation interbedded with volcanic rocks. The deposit formed by hydrothermal fluids which were transported through the fault that cuts the whole sequence and reacted with the magnetite of the iron formation. The process consumed magnetite and converted it to form the chalcopyrite of the copper ore (Huhn et al., 1999). For this reason, the iron formation pinches out where the deposit has its maximum thickness. Although the rock layers in Cristalino dip 50° to west, our synthetic models have vertical bodies for simplicity.

The synthetic susceptibility, density, and conductivity models corresponding to the synthetic geological model were used to forward model magnetic, gravity gradient, and DC resistivity data, respectively. Without loss of generality, we simulate ground surveys. The magnetic and gravity gradient data are co-located. The data separation is 50 m in the east direction, 75 m in the north direction, and 1.5 m above ground. The Earth's magnetic field was assumed to be the same as that field in the low-latitude region with field

Geology differentiation using machine learning

strength of 25,000 nT and zero inclination and declination (similar to the field in northern Brazil). The DC resistivity data follow the flat topography and has line spacing of 100 m in the north direction and station spacing of 25 m in the east direction. The survey configuration is dipole-dipole array with a 50-m electrode separation and 8 n-spacings. Uncorrelated Gaussian noise was added to the magnetic, gravity gradient, and resistivity data. We used the 3D potential field inversion algorithm developed by Li and Oldenburg (1996, 2003) 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 was inverted using the same mesh of cubic cells of 50x50x50m to ensure the spatial compatibility.

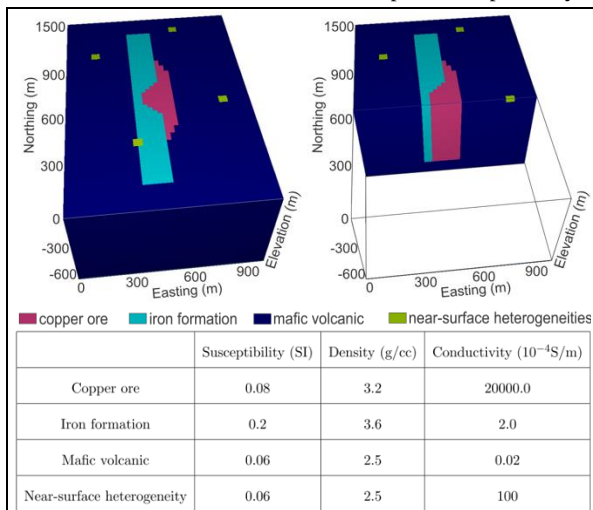


Figure 1: Synthetic geologic model based on Cristalino deposit showing the geologic units and their physical property values. The conductivity and density values were based on Telford et al. (1990). The susceptibility values were based on Clark and Emerson (1991).

The recovered susceptibility model (Figure 2a) shows two magnetic bodies that are associated with the two segments of the iron formation unit. The recovered density model (Figure 2b) also shows two main anomalies that are coincidental with the two segments of the iron formation. In addition, there is one anomaly of moderate density values associated with the copper ore. The recovered conductivity model (Figure 2c) has the main anomaly located in the central part of the model, which is spatially coincidental with the copper ore. The other anomalies of high conductivity over the area are related to the conductive overburden and are limited to the shallow layer only.

Little a priori geologic information is usually available in greenfield exploration under cover. Therefore, the construction of 3D models from geophysical data is through

minimally constrained inversions, and the resulting models are strongly influenced by the imposed smoothness of physical property values. For this reason, the change of physical properties between different units do not show sharp boundaries, but instead it is gradual and leads to inversion artifacts (Figure 2d). Given this common scenario for explorationists and the need of constructing integrated models for geology differentiation, we apply unsupervised machine learning (ML) to explore the structure contained in the model values and identify meaningful relations between physical properties to map regions of different geologic units. In unsupervised ML, clustering is the process of identifying patterns by grouping similar objects according to their attribute values. In our study the objects are the cells of the 3D models and the attributes the physical property values of each cell (susceptibility, density, and conductivity). The objective is to find the best grouping of attributes, or segmentation of crossplot, that corresponds to the units present in the geologic model. Ultimately, we are looking for the segmentation in the parameter domain, that corresponds to geological units in the space domain.

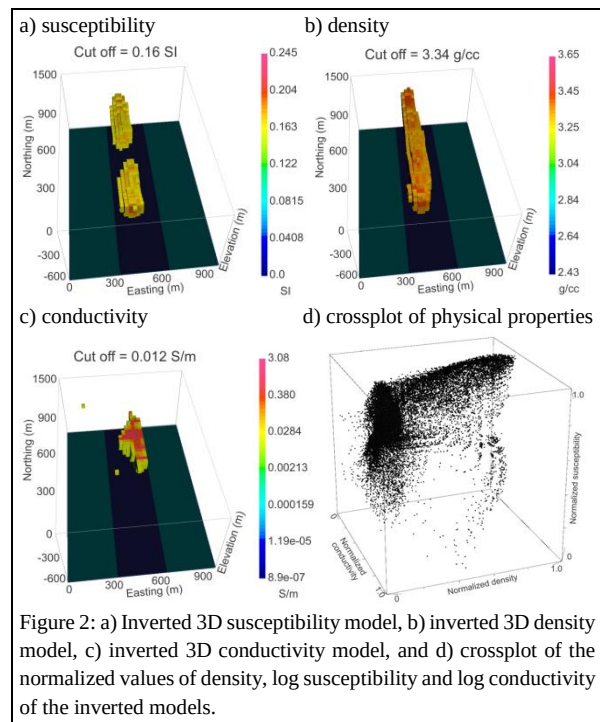
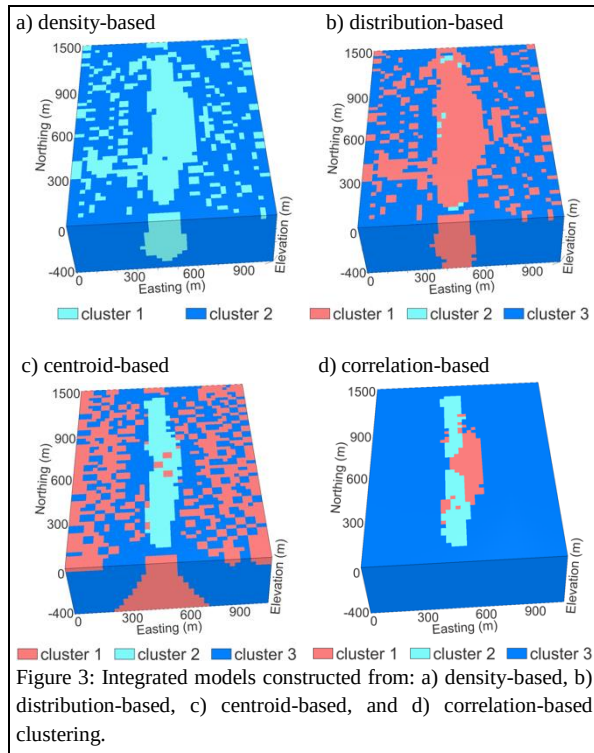


Figure 2: a) Inverted 3D susceptibility model, b) inverted 3D density model, c) inverted 3D conductivity model, and d) crossplot of the normalized values of density, log susceptibility and log conductivity of the inverted models.

The measure of similarity among objects is based on the similarity between the attributes of each object. In other words, the measure of similarity among cells of a model is based on the similarity between the physical properties of each cell. Clustering algorithms measure the similarity between attributes based on the distance between them in the

Geology differentiation using machine learning

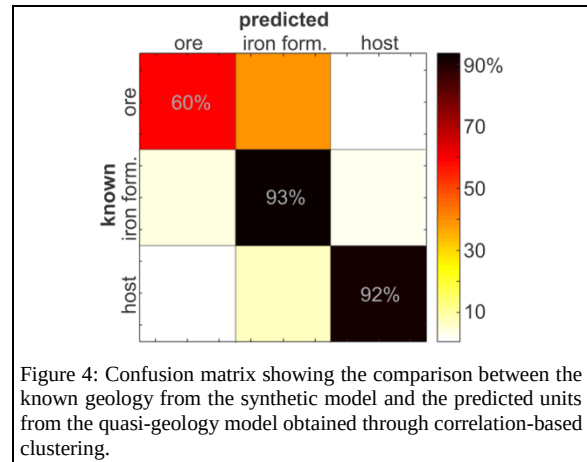
parameter domain. Different algorithms will measure this distance based on different metrics. Therefore, the choice algorithm should be compatible with the characteristics of the data being segmented. In our study, density-based clustering (Figure 3) could not identify clusters that correspond to the geologic units, and perform geology differentiation, because their classification is based on the distance between groups. Therefore, these methods require that different groups should be separated by a gap, and this is not a characteristic of the physical property models, where the change is gradual. Another option of measure of similarity is the statistical distribution of the clusters, which is the basis for distribution-based clustering algorithms. The statistical distribution needs to be known a priori, otherwise it becomes a strong assumption for the data. Here, the assumption of a Gaussian distribution worked well to identify the volcanic host rock and iron formation of the synthetic model, because their recovered physical property distributions are close to Gaussian. On the other hand, the assumption of a Gaussian distribution did not work for identifying the copper ore unit that has a bimodal distribution and, as a result, the unit became very noisy.



The application of centroid-based clustering also requires another strong assumption, because its good performance depends on the sphericity of clusters in the data. Therefore, the result will not be reliable if the clusters have linear distributions. In our study, the cluster corresponding to the

copper ore only identifies its core and incorporates inversion artifacts. On the other hand, correlation-based clustering has shown the best result in mapping all three geologic units. It successfully finds the subspaces of maximum correlation between physical properties for each geologic unit and is minimally influenced by inversion artifacts.

The confusion matrix (Figure 4), which compares each predicted cell of the 3D model with the known unit which they belong to in the synthetic model, shows that 60% of the copper ore cells were classified as ore, while 35% as iron formation. Some copper ore cells were classified as iron formation because the smooth magnetic inversion overlaps the copper ore in the interface between units, where the ore is thin, and the susceptibility parameter dominates in the classification. A total of 93% of the iron formation and 92% of the volcanic host rock cells were correctly predicted by the quasi-geology model.



Cristalino 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 5). This sequence is dipping approximately 50° to southwest, parallel to the fault plane that acted as a conduit for hydrothermal fluids (Huhn et al., 1999).

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 in the same way the synthetic data in the previous section. Then,

Geology differentiation using machine learning

correlation-based clustering was applied to perform geology differentiation in the three models (Figure 6). The clusters derived through this process were then used to map cells in the inversion models into different geologic units to build a quasi-geology model and achieve the desired geology differentiation. When applying to the well-studied Cristalino IOCG deposit, the results show a good spatial correspondence with the known geology from drilling information, and the method is able to identify the spatial location and extent of the copper ore unit. Although no prior information from drilling is used, the quasi-geology model from the unsupervised clustering analyses shows a 62% to 64% spatial match with known 3D geological model.

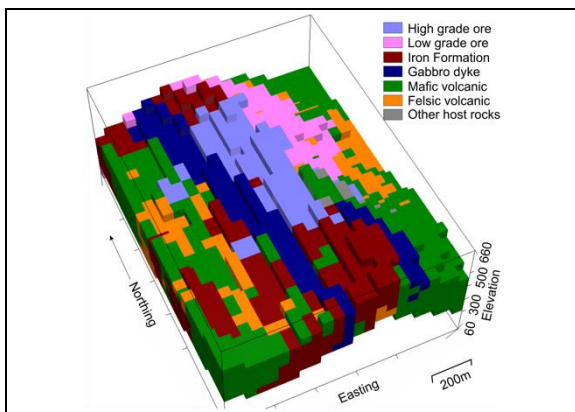


Figure 5: 3D geological model of Cristalino copper deposit constructed from 303 drillholes (adapted from Vale S.A., 2004).

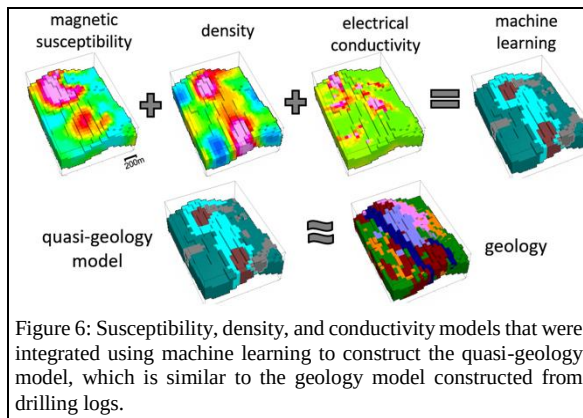


Figure 6: Susceptibility, density, and conductivity models that were integrated using machine learning to construct the quasi-geology model, which is similar to the geology model constructed from drilling logs.

Conclusions

In this work, we propose an objective geology differentiation method that supports integrated interpretation of multiple geophysical inversions in greenfield exploration. Considering the proposed method is entirely data-driven, the

high rate of match demonstrates that such geology differentiation is feasible in a complex geological setting such as the Cristalino copper deposit, where the target is obscured by the more anomalous responses of an adjacent iron formation. This work contributes to solving practical challenges of greenfield mineral exploration by providing effective unbiased integrated interpretation methods that produce directly interpretable quasi-geology models.

Acknowledgements

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. This work is supported in part by the Gravity and Magnetism Research Consortium (GMRC).

References

- 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.
- Li, Y., 2001, 3-D inversion of gravity gradiometer data: Presented at the 2001 SEG Annual Meeting, Society of Exploration Geophysicists.
- Li, Y., and D. W. Oldenburg, 1996, 3-D inversion of magnetic data: *Geophysics*, 61, 394–408.
- Li, Y., and D. W. Oldenburg, 2003, Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method: *Geophysical Journal International*, 152, 251–265.
- Li, Y., and D. W. Oldenburg, 2000, 3-D inversion of induced polarization data: *Geophysics*, 65, 1931–1945.
- NCL Brasil, 2005, Revision de la estimación de recursos del proyecto Cristalino: Technical report, Vale S.A. Internal Report.
- Vale S.A., 2004, Relatório de reavaliação de reservas estudo de pré-viabilidade projeto Cristalino: Technical report, Vale S.A. Internal Report.