INTEGRATED QUANTITATIVE INTERPRETATION
OF MULTIPLE GEOPHYSICAL DATA FOR
GEOLOGY DIFFERENTIATION

by

Aline Tavares Melo
A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Geophysics).

Golden, Colorado
Date ________________

Signed: _______________________
Aline Tavares Melo

Signed: _______________________
Dr. Yaoguo Li
Thesis Advisor

Golden, Colorado
Date ________________

Signed: _______________________
Dr. John Bradford
Professor and Head
Department of Geophysics
ABSTRACT

The future of mineral exploration depends on innovative methods of data integration and interpretation because new discoveries are becoming scarcer over the years. As brownfield exploration areas reach maturity and greenfield exploration faces increasingly deeper targets and targets hidden under cover, geophysics is becoming the primary exploration tool. When little a priori geological information is available, such as in greenfield exploration, multiple geophysical methods are necessary to improve interpretation and decrease exploration risk. However, it is challenging to deal with multiple geophysical methods in geologically complex areas. For this reason the main motivation of my thesis is to develop integrated quantitative interpretation methods of multiple geophysical data for geology differentiation. Multiphysics is fundamental for identifying geological units instead of just identifying isolated geophysical anomalies in different physical property models. It also allows uncertainties to be minimized if all the available data are properly integrated. Therefore, I first develop a method for geology differentiation based on spatially limited geological information and general relations of physical properties that can be applied to geophysical data over a large area. Then, in the absence of geological information, I incorporate more geophysical data and develop a method of geology differentiation by applying unsupervised machine learning (correlation-based clustering) for the construction of a quasi-geology model. Additionally, I develop a novel method to improve the construction of susceptibility models, the magnetic on-time transient electromagnetic (MoTEM) method. The use of more accurate physical property models improve the geology differentiation. The research I have developed contributes to solving practical challenges of greenfield mineral exploration by providing effective unbiased integrated interpretation methods that produce directly interpretable quasi-geology models.
# TABLE OF CONTENTS

**ABSTRACT** ........................................................................................................ iii

**LIST OF FIGURES** .......................................................................................... vii

**LIST OF TABLES** ............................................................................................ xv

**ACKNOWLEDGMENTS** ..................................................................................... xvi

**DEDICATION** ................................................................................................... xviii

**CHAPTER 1 INTRODUCTION** ........................................................................... 1

1.1 Geology differentiation of geophysical models using spatially limited geological information ....................................................... 2

1.2 Geology differentiation by applying unsupervised machine learning to multiple geophysical inversion models ......................... 3

1.3 Geology differentiation of susceptible targets using magnetic on-time transient electromagnetic (MoTEM) inversion .................. 5

**CHAPTER 2 GEOPHYSICAL INVERSIONS APPLIED TO 3D GEOLOGY CHARACTERIZATION OF AN IRON OXIDE COPPER-GOLD DEPOSIT IN BRAZIL** ................................................................................ 8

2.1 Introduction ................................................................................................... 9

2.2 Geological Setting ....................................................................................... 13

2.3 Geophysical data and inversion method .................................................... 19

2.4 2D geophysical inversions and geology characterization ......................... 23

2.5 3D geophysical inversions and geology characterization ......................... 29

2.6 Discussion .................................................................................................. 37

2.7 Conclusions ............................................................................................... 39
CHAPTER 3 GEOLOGY DIFFERENTIATION BY APPLYING UNSUPERVISED MACHINE LEARNING TO MULTIPLE GEOPHYSICAL INVERSIONS: METHODOLOGY

3.1 Introduction ................................................. 42

3.2 Synthetic model ........................................... 45
   3.2.1 Geophysical data ....................................... 47
   3.2.2 Inversion method ....................................... 50

3.3 Geology differentiation using clustering algorithms ................. 54
   3.3.1 Connectivity-based clustering ............ 58
   3.3.2 Density-based clustering .................... 58
   3.3.3 Distribution-based clustering ............ 61
   3.3.4 Centroid-based clustering .............. 61
   3.3.5 Correlation-based clustering .......... 63
   3.3.6 Discussion ........................................... 65

3.4 Quasi-geology model evaluation .......................... 67

3.5 Conclusions ................................ ............ 69

CHAPTER 4 GEOLOGY DIFFERENTIATION BY APPLYING UNSUPERVISED MACHINE LEARNING TO MULTIPLE GEOPHYSICAL INVERSIONS: CASE STUDY

4.1 Introduction ................................................. 72

4.2 Geological background of study site ........................ 79

4.3 Geophysical data and inversions ........................... 83

4.4 Geology differentiation ................................... 92
   4.4.1 Clustering analysis of inverted physical properties 93
LIST OF FIGURES

Figure 2.1  (a) Tectonic location of the Carajás Mineral Province at the southeastern margin of the Southern Amazon Craton, Brazil, (b) Geologic Map of the Carajás Mineral Province showing the area of the total gradient of the magnetic data in Figure 2.2 (yellow box), and study area (red box), Cristalino deposit, which is hosted by a volcanosedimentary sequence (modified from ). ................................. 14

Figure 2.2  Image of the total gradient of the magnetic data showing the study area, Cristalino deposit, the main mineral deposits and faults in the southern part of the Carajás Mineral Province. ................................. 15

Figure 2.3  Geologic cross section of line -1800 of Cristalino copper deposit showing the drill hole traces and chalcopyrite concentration of the hydrothermal zone hosted by the volcanic (mafic and felsic) and sedimentary (iron formation) rocks (modified from ). The red bar in the legend corresponds to the representation of the proportion of 10% of chalcopyrite. ................................. 15

Figure 2.4  Average modal distribution of the mineralized unit of Cristalino copper deposit. The category “other” includes apatite, titanite, muscovite, ilmenite, bornite, and chalcocite (modified from ). ................................. 16

Figure 2.5  Photographs of drill cores showing the ore texture types and abundance in the deposit: massive chalcopyrite (yellow) in breccia matrix, stockwork, stringers, and disseminated chalcopyrite. The dark minerals are composed mainly of magnetite and silicates. ................................. 17

Figure 2.6  Conceptual cross plot of electrical conductivity and magnetic susceptibility values for magnetite and chalcopyrite (based on ), showing the estimated relative trends of the mineralized rock with different ore textures and the iron formation trend, and the relative position of the geology differentiation classes. ................................. 18

Figure 2.7  Airborne magnetic data over the deposit area, showing the airborne (gray dotted lines) and ground (black dashed lines) surveys. Line -1800 of the ground survey was used for the 2D inversion and the airborne lines for the 3D inversion. ................................. 20
Figure 2.8 a) Magnetic profile of line -1800, field inclination of -3.5°, declination of -19°, and strength of 25,500 nT, b) resistivity pseudo-section (vertical exaggeration of 5x) for line -1800, dipole-dipole array with a dipole length of 60 m and 5 n-spacings, and c) topography of line -1800. The gray dotted rectangle shows the model extent.

Figure 2.9 Susceptibility model of the line -1800 showing the sharp boundary between the areas of high and low susceptibility.

Figure 2.10 Conductivity model of the line -1800. The estimated depth of investigation (DOI) is highlighted by the white line.

Figure 2.11 a) Scatterplot of susceptibility and conductivity values from the 2D inversions, the points are correspondent to the same cell location, b) classification based on the formulated conceptual behavior (Figure 2.6), and natural groupings in the data. Class 1 is associated to the iron formation, class 2 with the copper ore, class 3 with the host rock 1, and class 4 with the host rock 2.

Figure 2.12 Spatial distribution of the classes identified in the 2D scatterplot, showing the geology differentiation in the section. The dark gray line shows the estimated depth of investigation (D.O.I.) for the resistivity inversion.

Figure 2.13 Geology characterization based on the interpretation of the classes identified in the scatterplot of the 2D inversions, showing the association with specific geological units.

Figure 2.14 Geology characterization overlain by the available geologic information from drilling, showing the high spatial correspondence between the interpreted ore class and the high concentration of chalcopryite, and between the interpreted iron formation class and the magnetic rocks (iron formation and gabbro dyke).

Figure 2.15 a) Inverted 3D susceptibility model showing the magnetic anomalies associated mainly with iron formation, b) section of the susceptibility model at line -1800, c) inverted 3D conductivity model showing the high conductivity anomaly in the center and north-northwestern parts of the model, which is associated with copper ore, and d) section of the conductivity model at line -1800.
Figure 2.16 a) Scatterplot of susceptibility and conductivity from the 3D inversions, the points are correspondent to the same cell location, b) revised classification based on the 2D geology differentiation (Figure 2.11). Class 1 is associated to the iron formation, class 2 with the copper ore, class 3 with the host rock 1, and class 4 with the host rock 2.

Figure 2.17 a) Spatial distribution of the classes identified in the scatterplot of the 3D inversions, b) section of the classification of the model at line -1800, c) the interpretation of these classes based on the results of the 2D characterization, and d) section of the interpreted model at line -1800.

Figure 2.18 a) Geological map of Cristalino copper deposit (adapted from ), and b) top view of the 3D geology characterization volume overlain by the contours of the geological map, showing that the method was able to identify the core of the mineralized zone. The first layer of cells (50m), which is mainly associated with overburden, was removed.

Figure 3.1 Geological map of Cristalino copper deposit (adapted from ).

Figure 3.2 Synthetic geologic model based on Cristalino deposit showing the geologic units in the a) susceptibility and density models, b) conductivity model, and c) a section at north = 700 m in the conductivity model.

Figure 3.3 Data location of a) co-located magnetic and gravity gradient survey, and b) DC resistivity survey.

Figure 3.4 The total field anomaly produced by the model in Figure 3.2(a). The inducing field has direction I = 0° and D = 0° and a strength of 25000 nT. Uncorrelated Gaussian noise with standard deviation of 1% of the datum magnitude plus 1 nT was added to the data.

Figure 3.5 The simulated gravity gradient data observed above the model shown in Figure 3.2(a). Uncorrelated Gaussian noise with standard deviation of 1% of the datum magnitude plus 0.5% of the minimum of the component has been added to the data.

Figure 3.6 a) Synthetic geologic model overlain by the DC resistivity data locations and b) apparent resistivity pseudo-sections for north equals to 425, 725, 825, and 1025 m. Uncorrelated Gaussian noise with a standard deviation of 1% of the datum magnitude plus 0.5% of the minimum datum with maximum electrode separation was added to the data.
Figure 3.7  a) Inverted 3D susceptibility model showing the magnetic anomalies
associated with the iron formation, b) inverted 3D density model
showing the high density anomalies associated mainly with iron
formation and the intermediate density contrast anomaly associated
with the copper ore, and c) inverted 3D conductivity model showing
the high conductivity anomaly in the center of the model, which is
associated with copper ore. ........................................... 55

Figure 3.8  Crossplot of the normalized values of density, log susceptibility and log
conductivity of the inverted models. ............................... 56

Figure 3.9  Synthetic geologic model and crossplots of the normalized values of
density, log susceptibility and log conductivity of the inverted models
classified by the corresponding geologic unit. ..................... 57

Figure 3.10  Connectivity-based clustering dendrogram of the susceptibility, density,
and conductivity models showing that only one big cluster was formed
because of the lack of large distances between groups. ............... 59

Figure 3.11  Density-based clustering model and crossplots of the normalized values
of density, log susceptibility and log conductivity of the inverted models
classified by the corresponding cluster. ........................... 60

Figure 3.12  Expectation-maximization clustering model using Gaussian mixture
models (dotted gray lines), and crossplots of the normalized values of
density, log susceptibility and log conductivity of the inverted models
classified by the corresponding cluster. ............................ 62

Figure 3.13  K-means clustering model showing the segmentation Voroni-cells
(dotted gray lines), and crossplots of the normalized values of density,
log susceptibility and log conductivity of the inverted models classified
by the corresponding cluster. ........................................ 63

Figure 3.14  Correlation-based clustering model and crossplots of the normalized
values of density, log susceptibility and log conductivity of the inverted
models classified by the corresponding cluster. This result shows the
good correspondence between this model and the geologic model
(Figure 3.9). ............................................................. 65

Figure 3.15  Synthetic geology model and the quasi-geology model obtained from
correlation-clustering. ................................................. 67

Figure 3.16  Confidence matrix showing the comparison between the known geology
from the synthetic model and the predicted units from the
quasi-geology model obtained through correlation-based clustering. .... 68
Figure 3.17  Histogram plots for each physical property showing the range of values for the geological units of the synthetic model. ............................... 69

Figure 3.18  Histogram plots for each physical property showing the range of values for the clusters of the quasi-geology model obtained through correlation-based clustering. ............................... 69

Figure 4.1  (a) Tectonic location of the Carajás Mineral Province at the margin of the Amazon Craton, Brazil, (b) Geologic map of the Carajás Mineral Province showing the study area in the red box (modified from ; and ). 80

Figure 4.2  Geological map of Cristalino copper deposit showing the schematic section A-A’ in Figure 4.3 (adapted from ). ............................... 81

Figure 4.3  Schematic geological cross section showing the fault that acted as conduit for hydrothermal fluids that formed the high and low grade ore (adapted from ). ............................... 82

Figure 4.4  Airborne magnetic data over the deposit area. a) Total magnetic intensity data, b) regional data calculated from inversion-based regional-residual separation, and c) residual data, showing the airborne survey. ............................... 85

Figure 4.5  Airborne gravity gradient data over the deposit area, showing the five components used in the inversion. ............................... 86

Figure 4.6  DC resistivity acquisition lines over the terrain topography. ............................... 87

Figure 4.7  Inverted 3D susceptibility model showing the magnetic anomalies associated mainly with the iron formation. ............................... 89

Figure 4.8  Inverted 3D density contrast model showing the high density contrast anomalies associated mainly with iron formation and the intermediate density contrast anomaly associated with the high grade ore. ............................... 90

Figure 4.9  Inverted 3D conductivity model showing the high conductivity anomaly in the center of the model, which is associated with high grade ore. ............................... 91

Figure 4.10  Scatterplot of the normalized values of density, log susceptibility and conductivity recovered in the inverted models. ............................... 92

Figure 4.11  Flowchart of the correlation-based clustering with ORCLUS (arbitrarily ORiented projected CLUSter generation) algorithm showing the three main parts of the clustering process (based on ). ............................... 96
Figure 4.12 Objective function versus number of clusters, the red circle shows the point of maximum curvature $k = 4$, which is the optimum number of clusters. ................................. 98

Figure 4.13 3D spatial distribution of k-means clustering with $k = 4$ and the scatterplots of physical properties showing the clustering result. ............................. 99

Figure 4.14 3D spatial distribution of correlation-based clustering with $k = 4$ and the scatterplots of physical properties, showing less influence of irrelevant attributes and an improved spatial association with the geological units. ................................. 100

Figure 4.15 Histogram plots for each physical property showing the range of values for all classes of correlation-based clustering with $k=4$. ................................. 101

Figure 4.16 3D spatial distribution of correlation-based clustering with $k = 5$ and the scatterplots of physical properties, showing the subdivision of cluster 1. ................................. 102

Figure 4.17 Histogram plots for each physical property showing the range of values for all classes of correlation-based clustering with $k=5$. ................................. 103

Figure 4.18 Flowchart for the geology differentiation process. ................................. 104

Figure 4.19 3D geological model of Cristalino copper deposit constructed from 303 drillholes (adapted from ). ................................. 105

Figure 4.20 Histogram plots for each physical property showing the range of values for the simplified geological units in the 3D geological model. ................................. 106

Figure 4.21 3D plot of the result of correlation clustering with $k = 4$ compared to the high grade ore from the geological model. ................................. 106

Figure 4.22 Confusion matrix showing the comparison between the known geology from the 3D geological model and the predicted units from correlation-based clustering with $k=4$. ................................. 107

Figure 4.23 3D plot of the result of correlation clustering with $k = 5$ compared to the high grade ore from the geological model. ................................. 108

Figure 4.24 Confusion matrix showing the comparison between the known geology from the 3D geological model and the predicted units from correlation-based clustering with $k=5$. ................................. 109
Figure 5.1  
a) Dike with magnetic susceptibility equal to 0.1 SI in a non-magnetic background; b) synthetic magnetic data for the dike in a low latitude area of 0 inclination, 0 declination, and 25,000 nT field strength; and c) inversion result showing a clipped volume at 0.065 SI and the poor recovery of the shallow lateral edges of the dike.  

Figure 5.2  
Schematic diagram of the transmitter and receiver waveforms showing the secondary field signal due to eddy currents (response from conductivity $\sigma$), and the secondary field signal due to steady magnetic field (response from susceptibility $\kappa$).  

Figure 5.3  
Schematic view of the anomalous magnetic field ($\vec{B}_a$) induced by the Earth’s geomagnetic field ($\vec{B}_0$) near the Equator with 0 degrees inclination and declination.  

Figure 5.4  
Schematic view of the anomalous magnetic field ($\vec{B}_a$) induced by a current loop field ($\vec{B}_{0EM}$) near the Equator with 0 degrees inclination and declination.  

Figure 5.5  
Synthetic model with a magnetic body of 0.1 SI susceptibility and 2.0 S/m conductivity in a non-susceptible background of $2 \times 10^{-6}$ S/m conductivity showing the data location.  

Figure 5.6  
Transmitter waveform showing the 121 time steps used to compute the synthetic electromagnetic data.  

Figure 5.7  
Simulated electromagnetic data of on-time 1.3 ms from the synthetic model in Figure 5.5, showing that the effect of eddy currents is the predominant signal.  

Figure 5.8  
Simulated electromagnetic data of on-time 3.0 ms from the synthetic model in Figure 5.5, showing that as the eddy currents dissipate, the magnetic signal is predominant.  

Figure 5.9  
Simulated electromagnetic data of on-time 8.0 ms from the synthetic model in Figure 5.5, showing that the conductivity effect is negligible.  
The data in (1) was used in the MoTEM inversion.  

Figure 5.10  
a) Logarithmic of sensitivity of the datum highlighted in yellow and b) section view of the sensitivity of the same datum.  

xiii
Figure 5.11 Comparison of the kernel function \( G(z) \) directly beneath the observation point in Figure 5.10 showing the estimated depth weighting function given by \( W^2(z) = (z + z_0)^{-6} \) with \( z_0 = 12.5 \) m, and the sensitivity weighting function with \( \beta = 1 \). All curves are normalized for comparison.

Figure 5.12 a) Logarithmic of depth weighting, and b) logarithmic of sensitivity weighting used in the inversions of the anomalous body in Figure 5.5.

Figure 5.13 Susceptibility model obtained using depth weighting function, a) section through the showing the anomalous cube (dashed gray line), and b) susceptibility model with cutoff = 0.11 SI.

Figure 5.14 Susceptibility model obtained using sensitivity weighting function, a) section through the showing the anomalous cube (dashed gray line), and b) susceptibility model with cutoff = 0.17 SI.

Figure 5.15 a) Depth weighting for the function \( W^2(z) = (z + z_0)^{-6} \) used in the MoTEM inversion, and b) depth weighting for the function \( W^2(z) = (z + z_0)^{-3} \) used in the inversion of geomagnetic data showing the difference in decay.

Figure 5.16 Comparison of the kernel functions \( G(z) \) between the geomagnetic and MoTEM fields directly beneath an observation point showing the estimated depth weighting function given by \( W^2_{\text{Geomag}}(z) = (z + z_0)^{-3} \) and \( W^2_{\text{MoTEM}}(z) = (z + z_0)^{-6} \). All curves are normalized for comparison.

Figure 5.17 a) Susceptibility model obtained using the MoTEM method with depth weighting function, and b) susceptibility model using traditional geomagnetic data, both showing the anomalous cube (dashed gray line).

Figure 5.18 a) Section of the synthetic model of a dike, b) section of the susceptibility model obtained from inversion of geomagnetic data at 0 inclination, 0 declination, and 25,000 nT field strength, c) section of the susceptibility model obtained from inversion of MoTEM data, d) dike model at cut off = 0.09 SI, e) susceptibility model of (b) at cut off = 0.065 SI, f) susceptibility model of (c) at cut off = 0.085 SI, g) superposition of models (a), (b), and (c), h) superposition of models (b) and (c), and i) superposition of models (e) and (f).
LIST OF TABLES

Table 3.1 Physical property values for the synthetic geologic model. The conductivity and density values were based on , and the susceptibility values on . .............................................. 47

Table 4.1 Acquisition parameters of the magnetic, gravity gradient, and DC resistivity surveys. .......................................................... 84
ACKNOWLEDGMENTS

I would like to first thank God for the gift of life and for giving me the intelligence and strength to accomplish this project. Thanks for giving me an amazing mother who taught me to be strong and pursue my dreams. Thanks for an outstanding husband who completes me and makes my dreams also his dreams. Thanks for giving me a wonderful brother and family. Thanks for giving me an incredible advisor who taught me more than geophysics, who taught me to be a generous and caring person.

I want to express my infinite gratitude to my mother, Amélia, for her phenomenal support. She worked hard to give the education opportunities my brother and I have, thank you for all the sacrifices. I would like to thank my husband Carlos for everything, I would not have gotten here without you, your encouragement makes me go further. I love sharing life and being in this amazing journey with you. Thanks to my father José Manoel who is no long with us but would be very proud of me. To my brother, André, who always supported my decisions and cheered me up along the way.

I would like to express my gratitude to my advisor, Dr. Yaoguo Li, for his guidance, support, and understanding. Dr. Li encouraged and inspired me to do my best during my time at Colorado School of Mines. I am thankful for having the best mentor I could ask for. I want to thank Dr. Li and his wife Yu Mei for their generous hospitality making my husband and I welcome at their home. Thank you to my committee members Dr. Murray Hitzman, Dr. Kadri Dagdelen, Dr. Paul Sava, and Dr. Andrei Swidinsky for all advice and provoking questions. It has been an honor having the opportunity to learn from your experiences. Thanks to Dr. Terry Young, Dr. Roel Snieder, Dr. John Bradford, Michelle Szobody, Dawn Umpleby, and Joana Perez for everything you have done and do for the department and students. Thank you to Dr. Misac Nabighian for all help with research and for good stories about life.
I would not have come so far without of my professors from Universidade de Brasília, in Brazil. I would like to thank Dr. Augusto Pires, Dr. José Gaspar, and Dr. Reinhardt Fuck for the mentorship. I am also very thankful to all my fellow faculty members at Universidade Federal de Minas Gerais, in Brazil, for the encouragement to pursue my degree at Colorado School of Mines. A special thanks to Dr. Paulo Aranha and Dr. Antonio Romano for the support. Thanks to all friends I have from the time I worked for Vale and that have contributed with my formation as a geophysicist. Thanks to Alan King, Rob Eso, Cantidiano Freitas, Florivaldo Sena, Fernando Matos, and Brian Bengert for all geophysical discussions and incentive.

I would like to also thank the CGEM members both past and present for all the time we have spent together. Thanks to Dr. Jiajia Sun for being such a great friend and almost a co-advisor. Thank you Andy (Andy Andy Andy) McAlliley for helping me so much along the way and for being a wonderful friend. Thanks to my dear friends Andrea Balza and Qian Yin for making my days happier and lighter. Thanks Dr. Richard Krahenbuhl, Zhuo Liu, Patricia McQueen, Cici Martinez, Thomas Rapstine, Leon Foks, Liz Maag-Capriotti, and Joe Cappriotti for all good talks about research and life. A special thanks Mauricio Pedrassi, Paula Murillo, Emma Butler, Stephen Paskvich, and Elvan Aydin for the friendship along the way through graduate school.

Finally, I would like to thank CNPq, the National Council for Scientific and Technological Development of Brazil, for the sponsorship, Vale S.A. for providing data, and the sponsors of the Gravity and Magnetics Research Consortium (GMRC).
To God, to my husband, mother, and brother for all the love.
CHAPTER 1
INTRODUCTION

Mineral exploration has been facing many challenges associated with the downturn in the commodities cycle and lack of new discoveries. The decrease in the number of new discoveries is associated with many factors. In brownfield exploration (areas near producing mines), many mineral provinces are exhausting resources as they reach maturity. In greenfield exploration (new frontier areas), there is high risk involved because most of the outcropping and shallow deposits have already been found. Consequently, new targets are being discovered at increasingly greater depths or under cover. In this challenging scenario for exploration, the discovery of new mineral deposits is increasingly relying primarily on geophysics. Therefore, multiple geophysical methods are necessary to improve interpretation and decrease exploration risk, especially in greenfield exploration, which is the focus of this work, where little a priori geological information is available. However, it is challenging to integrate multiple geophysical data sets and models in geologically complex areas. For this reason, the main motivation of this thesis is to develop integrated quantitative interpretation methods of multiple geophysical data for geology differentiation of the subsurface.

The first research question I explore is how spatially limited geological information, such as the first few drillholes of a prospect, can be used to improve geology differentiation of geophysical data over a large area; aiming to support target assessment and drilling planning. This is important because the mineral exploration dynamics often requires an efficient and objective means of evaluating a prospect in early exploration stages. However, methods for improving interpretation are also necessary when not even sparse geological data are available. Therefore, in the second research question, I explore how unsupervised machine learning (clustering) can be effectively applied for geology differentiation of multiple physical property models in the absence of a priori geological information. The main outcome is an
integrated representation of the subsurface where the different units of the constructed quasi-
geology model are directly associated with specific geology units. This result contributes to
a geoscientific approach for integration.

The developed interpretation and integration methods are applied to a copper deposit
named Cristalino located in a low latitude area in northern Brazil. In such areas, the
Earth’s inducing magnetic field inclination, declination, and strength represent a challenge
for the construction of susceptibility models. Besides the naturally weaker inducing field,
when the magnetic body of interest has similar strike direction as the Earth’s magnetic field
declineation, the resulting anomalous field is weaker from the poor coupling. For this reason,
the susceptibility model does not have good recovery in the shallow parts. This scenario
inspired the use of an active source magnetic method to better recover information about
the magnetic body. Therefore, I propose the magnetic on-time transient electromagnetic
(MoTEM) method for imaging susceptibility. In the method, I invert the on-time part
of transient electromagnetic data to construct susceptibility models. Consequently, more
accurate physical property models will lead to the more accurate geology differentiation.
In the following, I discuss the research questions, the main past contributions, and the
advancements achieved by this work.

1.1 Geology differentiation of geophysical models using spatially limited geo-
logical information

Geology differentiation is the process of identification of geological units, here I accom-
plish this task using geophysical models. One category of application of geology differenti-
ation is on alteration mapping. In a seminal paper, Hanneson (2003), then Williams et al.
(2004) and Williams & Dipple (2007) use the relationship between density and susceptibility
model values to map alteration zones associated with different contents of magnetite and
hematite. Geology differentiation can also be used for lithology mapping (Kowalczyk et al.,
2010; Martinez et al., 2011; Fraser et al., 2012; Martinez & Li, 2015). Kowalczyk et al.
(2010) use the scatterplot of susceptibility versus density contrast values from separate in-
versions of magnetic and gravity data to define lithology classes on regional scale. Using a similar approach, Martinez et al. (2011) and Martinez & Li (2015) define lithotypes from the scatterplot of susceptibility versus density values derived from inversion of gravity gradient and magnetic data for an iron ore deposit. Other works on geology differentiation based on geophysically derived physical properties also focus on examining the clustering or grouping patterns among the physical properties values in a scatterplot (e.g., Bosch, 1999; Bosch et al., 2002; Bauer et al., 2003; Bedrosian et al., 2007).

The work I have developed is in a different category of geology differentiation, the focus is on mapping copper ore. I propose to identify classes in the scatterplot of physical properties to build a 3D quasi-geology model by grouping the estimated susceptibility and conductivity values from independent inversions of magnetic and DC resistivity data from a copper deposit. In contrast to the work by Kowalczyk et al. (2010), which focuses on the regional scale, and by Martinez & Li (2015), which focuses on classifying the units of an iron deposit using generic or site specific a priori geologic information, I combine the expected physical property values from literature and sparse site specific drillhole information into the differentiation of the mineralized unit. The method I am proposing enables a quick and integrated evaluation of targets, and is suitable for prospects in the initial stages of exploration. In Chapter 2, I demonstrated that the quasi-geology model obtained through this method can serve as a training example for us to learn the relation between the Cristalino copper deposit and its geophysical signature. The knowledge learned through the study can potentially be extended to exploration of covered or deep deposits in the same mineral province and possibly to other regions around the globe.

1.2 Geology differentiation by applying unsupervised machine learning to multiple geophysical inversion models

In the absence of geological information, multiple geophysical methods are used to increase interpretation confidence. However, all the available data, and physical property models, need to be properly integrated to make an impact on minimizing uncertainties in
geology differentiation. The incorporation of multiple models increases the complexity of pattern recognition and interpretation, therefore machine learning (ML) methods have been applied in many geosciences problems. Machine learning is the area that develops algorithms capable of learning and enabling computers to make human-like decisions. If the learning process uses input information with known output classification for training the algorithms, it is defined as supervised learning. If there is no labeled data for training, unsupervised learning algorithms identify patterns without classifying them. For this reason, these algorithms focus on exploring the data structure to find insights that are not obvious to human eyes (e.g., clustering algorithms). In greenfield exploration, there are no labeled examples for supervised ML; consequently, unsupervised ML is the choice for identifying patterns associated with new targets.

Machine learning has been applied for facies classification, from well log data and seismic attributes, in petroleum exploration for several decades (e.g., Serra & Abbott, 1982; Wolf & Pelissier-Combescure, 1982; Delfiner et al., 1987; Baldwin et al., 1990; Rogers et al., 1992; Ye & Rabiller, 2000; Schlanser et al., 2016; and Abreu et al., 2016). Applications also focus on the identification of seismic facies using seismic attributes and velocity models (e.g., Meldahl et al., 1999; Barnes & Laughlin, 2005; Strecker & Uden, 2002; Coléou et al., 2003; Gao, 2007; Roy et al., 2014; Zhao et al., 2015; and Qi et al., 2016). In geological mapping, ML tools have been applied to airborne geophysical data focusing on the construction of 2D pseudo-lithology maps for interpretation. For example, Paasche & Eberle (2009) apply fuzzy c-means clustering, Eberle & Paasche (2012) apply fuzzy Gustafson-Kessel clustering, and Carneiro et al. (2012) apply self-organizing maps to airborne data to construct pseudo-lithology maps. Other works also focus on 2D geological mapping applying ML to multiple geophysical data (e.g., Ranjbar et al., 2001; Martelet et al., 2006). Geological mapping can also be achieved using 3D physical property models. For example, Paasche et al. (2006) apply fuzzy c-means clustering to identify sedimentary units from physical property models and well log data. In mineral exploration, Barnett & Williams (2006) use known gold deposits
to train a neural network and construct a favourability map in regional scale using multiple data sets. Mahmoodi et al. (2014) apply fuzzy c-means to down hole data from a Ni deposit to characterize rock types and mineralization. Caté et al. (2017) apply supervised machine learning algorithms to predict the presence of gold, in drill cores, from geophysical logs.

Overall, much of the existing works using ML either focuses on large-scale applications such as regional geology and prospectivity mapping, or on formation scale such as lithofacies classification in drill holes. Only limited work is available on deposit scales, where ultimately drilling targets must be chosen. For this reason, in Chapter 3, I examine the performance of different unsupervised ML algorithms aiming to make advances in the differentiation of geologic units on deposit scales, and in the prediction of potential drilling targets in a greenfield mineral exploration scenario. I identify the use of correlation-based clustering for geology differentiation because of its capability of identifying patterns related to geological units among the inversion artifacts present in smooth minimally constrained inversions. Therefore, with the resulting quasi-geology model, the geoscientist can go further with interpretation and associate it with a geological setting, and interpret each identified unit. In Chapter 4, I further explore the proposed method using physical property models from Cristalino copper deposit. I propose a process of extracting information from multiple sources of data with an unbiased, quantitative, and integrated method, that empowers the geoscientist in the decision making.

1.3 Geology differentiation of susceptible targets using magnetic on-time transient electromagnetic (MoTEM) inversion

The result of integration of multiple physical property models is highly dependent on the geophysical data used in the inversion. For example, the measured magnetic field in low latitude areas is weak when the magnetic body has a direction similar to the Earth’s magnetic field declination. The low measured anomaly is the result of the lack of coupling between the Earth’s field and the target being investigated. In those regions, the weak magnetic anomaly poses a problem to constructing accurate susceptibility models because
it is difficult to correctly recover the edges and top of the target in the inversion process. Consequently, any integrated interpretation that uses this type of magnetic data will be subject to inaccuracies in the susceptibility model, which results from the limitations in the data.

Alternatively, the primary inducing field can be an a.c. field generated by a transmitter coil in place of the Earth’s main field, which is the case for the frequency-domain electromagnetic method (FEM). In this type of survey the conductivity and susceptibility signals are mixed in the data. There are different approaches to explore the susceptibility signal in FEM data such as, direct magnetite mapping (e.g. Fraser, 1973; Fraser, 1981) and inversion of magnetic susceptibility given a known conductivity model (e.g. Zhang & Oldenburg, 1997). Additionally, simultaneous mapping of and inversion for conductivity and magnetic susceptibility has been proposed by different authors (Beard & Nyquist, 1998; Zhang & Oldenburg, 1999; Huang & Fraser, 2000; Farquharson et al., 2003; Huang & Fraser, 2003; Sasaki et al., 2010), and show improved results for both conductivity and susceptibility models. However, it is difficult to separate the conductivity and susceptibility responses in the measured signal. For this reason, Noh et al. (2017) use sufficiently low frequencies that can image susceptibility without the influence of the electromagnetic induction response in a frequency domain electromagnetic survey.

Limitations of the FEM methods in the detection of weak conductors is the main reason why mining exploration shifted to transient electromagnetic (TEM) methods. This substitution caused a loss of versatility since FEM can be used for mapping both conductivity and susceptibility. In situations where magnetic units are also highly conductive, they can cause small variations in the late time channels of TEM surveys and a method for imaging both susceptibility and conductivity simultaneously was proposed by Zhdanov & Pavlov (2001). However, the focus of TEM is on the off-time data, which makes electrical conductivity the only property of interest, and the on-time measurements are not used. Although TEM data have the potential for mapping conductivity and susceptibility through the use of both off-
and on-time data, this potential has not been explored to date. Therefore, I propose magnetic on-time transient electromagnetic (MoTEM) method to image magnetic susceptibility. The method uses the on-time part of the waveform produced by the transmitter of transient electromagnetic (TEM) systems as the inducing field in place of the Earth’s geomagnetic field. Therefore, it is similar to an active magnetic method at near-zero frequency because the inducing field is assumed to be static. The method also assumes the transmitter waveform is long enough to allow the dissipation of eddy currents until the electrical conductivity effect becomes negligible. Then, the on-time data is inverted using the same framework of 3D inversion of geomagnetic data. In Chapter 5, I show that the inversion of synthetic on-time electromagnetic data using the MoTEM method successfully recovers the susceptibility of an anomalous body in an non-magnetic background. Therefore, demonstrating the feasibility of the method and its potential for improving susceptibility models and geology differentiation.
CHAPTER 2

GEOPHYSICAL INVERSIONS APPLIED TO 3D GEOLOGY CHARACTERIZATION
OF AN IRON OXIDE COPPER-GOLD DEPOSIT IN BRAZIL

Mineral exploration dynamics often requires an efficient and objective means of evaluating a prospect in early exploration stages, when only few drill holes have been drilled. In the case of deep prospects or prospects under cover, this evaluation will be mostly based on geophysical data. To develop an objective interpretation method capable of combining all the information available, we present this work as an integrated interpretation scheme of geophysical models and sparse geological data. The proposed method is based on the relationship between recovered physical properties obtained from 2D and 3D inversions, aiming to find patterns associated with geological units, such as iron formation, copper ore, and host rock. The interpretation is guided by theoretical relations of the minerals of interest (chalcopyrite and magnetite) and the sparse geologic information available. It is suitable for prospects in initial stages of exploration when only limited amount of mineralogical information is available from, say, one drill hole. We have demonstrated the success of the method using magnetic and DC resistivity data from Cristalino iron oxide copper gold (IOCG) deposit, located in northern Brazil, which is covered by a thick soil overburden. The theoretical behavior of the physical properties of chalcopyrite and magnetite was first combined with the rock types identified in the drill cores to find groups or classes associated with different amounts of these minerals. Then, these relative relations between units were applied to define four classes in the scatterplot of recovered susceptibility and conductivity values from 2D inversions. These four classes are associated with iron formation, copper ore, and two types of host rocks. After the validation with the known geology, the same interpretation scheme was applied to the scatterplot of recovered susceptibility and conductivity values from 3D inversions. The final interpreted volume allows the explorationist to have an approximate
estimate of the copper body extent.

2.1 Introduction

The main tools for mineral exploration include geological mapping, geochemistry, and geophysics. However, geological mapping and geochemistry have limited application in areas with extensive overburden cover, such as in Brazil and Australia, which results in limited understanding of their geology from direct observations. For these reasons, the discovery of new mineral deposits increasingly relies primarily on geophysics. In addition, new deposits are being found at progressively greater depths over the years. Consequently, improved and accurate methods of integrated interpretation of geophysical information are becoming more important, since a better understanding of prospects decreases exploration risk. A negative deep drill hole can have a huge impact on an exploration project, both economically and psychologically.

Greenfield exploration projects require large amounts of time and financial investment. The progress of a project heavily depends on economic factors, such as the global cycle of commodities. Economic variations can quickly change the budget and, consequently, the status of a project from active to inactive. For this reason, management often needs a rapid and objective way of determining the potential of a prospect in its early stages of exploration; in the case of a deposit under cover, it would be when only geophysical data and few drill holes are available. The response of a single geophysical method applied to a single target is well understood, but the complexity of interpretation increases when multiple methods are applied to a complex geology, and data integration becomes a key factor. Based on the need of a practical evaluation method capable of combining all the data, we present here an integrated interpretation scheme of geophysical models and sparse geologic data that identified the mineralized zone. The proposed method is based on the analysis of the relationships between recovered physical properties, from inversion, aiming to find patterns associated with different geological units.
Geology characterization is the process of identification of geological units, in this paper we accomplish this task using geophysical models. One category of application of geology characterization is on alteration mapping. In their seminal papers, Hanneson (2003) and Williams et al. (2004) use the relationship between density and susceptibility model values to map alteration zones associated with different contents of magnetite and hematite. Geology characterization can also be used for lithology mapping (Kowalczyk et al. (2010); Martinez et al. (2011); Fraser et al. (2012); Martinez & Li (2015)). Kowalczyk et al. (2010) use the scatterplot of susceptibility versus density contrast values from separate inversions of magnetic and gravity data to define lithology classes on regional scale. Using a similar approach, Martinez et al. (2011) and Martinez & Li (2015) define lithotypes from the scatterplot of susceptibility versus density values derived from inversion of gravity gradient and magnetic data on a deposit scale. The authors achieve lithology grouping (or equivalently, classification) in two different ways, either applying generic ranges extracted from literature or applying site specific ranges obtained from a priori geologic information. Other works on geology differentiation based on geophysically derived physical properties also focus on examining the clustering or grouping patterns among the physical properties values in a scatterplot (e.g., Bosch, 1999; Bosch et al., 2002; Bauer et al., 2003; Bedrosian et al., 2007).

Our present work is in a different category of geology characterization, and the focus is on mapping mineralized rock or ore. We identify classes in the scatterplot of physical properties to build a 3D pseudo-geology model by grouping the estimated susceptibility and conductivity values from independent inversions of magnetic and DC resistivity data from a copper deposit. In contrast to the work by Kowalczyk et al. (2010), which focuses on the regional scale, and by Martinez & Li (2015), which focuses on classifying the units of an iron deposit using generic or site specific a priori geologic information, our work combines the expected physical property relationships from textbooks and the rock types identified from sparse drill holes, into the differentiation of the mineralized unit. The method we are proposing enables a quick and integrated evaluation of targets, and is suitable for prospects
in the initial stages of exploration. The method is performed in four steps. First, the theoretical relations for the physical properties of the minerals associated to the rocks being investigated are established based on published reference values. In the second step, the established relations are applied to the scatterplot of the physical properties recovered from 2D inversions of the cross-section where the first drill holes were drilled. In the third step, the result of the classification is compared and validated against the drill hole information. In the last step, the classes defined and validated by geology information are applied to the scatterplot of the physical properties recovered from the 3D inversions. The final result will allow the interpreter to evaluate the prospect extent, distribution, and overall potential. The application of the classification to the scatterplot of physical properties recovered from the 2D inversions is an important step because the scatterplot shows less spreading of physical properties, consequently better defines groups. Additionally, it can be validated with the sparse geological data from drillcore to gain confidence to be applied to the 3D models.

We present the methodology and results of geology characterization based on inversions of magnetic and DC resistivity data at an iron oxide copper gold (IOCG) deposit named Cristalino in northern Brazil. The distribution patterns (i.e., clustering features) among the recovered susceptibility and conductivity values indicate that geophysical data inversions provide useful information for differentiating between different geological units such as iron formation and mineralized rock, thereby, characterizing the subsurface. A specific lithology type is not inferred for the region of mineralization due to the lack of a direct correlation between physical properties and lithology in an area of intense hydrothermal alteration such as in Cristalino. Instead, the correlation is more indicative of the mineralized zone resulting from hydrothermal alteration processes. The pseudo-geology model obtained this way can serve as a training example for us to learn the relation between an IOCG deposit and its geophysical signature. The workflow established in this case study can potentially be extended to exploration of covered deposits in the same mineral province and possibly to other regions in the globe.
Geophysical inversions have been demonstrating great success for understanding IOCG deposits. Leão-Santos et al. (2015) identify massive magnetite from hydrothermal alterations associated with the high-grade ore with the use of the magnetic amplitude inversion to an IOCG deposit in Carajás Mineral Province, Brazil. In the same province, Souza et al. (2015) use concatenated scheme with standard Euler deconvolution, total-field magnetic anomaly modeling, and magnetic amplitude inversion to identify a magnetite layer associated with copper ore in an IOCG greenfield exploration target. In South Australia, Zhdanov et al. (2012) apply joint inversion with Gramian constraints of gravity and magnetic data acquired over the Carrapateena IOCG deposit to interpret lithology and alteration patterns. Austin & Foss (2012) use inversion and forward modeling of magnetic and gravity data to generate and test models over IOCG deposits in the Gawler and Mount Isa provinces, Australia, confirming the importance of these methods for exploration of this type of deposit. Funk (2013) shows that magnetic, chargeability, and resistivity anomalies are important exploration vectors for IOCG deposits in Gawler Province but do not directly detect the mineralized rock. This is not the case in Cristalino, where the substitution of magnetite by chalcopyrite was extensive enough to form a low susceptibility zone associated with a high conductivity zone.

In the following, we first provide a brief geological background on IOCG deposits and specifically on Cristalino deposit. We then describe the geophysical data sets including magnetic and DC resistivity data. Next, we summarize the expected distribution behaviors of the physical property values for the deposit by combining theoretical relations with *a priori* information from the first drill holes. We will then perform joint interpretation of the geophysical inversion results, first with 2D inversions, which will be validate against the known geology, and then extend to the 3D inversions to evaluate the prospect potential. The final output is a 3D pseudo-geological model which successfully delineates the mineralized unit.
2.2 Geological Setting

IOCG deposits usually occur along fault splays off crustal scale extensional faults. They are associated with diverse rock types, resulting in a wide variety of deposit styles and mineralogy (Hitzman, 2000). The deposits are formed by magmatic-hydrothermal processes, are structurally controlled, and have copper and gold as the main economic elements (Groves et al., 2010). The primary mineralogical characteristic of all deposits in this class is the abundance of hydrothermal iron-oxide, either magnetite or hematite. In addition, the primary economic characteristic is the presence of copper, either as chalcopyrite or bornite. The most common texture of the ore is breccia, associated with stockwork, veins and veinlets (Hitzman, 2000).

The geophysical response of IOCG deposits is complex. If a deposit is associated with magnetite, such as at Cristalino, it will have a strong magnetic anomaly associated with it. However, the mineralized rocks can be weakly magnetic, as chalcopyrite is formed from the reaction between magnetite and the last pulses of the hydrothermal fluid. Consequently, the magnetic susceptibility of the mineralized unit will be dependent on the amount of magnetite left from the hydrothermal reaction. Both minerals, magnetite and chalcopyrite are conductive. Although chalcopyrite usually is more conductive, its general range overlaps the magnetite conductivity range. For IOCG deposits, the texture of mineralization is expected to be variable and it is common to have a breccia zone that gradually becomes stockwork, then veins, veinlets or stringer, and being disseminated. The breccia zone is expected to be the most conductive region because the brecciation process forms a matrix composed of massive chalcopyrite.

Cristalino copper gold deposit, located in northern Brazil (Figure 2.1(a)), is hosted by a splay of the Carajás fault in the southeastern part of Carajás Mineral Province (Figure 2.1(b)), both have been labeled on the image of the total gradient of the magnetic data (Figure 2.2). This splay fault strikes northwest, dipping approximately 50° southwest, and it acted as a conduit for hydrothermal fluids. The copper and gold ore formed by hydrothermal
alteration of a volcano-sedimentary sequence, consisting of mafic and felsic volcanic rocks interlayered with iron formation and intruded by a younger magnetic gabbro dyke (Figure 2.3). This hydrothermally-altered region is referred to here as the mineralized unit or zone, which is mainly composed of amphibole, quartz, feldspar, biotite, chalcopyrite, and magnetite, according to the modal distribution shown in Figure 2.4. The splay fault mainly cuts through the iron formation layer, which is composed by magnetite and quartz, and has hematite as a secondary mineral. The iron formation is the unit that reacted most with hydrothermal fluids to form the copper ore. The fluid reacted with the magnetite from the iron formation to form chalcopyrite (Huhn et al., 1999). The deposit is covered by a thick soil profile whose thickness varies between 40 and 60 m (Vale S.A., 2004), but it is not represented in the geological model used in this work.

Figure 2.1: (a) Tectonic location of the Carajás Mineral Province at the southeastern margin of the Southern Amazon Craton, Brazil (Almeida et al., 1981), (b) Geologic Map of the Carajás Mineral Province showing the area of the total gradient of the magnetic data in Figure 2.2 (yellow box), and study area (red box), Cristalino deposit, which is hosted by a volcanosedimentary sequence (modified from DOCEGEO, 1988; Grainger et al., 2008).
Figure 2.2: Image of the total gradient of the magnetic data showing the study area, Cristalino deposit, the main mineral deposits and faults in the southern part of the Carajás Mineral Province.

Figure 2.3: Geologic cross section of line -1800 of Cristalino copper deposit showing the drill hole traces and chalcopyrite concentration of the hydrothermal zone hosted by the volcanic (mafic and felsic) and sedimentary (iron formation) rocks (modified from Vale S.A. (2004)). The red bar in the legend corresponds to the representation of the proportion of 10% of chalcopyrite.
The category “other” includes apatite, titanite, muscovite, ilmenite, bornite, and chalcocite (modified from Vale S.A. (2004)).

The deposit is estimated at 482 Mt at 0.65% Cu and 0.06 g/t Au (NCL Brasil, 2005). The chalcopyrite occurs in the form of stockwork, stringers, breccias, or dissemination in the host rock; and can be filling fractures that cut the sequence (Huhn et al., 1999) (Figure 2.5). Cristalino is a deposit that has many challenging characteristics for interpretation: it is under a thick overburden, the mineralized zone is hosted by a fault zone that displaces and reacts with a iron formation layer, and it has four different ore textures.

The analysis of the theoretical behavior of susceptibility and conductivity for chalcopyrite and magnetite from generic values in the literature (Telford et al., 1990) shows that the susceptibility is distinct between the two minerals, magnetite is four to five orders of magnitude more susceptible than chalcopyrite (Figure 2.6). Although chalcopyrite usually is more conductive, its general range overlaps with that of magnetite conductivity. A visual evaluation of the rock samples from the drill cores in the field allows us to relatively place the main geological units in the plot of generic values (Figure 2.6). The iron formation and
mineralized units are represented by the ellipsoids, and we can represent its mineralogical variability trend by the arrows in the plot. In Figure 2.6, the brown dotted line represents a trend for iron formation and red dashed line for mineralized unit. The units can be placed in any point belonging to these lines according to its mineralogical variability of chalcopyrite and magnetite, which is associated with the relative position to the fault zone and intensity of hydrothermal activities. Since petrophysical measurements are not available in this area, which is a common situation for many remote exploration regions, this plot represent only relative relations derived from the theoretical reference values. The relative relations defined this way for the ore, iron formation and host rock classes will be applied to the physical properties recovered from the geophysical inversions also in a relative manner, but guided by the clusters present in the models.

In this case study, we are looking for the geophysical responses of the main ore mineral, chalcopyrite, which was formed hydrothermaly through the reaction with pre-existing mag-
netite. The pre-existing magnetite has two sources: sedimentary and hydrothermal. The sedimentary magnetite from the iron formation was already in the system when the hydrothermal process started, and the first pulses of hydrothermal fluid formed magnetite and altered the iron formation and volcanic rocks. Therefore we are also looking for the geophysical signature of the units that contain magnetite to better understand the whole system. The magnetite is mainly present in the iron formation, but there are variable amounts in the mineralized unit, depending on the effectiveness of the last pulses of hydrothermal fluids to replace magnetite by chalcopyrite. The host rocks do not have significant amounts of these minerals. The chalcopyrite and magnetite show anomalous conductivity values and the magnetite shows anomalous susceptibility values. For this reason, we use DC resistivity and magnetic data to differentiate the units with these minerals.
2.3 Geophysical data and inversion method

The study area is located in a low latitude region near the magnetic equator, with field inclination of -3.5°, declination of -19°, and strength of 25,500 nT. Two magnetic data sets were used in this study, the ground data for the 2D susceptibility inversion and the airborne data for the 3D inversion. The ground magnetic data were acquired on the same lines as in the DC resistivity survey and for the line used in this study, line -1800 (Figure 2.3 and Figure 2.7, the noise level was not highly variable between stations. However, the difference in noise level between lines and the spikes produced by magnetic soil caused many problems to fit the data in the 3D inversion. For this reason, the airborne magnetic data were used for the 3D inversion because the higher distance from the ground naturally filters the high frequency variation associated with shallow sources (i.e., magnetic overburden), which are not the focus of this work.

The ground magnetic data were acquired in 1998 over east-west lines, spaced at 200 m, and 10 m station-spacing (Figure 2.8(a)). The airborne magnetic data were acquired in 2000 in a draped survey over east-west lines, spaced 250 m, and a 200-m average terrain clearance (Figure 2.7(a)). The magnetic data are comprised of two main magnetic anomalies with overlapping patterns. The DC resistivity data were acquired in 1999 on the same lines of the ground magnetic survey. A dipole-dipole array with a dipole length of 60 m and 5 m-spacings (levels) were used to acquire the data (Figure 2.8(b)).

Areas with strong hydrothermal alteration, such as Cristalino, are commonly associated with remanent magnetization. This effect becomes even more evident close to the magnetic equator because of the weaker inducing field compared with other regions of the globe. Leão-Santos et al. (2015) show that the presence of remanent magnetization changes the anomaly geometry and amplitude of an IOCG deposit in Carajás, imposing a challenge for the construction of a susceptibility model. In order to investigate the presence of remanent magnetization in the Cristalino deposit, we first compared the anomaly patterns of the field data with synthetic data of a dipping magnetic body (simulating the expected iron
Figure 2.7: Airborne magnetic data over the deposit area, showing the airborne (gray dotted lines) and ground (black dashed lines) surveys. Line -1800 of the ground survey was used for the 2D inversion and the airborne lines for the 3D inversion.
Figure 2.8: a) Magnetic profile of line -1800, field inclination of -3.5°, declination of -19°, and strength of 25,500 nT, b) resistivity pseudo-section (vertical exaggeration of 5x) for line -1800, dipole-dipole array with a dipole length of 60 m and 5 n-spacings, and c) topography of line -1800. The gray dotted rectangle shows the model extent.

formation layer). The comparison shows that the same anomaly geometry is present in both datasets showing a magnetic low in the center of the anomalous body with magnetic highs in the northern and southern edges. This result indicates that the direction of remanence is close to the inducing field, which was confirmed by applying the cross-correlation-based magnetization direction estimation method (Dannemiller & Li, 2006). Through this method we estimated a total magnetization direction with an inclination of 0° and declination of -18°. These values are close to the inducing field inclination of -3.5° and declination of -19°. Since the direction of total magnetization is sufficiently close to the direction of the inducing field, we performed the magnetic inversion by using the inducing field direction as the magnetization direction. The recovered model, therefore, represents an effective susceptibility that includes the remanence effect. Consequently, the recovered model will have higher-than-expected susceptibility values, but the spatial distribution of the effective susceptibility is valid. For the geology differentiation presented here, which is based on the patterns of
recovered physical properties in the cross plot, the effective susceptibility is sufficient.

We used the 3D inversion algorithm developed by Li & Oldenburg (1996) to invert the magnetic data and Li & Oldenburg (2000a) to invert the DC resistivity data. Both inversions consider the following relationship:

\[ \mathbf{d} = \mathbf{F} \mathbf{m}, \]  

(2.1)

where \( \mathbf{d} = [d_1, d_2, ..., d_N]^T \) is the data vector (magnetic or DC resistivity), \( \mathbf{F} \) is the forward operator, and \( \mathbf{m} = [m_1, m_2, ..., m_M]^T \) is the model vector containing the physical property (susceptibility or logarithmic conductivity) of each cell in the model, with \( N \) being the number of data and \( M \) the number of model cells. For magnetic inversion, the forward operator \( \mathbf{F} \) is a simple linear system:

\[ \mathbf{d} = \mathbf{G} \mathbf{m} \]  

(2.2)

where \( \mathbf{G} \) is the \( N \times M \) sensitivity matrix, which contains the physical relationship between each cell in the model with respect to each datum. The inverse solution is obtained using Tikhonov regularization by solving the following constrained minimization problem:

\[ \phi = \phi_d + \beta \phi_m, \quad \text{subject to} \quad \mathbf{b}_l \leq \mathbf{m} \leq \mathbf{b}_u, \]  

(2.3)

where \( \phi \) is the objective function, \( \phi_d \) is the data misfit function, \( \phi_m \) is the model objective function, \( \beta \) is the regularization parameter, and \( \mathbf{b}_l \) and \( \mathbf{b}_u \) are the upper and lower bounds, respectively, of the model values. Here we used \( \mathbf{b}_l = 0 \) and a large \( \mathbf{b}_u \) to simulate a situation with no upper bound for the magnetic inversions. The DC inversion was done using the logarithm of conductivity values. Therefore, the bound constraints are naturally taken care of by the use of the logarithm of conductivity as the unknowns.

The inversion process requires the standard deviation of the noise (\( \epsilon \)) in the data, but this value is usually not known for field data. Therefore, to ensure an adequate estimation of the standard deviation, we used the following steps: first assumed \( \epsilon = 1 \) to run the first inversion and used the discrepancy principle (Parker, 1994) to obtain an initial estimate of the regularization parameter (\( \beta \)) magnitude. After running inversions using a large range
of βs, we applied the L-curve criterion (Hansen, 1992) to select the optimum β. The misfit corresponding to this optimum β was then used to estimate the adjusted ε from the misfit function:

\[ \phi_d = \sum_{i=1}^{N} \left( \frac{d_{i}^{\text{obs}} - d_{i}^{\text{pre}}}{\epsilon_i} \right)^2 \]  

(2.4)

\[ \epsilon_{\text{adjusted}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (d_{i}^{\text{obs}} - d_{i}^{\text{pre}})^2} \]  

(2.5)

where \(d_{i}^{\text{obs}}\) is the observed data and \(d_{i}^{\text{pre}}\) is the predicted data corresponding to the optimum β. This relation is obtained by assuming the target misfit \(\phi_d = N\) (discrepancy principle) in the misfit function in equation 3.9.

In the following step, we built a new Tikhonov curve using \(\epsilon_{\text{adjusted}}\) as the estimated standard deviation of noise. The procedure of adjusting the ε value was repeated a second and third times to fine tune the estimation, and the third L-curve was used to select the final model. The final susceptibility model was selected using the work flow previously described and the final adjusted standard deviation was estimated to be 20 nT, which is 0.9% of the magnetic data range.

For the magnetic inversion, a constant value was assigned for ε to ensure consistency in the model misfit function. However, this is not recommended for DC data because the signal level depends on the transmitter-receiver separation. Therefore, one constant value would not be appropriate for the different signal levels and associated error. For this reason, the minimum noise level was defined as 0.40567 mV/A. The initial ε was this minimum voltage plus 1% of the datum, and this value was then adjusted to 2% after using the procedure described to estimate \(\epsilon_{\text{adjusted}}\).

2.4 2D geophysical inversions and geology characterization

The 2D inversions of the magnetic and DC data were first presented by Melo et al. (2015). The 2D assumption is a reasonable one given the geology strikes approximately north-south
and lines are perpendicular to the mineralized zone and to the topographic ridges. The recovered susceptibility model for line -1800 (Figure 2.9) identified a high susceptibility zone, which is partially associated with the mineralized rock, but mainly with the iron formation and gabbro dyke on the west side of the line. The area of low susceptibility is mainly coincident with the mineralized rock. One of the significant features is the delineation of a sharp boundary between the magnetic rocks and the mineralized zone despite the use of the classic regularized smooth inversion.

![Figure 2.9: Susceptibility model of the line -1800 showing the sharp boundary between the areas of high and low susceptibility.](image)

In the recovered conductivity model for line -1800 (Figure 2.10), the high conductivity anomaly near the center of the line is associated with the mineralized zone. The high conductivity is mainly contained in the region of low susceptibility anomaly. In the magnetic susceptibility section, the main anomaly is associated with a specific lithology type, the iron formation. On the other hand, the chalcopyrite, which comes from the hydrothermal overprint on the preexisting lithology, replaces the magnetite in the iron formation and iron-silicates in the volcanic rocks. Therefore, the conductivity section better represents the hydrothermal alteration than the lithology. The shallow anomalies of high conductivity values are associated with conductive overburden. The depth of investigation (D.O.I.) of
the DC resistivity data was estimated altering the reference models in the inversions and identifying the region of similarity between features in the models (Oldenburg & Li, 1999).

Figure 2.10: Conductivity model of the line -1800. The estimated depth of investigation (DOI) is highlighted by the white line.

The first phase of the geology characterization method presented in this case study is based on 2D inversions because the patterns in the physical properties can be more readily identified. In addition, the method takes advantage of the initial geologic information available from the first drill holes drilled in the area, even though petrophysical measurements are not available to be incorporated in the inversion. We are looking for the conductivity and susceptibility anomalies caused by chalcopyrite and magnetite. However, these minerals do not occur purely, but as components of the geological units. The magnetite is associated with quartz and hematite in the iron formation and, in the mineralized unit, the chalcopyrite is associated with magnetite, hematite, and quartz near the iron formation, and with silicates and carbonates near the volcanic host rock. Consequently, it is hard to specify the conductivity and susceptibility of the geological units without petrophysical analysis, which are commonly not available at this stage of an exploration project in remote areas. Knowing that chalcopyrite and magnetite are the minerals of interest, however, we can use the conceptual absolute behavior of these minerals (e.g., Telford et al., 1990) to establish relative
behaviors for the geological units in our study area.

As previously described in the geological setting section, the mineralized zone of the Cristalino deposit has four different textures. The conductivity and susceptibility behaviors will vary according to the relative amount of chalcopyrite and magnetite associated with each type of texture, thus defining a relative trend for the whole mineralized unit (Figure 2.6). The massive unit is the one with the largest volume of chalcopyrite, and smallest volume of magnetite; therefore, it has the highest values of conductivity associated with the lowest values of susceptibility. As the texture of the mineralized units become stockwork and stringers, the amount of chalcopyrite decreases and that of magnetite increases. The disseminated unit is expected to have the highest susceptibility and slightly lower conductivity values because the chalcopyrite grains are disseminated in the iron formation, which is a magnetite-rich unit. Another trend that can be defined is for the iron formation unit (brown dotted line in Figure 2.6). This trend is estimated to have a constant susceptibility near the middle of the conceptual range and variable conductivity. The conductivity of the iron formation unit will increase near the fault zone (conduit of hydrothermal fluids) because of the higher amount of hydrothermal chalcopyrite in this part of the unit.

The objective of relating generic values of susceptibility and conductivity from literature and the rock samples of this prospect is to identify patterns of relative behavior between them. This behavior will serve as a guide for the interpretation of the physical properties recovered from inversions. This is a suitable assumption since geologic interpretation of inverted physical properties from unconstrained inversions is only possible in a relative sense, because the models exhibit reduced contrasts and less variability than true values (e.g. Lelièvre et al., 2009; Sun & Li, 2015).

Susceptibility versus conductivity model values were cross plotted on a log-log scale (Figure 2.11a) with the objective of identifying different classes. The conceptual behavior previously discussed was applied to identify the patterns associated with classes 1 and 2 (Figure 2.11(b)). Classes 3 and 4 were selected based on preferential groupings in the remaining
recovered physical property values. Class 1 comprises the zone with highest susceptibility and with intermediate conductivity values. Class 2 is well defined as a zone of high conductivity and variable susceptibility. Both class 3 and class 4 spread over a wide range of susceptibility and conductivity values, but class 3 appears to cluster at low conductivity values and class 4 at intermediate conductivity values. No transformation was done between absolute values from literature and the recovered values from inversion. The classification process applied the relative position of the classes to the scatterplot, and adjustments were made respecting natural boundaries present in the inverted model values.

Figure 2.11: a) Scatterplot of susceptibility and conductivity values from the 2D inversions, the points are correspondent to the same cell location, b) classification based on the formulated conceptual behavior (Figure 2.6), and natural groupings in the data. Class 1 is associated to the iron formation, class 2 with the copper ore, class 3 with the host rock 1, and class 4 with the host rock 2.

Each of the points in the scatterplot is associated with a model cell in the spatial domain. We therefore can express the classification results (Figure 2.11(b)) in spatial domain (i.e., in section view) for a more intuitive geology interpretation. The classes show spatial consistency within each group (Figure 2.12). The comparison between this result and the information
from the drill holes allows the interpreter to associate class 1 with the magnetic units (iron formation and gabbro dyke) and class 2 with the copper ore, similar to the conceptual behavior shown in Figure 2.6. Classes 3 and 4 show some separation in the scatterplot and are reflecting variations in the volcanic rocks that are not obvious in the drill hole logging, but are likely reflecting the hydrothermal alteration in these units. Here we assigned class 3 to be associated with one type of host rock and class 4 with a second type of host rock (Figure 2.13). Further investigation of the geochemical and physical property variations would clarify this relation. The deep weathering profile in the Amazon region overprinted the magnetic response of the iron formation layer near the surface, for this reason the magnetic anomaly appears deeper than it should be. Because the geological model (Figure 2.3) was constructed without the overburden layer, we kept our interpretations consistent with the understanding regarding geology. Since the iron formation is parallel to the volcanic rocks, we interpreted a continuous layer up to the surface for class 1. This model is considered a pseudo-geology model because it is derived from geophysical models and extrapolates the region of known geology (drill cores).

Figure 2.12: Spatial distribution of the classes identified in the 2D scatterplot, showing the geology differentiation in the section. The dark gray line shows the estimated depth of investigation (D.O.I.) for the resistivity inversion.
The resulting 2D pseudo-geology model proved to be highly consistent with the geologic section constructed from drill cores and with the location of the mineralized unit (Figure 2.14). The model identified important geological-geophysical relationships for a complex area. The defined relationships have potential to be used as exploration criteria, allowing explorers to extend them to 3D to increase the comprehension of undrilled areas. The application to 3D-inversions is going to give us an estimation of the location, geometry, and extension of the mineralized rock. The identification of these geological-geophysical relationships and its application in 3D represents a step forward in order to make new discoveries.

2.5 3D geophysical inversions and geology characterization

We now extend the approach described in the preceding section to 3D inversion results, since the geology in the deposit area is inherently 3D. The study area extends 1.1 km in the east-west direction, and 1.5 km in the north-south direction. The data from both geophysical methods were inverted using the same mesh to ensure the spatial compatibility between the models. The mesh is composed of cubic cells of 50 x 50 x 50 m, and padding cells were used in the north and east directions, as well as, at depth. Additionally, for the magnetic...
Figure 2.14: Geology characterization overlain by the available geologic information from drilling, showing the high spatial correspondence between the interpreted ore class and the high concentration of chalcopyrite, and between the interpreted iron formation class and the magnetic rocks (iron formation and gabbro dyke).

For the magnetic data, we removed the International Geomagnetic Reference Field (IGRF) and performed a regional-residual (Figure 2.7(b)) separation in the data using an inversion-based method (Li & Oldenburg, 1998). The inversion-based regional-residual separation consists of performing an inversion over an area larger than the area of interest, once the susceptibility model is obtained, the area of interest is scooped from the model and these cells are set to zeros. A forward calculation of the magnetic field is then performed over the modified susceptibility model (i.e., the regional susceptibility model) and the calculated data is considered the regional field, which is then removed from the observed magnetic data, and the resulting residual data are used for the inversion (Figure 2.7(c)). For Cristalino, the padding cells of the mesh extended 3 km beyond the 1.1 x 1.5 km of the study area in all directions, increasing the size gradually from 50 to 800 m. A similar approach
to estimating data standard deviations and choosing the optimal regularization parameter applied in 2D was also applied to the 3D problem to produce a magnetic susceptibility model (Figure 2.15(a)).

The recovered susceptibility model shows two magnetic bodies dipping approximately 50° to the southwest. The northern body has a larger volume than the one to the south. The large susceptibility values in the recovered model are deemed to be due to remanence in the magnetic rocks. As previously discussed, the remanence and inducing field directions are similar, for this reason the recovered model includes the remanence effect and represents the effective susceptibility. Therefore, even if the magnitudes differ from realistic values, the spatial distributions are valid, especially considering that the geology characterization method presented here is based on the relative relationship of physical properties. Remanent magnetization is not believed to be affecting the geometry of the recovered model, which is highly consistent with the known geology. The recovered conductivity model (Figure 2.15(b)) has the main anomaly located in the central and north-northwestern parts of the model. The other high-conductivity anomalies over the area are related to the conductive overburden and are limited to the near surface only. The same depth of investigation curve estimated for the 2D data was applied for the 3D model, which was clipped at an average depth of 250 m below the topographic surface.

Similar to the 2D geology characterization, the susceptibility versus conductivity values from the 3D models were cross-plotted on a log-log scale (Figure 2.16(a)), and the same conceptual behavior was applied to identify patterns. The scatterplot from the 3D inversion results shows more smoothing and spreading of the recovered physical property values throughout a wider range than for the 2D inversions. Despite the difference in the susceptibility and conductivity ranges, the same conceptual behavior can be applied to define the class associated with the magnetic units (Figure 2.16(b)). No transformation was applied to the boundaries of classification, similarly to the 2D classification, the classes defined based on the conceptual behavior of the minerals were applied respecting natural clusters in the
Figure 2.15: a) Inverted 3D susceptibility model showing the magnetic anomalies associated mainly with iron formation, b) section of the susceptibility model at line -1800, c) inverted 3D conductivity model showing the high conductivity anomaly in the center and north-northwestern parts of the model, which is associated with copper ore, and d) section of the conductivity model at line -1800.
model values.

Figure 2.16: a) Scatterplot of susceptibility and conductivity from the 3D inversions, the points are correspondent to the same cell location, b) revised classification based on the 2D geology differentiation (Figure 2.11). Class 1 is associated to the iron formation, class 2 with the copper ore, class 3 with the host rock 1, and class 4 with the host rock 2.

The spatial distribution of the classification of the 3D models (Figure 2.17(a)) and the interpretation of the classes (Figure 2.17(b)) shows the estimated distribution of the copper ore unit (class 2). The interpreted ore unit is mainly concentrated in the center and north-northwestern parts of the model, where the cells belonging to this class are continuous in depth. The cells in the west are limited to the first layer and are probably associated with conductive overburden. The result of the geology characterization also shows that class 1, associated with the iron formation (and gabbro dyke) unit, is not continuous, and pinches out in the center of the model, where the ore unit has its maximum thickness. The difference in the distribution of the host rocks 1 and 2 (classes 3 and 4) might be reflecting differences in the mineralogy of the volcanic rocks, as well as in the hydrothermal alteration associated with the mineralization event.
Figure 2.17: a) Spatial distribution of the classes identified in the scatterplot of the 3D inversions, b) section of the classification of the model at line -1800, c) the interpretation of these classes based on the results of the 2D characterization, and d) section of the interpreted model at line -1800.
The case study presented here is about a known copper deposit, which was extensively drilled. For this reason, we can compare the result of the 3D geology characterization (Figure 2.17(b)) with the geological map of the same area. Because the deposit is under cover, the geological map was built with the drilling information using the first rock unit present under the overburden. This map was not used in the 3D geology characterization process, but was used only to validate the interpretation and confirm the feasibility of our geology characterization method. The geological map (Figure 2.18(a)) shows the distribution of the mineralized unit, which has its maximum thickness where the iron formation pinches out due to the reaction between magnetite and hydrothermal fluids that formed the chalcopyrite. For comparison with the geological map, the first layer (50 m) of the 3D geology characterization volume was removed because it is heavily influenced by the overburden conductivity response (Figure 2.18(b)). The comparison between the ore class in the 3D geology characterization and the mapped ore shows that the characterization was able to identify the core of the mineralized unit. Its smaller extent is probably associated with differences in the ore texture, since the stringer and disseminated chalcoprytes are much less conductive than massive and stockwork chalcopyrite. In this case, induced polarization would be an important complementary method for identifying this types of mineralization. The north-northwestern part of the 3D geology characterization that is associated with the ore class does not have association with mapped mineralized unit. However, this portion, which is spatially coincident with the fault, is associated with hydrothermal chalcopyrite inside the iron formation, as observed in the drill holes of line -1800. The distribution and volume of class 2 can be used for planning the next drill holes or for comparison with the potential volume of mineralized rock in other targets. The comparison between the unit classified as ore and the drill hole logs indicates that the minimum amount of chalcopyrite required for this method to identify a cell as ore is 2.5%. This equates to roughly 1% Cu and indicates that the method is effective at mapping the higher-grade zones (the average resource grade is 0.65% Cu).
Figure 2.18: a) Geological map of Cristalino copper deposit (adapted from Vale S.A., 2004), and b) top view of the 3D geology characterization volume overlain by the contours of the geological map, showing that the method was able to identify the core of the mineralized zone. The first layer of cells (50m), which is mainly associated with overburden, was removed.
Most of the northern part of the iron formation class is spatially correlated with the mapped iron formation, especially considering we used unconstrained smooth inversion to build the models. On the other hand, the southern part of the iron formation class is smaller than the actual mapped iron formation. This difference may be attributed to variations in magnetite-over-hematite ratio inside the iron formation layers. To avoid misinterpretation due to these variations, gravity would be an important complementary method to better recover the iron formation response. The host rock-1 and host rock-2 classes are probably associated with hydrothermal variations in the volcanic rocks, which can only be confirmed by a detailed geochemical study.

2.6 Discussion

The 3D geology characterization scheme presented here is shown to be feasible for characterization of the mineralized rock. The interpretation mainly identified the core of the ore unit, which is associated with the most conductive parts related with stockwork and massive chalcopyrite. Additionally, it was able to identify the iron formation unit, which has high spatial correlation with the known geology. Variations in texture and proportions of chalcopyrite and magnetite through the deposit might affect the classification of the mineralized unit. If the area has high susceptibility, but also high conductivity it is classified as ore in the scheme presented in this work. However, if an area has high susceptibility but the texture of the ore is dominantly disseminated, which decreases the conductivity, it will be classified as iron formation.

The differences between the characterization result and the known geology, presented here, are due to the limitations of the DC resistivity and magnetic methods in this geological setting. The DC resistivity method is unable to detect disseminated chalcopyrite, which is one of the ore textures present in the deposit. The magnetic method identify anomalous magnetic areas that can be associated with ore containing high amount of magnetite, but end up being classified as iron formation. Therefore, the addition of other geophysical methods, such as induced polarization and gravity would increase the interpretation accuracy.
Incorporating more data sets would also increase the complexity of pattern recognition and interpretation, therefore automated methods would be necessary to facilitate interpretation.

The inversion of magnetic data in this case deserves additional attention. Many geological units and IOCG deposits in the study region are known to have significant remanent magnetization. It is now well understood that appropriate processing and inversion are required to deal with such situations if the remanent magnetization direction deviates significantly from the induced component. Fortunately, the Cristalino deposit is a special case where the remanent magnetization is closely aligned with the inducing field at the deposit scale, so no special treatment was required. In general, however, we believe that inversion algorithms capable of dealing with the influence of remanent magnetization must be used if evidence and data analyses suggest that the total magnetization direction is different from the inducing field direction (e.g., Li et al., 2010; Ellis et al., 2012; Fullagar et al., 2008; Lelièvre & Oldenburg, 2009; Li & Sun, 2016; and Fournier et al., 2016).

If petrophysical information is available, the interpreter should incorporate this information in the inversion process to obtain more accurate models, for example, by applying constraints using guided fuzzy c-means clustering (Sun & Li, 2015). However, petrophysical data takes time to be acquired and is often not available in the first stages of exploration. Nevertheless, explorationists still need an objective interpretation method that combines all information available. For this reason, in such situations we propose the following workflow:

1. visually log the core samples available to define rock types and relative proportion of the minerals of interest,
2. formulate the conceptual behavior expected for those minerals,
3. construct 2D unconstrained geophysical models with the data available,
4. apply the expected behavior patterns to the scatterplot of the recovered physical properties,
5. validate the 2D interpretation against the available geologic information,
6. apply the interpretation to the 3D unconstrained geophysical models, and

7. analyze the mineralization potential.

2.7 Conclusions

We have developed a geology characterization method of 3D geophysical inversions capable of identifying copper mineralization. This case study has shown the results of combining different geophysical methods to define geological units. The geology characterization method presented here is a powerful tool for quick and integrated evaluation of targets with geophysical data and sparse geologic information. The method can be applied in a large variety of situations, but is especially suited in the first stages of exploration of deep targets or targets under cover, such as Cristalino.

For Cristalino IOCG copper deposit, we have shown in the 2D inversion that the high susceptibility area, corresponding to the iron formation and gabbro dyke, is separated by a sharp boundary from the low susceptibility area, corresponding mostly to the ore deposit, and from the intermediate susceptibility host rocks. The conductivity values recovered from the DC resistivity inversion show a good correspondence with the mineralization. The same behavior is observed in the 3D inversions, but the models are smoother and the separation of the groups in the scatterplot is not so obvious, making the 2D inversion a necessary step.

This geology characterization method has shown its feasibility for identification and characterization of the mineralized unit. It proved to be a powerful interpretation tool, which can help understand exploration areas in an integrated manner. It also provide means of estimating the mineralized rock extension and comparing targets. Integrated interpretation increases the knowledge of an area while decreasing the risk on decision making.

2.8 Acknowledgments

A. T. Melo acknowledges the CNPq (Brazilian National Council for Scientific and Technological Development) for the scholarship. We thank Vale S.A. for providing the data used in this study and for permission to publish results obtained from them. This work is supported
in part by the Gravity and Magnetics Research Consortium. We thank the N. Williams, M. Leão-Santos, P. Kowalczyk, one anonymous reviewer, and R. Smith, for all their valuable suggestions that helped to improve the manuscript.
CHAPTER 3
GEOLOGY DIFFERENTIATION BY APPLYING UNSUPERVISED MACHINE LEARNING TO MULTIPLE GEOPHYSICAL INVERSIONS: METHODOLOGY

Greenfield exploration in areas under cover is challenging because little a priori geological information is available to improve geophysical models. To overcome this difficulty, multiple geophysical methods are often employed in order to increase the subsurface understanding. Therefore, effective quantitative methods for integrating multiple inverted physical property models are necessary to increase the value of information and take the interpretation further into geology differentiation of different units. For this reason, applications of machine learning are growing in geosciences 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 complete and accurate, 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 on new targets. Unfortunately, the lack of training data for greenfield exploration only allows the application of unsupervised machine learning, which has the task of exploring hidden structures in the data. Therefore, considering a greenfield exploration scenario, where specific a priori geologic information is unavailable, we perform independent minimally constrained inversions of magnetic, gravity gradient, and DC resistivity data over a synthetic geologic model and investigate the use of unsupervised ML in geology differentiation. The synthetic model 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. The performance of connectivity, density, distribution, centroid, and correlation-based clustering methods on
the identification of the three geologic units is evaluated. We show that correlation-based clustering gives 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.

3.1 Introduction

The human mind works in amazing ways and is capable of incredible things. It is even capable of building machines to make decisions similarly to humans, the artificial intelligence. The applications of artificial intelligence are growing in geosciences, and one particularly interesting example is the partnership between IBM’s supercomputer Watson and the mining company Goldcorp to select and analyze exploration targets in the Red Lake Complex in Canada. Mark Fawcett from IBM says “We are not replacing geologists, we are making the geologist infinitely more powerful by being able to go through information that they can use to make decisions” (Strong, 2017). However, similar to the human learning process, the algorithms need time to learn when they start in a new field. Although these tools have demonstrated efficiency in other areas, such as medical industry by helping diagnose complex cancers, they are still in infancy in the mining industry and will need to go through many learning stages to reach efficiency in mining exploration (Strong, 2017).

Within artificial intelligence, machine learning (ML) is the area that develops algorithms capable of learning and enabling computers to take human-like decisions or more. The algorithms recognize patterns without being explicitly programmed as they experience different situations. If the learning process uses input information with known output classification, it is defined as supervised learning. In supervised learning, large data sets are used in the learning process to improve classification and decrease the error rate in classifying data with unknown outputs (e.g., neural networks and support vector machine). In unsupervised learning, there is no input information with known output results for training. For this reason, these algorithms focus on exploring the data structure to find insights that are not obvious to human eyes (e.g., clustering algorithms and self-organizing maps). In mining exploration,
supervised ML is better suit for brownfield districts where the data from mines can be used for training the algorithms and targeting new prospects. On the other hand, in greenfield exploration, unsupervised ML is powerful for selecting new targets. However, it is important to emphasize that machine learning tools do not replace the human decisions, but may remove human bias from the analysis of data and provide speed and accuracy. The interpretation still in the geoscientist’s hands.

Because new ore discoveries have become increasingly expensive and risky as they tend to be deeper, all available information needs to be effectively integrated and evaluated in order to decrease the exploration risk. Therefore, an increasing amount of geologic, geochemical, and geophysical data are being acquired in mineral exploration to diminish the associated risks. The amount of data being acquired requires effective methods for extracting information from them. For this reason, machine learning methods have shown a strong potential to improve the rate of new discoveries of mineral deposits, since it can quickly evaluate large amounts of data quantitatively to discover hidden structures in the data.

Machine learning has been applied to the interpretation of geophysical data in petroleum exploration for several decades. Commonly used methods for facies classification from well log data and seismic attributes include hierarchical clustering (e.g., Serra & Abbott, 1982), modal distribution analysis (e.g., Wolf & Pelissier-Combescure, 1982), k-means clustering and discriminant analysis (e.g., Delfiner et al., 1987), neural networks (e.g., Baldwin et al., 1990; Rogers et al., 1992), graph-based clustering (e.g., Ye & Rabiller, 2000), and many others (e.g., Schlanser et al., 2016; Abreu et al., 2016). The identification of seismic facies using seismic attributes and velocity models have relied on the application of neural network (e.g., Meldahl et al., 1999), k-means and fuzzy c-means clustering (e.g., Barnes & Laughlin, 2005), self-organizing maps (e.g., Strecker & Uden, 2002), in different geological settings (e.g., Coléou et al., 2003; Gao, 2007; Roy et al., 2014; Zhao et al., 2015; and Qi et al., 2016).

In geological mapping, ML tools have been applied to airborne geophysical data, such as radiometric, magnetic, gravity, and electromagnetic, focusing on the construction of 2D
pseudo-lithology maps for interpretation. For example, Paasche & Eberle (2009) apply fuzzy c-means clustering to radiometric, magnetic and gravity data to produce a zoned geophysical map that outlines geological units. Eberle & Paasche (2012) apply fuzzy Gustafson-Kessel clustering to satellite imagery, airborne radiometric, and regional geochemical data to construct a pseudo-lithology map. Carneiro et al. (2012) apply self-organizing maps to magnetic and radiometric data to extract geophysical signatures associated with lithology types and produce a pseudo-geologic map over an area with gold deposits in the Amazon region. Other works also focus on 2D geological mapping applying ML to multiple geophysical data (e.g., Ranjbar et al., 2001; Martelet et al., 2006). Geological mapping can also be achieved using 3D physical property models. For example, Paasche et al. (2006) apply fuzzy c-means clustering to identify sedimentary units from physical property models and well log data.

In mineral exploration, Barnett & Williams (2006) use known gold deposits to train a neural network and construct a favourability map in regional scale using multiple data sets. Mahmoodi et al. (2014) apply fuzzy c-means to down hole data from a Ni deposit to characterize rock types and mineralization. Caté et al. (2017) apply supervised machine learning algorithms to predict the presence of gold, in drill cores, from geophysical logs.

Overall, much of the existing works using ML either focuses on large-scale applications such as on regional geology and prospectivity mapping, or on formation scale such as lithofacies classification in drill holes. Only limited work is available on deposit scales, where ultimately drilling targets must be chosen. For this reason, and aiming to make advances on the applications of ML for the differentiation of geologic units in deposit scale and prediction of potential drilling targets in greenfield mineral exploration, we examine different unsupervised ML algorithms. We focus on unsupervised learning, specifically clustering algorithms, because of the potential in greenfield exploration and areas under cover, where minimal priori information is available.

We evaluate the performance of connectivity, density, distribution, centroid, and correlation-based clustering methods on differentiating geologic units using three physical property
models derived from geophysical inversions of data sets from a synthetic geologic model. The synthetic model was inspired by an iron-oxide copper gold (IOCG) deposit. We show that for the type of geophysical models commonly used when little a priori information is available (e.g., areas under cover where outcrops and drillholes are not present), which are often minimally constrained smooth models, the ML algorithm with best performance is correlation-based clustering, which could identify the three geologic units, including the copper ore with minimal influence of inversion artifacts.

In the following, we first build a 3D synthetic geological model and compute magnetic, gravity gradient, and DC resistivity data. We next invert the synthetic geophysical data to build models of susceptibility, density, and conductivity from minimally constrained inversions. Then we give an overview of the theoretical basis of the clustering algorithms evaluated for geology differentiation, and discuss the best result for the geologic model used in the study.

3.2 Synthetic model

The synthetic geologic model constructed for this study is based on Cristalino iron oxide copper gold (IOCG) deposit, in northern Brazil (Figure 3.1). The copper deposit is hosted by iron formation interbedded with volcanic rocks. The deposit formed by hydrothermal fluids which were conducted by the fault that cuts the whole sequence and reacted with the magnetite of the iron formation. The process consumes magnetite and converts 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.

Based on the characteristics of the main geologic units present in Cristalino: i) copper ore, ii) iron formation, and iii) mafic volcanic host rock, we build a synthetic model (Figure 3.2) with similar geometry and physical property values equivalent to the same type of rocks in Cristalino (Table 3.1). The conductivity and density values were based on Telford et al. (1990), the conductivity of the copper ore was not taken from the chalcopyrite values, but from the values for copper ore (page 289). The susceptibility values were based on Clark &
Emerson (1991) because they have measurements for iron formation specifically. Although Cristalino sequence dips 50° to west, our synthetic models have vertical bodies for simplicity (Figure 3.2(c)). For the conductivity model, few shallow conductive cells were added to represent near-surface heterogeneity and evaluate the effect on near surface geologic noise on the geology differentiation.

![Geological map of Cristalino copper deposit](image)

Figure 3.1: Geological map of Cristalino copper deposit (adapted from Vale S.A., 2004).

The maximum thickness of the iron formation layer is 175 m and it gradually becomes thinner in the center, where the thickness is 50 m. The copper ore body has the opposite
behavior, it is thinner in the northern and southern borders (from 25 to 75 m) and has its maximum thickness in the center (175 m), where the iron formation has its minimum thickness. Both geological bodies have maximum depth extension of 500 m.

Table 3.1: Physical property values for the synthetic geologic model. The conductivity and density values were based on Telford et al., 1990, and the susceptibility values on Clark & Emerson, 1991.

<table>
<thead>
<tr>
<th></th>
<th>Susceptibility (SI)</th>
<th>Density (g/cc)</th>
<th>Conductivity (10⁻⁴ S/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper ore</td>
<td>0.08</td>
<td>3.2</td>
<td>20000.0</td>
</tr>
<tr>
<td>Iron formation</td>
<td>0.2</td>
<td>3.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Mafic volcanic</td>
<td>0.06</td>
<td>2.5</td>
<td>0.02</td>
</tr>
<tr>
<td>Near-surface heterogeneity</td>
<td>0.06</td>
<td>2.5</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 3.2: Synthetic geologic model based on Cristalino deposit showing the geologic units in the a) susceptibility and density models, b) conductivity model, and c) a section at north = 700 m in the conductivity model.

#### 3.2.1 Geophysical data

The synthetic susceptibility, density, and conductivity models (Figure 3.2) were used for forward modeling of magnetic, gravity gradient, and DC resistivity data, respectively. The magnetic and gravity gradient data are co-located (Figure 3.3(a)). The data separation is
50 m in the east direction, 75 m in the north direction, and 40 m above ground. The Earth’s magnetic field was considered the same as the field in the low-latitude region with field 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 of 0 m elevation with line spacing of 100 m in the north direction and station spacing of 25 m in the east direction (Figure 3.3(b)). The survey configuration is dipole-dipole array with 50 m of electrode separation and 8 n-spacings.

The forward magnetic modeling considers a small magnetic susceptibility ($\kappa << 1$), as is the common case with materials in mineral exploration. Therefore, to the first order approximation, the magnetization $\vec{J}$ is proportional to the susceptibility, and is given by the product of the inducing magnetic field $\vec{H}_0$ with the susceptibility:

$$\vec{J} = \kappa \vec{H}_0 \quad \text{(3.1)}$$

This formulation ignores the self-demagnetization effect and the presence of remanent magnetization. The anomalous field produced by the distribution of magnetization $\vec{J}$ is given by
the following equation that contains a dyadic Greens function:

\[ \vec{B}_a(\vec{r}) = \frac{\mu_0}{4\pi} \int_{\Delta V} \nabla \nabla^T \frac{1}{|\vec{r} - \vec{r}_0|} \cdot \vec{J}(\vec{r}_0) dv, \]  

(3.2)

where \( \vec{r} \) is the observation point location, \( \vec{r}_0 \) is the source location, and \( V \) represents the volume of magnetization. It was added uncorrelated Gaussian noise with a standard deviation of 1% of the datum magnitude plus 1 nT to the anomalous field data (Figure 3.4).

Figure 3.4: The total field anomaly produced by the model in Figure 3.2(a). The inducing field has direction \( I = 0^\circ \) and \( D = 0^\circ \) and a strength of 25000 nT. Uncorrelated Gaussian noise with standard deviation of 1% of the datum magnitude plus 1 nT was added to the data.
The gravity gradient forward modeling assumes that we have a distributed density contrast \( \rho(\vec{r}_s) \) inside a volume \( V \) as the source of our measurements on the surface. The gravity gradient is given by:

\[
T(\vec{r}) = \gamma \int_{\Delta V} \rho(\vec{r}_s) \nabla \nabla^T \frac{1}{|\vec{r} - \vec{r}_s|} d\vec{r}',
\]

where \( \gamma \) is the gravitational constant. The gradient tensor is symmetric and has a zero trace by the potential field theory. For this reason, there are five independent components in the tensor at each observation location and the gradient tensor \( T(\vec{r}) \) can be written as:

\[
T(\vec{r}) = \begin{pmatrix}
T_{xx} & T_{xy} & T_{xz} \\
T_{yx} & T_{yy} & T_{yz} \\
T_{zx} & T_{zy} & T_{zz}
\end{pmatrix}.
\]

(3.4)

For each component, it was added uncorrelated Gaussian noise with a standard deviation of 1\% of the datum magnitude plus 0.5\% of the minimum of the component (Figure 3.5).

The DC resistivity forward modeling operator is defined by the equation:

\[
\nabla \cdot (\sigma \nabla \phi) = -I \delta(\vec{r} - \vec{r}_s),
\]

(3.5)

with appropriate boundary conditions, where \( \sigma \) is the electrical conductivity (Figure 3.2(b)), \( \phi \) is the measured potential in the absence of induced polarization, and \( I \) is the input current. The solution is obtained by solving 3.5 using a finite volume method. It was added to the data uncorrelated Gaussian noise with a standard deviation of 1\% of the datum magnitude plus 0.5\% of the minimum datum with maximum electrode separation (Figure 3.6).

The forwarded simulated magnetic, gravity gradient, and DC resistivity data with added noise was used for constructing physical property models through the inversion methods explained in the following section.

### 3.2.2 Inversion method

The data corresponding to each geophysical method were independently inverted to build susceptibility, density, and conductivity models. No specific prior geologic information was used to constrain the inversions because our goal is to simulate exploration in greenfield
Figure 3.5: The simulated gravity gradient data observed above the model shown in Figure 3.2(a). Uncorrelated Gaussian noise with standard deviation of 1% of the datum magnitude plus 0.5% of the minimum of the component has been added to the data.
Figure 3.6: a) Synthetic geologic model overlain by the DC resistivity data locations and b) apparent resistivity pseudo-sections for north equals to 425, 725, 825, and 1025 m. Uncorrelated Gaussian noise with a standard deviation of 1% of the datum magnitude plus 0.5% of the minimum datum with maximum electrode separation was added to the data.

areas. We used the 3D potential field inversion algorithm developed by Li & Oldenburg (1996, 2003) to invert the magnetic data, Li (2001) to invert the gravity gradient data, and Li & Oldenburg (2000a) to invert the DC resistivity data. All three consider the following relationship:

\[ \mathbf{d} = F [\mathbf{m}], \]  

(3.6)

where \( \mathbf{d} = [d_1, d_1, ..., d_N]^T \) is the data vector (magnetic, gravity gradient, or DC), \( F \) is the forward operator, and \( \mathbf{m} = [m_1, m_1, ..., m_N]^T \) is the model vector containing the physical property (susceptibility, density, or logarithmic conductivity) of each cell in the model, being \( N \) the number of data and \( M \) the number of model cells. For gravity and magnetic the
forward operator $F$ becomes a simple linear system:

$$d = G \mathbf{m}$$  \hspace{1cm} (3.7)

where $G$ is the $N \times M$ sensitivity matrix, which contains the physical relationship between each cell in the model with respect to each datum.

The inverse solution is obtained by solving the following constrained minimization problem using Tikhonov regularization:

$$\text{min. } \phi = \phi_d + \beta \phi_m, \quad \text{subject to } b_l \leq \mathbf{m} \leq b_u,$$  \hspace{1cm} (3.8)

where $\phi$ is the objective function, $\phi_d$ is the data misfit function, $\phi_m$ is the model objective function, $\beta$ is the regularization parameter, and $b_l$ and $b_u$ are the lower and upper bounds, respectively, of the model values. For the magnetic inversion, we used $b_l = 0$ and a large $b_u$ to simulate a situation with no upper bound. For the gravity gradient inversion, we used $b_l = -1.0$ and $b_u = 4.0$ to allow a wide range of possible density contrast variations. The DC inversion was done using the logarithm of the conductivity values, which naturally takes care of the bound constraints. A zero-reference model was used for the magnetic and gravity gradient inversions and a best-fitting half-space as the reference model for the DC resistivity inversion (Li & Oldenburg, 2000a).

The standard deviations ($\epsilon$) of the Gaussian noise added to the data were used in the inversion process. Therefore, the models were obtained by using the discrepancy principle (Parker, 1994), where the target misfit equals the number of data ($\phi_d = N$):

$$\phi_d = \sum_{i=1}^{N} \left( \frac{d_i^{\text{obs}} - d_i^{\text{pre}}}{\epsilon_i} \right)^2 = N,$$  \hspace{1cm} (3.9)

where $d_i^{\text{obs}}$ is the observed data and $d_i^{\text{pre}}$ is the predicted data.

The study area comprises 1.0 km in the east-west direction and 1.5 km in the north-south direction. The data of each geophysical method were inverted using the same mesh to ensure the spatial compatibility among the models. The mesh is composed of cubic cells of 25 x 25 x 25 m, and padding cells were used in the north, south, west, and east directions as well.
as at depth. The padding cells of the mesh extended 2 km beyond the 1.0 by 1.5 km of the study area in all directions, increasing the size gradually from 50 to 500 m.

The recovered susceptibility model (Figure 3.7(a)) shows two magnetic bodies which are associated with the two segments of the iron formation unit. The recovered density model (Figure 3.7(b)) also shows two main anomalies that are coincident 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 3.7(c)) 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. The inversion results show that the iron formation has anomalies in two physical property models, susceptibility and density, while the copper ore has density and conductivity anomalies.

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. Given this common scenario for explorationists and the need of constructing integrated models for geology differentiation, we will evaluate the performance of different clustering algorithms in the task of identifying groups based on the inverted physical properties. The correspondence between the identified groups and geologic units will allow geology differentiation in an quantitative integrated way.

3.3 Geology differentiation using clustering algorithms

The crossplot of physical properties is a common tool for geology differentiation from geophysical models. This tool is frequently used with two physical properties for direct identification of the groups corresponding to different geological units (e.g. Bosch, 1999; Bosch et al., 2002; Bauer et al., 2003; Bedrosian et al., 2007; Kowalczyk et al., 2010; Martinez
Figure 3.7: a) Inverted 3D susceptibility model showing the magnetic anomalies associated with the iron formation, b) inverted 3D density model showing the high density anomalies associated mainly with iron formation and the intermediate density contrast anomaly associated with the copper ore, and c) inverted 3D conductivity model showing the high conductivity anomaly in the center of the model, which is associated with copper ore.

et al., 2011; Martinez & Li, 2015; Melo et al., 2017). However, the use of more than two geophysical models is increasing in mineral exploration because the association of multiple
physical properties reduces the risk of using models are not well constrained for selecting drilling targets. The simultaneous interpretation of multiple models increases the complexity of the process of, for example, finding patterns in the 3D crossplot of the inverted physical property models in Figure 3.8. Therefore, automated methods are needed for the segmentation of the crossplots into groups that correspond to geologic units. For this reason, 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. For classification purposes, a linear transformation may be applied and the physical properties scaled to vary in a range from 0 to 1, where 0 is the smallest value of the model and 1 is the largest value. This procedure avoid the influence of different scales of the physical properties in the measures of distance in the clustering algorithms. The result of the segmentation through the application of unsupervised ML to the models is expected to produce groups similar to the original geologic units, resulting in the expected segmentation in Figure 3.9.

Figure 3.8: Crossplot of the normalized values of density, log susceptibility and log conductivity of the inverted models.
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 the data, or segmentation of crossplot, that corresponds to the units present in the geologic model. Therefore, we will discuss the most widely used clustering algorithms, as there are possibly tens of published algorithms. We evaluated the performance of connectivity, density, distribution, centroid, and correlation-based clustering on differentiating geologic units using the susceptibility, density, and conductivity models derived from geophysical inversions.

Figure 3.9: Synthetic geologic model and crossplots of the normalized values of density, log susceptibility and log conductivity of the inverted models classified by the corresponding geologic unit.
3.3.1 Connectivity-based clustering

Connectivity-based clustering is also known as hierarchical clustering, and is based on the main idea that objects are more related to nearby objects than to objects farther away. In our study, the objects are the cells of the inverted models which have recovered susceptibility, density, and conductivity values associated. Therefore, the groups, or clusters, are formed based on the distance between these attributes in each cell. The method builds a hierarchy of clusters based on the distance between objects, and presents their linkage in a dendrogram (i.e., tree diagram).

In our study we use the algorithm CLINK (complete-linkage clustering) (Defays, 1977). CLINK uses an agglomerative process where each object starts as one cluster and pairs of clusters are merged as they move up the hierarchy. At each iteration, the two clusters separated by the shortest distance are merged. The distance function can vary depending on the objective of the partitioning, and in our study we used the Euclidean distance. The result of applying CLINK to the physical property models of this study shows that the method was not able of separating clusters, but instead gradually joined all the objects in one single cluster, as shown by the dendrogram in Figure 3.10. All the cells were joined into one big cluster because there is not enough separation between the attributes of the geologic units. The lack of separation happens due to the smooth variations in the physical property models. Therefore, the method is not appropriate for the segmentation of the results of poorly constrained inversions.

3.3.2 Density-based clustering

Density-based clustering defines clusters as connected regions of high density of objects. The method measures the “reachability” between objects based on a density criterion defined by a minimum number of objects within a radius Kriegel et al. (2011). Therefore, the clusters can have arbitrary shapes. We first evaluated the algorithm OPTICS (Ordering points to identify the clustering structure) (Ankerst et al., 1999), which defines the radius of linkage
between objects based on the distance among them. The minimum number of points to form a cluster of 1,000 was used to avoid overfitting the physical property data that has 29,009 objects (cells). Similarly to the connectivity-based clustering, the result shows only one cluster because of the lack of separation between groups.

The second test used the algorithm DBSCAN (Density-based spatial clustering of applications with noise) (Ester, 1996), where the minimum number of objects and radius are defined a priori. We defined 1,000 as the minimum number of objects to form a cluster, and after several tests, the radius of “reachability” as 0.08. Where each normalized physical property varies from 0 to 1 and the radius corresponds to the minimum distance to form a cluster. When the radius is smaller than 0.08 many small clusters are formed, and if it is larger than 0.08 only one big cluster is formed. The result of DBSCAN clustering (Figure 3.11) shows two groups, cluster 1 is mostly associated with the iron formation and copper ore in the shallow layer and with the copper ore in the deeper layers. Cluster 2 is
mostly associated with the mafic volcanic host rock, but in the deeper parts it also includes the iron formation. In fact, DBSCAN identifies cluster 1 as noise, meaning that the objects belonging to this cluster have distances larger than 0.08. The spatial distribution of the cells (Figure 3.11) shows that cluster 1 groups the largest physical property values, associated with the anomalies, and the smallest values, which are associated with inversion artifacts from smoothness. Therefore, density-based clustering algorithms are not appropriate for the segmentation of data from smooth geophysical models due to the lack of separation between the physical property values of different geologic units.

Figure 3.11: Density-based clustering model and crossplots of the normalized values of density, log susceptibility and log conductivity of the inverted models classified by the corresponding cluster.
3.3.3 Distribution-based clustering

In distribution-based clustering, clusters are defined as objects belonging to the same statistical distribution. The expectation-maximization (EM) (Dempster & Rubin, 1977) algorithm iteratively estimates the maximum likelihood of parameters of statistical models based on the distribution of the objects. One prominent method is the expectation-maximization using Gaussian mixture models, where the objects are modeled with a fixed number of Gaussian distributions (number of cluster), which are randomly initialized and iteratively optimized to better fit the objects.

The result of EM clustering (Figure 3.12) shows that the method was able to identify one cluster spatially associated with the copper ore (cluster 1), one corresponding to the iron formation (cluster 2), and another with the mafic host rock (cluster 3). The distribution of cluster 2 and 3, in the model domain, approximate a Gaussian distribution and the clustering result represents well the iron formation and host rock. On the other hand, the distribution of the copper ore cluster is not Gaussian, but bimodal. Therefore, the Gaussian model fitted to form cluster 1 includes many cells with physical property values associated with inversion artifacts. Consequently, the copper ore is mapped together with many cells that should have been mapped as iron formation and volcanic host rock. Although the result look reasonable, despite the amount of cells misclassified as ore, there may be no concisely defined statistical model for field data and assuming Gaussian distributions is a strong assumption.

3.3.4 Centroid-based clustering

The centroid-based clustering methods assumes as similarity criteria the distance between the points and the cluster center (e.g. mean, median, meoid). The clustering can be hard, when each object belongs to a cluster or not (e.g. k-means) (MacQueen, 1967), or soft, which calculates the likelihood of each object to belong to each cluster to a certain degree (e.g., fuzzy c-means) (Dunn, 1973).
Figure 3.12: Expectation-maximization clustering model using Gaussian mixture models (dotted gray lines), and crossplots of the normalized values of density, log susceptibility and log conductivity of the inverted models classified by the corresponding cluster.

K-means clustering partitions the objects into \( k \) clusters in which each object belongs to the cluster with the nearest mean, this results in a partitioning of the data space into Voronoi cells and favors spherical clusters. For a fixed number of clusters, it is an optimization problem with the objective of finding \( k \) cluster centers and assign the objects to the nearest group, such that the distances from the cluster centers are minimized. K-means clustering minimizes the objective function (MacQueen, 1967):

\[
E = \sum_{i=1}^{k} \sum_{p \in C_i} \| p - \mu_i \|^2
\]  

(3.10)

where \( k \) is the number of clusters, \( p \) is a point in a cluster \( C_i \), and \( \mu_i \) is the mean of cluster \( C_i \).
The result of k-means clustering of our inversion models (Figure 3.13) shows that cluster 1 corresponds mostly to the inversion artifacts, but can also map the core of the copper ore unit. Cluster 2 identifies well the iron formation, but also unites the copper ore unit into this group. Cluster 3 is mostly associated with the mafic host rock.

K-means clustering is one of the simplest unsupervised learning algorithms and a good initial approach to explore most obvious structures in the data. However, it favors spherical clusters and the groups in the physical properties crossplot are not spherical.

Figure 3.13: K-means clustering model showing the segmentation Voroni-cells (dotted gray lines), and crossplots of the normalized values of density, log susceptibility and log conductivity of the inverted models classified by the corresponding cluster.

3.3.5 Correlation-based clustering

Correlation-based clustering was developed for clustering high-dimensional data. When multiple dimensions are used: i) the clusters are difficult to visualize, ii) the concept of
distance becomes less precise, iii) some attributes are more relevant for some clusters than others, and iv) some attributes are likely to be correlated in arbitrarily oriented subspaces (Kriegel et al., 2009) Given these characteristics data, four main types of algorithms exist for clustering of high-dimensional data. One is subspace-clustering, where the axis of a defined subspace are used to identify clusters, but not all clusters might exist in the subspace. Projected clustering determines clusters for a specific subset of dimensions based on a projected distance, and define the set of relevant attributes for each cluster. In correlation clustering, correlations among the attributes of objects guide the clustering process and the correlations may be different in different clusters. The hybrid clustering approaches use subspaces, projection, and correlation in the clustering process.

In our study we used the algorithm ORCLUS (arbitrarily ORiented projected CLUSter generation) (Aggarwal & Yu, 2000; Schubert et al., 2015) because it is a hybrid approach that identify arbitrary subspaces based on the correlation of the objects. Since its main objective is to look for correlations between attributes of objects, here we refer to it a correlation-based clustering. The algorithm initializes with a large number of initial clusters and uses the small eigenvectors (small variance) of the covariance matrix of the objects within each cluster to find a set of vectors that define the subspace of each cluster. Then evaluates pairs of clusters and decide if two clusters fit into the same pattern of behavior, if so they are merged into a single cluster. The merging decision is a two step process, first it finds the subspace that defines the pair of clusters. Then, it projects the objects into this subspace and compute the distances of these objects to the centroid of the cluster, if the distance is small, the two clusters are merged into one. The algorithm iteratively merges clusters based on their projected distances until the user input number of clusters is reached. The main idea is to transform each group of the data into a new coordinate system in which the second order correlations are minimized.

The model resulting from correlation-based clustering (Figure 3.14) has three clusters identical to the synthetic geological model (Figure 3.9). The segmentation of the physical
properties successfully mapped the copper ore as cluster 1, the iron formation as cluster 2, and the mafic volcanic host rock as cluster 3. The inversion artifacts minimally influence the clustering process and the algorithm was able to find the segmentation guided by the correlation of attributes correspondent to the geologic units. In addition, the high conductivity cells correspondent to the overburden are not classified as ore because they lack association with moderate density anomaly, a characteristic of the copper ore.

Figure 3.14: Correlation-based clustering model and crossplots of the normalized values of density, log susceptibility and log conductivity of the inverted models classified by the corresponding cluster. This result shows the good correspondence between this model and the geologic model (Figure 3.9).

3.3.6 Discussion

Different clustering algorithms use different measures of similarity between objects to define a cluster. Therefore, the choice algorithm should be compatible with the characteristics
of the data being segmented. In our study, we consider situations with little a priori information available to constrain geophysical models. For this reason, we use multiple geophysical methods to reduce the uncertainty in interpretations. However, because of the absence of specific information to constrain the inversions, L2 smooth inversions are applied to construct physical property models. As a result, the models are subject to the presence of inversion artifacts and lack of clear separation between groups of properties associated with specific geology units. These characteristics represent a barrier for clustering algorithms that use the distance of separation between objects to define groups. For instance, connectivity and density-based clustering 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, 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. In this paper, we showed that assuming 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 main point is that the statistics of the physical properties of geologic units in the field is not known unless petrophysical studies have been conducted, which is not the case in greenfield exploration. The application of centroid-based clustering also requires another strong assumption, because its good performance depends on the sphericity clusters in the data. Therefore, the result will not be accurate 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.

### 3.4 Quasi-geology model evaluation

The resulting model from correlation-based clustering is similar to the synthetic geology model (Figure 3.15). This clustering method has the capability of identifying the correlation among susceptibility, density, and conductivity model values correspondent with each unit, the copper ore, iron formation, and volcanic host rock. Therefore, we refer to it as a quasi-geology model, because it represents the subsurface geology but was not built from drilling information.

![Figure 3.15: Synthetic geology model and the quasi-geology model obtained from correlation-clustering.](image)
The confidence matrix (Figure 3.16), 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 are 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 cells and 92% of the volcanic host rock cells were correctly predicted by the quasi-geology model.

The histogram plots of each physical property for all geology units (Figure 3.17) show that the distributions of properties are similar with the quasi-geology model distributions (Figure 3.18). The susceptibility distribution of the copper ore (Figure 3.17) has two groups: one of high to moderate and another of moderate to low susceptibility. Cluster 1 of correlation-based clustering reproduces the same two susceptibility groups (Figure 3.18). Additionally, the other clusters have the same physical property patterns than the corresponding geology units.

Figure 3.16: Confidence matrix showing the comparison between the known geology from the synthetic model and the predicted units from the quasi-geology model obtained through correlation-based clustering.
3.5 Conclusions

We have examined clustering, a class of methods of unsupervised machine learning, to explore the structure and extract information from multiple geophysical inversion models in a quantitative and integrated manner. This study focused on evaluating different clustering algorithms on the geology differentiation of minimally constrained geophysical models, which are commonly used in greenfield exploration when little a priori information is available. We showed that clustering methods with different metrics are influenced in different degrees by smoothness and inversion artifacts in the segmentation of crossplots. With the synthetic example based on a real copper deposit, we demonstrate that correlation-based clustering can find the maximum correlation subspaces of each geological unit and, therefore, performing geology differentiation with a high degree of accuracy. The quasi-geology model built from
correlation-based clustering showed success in finding patterns in a complex geologic setting with minimum influence of near-surface conductivity heterogeneities.

This work shows that the application of unsupervised machine learning is feasible on deposit scales for the identification of potential drilling targets. In addition, we move one step further our understanding in how different clustering algorithms explore the structure of the data for the type of models we use in greenfield exploration. Therefore, with this integrated model, geoscientists can go further with interpretation and associate it with a geological setting, and interpret the quasi-geology model. Because correlation-based clustering is designed for high-dimensional data, many more physical properties can be added (e.g., chargeability, conductivity from electromagnetic and magnetotelluric, velocity). The quasi-geology model can also incorporate information from multiple sources besides geophysical models to support targeting drilling areas. The process of extracting information from multiple sources of data with an unbiased, quantitative, and integrated method empowers geoscientists in the decision making.
CHAPTER 4

GEOLOGY DIFFERENTIATION BY APPLYING UNSUPERVISED MACHINE LEARNING TO MULTIPLE GEOPHYSICAL INVERSIONS: CASE STUDY

Geophysical inversions in mineral exploration have become fundamental to the understanding of physical property distribution in the 3D subsurface and as a means of quantitative geophysical interpretation. 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 increase the reliability of interpretation and reduce drilling risks, methods are needed to connect geophysical inversions to subsurface geology. We developed a method for applying unsupervised machine learning (clustering analyses) to identify geological units from multiple inverted physical property models. Independent inversions of magnetic, gravity gradient, and DC resistivity data in an iron oxide copper gold (IOCG) deposit were performed to construct 3D models of susceptibility, density, and conductivity, respectively. We then applied unsupervised clustering analyses to the inverted physical properties to identify associations corresponding to geological units. We choose to use correlation-based clustering, which employs a robust measure of similarity for the model value distributions. This method focuses on finding similarities or correlations, between points in the parameter space of inverted physical properties using the variance of clusters. To identify the optimum number of clusters, we first use the L-curve criterion from inverse theory and k-means clustering. The clusters derived through this process are then used to map cells in the inversion models into different geologic units to built 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. Given that 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.

4.1 Introduction

The increasing use of geophysical inversions in mineral exploration has resulted in great success in selecting and understanding prospects for drilling. These prospects often become associated with new ore discoveries, and the chance of success increases with the amount of geological knowledge available to incorporate in the inversion and interpretation. In brownfield exploration, the a priori geological information from producing mines and known prospects is used to improve geophysical models of surrounding targets. On the other hand, such a priori information is rarely available for greenfield exploration prospects. In such cases, one may have only regional geologic maps that do not contain much information if the area is covered by overburden. Extracting geological meaning from geophysical inversions is challenging in both brownfield and greenfield exploration. However, interpretation becomes more difficult in areas with little auxiliary geological knowledge. Consequently, the geophysical models are not well constrained, which increases the risk when used directly to select drilling targets. To deal with this challenge, multiple geophysical methods and associated models are needed. The simultaneous interpretation of multiple physical property models also increases the complexity of the process and, therefore, automated and objective means are needed. For this reason, we apply machine learning (ML) to identify meaningful relations between physical properties and map regions of different geologic units within the model domain, i.e., to carry out geology differentiation.

We define geology differentiation or characterization as the interpretation of geophysical data and models to identify associations between inferred physical properties and different geologic units. In general definition, geology differentiation can be performed in the data
or model domain. To our knowledge, Garland (1951) was the first to formally characterize different geological units in the data domain. The author calculated the ratio of magnetization magnitude to density from magnetic and gravity data through Poisson's relation. Kanasewich & Agarwal (1970) explore this method using wavenumber-domain properties of gravity and magnetic data. Using a similar approach, Dransfield et al. (1994) calculate the ratio of apparent susceptibility to density and build a pseudo-lithology map for interpretation. Those approaches are limited to results in 2D maps, and the depth dimension is only poorly explored through analysis of different frequency content in the data. More recently, there has been many works on using machine learning approaches to mapping geology using airborne and remote sensing data. To avoid repetition, we will review these works later in this section.

With the routine use of geophysical inversions, interpretations have transitioned from the data domain to the model domain. Correspondingly, geology characterization is better achieved in the model domain. When using multiple geophysical models, the differentiation process consists of two steps. The first step divides or segments the crossplot of physical properties to identify different combinations of physical property value ranges that could correspond to different geological units. The second step maps each spatial region of the model domain into different geologic units based on the inverted physical property values and the segmentation identified in the first step.

There have been many works in this area. These works in the model domain can be grouped into three categories based on the amount of a priori information used in the first step that segments the physical property crossplot: (i) site specific information, (ii) general geologic information, or (iii) no a priori information. Once the differentiation is done, one may further characterize the identified geologic units based on the the information available. For instance, specific lithology or alteration types can be identified based on site specific information to achieve geology characterization, or we can infer what each identified unit might be based on general geologic knowledge and setting.
In the category of methods that require site specific a priori information, Bosch (1999) presents a formulation to invert for lithology type using density and susceptibility models jointly. The method is called lithologic tomography and requires petrophysical, geostatistical, and structure information to constrain the lithology models, for these reasons it is only applied to areas with high geological knowledge. Bauer et al. (2003) define physical property classes based on natural clusters formed by the inverted P velocity and the Poisson’s ratio, and associate these classes with petrophysical data of the study area to differentiate lithologies. Guillen et al. (2004) and Lane & Guillen (2005) extend the scheme developed by Bosch (1999) from 2D to 3D calling the method litho-inversion, and apply it to field data, exploring different initial models. Fullagar et al. (2004) invert for density and susceptibility variations inside pre-defined lithological units, and use this information to refine the lithological contacts. This method requires good initial models and is suitable for refinement of the parts of a geological model with no drilling information adjacent to areas with many drillholes. In one of their two approaches, Martinez & Li (2015) perform an end-member analysis based on a known geologic cross-section to establish a mapping of inverted density and susceptibility to different types of iron formation and then apply the mapping to 3D models to achieve a 3D lithology characterization. Melo et al. (2015, 2017) identify copper ore through the association between geological units, identified in few drillholes, and patterns in the cross-plot of susceptibility versus conductivity model values. The characterization is guided by the theoretical relation between conductivity and susceptibility ranges for magnetite and chalcopyrite.

In the category in which only general a priori physical property information is available, Hanneson (2003) uses empirical relations between physical properties and the percentage of minerals to classify the mineralogical composition of susceptibility and density models. Williams et al. (2004) use these same empirical relations and known geology information to identify alteration zones from inverted density and susceptibility models obtained from constrained separate inversions on a regional scale. Further exploring this method Williams
Dipple (2007) apply a mineralogy unmixing technique through linear programming to estimate mineral abundances and define alteration zones. Martinez et al. (2011) and Martinez & Li (2015) apply specific ranges of density and magnetic susceptibility obtained from literature to the crossplot of physical properties to define classes and then assign lithology types to each class. These methods are highly relevant for geology differentiation in greenfield exploration, since only general a priori information is required for physical property values.

When there is a lack of directly usable prior information linking physical properties to different geologic units, a variety of approaches have been used. Bedrosian et al. (2007) apply a non-linear least-squares fitting to the probability density function of the crossplot of resistivity and velocity model values from magnetotelluric and seismic data so that different classes and, consequentially, lithology types, can be identified. Kowalczyk et al. (2010) directly partition the crossplot of density versus susceptibility, derived from inversions, to define classes of different lithologies, and apply this classification to produce a 3D regional geological model without inferring lithological types. Fraser et al. (2012) apply self-organizing maps (SOM) to physical property models from multiple geophysical inversions to produce a pseudo-geological model. Devriese et al. (2017); Fournier et al. (2017); Kang et al. (2017) use density, susceptibility, conductivity and chargeability models derived only from airborne geophysical data to build a model of two kimberlite pipes, which are referred to as the petrophysical model.

Regardless the amount of a priori information available, crossplots of physical properties are a common starting point used in the interpretation of geophysical models to characterize or differentiate between geological units. Segmenting the crossplots manually is feasible when two physical properties are being used. When more physical properties are added to the interpretation, the increased dimensionality of the crossplots makes interpretation through direct segmentation difficult. To overcome this difficulty, machine learning (ML) methods have long been applied to extract geological information from multiple types of geophysical data (e.g., well log data, seismic attributes, airborne geophysical data).
Machine learning is a large area within artificial intelligence, which is responsible for the learning part involved in developing algorithms that enable computers to take human-like decisions. It involves the development and application of algorithms that can extract information from data without being explicitly programmed, or equivalently, automatic pattern recognition. The learning process can be supervised or unsupervised. In supervised learning a known dataset is used to train the algorithm with labels and build the ability to classify unknown data (e.g., neural networks, support vector machine). In unsupervised learning, there is no known dataset for training. Therefore, the algorithm explores the structure of the data to discover meaningful information (e.g., clustering algorithms, self-organizing maps, generative topographic mapping). In this paper, we focus on unsupervised learning applied to physical property models derived from multiple geophysical data sets because we are interested in a method capable of finding patterns associated with geological units in areas under cover, where outcrops and drillholes are sparse or unavailable.

Machine learning has been applied to the interpretation of geophysical data in petroleum exploration for several decades. For instance, well log data and seismic attributes are analyzed using machine learning approach for facies classification. Commonly used methods include the principal component analysis and hierarchical clustering (e.g., Serra & Abbott, 1982), modal distribution analysis (e.g., Wolf & Pelissier-Combesure, 1982), k-means clustering and discriminant analysis (e.g., Delfiner et al., 1987), neural networks (e.g., Baldwin et al., 1990; Rogers et al., 1992), graph-based clustering (e.g., Ye & Rabiller, 2000), and many others, resulting in a wide range of applications (e.g., Schlanser et al., 2016; Abreu et al., 2016). The identification of seismic facies using seismic attributes and velocity models have relied on the application of neural network (e.g., Meldahl et al., 1999), k-means and fuzzy c-means clustering (e.g., Barnes & Laughlin, 2005), self-organizing maps (e.g., Strecker & Uden, 2002), generative topographic mapping (e.g., Wallet et al., 2009), in different geological settings (e.g., Coléou et al., 2003; Gao, 2007; Roy et al., 2014; Zhao et al., 2015; and Qi et al., 2016).
In geological mapping, ML tools have been applied to airborne geophysical data, such as radiometric, magnetic, gravity, and electromagnetic, focusing on the construction of 2D pseudo-lithology maps for interpretation. For example, Paasche & Eberle (2009) apply fuzzy c-means clustering to airborne radiometric and magnetic as well as ground-based gravity data to produce a zoned geophysical map that outlines geological units. Eberle & Paasche (2012) apply fuzzy Gustafson-Kessel clustering to satellite imagery, airborne radiometric, and regional geochemical data to construct a pseudo-lithology map. Carneiro et al. (2012) apply self-organizing maps to magnetic and radiometric data to extract geophysical signatures associated to lithology types and produce a pseudo-geologic map over an area with gold deposits in the Amazon region. Other works also focus on 2D geological mapping applying ML to multiple geophysical data (e.g., Ranjbar et al., 2001; Martelet et al., 2006). Geological mapping can also be achieved using 3D physical property models. For example, Paasche et al. (2006) apply fuzzy c-means clustering to identify sedimentary units from physical property models and well log data.

In mineral exploration, Barnett & Williams (2006) use known gold deposits to train a neural network and construct a favourability map in regional scale using multiple data sets. Mahmoodi et al. (2014) apply fuzzy c-means to down hole data from a Ni deposit to characterize rock types and mineralization. Caté et al. (2017) apply supervised machine learning algorithms to predict the presence of gold, in drill cores, from geophysical logs acquired in a volcanogenic massive sulfide deposit in Canada. Unfortunately, acquisition of down hole physical property data is not a common procedure in mining. Although, ML algorithms are being extensively applied for reservoir characterization and facies recognition from seismic data, little research has been focusing on applying ML algorithms for characterization of mineral deposits using physical properties recovered from geophysical inversions.

Overall, much of the existing works using ML either focus on large-scale applications such as on regional geology and prospectivity mapping, or on formation scale such as lithofacies classification in drill holes. Only limited work is available of deposit scales, where ultimately
drilling targets must be chosen. As in any form of geology mapping, the result of geology differentiation carries with it inherent uncertainties and it must be assessed or quantified. Aiming to make advances in this field and face the above three challenges, namely, ML-based geology differentiation at deposit scale, prediction of potential drilling targets, and assessment of uncertainty, we apply a unsupervised learning algorithm (correlation-based clustering) at an IOCG deposit to find patterns in three physical property models recovered from geophysical inversions and identify copper ore.

We choose to use an algorithm known as ORCLUS (arbitrarily ORiented projected CLuster generation) by Aggarwal & Yu (2000) to identify meaningful relations among physical properties. ORCLUS is a powerful algorithm originally designed to cluster high dimensional data. Its fundamental principle of looking for correlations in the data in arbitrarily oriented sub-spaces maximizes the influence that specific physical properties have on the separation of each group, considering their proper characteristics. In other words, some points are correlated with respect to a given set of features and others are correlated with respect to different features. This method has certain advantages over other algorithms, such as, self-organizing maps (SOM) and k-means or fuzzy c-means clustering. For example, SOM is a vector quantization technique that implies dimension reduction of the data to output a 2D map, and consequently loss of information. K-means or fuzzy c-means clustering, on the other hand, separate groups based on the distance between the points and the center of the cluster, which implies on favoring spherical cluster. Clusters with different shapes such as linear or elliptical scattering can be identified with generalized clustering, but the shapes much be supplied a priori. However, the relationship between some physical properties obtained from separate inversions can be complex and have different shapes and orientation. In such cases, correlation-based clustering is expected to be more adaptive and yield good segmentation of the parameter space. The result from our study has shown a good correspondence between the integration of the physical property models and the geological model derived from drill holes.
In the following, we first build 3D models of susceptibility, density, and conductivity from minimally constrained inversions of magnetic, gravity gradient, and DC resistivity data, respectively, at Cristalino iron-oxide copper gold (IOCG) deposit in northern Brazil. We then determine the number of clusters, or different patterns, in the 3D crossplot of the above physical properties and construct an integrated geological model. We accomplish this by first applying k-means clustering to the physical property models and determine the optimal number of clusters using an L-curve criterion. Next, we refine the clusters by applying the correlation-based clustering analysis while fixing them at the optimal number. The final clusters from this two step process are used to map the inverted physical property models into a quasi-geology model. As a verification of the approach, we compare the model from our geology differentiation process with that constructed from a detailed drilling program, and show that the two have a high degree of correspondence despite the fact none of the information from the drilling is used in the differentiation process.

4.2 Geological background of study site

Our study focuses on Cristalino deposit, which is a world class IOCG deposit located in the Carajás Mineral Province, which is a highly mineralized metallogenic region in northern Brazil (Figure 4.1). The class of iron oxide copper gold (IOCG) deposits contain economic grades of copper and gold, and are associated with iron oxides such as magnetite or hematite, or both. They are formed by hydrothermal fluids that rise through deep crustal faults as conduits. The hydrothermal alteration happens in stages and is replacive. The sulfide mineralization is a late-stage alteration event and the sulfides (primarily chalcopyrite) generally replace magnetite formed in previous stages. IOCG deposits have a variety of morphologies, from stratabound sheets to irregular stockwork breccia zones associated with veins and veinlets (Hitzman et al., 1992). The structural control and diverse styles of mineralization result in a high degree of complexity in both mineralogy and geometry in the final deposits. Consequently, exploring for new IOCGs is challenging because fixed exploration models are not always appropriate or applicable. For this reason, Cristalino deposit serves
as a good example for our study.

Figure 4.1: (a) Tectonic location of the Carajás Mineral Province at the margin of the Amazon Craton, Brazil (Almeida et al., 1981), (b) Geologic map of the Carajás Mineral Province showing the study area in the red box (modified from DOCEGEO, 1988; and Grainger et al., 2008).

Cristalino deposit contains 482 Mt @ 0.65% Cu and 0.06 g/t Au (NCL Brasil, 2005). It is hosted by a splay of the Carajás Fault, which is a major crustal fault. This splay fault cuts through a volcano-sedimentary sequence composed of iron formation interlayered with mafic and felsic volcanic rocks (Figure 4.2 and Figure 4.3). This sequence is dipping approximately 50° to southwest, parallel to the fault plane that acted as the conduit for the hydrothermal fluids (Figure 4.3), and the whole sequence is intruded by a younger gabbro dyke. The main ore minerals are chalcopyrite and gold. The chalcopyrite occurs in the form of stockwork, stringers, breccias, and dissemination in the host rock (Huhn et al., 1999).
The hydrothermalism in Cristalino is different from other IOCG deposits because it overprints an area with iron formation containing a large volume of magnetite of sedimentary origin. As a consequence, in the first alteration stage the iron in the hydrothermal fluid mainly contributed to the recrystalization of existing magnetite and to the formation of iron silicates in the volcanic rocks (Huhn et al., 1999). The final hydrothermal stage, which involves formation of chalcopyrite, happened through the replacement of magnetite from the
iron formation, and iron silicates from the volcanic rocks (Lobato, 2000). For this reason, the iron formation unit is not continuous and pinches out where the copper ore is thicker (Figure 4.2). The ore zone is subdivided into high and low-grade zones. The high-grade ore is associated with the replacement of magnetite from the iron formation, and has some remaining magnetite in the interface with the iron formation. The low-grade ore is associated with the replacement of iron silicates from the volcanic host rocks and has little magnetite.

The structural control and diverse styles of mineralization of IOCG deposits result in a high degree of complexity (of mineralogy and geometry) in the final deposits. One can imagine how this complexity transfers to geophysical responses. IOCG deposits are likely to occur in areas of magnetic anomalies; although not necessarily coincident with specific anomalies, which will depend on the degree of the replacement of magnetite by chalcopyrite. For this reason, the magnetic method is an important geophysical method for IOGC exploration. Associated with the magnetic method, gravity methods are important for selecting specific targets because of the high density of the association of chalcopyrite and magnetite. Once more specific targets have been selected, they are evaluated for drilling by applying methods for direct detection of chalcopyrite, such as the resistivity method. For this reason, we focused on these three geophysical methods to study Cristalino deposit. At Cristalino,
the nearby iron formation unit has the most anomalous density and susceptibility anomalies while the copper ore have moderate values of both physical properties. These moderate values are sufficient to distinguish the ore unit from the host rock, but its high conductivity is important to differentiate from the iron formation.

4.3 Geophysical data and inversions

The study area comprises 1.1 km in the east-west direction, 1.5 km in the north-south direction, and is centered over the Cristalino copper deposit. The geological characterization scheme presented in this study uses 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. Although there is a lot of geological information available from drilling in this area, this information was not used to constrain the inversions because our goal is to simulate exploration in greenfield areas. The magnetic and DC resistivity data, and corresponding susceptibility and conductivity models, were first presented by Melo et al. (2015, 2017) in a limited study. Subsequently, a set of full-tensor gravity gradient data was inverted to obtain a 3D model of density contrast. The feasibility of integrating the density model with susceptibility and conductivity models was first presented by Melo & Li (2016) using k-means clustering analysis. We present the essentials of the three data sets and corresponding inversions in this section.

The acquisition parameters of all three data sets used in this work are specified in (Table 4.1). For the airborne magnetic data we removed the International Geomagnetic Reference Field (IGRF) and performed a regional-residual separation in the data (Figure 4.4) using an inversion-based method (Li & Oldenburg, 1998). This step was applied due to the strong magnetic gradient in the region. The study area is located in a low latitude near the Equator, with field inclination of -3.5°, declination of -19°, and strength of 25,500 nT. The magnetic data comprises two main magnetic anomalies (Figure 4.4(a)) with overlapping patterns. Melo et al. (2017) report the presence of magnetic remanence in the data. However, the anomaly pattern and the estimation of the direction of remanence (inclination of 0° and declination
of $18^\circ$) demonstrate that the direction of total magnetization is sufficiently close to the direction of the inducing field. Therefore, they perform the magnetic inversion by using the inducing field direction as the magnetization direction. Consequently, the recovered model represents an effective susceptibility that includes the remanence effect and will have higher than expected susceptibility values, but the spatial distribution of the effective susceptibility is valid. For the geology differentiation presented here the effective susceptibility is sufficient.

Table 4.1: Acquisition parameters of the magnetic, gravity gradient, and DC resistivity surveys.

<table>
<thead>
<tr>
<th>Method</th>
<th>Acquisition year</th>
<th>Line spacing</th>
<th>Line direction</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetic</td>
<td>2000</td>
<td>250 m</td>
<td>east-west</td>
<td>draped, average terrain clearance 200 m</td>
</tr>
<tr>
<td>Gravity Gradient</td>
<td>2007</td>
<td>200 m</td>
<td>east-west</td>
<td>draped, average terrain clearance 200 m</td>
</tr>
<tr>
<td>DC</td>
<td>1999</td>
<td>200 and 400 m</td>
<td>east-west</td>
<td>60 m dipole-dipole and 5 m spacings</td>
</tr>
</tbody>
</table>

Similarly to the magnetic data, the gravity gradient data show two main anomalies (Figure 4.5), which are more intuitively shown by the $T_{zz}$ component of the data. Another interesting feature is the anomaly with intermediate amplitude values between the two main anomalies. The DC resistivity lines are approximately perpendicular to the topographic ridge in the area (Figure 4.6), which is coincident with the structure that hosts the mineralization. There are two types of high conductivity anomalies, one that has large volume and comprises shallow to deep levels, and anomalies of small volume near the surface.

The data of each geophysical method were inverted using the same mesh to ensure the spatial compatibility between the models. The mesh is composed of cubic cells of 50 x 50 x 50 m, and padding cells were used in the north, south, west, and east directions as well.
as at depth. The padding cells of the mesh extended 3 km beyond the 1.1 x 1.5 km of the study area in all directions, increasing the size gradually from 50 to 800 m. We used the 3D potential field inversion algorithms developed by Li & Oldenburg (1996, 2003) to invert the magnetic data, Li (2001) to invert the gravity gradient data, and Li & Oldenburg (2000a) to invert the DC resistivity data. All three consider the following relationship:

$$d = F \left[ m \right]$$  \hspace{1cm} (4.1)

where \(d = [d_1, d_1, ..., d_N]^T\) is the data vector (magnetic, gravity gradient, or DC), \(F\) is the forward operator, and \(m = [m_1, m_1, ..., m_M]^T\) is the model vector containing the physical property (susceptibility, density, or logarithmic conductivity) of each cell in the model, being \(N\) the number of data and \(M\) the number of model cells. For gravity and magnetic the forward operator \(F\) becomes a linear system:

$$d = G \ m$$  \hspace{1cm} (4.2)
where $G$ is the $N \times M$ sensitivity matrix, which contains the physical relationship between each cell in the model with respect to each datum.

The inverse solution is obtained by solving the following constrained minimization problem using Tikhonov regularization:

$$
\min \phi = \phi_d + \beta \phi_m, \quad \text{subject to} \quad b_l \leq \mathbf{m} \leq b_u
$$

(4.3)

where $\phi$ is the objective function, $\phi_d$ is the data misfit function, $\phi_m$ is the model objective function, $\beta$ is the regularization parameter, and $b_l$ and $b_u$ are the upper and lower bounds,
Figure 4.6: DC resistivity acquisition lines over the terrain topography.

respectively, of the model values. For the magnetic inversion, we used $b_l = 0$ and a large $b_u$ to simulate a situation with no upper bound. For the gravity gradient inversion, we used $b_l = -0.5$ and $b_u = 2.0$ to allow a wide range of physically possible density contrast variations. The DC inversion was done using the logarithm of the conductivity values, which naturally takes care of the bound constraints. A zero-reference model was used for the magnetic and gravity gradient inversions and a best-fitting half-space as the reference model for the DC resistivity inversion (Li & Oldenburg, 2000a).

The inversion process requires the standard deviation of the noise in the data. However, this value is usually not known for field data. To ensure an adequate estimation of the standard deviation, we used the method described by Melo et al. (2017). The method
consists of first assuming a standard deviation of 1 and running a set of inversions using regularization parameters ($\beta$) over a large range of values. Then, apply the L-curve criterion (Hansen, 1992) to the Tikhonov curve of misfit versus model norm values and select the point of maximum curvature of the curve as the optimum $\beta$. The misfit corresponding to this optimum $\beta$ is then used to estimate the adjusted standard deviation. In the following step, a new Tikhonov curve is built using the adjusted standard deviation as the noise estimation. The procedure of adjusting the standard deviation value is repeated a second time to fine tune the estimation, and the L-curve of the third set of inversions was used to select the final model.

The estimated standard deviation of the final susceptibility model is 20 nT, which corresponds to 0.9% of the magnetic data range. The recovered susceptibility model (Figure 4.7) shows two magnetic bodies dipping approximately 50° to the southwest and associated with the iron formation. The northern body has a larger volume than the one to the south, and this difference is possibly associated with differences in the magnetite and hematite contents in the iron formation, or the presence of hydrothermal magnetite. Additionally, the eastern area of the northern anomaly is partially associated with the copper ore. The large recovered susceptibility values are judged to be due to remanent magnetization in the magnetic rocks. As demonstrated by Melo et al. (2017), the direction of remanent magnetization is similar to the direction of the inducing field, so the total magnetization direction is well approximated by the inducing field direction. Therefore, the inversion using this magnetization direction yields an effective susceptibility model.

For the density inversion, the estimation of the standard deviation of the noise was done for each component and the final values for the components are: $T_{xx} = 1.79$ Eo, $T_{xy} = 1.44$ Eo, $T_{xz} = 2.42$ Eo, $T_{yy} = 2.67$ Eo, $T_{yz} = 3.00$ Eo, and $T_{zz} = 2.96$ Eo. The standard deviation estimations are between 2.11% and 2.9% of the data range of the components. These standard deviation values are reasonable because the error has several different sources, such as aircraft movement, acquisition system, and pre-processing steps (Martinez et al., 2018).
Figure 4.7: Inverted 3D susceptibility model showing the magnetic anomalies associated mainly with the iron formation.

Additionally, the digital elevation model used for the terrain correction was from SRTM (Shuttle Radar Topography Mission) with spatial resolution of 90 m, which is not appropriate to capture and remove the high frequency signal from roughness in the terrain. The recovered density contrast model (Figure 4.8) shows two main anomalies. The high density contrast anomaly in the north is spatially coincident with the magnetic anomaly, and the one in the south is only partially coincident with the other magnetic anomaly. This difference is probably related to a larger concentration of hematite in the southern anomaly. In the area between the two high density anomalies is an intermediate density body that is spatially associated with the copper ore.

For the gravity gradient and magnetic data inversions, an estimated constant value of standard deviation was used for each component including the total-field anomaly as is the common practice. For DC resistivity data, however, the noise characteristics are different. Because the signal level depends on the separation between current and potential electrodes, one constant standard deviation would not be appropriate for data from different electrode separations. The noise level is estimated to be 0.406 mV/A plus 2% of the absolute value of the normalized potential differences. The recovered conductivity model (Figure 4.9) has
Figure 4.8: Inverted 3D density contrast model showing the high density contrast anomalies associated mainly with iron formation and the intermediate density contrast anomaly associated with the high grade ore.

the main anomaly located in the central part of the model, which is spatially coincidental with the know 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. The depth of investigation (DOI) of the DC resistivity data was estimated by altering the reference models in the inversions and identifying the region of similarity between features in the models (Oldenburg & Li, 1999).

The three physical property models show different degrees of smoothness. The susceptibility model shows broad features because the data are from a low latitude region and magnetic declination is nearly parallel to the the dominant structure direction. The conductivity model appears to be diffused especially in the north-south direction, which is due to the wide line spacing. The density model has the sharpest contacts. Thus, there are certain amount of fuzziness in the physical property values related to artifacts of smooth inversion.

For an interpreter who is not familiar with the specific geology of the deposit (which could be the case in greenfield exploration), the interpretation of the susceptibility and density models is mainly focused on the two anomalous bodies that are partially coincident. However, the conductivity model shows a different structure, where the anomalous body is located
between the two susceptibility and density anomalies. In addition, this high conductivity anomaly appears to be associated with a moderate density. The interpretation of inverted physical properties commonly stops here at this qualitative analysis, which does not fully support decision making for drilling. This is the point where geology differentiation take the interpretation one step further by transforming qualitative interpretations of geophysical models into a model that is a direct representation of geology. A natural path for identifying the association of physical properties with different geologic units is to cluster or segment the crossplot of physical properties and map the multiple physical property models into a single model consisting of predicted geological units. We term this model as the quasi-geology model.

The patterns in the crossplot cannot be readily identified in the crossplot when more than two inverted models are involved. The crossplot in Figure 4.10 illustrate this difficulty. Thus, automated methods capable of finding even subtle associations are necessary. We apply unsupervised ML to explore the structure of the data aiming to find associations that allow a more precise geology differentiation.
4.4 Geology differentiation

We now proceed to develop an automated geology differentiation method and demonstrate it using the physical property models at Cristalino deposit described in the preceding section. We remark that the integration and interpretation of multiple physical property models is still largely qualitative in mineral exploration. Consequently, the process is heavily dependent on the experience of the interpreter, which imposes limitations on the process of extracting joint information from the models. Automatic methods can enable us to fully identify and extract the content from multiple sources of physical properties quantitatively.

For our study, we assume that there is minimal amount of geological information available independent of the geophysical data sets and corresponding inverted models. Our focus is on unsupervised machine learning (ML) because it explores the structure of the data to discover meaningful information and has the potential of finding associations of physical properties that reflect geological variations. Although Cristalino deposit has been drilled extensively,
we have chosen this site due to the amount of geologic information available to compare and validate our geology differentiation results. We do not use the geological information in the inversion and differentiation steps.

4.4.1 Clustering analysis of inverted physical properties

Within unsupervised ML field, clustering is the process of identifying patterns by grouping similar objects, which in our case are the model cells with different combinations of inverted physical property values. The definition of similarity is dependent upon the data object and the purpose of the analysis. For this reason, there are many clustering methods with different measures of similarities. For example, the centroid-based clustering methods assumes as similarity criteria the distance between the points and the cluster center (e.g. k-means, k-meoids, fuzzy c-means) (Bezdek, 1981; Dunn, 1973; MacQueen, 1967). Therefore, these methods perform well on spherical clusters. The distribution-based method uses statistical distributions as similarity criteria. For example, in the Gaussian mixture models using expectation maximization algorithm, each cluster is modeled by a normal distribution (Dempster & Rubin, 1977). This type of methods requires known statistical distribution a priori, which requires a strong assumption for field data. Density-based methods define clusters as connected dense regions (e.g. DBSCAN and OPTICS) and can find arbitrary cluster shapes (Ankerst et al., 1999; Ester, 1996). These methods measures the distance and reachability between objects and, consequently, the clusters need to be separated by distance to begin with. As a result, clusters from smooth models, which have a core of high density of points that gradually diminishes and merges with other clusters, may end up as one big cluster because the groups do not have sufficient separations in the crossplots.

In general, the clusters patterns in the independently inverted physical property models are not spherical, nor do they have clear separations, as demonstrated by the crossplot of susceptibility, density, and conductivity (Figure 4.10) at Cristalino deposit. Furthermore, we could not know a prior or assume particular statistical distributions for the clusters. Therefore, among the clustering techniques available, the one that has a similarity measure better
suited to our problem is the correlation-based clustering, because it looks for correlation between attributes and can identify clusters with arbitrary orientations. We chose a technique that could find appropriate combinations of physical properties that represented geological units while respecting which physical property was more diagnostic for each group. This technique explores the data structure by looking for subspaces that best represents each cluster based on the correlation between the points belonging to the cluster. The flexibility in the subspace orientation allows the algorithm to find the combination of physical properties that is more relevant to a given cluster while minimizing the influence of the physical properties that do not help differentiating the cluster. Therefore, it allows the identification of distinct clusters based on the different combinations of physical properties characteristic of each cluster. Each of these combinations with its respective cluster represent a different geological unit present in the area.

The correlation-based clustering algorithm operates by finding the subspace that best represents 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. The algorithm requires the specification of the number of clusters and an initial guess of clustering. We have developed a hybrid approach by which we first use the k-means clustering to estimate an optimal number of clusters for use in the correlation clustering. The result of the initial k-means analysis also provides the initial guess for the final correlation-based clustering. For completeness, we provide a brief description of the two clustering techniques, and develop a method for estimating the optimal cluster number, and then the application of the correlation-based clustering to the Cristalino deposit.

4.4.2 K-means and correlation-based clustering

K-means clustering minimizes the objective function (MacQueen, 1967):

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} \| p - \mu_i \|^2$$

(4.4)
where \( k \) is the number of clusters, \( p \) is a point in a cluster \( C_i \), and \( \mu_i \) is the mean or center of cluster \( C_i \). For classification purposes, a linear transformation may be applied and the physical properties scaled to vary in a range from 0 to 1, where 0 is the smallest value of the model and 1 is the largest value. K-means clustering is one of the simplest unsupervised learning algorithms and a good initial approach to explore most obvious structures in the data. However, the clusters in the physical properties crossplot are not spherical, so we use correlation-based clustering to refine the final clustered model. We have implemented the algorithm ORCLUS (arbitrarily ORiented projected CLUSTER generation) (Aggarwal & Yu, 2000; Schubert et al., 2015) for this study.

The algorithm (Figure 4.11) initializes with a large number of initial points or seeds (around 20 times larger than the final number of clusters) 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 to diagonalize the matrix and find a set of vectors that define the subspace of each cluster. Only the orthonormal eigenvectors with the least spread (small eigenvalues) are chosen because they represent smaller variances, or the parameters with higher degrees of similarity. The algorithm then evaluates pairs of clusters and decide if two clusters fit into the same pattern of behavior. If they do, the pair is merged into a single cluster. The merging decision is a two step process, first it computes the covariance matrix of the union of pair of clusters and its eigenvalue decomposition, then the eigenvectors corresponding to the smallest eigenvalues are used to find the least spread subspace for the points. This subspace represents the directions which the parameters are similar. In the second merging step, the points of the two clusters being evaluated are projected to the subspace defined in the previous step, and the distances of these points to the centroid of the projected cluster are measured. This distance is used to calculate the projected energy, which is the summation of the mean square of the distances. The decision to merge clusters is based on the similarity in the direction of their subspace vectors. The projected energy act as an implicit objective function of the cluster, and if it is
small, the two clusters are merged into one. The algorithm iteratively adjusts the clusters by ignoring the eigenvectors with large variance, and merges clusters based on their projected distances until the user input number of clusters is reached. The essence is to transform each group of the data into a new coordinate system in which the second order correlations are minimized.

Figure 4.11: Flowchart of the correlation-based clustering with ORCLUS (arbitrarily ORiented projected CLUSter generation) algorithm showing the three main parts of the clustering process (based on Aggarwal & Yu (2000)).
4.4.3 Optimal number of clusters

Clustering requires the number of clusters as an input. In our geology differentiation, specifying the number of clusters is equivalent to specifying the number of distinct geology units that can be identified from the inverted physical property models. When we do not assume prior knowledge as is the case with our investigation, this number must be determined from the physical property models. We use a common method for estimating the number of clusters in computer science, known as the elbow method, to achieve this (Raschka, 2016). This method is similar to the L-curve criterion used in the linear inversion theory (Hansen, 1992). For simplicity, we based the estimation on the k-means clustering because it is the simplest and fastest to evaluate.

The method is based on evaluating the variability of the clustering performance as a function of the number of clusters. This variability is associated with the measure of similarity of the clustering method and is commonly represented by some measure of distance between data objects and cluster centers. An objective function that consists of these distances is minimized to identify the cluster centers and the data objects belonging to each cluster. For instance, the objective function $E$ in k-means clustering is defined in equation 4.4. With a too small number of clusters, the objective function will be large. On the other hand, the objective function is zero when the number of clusters is equal to the number of data. The optimal number is chosen to be the value beyond which adding more clusters does not improve modeling of the data or, equivalently, reduce the variability significantly.

As our objective is to find the number of clusters $k$ that minimizes the objective function while keeping meaningful number of clusters, we choose as the optimal number of clusters the point of maximum curvature on the curve of $E$ as a function of $k$. For the physical property models at Cristalino, this optimal value is $k = 4$ (Figure 4.12). The curve shows that adding more than 4 clusters has a small impact in reducing the k-means objective function. Once the optimal $k$ is defined, we test the consistency of the clustering results with different random initialization of cluster centers that resulted in clusters with 98% of similarity.
Figure 4.12: Objective function versus number of clusters, the red circle shows the point of maximum curvature $k = 4$, which is the optimum number of clusters.

The spatial distribution of the k-means clustering with $k = 4$ (Figure 4.13) shows that the clusters have spatial continuity and geologically reasonable patterns. Cluster 1 has moderate to high susceptibility, moderate density, and mostly high and partially moderate conductivity. This combination of physical properties could be related to a zone with sulfide and magnetite. Therefore, this cluster could be interpreted as being potentially associated with the ore unit. Cluster 2 has relatively high susceptibility, high density, and moderate conductivity, which is a combination of physical property values expected for iron formation or ultramafic rocks. Cluster 3 has low to moderate susceptibility, low density, and low to moderate conductivity, where the cells with larger values of conductivity are in the shallow layers. The location of cluster 3 and its range of physical properties show that it is likely associated with host rock. Cluster 4 is the smaller cluster and contains only few model cells. It has low susceptibility, low density and low to moderate conductivity, and captures the smallest values associated with the smoothness in the models, which are judged to related to the artifacts from smooth geophysical inversion and may not carry much information. In data classification, such data objects are referred to as to as irrelevant attributes, which is also a fitting term in our study. This analysis yields the optimal number of clusters and also provides an initial guess.
Figure 4.13: 3D spatial distribution of k-means clustering with $k = 4$ and the scatterplots of physical properties showing the clustering result.

4.4.4 Results at Cristalino Deposit

After defining the optimum number of clusters, we applied the correlation-based clustering, which is less susceptible to irrelevant attributes because of its measure of similarity based on correlation. The spatial distribution of the correlation-based clustering with $k = 4$ (Figure 4.14) shows the same general cluster distribution as k-means clustering, but with 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 with $k = 4$ classifies some cells as belonging to cluster 1 (potentially corresponding to ore) while
correlation-based clustering do not. These cells lack any association with the corresponding geophysical signature of ore, which are high conductivity, intermediate density, and moderate susceptibility. The result is less influenced by isolated anomalies, when only one physical property shows anomaly.

Figure 4.14: 3D spatial distribution of correlation-based clustering with $k = 4$ and the scatterplots of physical properties, showing less influence of irrelevant attributes and an improved spatial association with the geological units.

The histogram plots of all clusters for each physical property (Figure 4.15) summarizes the ranges that show correlation for each cluster. There is overlapping of susceptibility ranges between clusters 1, 2, and 3, while cluster 4 is clearly associated with the lowest values. The overlapping is a consequence of the smooth transition between anomalies in
the inverted models, which is a common characteristic of such physical property models. Although there is a superposition of values, the histogram plot shows that, for susceptibility, cluster 1 is mostly associated with moderate-high values, cluster 2 with highest values, and cluster 4 with moderate-low susceptibilities. The density model has sharper contacts; as a consequence, the clustering shows more defined ranges for clusters 1, 2, and 3, while cluster 4 gets a small subset in the same range as cluster 3. Similar to the susceptibility model, the conductivity model is also smooth. In addition, the conductivity model clearly has two groups of anomalies: shallow ones spread over the mesh and a continuous anomaly that reach greater depths and has a defined geological strike. The shallow anomalies are potentially associated with conductive overburden, while the other anomaly with sulfides. The histogram plot for conductivity shows that cluster 1 has the highest conductivity values, cluster 2 mostly the moderate values, cluster 3 low to moderate values, and cluster 4 gets a small subset in the same range as cluster 3.

![Histogram plots for each physical property showing the range of values for all classes of correlation-based clustering with \( k = 4 \).]

The proposed method for selecting the optimum number of clusters proved to give an estimation of \( k \) that is highly geologically feasible, but this number should be taken as one possible estimation, which was chosen to avoid under- or over-fitting the physical property values. We need to explore the correlation-based clustering using \( k \) values that vary from the estimated one. Through such explorations, we may identify clusters that are correlated to geological features, but were not identified by k-means clustering due to simplistic similarity
measure. The result of applying correlation-based clustering with \( k = 5 \) (Figure 4.16) shows that cluster 1 becomes subdivided into two. The histogram plots for five clusters (Figure 4.17) is similar for that for four clusters. Only cluster 1 is altered, while the other clusters remain the same. The histograms of susceptibility and density show that the ranges stay the same, while for conductivity two different groups are created. Clustering with \( k=5 \) found a new correlation where the governing attribute is conductivity and the new cluster 1 is associated with the most anomalous conductivity values.

![Figure 4.16: 3D spatial distribution of correlation-based clustering with \( k = 5 \) and the scatterplots of physical properties, showing the subdivision of cluster 1.](image)

The result of correlation-based clustering demonstrates that we can take one step further from conventional qualitative interpretation of inverted physical property models by apply-
Figure 4.17: Histogram plots for each physical property showing the range of values for all classes of correlation-based clustering with $k=5$.

ing clustering to the segmentation of the crossplot of physical properties. Now we have a 3D integrated model with units that were quantitatively defined by their intrinsic associations of physical properties. This quasi-geology model can be interpreted as belonging to a geological setting where there is a unit that hosts two anomalous units, one potentially associated with sulfides and another with magnetite. The high susceptibility and density of the second unit raises two strong possibilities: iron formation of ultramafic rocks. The quasi-geology model allows the interpreter to evaluate the units within the exploration geological setting and decide where to drill objectively. Figure 4.18 summarizes the process. We first construct physical property models by inverting geophysical data sets. Next we perform the geology differentiation by clustering or segmenting the crossplot of physical properties using unsupervised machine learning. The final step maps the physical property models into different geologic units based on the clustering to build a quasi-geology model.

4.5 Evaluation of the quasi-geology model

We now proceed to assess the reliability and, thereby, the value of information in the quasi-geology model through geology differentiation by comparison with the 3D geological model of Cristalino built from 303 drillholes (Vale S.A., 2004). This detailed geological model captures variations in the geology within meters. However, the geophysical data does not have the resolution to image such small variations. We down-sampled the 3D geological
model to the same mesh used in the geophysical inversions. In the geological model, each unit was represented by a closed volume. These volumes were discretized and interpolated to 50-m cubic cells used in the inversion meshes. The result is a simplified geology that represents the geometry and volume of the main units (Figure 4.19). In order to simplify the comparison between drilling derived geological model and quasi-geology model through clustering, the known geological units were grouped into three main categories: i) ore unit, which comprises the high grade ore, ii) the iron formation unit, which comprises iron formation and the gabbro dike, and iii) the host rock unit, composed of the mafic volcanic, felsic volcanic, low grade ore, and other host rocks. The low grade ore appears as host rock because its signature in the geophysical models is similar to that of the host rock signature and could not be distinguished. The histogram plots of all geology units for each physical property (Figure 4.20) shows that the overall distributions are similar to the clustering histograms (Figure 4.15 and Figure 4.17). However, the main difference is that the quasi-geology model has fewer cells as host rock than there are in the geological model. This difference happens because the hydrothermal copper deposit, which occur mainly as stockwork veins (Melo et al., 2017), is hosted by a volcano-sedimentary sequence composed of iron formation interlayered with mafic and felsic volcanic rocks. The geophysical data and, consequently,
the inverted models, do not have the resolution to differentiate the interlayering of iron formation with the volcanic rocks and mineralized veins. Therefore, the region composed mostly by iron formation, but with intercalations of volcanic rocks, will be highly magnetic in the susceptibility model and classified as one cluster. Thus, the differences between the geological model and the quasi-geology model are due to limitations in the geophysical data resolution and inversions with no site specific constraints.

Figure 4.19: 3D geological model of Cristalino copper deposit constructed from 303 drillholes (adapted from Vale S.A., 2004).

A visual comparison of spatial patterns in the geological model and the quasi-geology model with \( k=4 \) (Figure 4.21) shows the following correspondence: i) high grade ore and cluster 1, ii) iron formation and cluster 2, and iii) host rocks and cluster 3. The irrelevant attributes associated with cluster 4 has no correspondence in the geological model.

To quantify the comparison, we compute the confusion matrix shown in Figure 4.22, which describes the performance of a classification model by showing the proportion of known cells that were correctly predicted, and the proportion predicted as other classes.
Figure 4.20: Histogram plots for each physical property showing the range of values for the simplified geological units in the 3D geological model.

Figure 4.21: 3D plot of the result of correlation clustering with $k = 4$ compared to the high grade ore from the geological model.

There is a 62% to 64% match between the three known geology units from drilling and those in the quasi-geology model. 62% of the known ore was classified as such, also 62% of the host rock was classified correctly (cluster 3), and 64% of the iron formation was also correctly predicted (cluster 2). The main spatial differences are that 26% of the known host rock was predicted as ore because of the interlayering between these units, the host rock was
misclassified. 28% of the known ore region was predicted as iron formation because parts of the ore rich in magnetite are highly magnetic, therefore, were classified as iron formation. 18% of the known iron formation was predicted as host rock probably because of the small iron formation bodies that occur in the west of the geological model but does not have significant expression in the magnetic and gravity gradiometry data. In addition, 14% of the known iron formation was classified as ore, probably because of the presence of chalcopyrite veins that increase the conductivity.

![Confusion matrix](image)

Figure 4.22: Confusion matrix showing the comparison between the known geology from the 3D geological model and the predicted units from correlation-based clustering with \( k = 4 \).

When \( k = 5 \) in correlation-based clustering, cluster 1 becomes smaller in spatial extent and has a higher spatial correspondence with the high-grade ore (Figure 4.23). This cluster is mainly associated with large conductivity values, showing that the method was able to identify a cutoff value that has higher visual association with the copper ore conductivity. The confusion matrix (Figure 4.24) that compares the known geology with the quasi-geology model with \( k = 5 \) shows a lower match than the results with \( k = 4 \). 48% of the ore is predicted
as ore, and 57% of both, host rock and iron formation, are correctly predicted. The main
difference is that 14% of the ore is classified as this new cluster (cluster 5 or host rock 2).
This percentage difference is the primary change between the two quasi-geology models. Five
clusters captures transitional physical property values in the inversion, and one third of it
corresponds to known host rock, one third to ore, and one third to iron formation. Although
the five-cluster model has a lower performance compared with the four-cluster model, the
results for the three units of interest are not drastically different. The consistency in the of
identified ore and iron formation in the two models gives us confidence that these regions
identified for these two units are credible and does not critically depend on the choice of the
cluster number.

Figure 4.23: 3D plot of the result of correlation clustering with \( k = 5 \) compared to the high
grade ore from the geological model.
4.6 Conclusion

In this work, we propose an objective geology differentiation method that supports integrated interpretation of multiple geophysical inversions in greenfield exploration. We use independent susceptibility, density, and conductivity models over the IOCG Cristalino copper deposit to identify different geologic units through unsupervised machine learning. We apply correlation-based clustering, which takes advantage of the correlation between physical properties for different geologic units. We show that the iron formation has the most anomalous values of susceptibility and density, and moderate values of conductivity. The high grade ore has susceptibility varying from moderate to high, moderate density, and moderate-high conductivity values. Thus, the combination of these physical property ranges is necessary to identify the ore unit, which is our main target. Otherwise, it would not stand out from the anomalies associated with the iron formation. Therefore, the proposed method
successfully identified the physical property ranges of each geologic unit and differentiates between them. We show that k-means clustering is a good initial approach to explore the most obvious structures in the data. We demonstrate how this simple clustering method supports the identification of the optimal number of clusters through the application of the L-curve criterion to its objective function for different numbers of clusters. However, the results are heavily influenced by irrelevant attributes, and we show that correlation-based clustering (ORCLUS algorithm) identified the copper ore, iron formation, and host rock with a high degree of similarity with the geologic model built from drill hole logging data. The main challenge is that the geophysical inversion cannot capture the small variations of the interlayering between the host rock with iron formation cut by ore veins. Therefore, the quasi-geology model yield a good result given the geophysical models available. Sharper geophysical models can potentially increase the clustering performance (e.g. joint inversion or cooperative inversion models), because correlation-based clustering will find more accurate physical property associations, consequently, more accurate geological differentiation. This method is entirely driven by the data (and models), and proved to work in a complex geological setting such as Cristalino. The work flow presented here can be applied to any type of target in greenfield and brownfield exploration to increase the understanding and confidence in the drilling planning stages using a quasi-geology model.

4.7 Acknowledgments

A. T. Melo acknowledges the CNPq (Brazilian National Council for Scientific and Technological Development) for the scholarship. We thank Vale S.A. for providing the data used in this study and for permission to publish results obtained from them. This work is supported in part by the Gravity and Magnetics Research Consortium. We also thank all CGEM members, especially Jiajia Sun for discussions.
CHAPTER 5
MAGNETIC ON-TIME TRANSIENT ELECTROMAGNETIC (MoTEM) METHOD FOR IMAGING MAGNETIC SUSCEPTIBILITY

We present the magnetic on-time transient electromagnetic (MoTEM) method to image subsurface magnetic susceptibility. The method uses the on-time part of the waveform produced by the transmitter of transient electromagnetic (TEM) systems as the inducing magnetic field. Therefore, it is similar to an active magnetic method at near-zero frequency because the inducing field is assumed to be static. The method also assumes that the transmitter waveform is long enough to allow the dissipation of eddy currents until the electric conductivity effect becomes negligible. Then, the on-time data is inverted using the same framework of 3D inversion of geomagnetic data. We first develop an algorithm to compute the sensitivity matrix using the field of the current loop as the inducing field in place of the Earth’s main magnetic field. Then, we modify the weighting function that counteracts the natural decay of the kernel, which is the inverse of the distance raised to the power of six by testing both depth and sensitivity weighting. Inversion of synthetic on-time electromagnetic data using the MoTEM method successfully recovers the susceptibility of an anomalous body in an non-magnetic background. Therefore, we have demonstrated the feasibility of the method and its potential for improving susceptibility models.

5.1 Introduction

Magnetic properties of the subsurface are important for detecting geological structures associated with mineralization, and for direct targeting when the ore has magnetic minerals. The magnetic susceptibility of the subsurface is most commonly obtained from inversion of data obtained from measurements of the anomalous field induced by the Earth’s geomagnetic field. Magnetic surveys measure the total field, which includes the effect of both induced
and remanent magnetization. The geomagnetic field is small in low latitude regions, consequently, the measured subsurface response is weak. When the magnetic body (e.g. an iron formation layer) has a similar geologic strike with the direction of the Earth’s magnetic field, there is a poor coupling with the lateral edges of the magnetic bed (Figure 5.1(a) and (b)). For this reason, the resulting inverted susceptibility model has a poor recovery of the shallow part (Figure 5.1(c)). This is a common problem in regions near the Equator, where the weak magnetic anomaly is a problem for constructing accurate susceptibility models in the inversion process. This is the case of many geological bodies in Carajás Mineral Province in northern Brazil, a prolific mineral district with mineralization associated with hydrothermalism.

Figure 5.1: a) Dike with magnetic susceptibility equal to 0.1 SI in a non-magnetic background; b) synthetic magnetic data for the dike in a low latitude area of 0 inclination, 0 declination, and 25.000 nT field strength; and c) inversion result showing a clipped volume at 0.065 SI and the poor recovery of the shallow lateral edges of the dike.

Alternatively, the magnetic susceptibility can be obtained from frequency-domain electromagnetic (FEM) data. In this type of survey, a transmitter generates a primary field,
which induces currents in electrically conductive bodies in the ground (Faraday’s law). The induced currents create a secondary magnetic field, which is measured by the receiver in the presence of the primary field. For this reason, the method is primarily designed to detect the secondary magnetic field related to variations in conductivity ($\sigma$) and commonly considers that the ground has the free space magnetic permeability ($\mu_0$). However, the magnetization phenomena occur when the medium has $\mu > \mu_0$, and the secondary field generated by the magnetic dipoles are in-phase with the primary field but has opposite sign. There are different approaches to explore this information in FEM data such as, direct magnetite mapping (e.g. Fraser, 1973, and Fraser, 1981) and inversion of magnetic susceptibility (e.g. Zhang & Oldenburg, 1997). The field from magnetization decreases the amplitude of the anomaly caused by eddy-currents if the body of high conductivity also has high magnetic susceptibility ($\kappa = \frac{\mu}{\mu_0} - 1$). Thus, not taking into account the correct magnetic susceptibility can lead to erroneous modeling of conductivities. To avoid this problem, simultaneous mapping of and inversion for conductivity and magnetic susceptibility from FEM data is proposed by different authors (Beard & Nyquist, 1998; Zhang & Oldenburg, 1999; Huang & Fraser, 2000; Farquharson et al., 2003; Huang & Fraser, 2003; Sasaki et al., 2010) that have shown improved results for both conductivity and susceptibility models. However, the conductivity and susceptibility responses are mixed in the measured signal. Noh et al. (2017) show that a FEM survey using an active source at sufficiently low frequencies can image susceptibility without the influence of the electromagnetic induction response for shallow targets ($<300$ m).

As conductors become deeper or weaker, the secondary field becomes small or and difficult to be detected in the presence of the much stronger primary magnetic field. For this reason, mineral exploration mainly uses transient electromagnetic (TEM) methods because it can detect smaller secondary magnetic fields in the absence of primary field. In TEM methods, the transmitter primary field is ramped up and then turned off to induce currents in the ground. Then, the receiver measures the decay of the secondary magnetic field during the
off-time in the absence of the primary field. This method is designed for mapping electrical conductivity because the main phenomenon is the decay of eddy-currents. Magnetic minerals in a conductive medium affect TEM off-time measurements, making the field decay slower in the late times than it would for purely conductive rocks. Zhdanov & Pavlov (2001) present a method for modeling both susceptibility and conductivity simultaneously using the late channels. Their method is a 1D imaging technique that builds models containing thin sheets with anomalous conductivity and magnetic susceptibility, and is suitable in the case of strong magnetic and conductive anomalies. Modeling the late times in TEM is similar to modeling low frequencies in FEM. However, Noh et al. (2017) make the frequencies low enough to avoid the influence of conductivity while Zhdanov & Pavlov (2001) model both conductivity and susceptibility. Additionally, both approaches do not consider the induced polarization effect that occurs in low frequencies. Traditionally, the focus of TEM is on the off-time data, which makes electrical conductivity the property of interest, and the on-time measurements are not used. However, the on-time part of the data carries important information about the subsurface magnetic susceptibility because it is similar to a magnetic survey with an active source.

Therefore, we develop the magnetic on-time transient electromagnetic (MoTEM) method to image subsurface magnetic susceptibility. The method uses the on-time part of the waveform produced by the transmitter of transient electromagnetic systems as the inducing field.

5.2 MoTEM method

Although TEM data have the potential for mapping conductivity and susceptibility through the use of both off and on-time data, the potential for mapping susceptibility from the on-time data has not been explored. For the MoTEM method, the transmitter waveform (Figure 5.2) is constant during the on-time. The transmitter primary field will induce eddy currents in the ground, and in the first stage the secondary field measured by the receiver will mainly respond to conductivity. After the dissipation of the eddy currents, the field becomes nearly static and the secondary field measured by the receiver will mainly respond
to susceptibility (Figure 5.2), and this is the signal of interest for the MoTEM method. Furthermore, the main advantages of using on-time electromagnetic data for modeling magnetic susceptibility are: (i) the data are not influenced by remanent magnetization because the inducing field alternates, (ii) the inducing dipole has constant direction over the survey, and as a consequence the anomaly shapes are independent of latitude, and (iii) the anomalies are spatially better defined, with less overlap from anomalies nearby (Fraser, 1973).

One TEM survey has the potential of giving information about two physical properties, but has been limited to only one. Therefore, it is important to increase the value of the acquired data by extracting as much information as possible. The on-time TEM data can be valuable for extracting susceptibility of magnetic bodies and can be used in combination with geomagnetic data to detect and estimate the presence of remanence. In the proposed method, we invert the on-time TEM data in the same manner in which geomagnetic magnetic data are inverted. Instead of considering the Earth’s static field in the forward operator, however, we consider the static field from the EM transmitter loop. The assumption of a static inducing field is the same as assuming a near-zero frequency, which differs from the low frequency assumption of Noh et al. (2017). The MoTEM method can take advantage of existing airborne and ground TEM systems and can be applied to both. Therefore, with MoTEM we propose full waveform acquisition and modeling of two physical properties, susceptibility from the on-time, and conductivity from the off-time part of the waveform. In this paper, we demonstrate the feasibility of the method with a synthetic model. In the following we show the sensitivity calculation, then the inversion method. Next, we show the results for the synthetic model and demonstrates the feasibility of the MoTEM method.

5.3 Sensitivity calculation

In a traditional magnetic survey, the magnetic data is acquired by measuring with a magnetometer (receiver) the total field that is the sum of the secondary magnetic field ($\vec{B}_a$) induced in the ground by the Earth’s primary geomagnetic field ($\vec{B}_0$) (transmitter) and the inducing field. The aim is to image the subsurface susceptibility. For mineral explo-
ration applications, the covered area is small enough that the geomagnetic field strength, inclination, and declination can be considered constant. The forward magnetic modeling commonly adopts the Born approximation, which assumes that the magnetic susceptibility is small ($\kappa << 1$), as is the common case with materials in mineral exploration. Therefore, to the first order approximation, the magnetization $\vec{J}$ is proportional to the susceptibility $\kappa$, and is given by:

$$\vec{J} = \frac{\kappa}{\mu_0} \vec{B}_0$$  \hfill (5.1)

This formulation ignores the self-demagnetization effect and the presence of remanent magnetization. The anomalous field produced by the distribution of magnetization $\vec{J}$ is given by the following equation that contains a dyadic Greens function:

$$\vec{B}_a(\vec{r}_{rx}) = \frac{\mu_0}{4\pi} \int_{\Delta V} \nabla \nabla^T \frac{1}{|\vec{r}_{rx} - \vec{r}_s|} \cdot \vec{J}(\vec{r}_s) dv,$$  \hfill (5.2)

where $\vec{r}_{rx}$ is the receiver observation location, $\vec{r}_s$ is the subsurface source location, and $V$ represents the volume of magnetization. Therefore, the anomalous field obtained from
processing the measured data is equal to:

\[
\vec{B}_a(\vec{r}_{rx}) = \frac{1}{4\pi} \int_{\Delta V} \vec{B}_0 \kappa \nabla \nabla^T \frac{1}{|\vec{r}_{rx} - \vec{r}_s|} dv,
\]

(5.3)

for constant susceptibility within a volume of source region, which in this case is each cell of the discretized subsurface (Figure 5.3).

![Schematic view of the anomalous magnetic field induced by the Earth's geomagnetic field near the Equator with 0 degrees inclination and declination.](image)

Figure 5.3: Schematic view of the anomalous magnetic field \((\vec{B}_a)\) induced by the Earth’s geomagnetic field \((\vec{B}_0)\) near the Equator with 0 degrees inclination and declination.

However, there is a poor coupling between elongated magnetic bodies and the geomagnetic field when the strike direction of the body is close to the Earth’s magnetic field direction (e.g. Figure 5.1). For this reason, in MoTEM we propose to use the magnetic field produced by a current loop \((\vec{B}_{0EM})\) in place of the geomagnetic field \((\vec{B}_0)\) as an inducing field (Figure 5.4). The MoTEM field is induced during the on-time part of TEM waveform systems. Therefore, it takes advantage of existing equipment to measure the response of conductivity and susceptibility in the same survey, a full-waveform acquisition. However, the transmitter waveform needs to be long enough to the eddy-currents dissipate and reach steady state.
during the on-time (Figure 5.2). When this state is reached, Ampere’s law is governing and there are no electrical fields associated with a changing magnetic field (Faraday’s law). Therefore, the data measured during this time interval is not sensitive to conductivity but only to susceptibility.

![Diagram](image.png)

Figure 5.4: Schematic view of the anomalous magnetic field ($\vec{B}_a$) induced by a current loop field ($\vec{B}_{0EM}$) near the Equator with 0 degrees inclination and declination.

In the sensitivity calculation, we compute the current-loop field ($\vec{B}_{0EM}$) for each cell center for a given transmitter position. Then, $\vec{B}_{0EM}$ replaces $\vec{B}_0$ in equation 5.3, which is used to compute the anomalous magnetic field ($\vec{B}_a$) produced by the cell ($\vec{r}_s$) at a given receiver position ($\vec{r}_{rx}$). Therefore, the measured anomalous field decay is proportional to the inverse of the distance to the power of 6 ($1/r^6$), because the primary field decays at $1/r^3$ from the transmitter current loop to the cell, induces a secondary magnetic field that decays $1/r^3$ from the cell to the receiver. For this reason the measured fields are small and subject to depth limitations and equipment resolution. The current-loop field is computed using the
Biot-Savart law:

\[
\vec{B}_{0EM}(\vec{r}_s, \vec{s}, \vec{r}_{tx}) = \frac{\mu_0 I}{4\pi} \int_{\Delta V} \frac{d\vec{l} \times (\vec{r}_s - \vec{r}_{tx})}{|\vec{r}_s - \vec{r}_{tx}|^3} dv, \tag{5.4}
\]

where, \( I \) is the current in the transmitter, \( d\vec{l} \) is a loop segment, \( \vec{s} \) is the loop size and orientation, and \( \vec{r}_{tx} \) the transmitter location. The scheme presented is then used for the computation of the sensitivity matrix, which is used in the inversion of the TEM on-time data in the same way geomagnetic data are inverted. The anomalous field induced by the controlled source electromagnetic field is then computed by:

\[
\vec{B}_a(\vec{r}_{rx}) = \frac{1}{4\pi} \vec{B}_{0EM}(\vec{r}_s, \vec{s}, \vec{r}_{tx}) \int_{\Delta V} \kappa \nabla \nabla^T \frac{1}{|\vec{r}_{rx} - \vec{r}_s|} dv, \tag{5.5}
\]

for each cell of the model.

### 5.4 Inversion

We used the 3D potential field inversion algorithm developed by Li & Oldenburg (1996) to invert the magnetic data, which consider the following relationship:

\[
d = F [m], \tag{5.6}
\]

where \( d = [d_1, d_1, ..., d_N]^T \) is the magnetic data vector, \( F \) is the forward operator, and \( m = [m_1, m_1, ..., m_M]^T \) is the model vector containing the susceptibility of each cell in the model, \( N \) the number of data and \( M \) the number of model cells. For MoTEM data, the forward operator \( F \) becomes a linear system:

\[
d = G \, m \tag{5.7}
\]

where \( G \) is the \( N \times M \) sensitivity matrix, which contains the physical relationship between each cell in the model with respect to each datum, and was computed using \( \vec{B}_{0EM} \) from a steady state current loop as the inducing field.

The inverse solution is obtained by solving the following constrained minimization problem using Tikhonov regularization:

\[
\phi = \phi_d + \beta \, \phi_m, \quad \text{subject to} \ b_l \leq m \leq b_u, \tag{5.8}
\]
where $\phi$ is the objective function, $\phi_d$ is the data misfit function, $\phi_m$ is the model objective function, $\beta$ is the regularization parameter, and $b_l$ and $b_u$ are the upper and lower bounds, respectively, of the model values.

The inversion process described by Li & Oldenburg (1996) applies a depth weighting to counteract the natural decay of the magnetic field and overcome the tendency of putting the model structures near the surface. For inversion of geomagnetic data, they propose using the function $(z+z_0)^{-3}$, where $z$ is the depth of the model cells and $z_0$ is obtained by matching this function with the kernel function beneath the observation point. This function is consistent with the decay of the magnetic field by inverse distance cubed, and the depth weighting function is:

$$W(z) = \frac{1}{(z+z_0)^{\frac{3}{2}}} \quad (5.9)$$

Considering that the function is consistent with the decay of the magnetic field, in our MoTEM experiment we use:

$$W(z) = \frac{1}{(z+z_0)^{\frac{6}{2}}} \quad (5.10)$$

Additionally, we tested the sensitivity weighting described by Li & Oldenburg (2000b) because it depends on the overall sensitivity of the entire data to a particular cell in the model. The sensitivity weighting function is defined as:

$$W_j = \left( \sum_{i=1}^{N} G_{ij}^2 \right)^{\frac{\alpha}{2}} \quad j = 1, \ldots, M \quad (5.11)$$

where $0.5 < \alpha < 1.5$. The value of $\beta$ is usually 1.0, and larger $\alpha$ means stronger weighting.

5.5 Synthetic example

The synthetic model used to demonstrate the feasibility of the MoTEM method is a magnetic cube of 0.1 SI susceptibility in a non-magnetic background (Figure 5.5) and 2.0 S/m conductivity in a background of $2.0 \times 10^{-6}$ S/m conductivity. The cube is $100 \times 100 \times 100$ m, and its top is at 25 m depth, in a core mesh of $500 \times 500 \times 200$ m discretized
in 12.5 m cubic cells. The observation spacing is 25 m in the north and east directions. Electromagnetic data was simulated using a transmitter current of 170 A in a circular loop of 13 m radius at 1 m above the ground. Each data location (Figure 5.5) represents one pair of receiver B field measurement in the center of a transmitter loop. The waveform used by the transmitter to produce primary field (Figure 5.6) has a 1 millisecond ramp on, 7 milliseconds of steady current, 1 millisecond ramp off, 5 milliseconds of off-time, and is discretized into 121 time steps.

Figure 5.5: Synthetic model with a magnetic body of 0.1 SI susceptibility and 2.0 S/m conductivity in a non-susceptible background of $2.0 \times 10^{-6}$ S/m conductivity showing the data location.

Figure 5.6: Transmitter waveform showing the 121 time steps used to compute the synthetic electromagnetic data.
Synthetic data was simulated for three scenarios: 1) both conductivity and susceptibility anomalies, 2) only conductivity anomaly, and 3) only susceptibility anomaly with constant half-space conductivity, equal to the background conductivity. The simulated data show that in the first on-time channels the eddy currents and the response from conductivity is the predominant signal (Figure 5.7). As the eddy currents dissipate, the response from magnetic susceptibility becomes relatively stronger (Figure 5.8), until the conductivity effect becomes negligible (Figure 5.9) and this data was used for the MoTEM inversion. The time length necessary for the eddy currents to dissipate will vary depending on the depth, geometry, and shape of the conductive bodies underground.

![Figure 5.7: Simulated electromagnetic data of on-time 1.3 ms from the synthetic model in Figure 5.5, showing that the effect of eddy currents is the predominant signal.](image)
Figure 5.8: Simulated electromagnetic data of on-time 3.0 ms from the synthetic model in Figure 5.5, showing that as the eddy currents dissipate, the magnetic signal is predominant.

For this experiment, Figure 5.10 shows the log sensitivity of the datum highlighted in yellow with respect to the cells of the model, which is one column of the sensitivity matrix $G$. Figure 5.11 shows the comparison of the depth weighting ($W_{\text{depth}}^2(z)$) and sensitivity weighting ($W_{\text{sensitivity}}^2(z)$) functions with the kernel ($G(z)$) corresponding to the cells beneath the same datum. The two weighting functions have similar behavior (Figure 5.12) and both functions were tested in the inversion to counteract the natural decay of the fields. The sensitivity weighting carries more influence of the footprint of the observation location.

The synthetic data computed using both susceptibility and conductivity models (Figure 5.9) were used in the inversion. However, the conductivity effect is negligible in the data. The transmitter primary field was removed and uncorrelated Gaussian noise with standard

123
Figure 5.9: Simulated electromagnetic data of on-time 8.0 ms from the synthetic model in Figure 5.5, showing that the conductivity effect is negligible. The data in (1) was used in the MoTEM inversion.

Figure 5.10: a) Logarithmic of sensitivity of the datum highlighted in yellow and b) section view of the sensitivity of the same datum.
Figure 5.11: Comparison of the kernel function $G(z)$ directly beneath the observation point in Figure 5.10 showing the estimated depth weighting function given by $W^2(z) = (z + z_0)^{-6}$ with $z_0 = 12.5$ m, and the sensitivity weighting function with $\beta = 1$. All curves are normalized for comparison.

Figure 5.12: a) Logarithmic of depth weighting, and b) logarithmic of sensitivity weighting used in the inversions of the anomalous body in Figure 5.5.

deviation of 1% of the datum magnitude plus a floor equals to 0.01 nT was added to the anomalous field data. The standard deviations of the Gaussian noise added to the data were used in the inversion process. Therefore, the best model was obtained by assuming the discrepancy principle (Parker, 1994), where the target misfit equals the number of data ($\phi_d = N$). We used $b_l = 0$ and $b_u = 1$, and a reference model equals to zero. The inversion was able to successfully recover the anomalous body (Figure 5.13 and Figure 5.14); therefore, con-
firming the feasibility of the MoTEM method to map subsurface susceptibility. Both models, one with depth and the other with sensitivity weighting functions, recover the dimensions of the anomalous body. However, the model that uses depth weighting (Figure 5.13) is less affected by the weighting in its deeper parts. Inversions with sensitivity weighting with weaker weighting using the parameters $\alpha$ equals to 0.9, 0.8, and 0.7 in equation 5.11 were also tested. In these models the influence of the weighting is weaker in the bottom of the model, however, the anomalous body gets concentrated in its upper part and loose symmetry.

Figure 5.13: Susceptibility model obtained using depth weighting function, a) section through the showing the anomalous cube (dashed gray line), and b) susceptibility model with cutoff $= 0.11$ SI.

Figure 5.14: Susceptibility model obtained using sensitivity weighting function, a) section through the showing the anomalous cube (dashed gray line), and b) susceptibility model with cutoff $= 0.17$ SI.
5.6 Comparison with geomagnetic data

Although the proposed method uses an active magnetic field that results in a faster decay of the secondary measured field than the traditional magnetic data (Figure 5.15), the use of an appropriate function to counteract the fast decay allow the recovery of anomalous magnetic bodies in a similar manner (Figure 5.16). For comparison, we used the synthetic model in Figure 5.5 and calculated magnetic data using the Earth’s magnetic inducing field in the north pole region because of the similarity in direction with the field induced by the horizontal loop of our experiment. Therefore, the synthetic data was computed using inclination of 90 degrees, declination of 0 degrees, and field strength of 80,000 nT. Then, uncorrelated Gaussian noise with standard deviation of 1% of the datum magnitude plus a floor equals to 1.0 nT was added to the data. This data was inverted using Tikhonov regularization to minimize equation 5.8 and the recovered model (Figure 5.17) is similar to the MoTEM recovered model. However, the MoTEM susceptibility model has a stronger influence of the distortions cause by the fast decay. Nonetheless, the method obtained a result comparable to the traditional magnetic inversion for this target. The depth of investigation will depend on the transmitter current, loop dimensions, and distance from the ground.

Figure 5.15: a) Depth weighting for the function $W^2(z) = (z + z_0)^{-6}$ used in the MoTEM inversion, and b) depth weighting for the function $W^2(z) = (z + z_0)^{-3}$ used in the inversion of geomagnetic data showing the difference in decay.
Figure 5.16: Comparison of the kernel functions $G(z)$ between the geomagnetic and MoTEM fields directly beneath an observation point showing the estimated depth weighting function given by $W_{\text{Geomag}}^2(z) = (z + z_0)^{-3}$ and $W_{\text{MoTEM}}^2(z) = (z + z_0)^{-6}$. All curves are normalized for comparison.

Figure 5.17: a) Susceptibility model obtained using the MoTEM method with depth weighting function, and b) susceptibility model using traditional geomagnetic data, both showing the anomalous cube (dashed gray line).

When the same survey configuration is applied to simulate data for the dike in Figure 5.1(a) and the resulting data is used in the inversion scheme developed here, the resulting anomalous body recovers the shallow part of the dike with a high degree of accuracy (Figure 5.18). The depth of investigation for this survey configuration for the dike synthetic model is estimated to be 250 m.
5.7 Conclusion

We have developed the magnetic on-time transient electromagnetic (MoTEM) method to obtain subsurface susceptibility from TEM data. The method uses the well developed 3D inversion of magnetic data framework because we consider the inducing field is static and long enough to allow the dissipation of eddy currents. Therefore, the effect of conductivity is negligible and susceptibility is the physical property causing anomalies. The main differences from geomagnetic inversions are in the sensitivity computation and depth weighting function. The sensitivity computation uses the field from the current loop as the inducing field.
field instead of the geomagnetic field. The other difference is that the faster kernel decay \((1/r^6)\) of this method, than for magnetic data \((1/r^3)\), requires an appropriate depth weighting function to counteract its natural decay. Therefore, we tested two different functions, the depth weighting and sensitivity weighting, and the inverted models using both functions successfully recovered the magnetic body and confirmed the feasibility of the method. However, the depth weighting function is associated with a better recovery in the deeper parts of the model.

This study shows, through a synthetic example, that magnetic susceptibility can be inverted from on-time transient electromagnetic data. This is an important step towards the exploration of the potential of this method. However, additional studies are necessary to understand the dependency on subsurface properties for the magnetic field to reach steady state. Additionally, the depth of investigation needs to be explored due to its dependency on loop size, current, and distance from the ground. The results shown here are promising as an alternative method for modeling magnetic susceptibility in areas with poor coupling with the Earth’s magnetic field, and affected by magnetic remanence. The remanence can be estimated in combination with geomagnetic data and the joint inversion of both data has the potential for constructing more accurate susceptibility models. The MoTEM method allows full waveform acquisition and inversion of TEM data for modeling two physical properties. Another exciting possibility is that the use of multiple simultaneous current sources may produce an theoretically unique problem for the susceptibility inversion.

5.8 Acknowledgements

A. T. Melo acknowledges the CNPq (Brazilian National Council for Scientific and Technological Development) for the scholarship. We thank Vale S.A. for providing the data used in this study and for permission to publish results obtained from them. This work is supported in part by the Gravity and Magnetics Research Consortium. We also thank Rob Eso for discussions. This work is supported in part by the Gravity and Magnetics Research Consortium (GMRC).
CHAPTER 6

CONCLUSIONS

New mineral deposit discoveries are becoming scarcer over the years. For this reason, mineral exploration must depend on innovative methods of data interpretation to make new discoveries and supply the world’s demand of metals. The work I have developed in this thesis contributes to solving practical challenges of mineral exploration. The proposed methods provide means of constructing quantitative integrated quasi-geology models for interpretation.

In Chapter 2, I have developed a geology characterization method of 2D and 3D geophysical inversions capable of identifying the mineralization zone. The method is performed in four steps. First, the theoretical relations for the physical properties of the minerals associated to the rocks being investigated are established based on published reference values. In the second step, the established relations are applied to the scatterplot of the physical properties recovered from 2D inversions of the cross-section where the first drill holes were drilled. In the third step, the result of the classification is compared with and validated against the drill hole information. In the last step, the classes defined, and validated, by geology information, are applied to the scatterplot of the physical properties recovered from the 3D inversions. The final result will allow the interpreter to evaluate the prospect volume, distribution, and overall potential. Through a case study, I show the results of combining different geophysical methods to define geological units. The geology characterization method presented here is a powerful tool for quick and integrated evaluation of targets with geophysical data and sparse geologic information. The method can be applied in a variety of situations, but is especially suited in the first stages of exploration of targets at large depths or under cover.

This geology characterization method has shown its feasibility for identification and characterization of the mineralized unit. It proved to be a powerful interpretation tool, which
can help understand exploration areas in an integrated manner. It also provide means of estimating the mineralized rock extension and comparing targets.

When lacking a priori geological information and using multiple geophysical models, I propose to use machine learning to build quasi-geology models for interpretation. In Chapter 3, a class of methods of unsupervised machine learning known as clustering, is used to explore the structure and extract information from multiple geophysical inversion models in a quantitative and integrated manner. The study focused on evaluating different clustering algorithms on the geology differentiation of minimally constrained geophysical models, which are commonly used in greenfield exploration when little a priori information is available. I showed that clustering methods with different metrics are influenced to different degrees by smoothness and inversion artifacts when being applied to the segmentation of crossplots. With a synthetic example based on a real copper deposit, I demonstrate that correlation-based clustering can find the correlation subspaces of each geological unit, and therefore, perform geology differentiation with a high degree of accuracy. The quasi-geology model built from correlation-based clustering showed success in finding patterns in a complex geologic setting. This work shows that the application of unsupervised machine learning is feasible on the deposit scales for the identification of potential drilling targets. In addition, we move one step further our understanding on how different clustering algorithms explore the structure of the data for the type of models we use in greenfield exploration.

In Chapter 4, I use independently inverted susceptibility, density, and conductivity models over the Cristalino copper deposit to identify different geologic units through unsupervised machine learning. I apply correlation-based clustering, which takes advantage of the correlation between physical properties for different geologic units. I show that the iron formation has the most anomalous values of susceptibility and density, and moderate values of conductivity. The high grade ore has susceptibility varying from moderate to high, moderate density, and moderate-high conductivity values. Thus, the combination of these physical property ranges is necessary to identify the ore unit, which is our main target. Otherwise, it
would not stand out from the anomalies associated with the iron formation in the individual models. Therefore, the proposed method successfully identified the physical property ranges of each geologic unit and differentiates between them. I show that k-means clustering is a good initial approach to explore the most obvious structures in the data. I demonstrate how this simple clustering method supports the identification of the optimal number of clusters through the application of the L-curve criterion to its objective function for different number of clusters. However, the results are heavily influenced by irrelevant attributes, and I show that correlation-based clustering (ORCLUS algorithm) identified the copper ore, iron formation, and host rock with a high degree of similarity with the geologic model built from drill hole logging data. The main challenge is that the geophysical inversion cannot capture the small variations of the interlayering between the host rock with iron formation cut by ore veins. Therefore, the quasi-geology model yield a good result given the geophysical models available. The proposed method is entirely driven by the data (i.e., the inverted models), and proved to work in a complex geological setting such as Cristalino. The work flow presented here can be applied to any type of target in greenfield and brownfield exploration to increase the understanding and confidence in the drilling planning stages using a quasi-geology model.

The construction of a quasi-geology model using inverted susceptibility models in low latitude regions might suffer from inaccuracies in the shallow parts. Therefore, in Chapter 5, I have developed the magnetic on-time transient electromagnetic (MoTEM) method to obtain subsurface susceptibility from TEM data. The method uses the well consolidated framework of 3D inversion of magnetic data because we consider the inducing field is static and long enough to allow the dissipation of eddy currents. At that stage, the effect of conductivity is negligible and susceptibility is the physical property causing anomalies. The main differences from traditional magnetic inversions are in the sensitivity computation and depth weighting function. The sensitivity computation uses the field from the current loop as the inducing field instead of the geomagnetic field. The other difference is that the faster kernel decay ($1/r^6$) of this method, than for magnetic data ($1/r^3$), requires an appropriate
depth weighting function to counteract its natural decay. This study shows, through a synthetic example, that susceptibility can be obtained through the inversion of on-time transient electromagnetic data. The results shown here are promising as an alternative method for modeling magnetic susceptibility in areas with poor coupling with the Earth’s magnetic field, and affected by magnetic remanence. The remanence can be estimated in combination with geomagnetic data and the joint inversion of both data sets has the potential for constructing more accurate susceptibility models. The MoTEM method provides the theoretic basis for full waveform acquisition and inversion of TEM data for modeling two physical properties. This is an important step towards the exploration of the potential capability of this method for recovering more accurate models; and consequently more accurate geology differentiation results.

This thesis contributes to the development of integrated quantitative interpretation methods in frontier areas. The demonstration of the feasibility of using unsupervised machine learning for geology differentiation on the deposit scales and the development of a novel method, MoTEM, to recover susceptibility models are the main advancements of this work in greenfield exploration. The general approach of geology differentiation using multiple physical property models can be applied on a wide range of scales from 1 km covering a deposit to 1,000 km covering an entire mining terrane. The geologic units mapped through the differentiation approach can include identified lithology types, zones of mineralization, and different types of alteration. A surprisingly encouraging observation is that separately inverted models, when examined jointly, contain sufficient information for this approach to produce meaningful results.

Data image-based interpretation was dominated by anomaly “bump hunting” or similar qualitative approach in the early stage of exploration geophysics. Inversions have increased the quantitative level significantly and changed the paradigm from the data domain to model domain of physical properties, but a significant portion of the interpretation appears to have remained in the mode of “bump hunting” by focusing on anomalous physical property zones.
Combining multiple physical properties models, however, enable us to differentiate between lithologic units, alteration types, or mineralization zones; or even identify them. Integrated interpretation is the next step change in quantitative interpretation of geophysical data.

To advance to this next stage, it has been long recognized that we must produce geological models. Specifically, these geological models include 3D maps of different geology units. When we geophysicists can predict and map geology in such manners, we may be able to change the landscape of mineral exploration and hope to increase the discovery rate in the coming decades. Thus, geology differentiation is a new frontier.

6.1 Future research directions

I presented a method for geology differentiation using multiple minimally constrained independent inversions. However, improvements in the physical property models will positively impact geology differentiation. Therefore, joint inversion or cooperative inversion methods can potentially construct sharper models. Then, correlation-based clustering will find more accurate physical property associations, consequently, more accurate geological differentiation.

Another path that can be explored is the construction of a database with geophysical models of multiple data from different deposits around the world. This database can be used for training supervised machine learning algorithms. Therefore, the trained algorithms can potentially support prediction of targets in greenfield exploration areas.

Regarding the MoTEM method, additional studies are necessary to understand the dependency on subsurface properties for the magnetic field to reach steady state. Additionally, the depth of investigation needs to be explored to understand its dependency on loop size, current, and distance from the ground. Once those parameters are defined, field tests are necessary to confirm the method applicability.
REFERENCES CITED


Ester, Martin; Kriegel, Hans-Peter; Sander Jörg; Xu Xiaowei. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Pages 226–231 of: Simoudis, Evangelos; Han, Jiawei; Fayyad Usama M (ed), Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press.


Leão-Santos, Marcelo, Li, Yaoguo, & Moraes, Roberto. 2015. Application of 3D magnetic amplitude inversion to iron oxide-copper-gold deposits at low magnetic latitudes: A case study from Carajás Mineral Province, Brazil. *Geophysics, 80*(2), B13–B22.


Souza, Joo Paulo Gomes, Leão-Santos, Marcelo Henrique, de Oliveira Freitas, Cantidiano, Matos, Fernando Martins Vieira, & de Carvalho, Diogo Luiz Orpho. 2015. Applicability of standard Euler deconvolution, modeling, and amplitude magnetic data inversion in Greenfield programs: The Leite target case study, Carajás Mineral Province, Brazil. Interpretation, 3(3), T131–T143.


