2/28/2024

Integration of Sentinel-2 and ASTER Satellite Data for Gold Mineral Exploration in Northern Cameroon

NGORAN Serge Fondzefe
MASTERS GEOSPATIAL TECHNOLOGIES

Supervisor
Torsten Prinz

Co-Supervisor
Bayiga Elie Constantin

Co-Supervisor
Jose Francisco Ramos

Declaration

I, NGORAN Serge Fondzefe, declare that this thesis titled, **'Integration of Sentinel-2 and ASTER Satellite Data for Gold Mineral Exploration in Northern Cameroon'** and the work presented are my own. I affirm that:

This work was carried out following the regulations of the University of Münster.

I conducted all research and analysis unless indicated otherwise in the text.

This document has not been submitted for any degree or examination at this or any other university.

This work has not been published, in part or whole, except as outlined in the text.

All sources used have been cited and acknowledged appropriately following the academic writing standards of University of Münster.

This declaration is made with the understanding that violating these affirmations may result in disciplinary action according to the policies of the University of Münster concerning academic integrity.

NGORAN Serge Fondzefe

28/02/2024

Acknowledgement

First, I wish to thank God for the inspiration, the energy, and the capacity to delve into the GIS universe. I thank Professor Marco Painho for granting me that opportunity after interview to be a part of this prestigious program. I would like to express my deepest gratitude to my supervisor Dr Torsten, whose guidance and expertise were instrumental in shaping the direction and completion of my thesis work; professor Bayiga for accepting to embark on this thesis work with me providing materials and answers to every situation I brought up. Thank you to the entire teaching staff whose patience, knowledge, and encouragement have been invaluable throughout this journey. I am also profoundly thankful to my employer, Explorers 33, for their generosity in granting me access to critical data that significantly enhanced the quality and depth of my research. Their support has been pivotal in achieving the outcomes of this work. Additionally, I owe a great deal of appreciation to my classmates and friends, who have been incredibly supportive by offering advice and showing me how to process certain tasks, which were crucial for overcoming the challenges faced during my research. Their practical help and insights have been indispensable. I cannot overlook the unwavering moral support and encouragement provided by my family. Their belief in me and constant motivation have been a source of strength and resilience, driving me to persevere through the highs and lows of this academic endeavor. Lastly, Special thanks to my wife Claribell and daughter Lyona, their inspiration and support has been the bedrock of my success, and for that, I am eternally grateful.

Abstract

Africa is one of the richest continents in terms of mineral resources yet most of the people continue to live in poverty and technological advancement is very low. The mineral sector is one of the industries which the continent can take advantage of to haul itself out of poverty. The demand for mineral continues to increase. Searching for minerals to feed our sustainability goals in a sustainable and environmentally friendly manner is of utmost importance.

Remote Sensing is one of the technologies which can be leveraged to achieve our goals. It is a multidisciplinary domain that is at the helm of innovation whose application span almost every industry. Unfortunately, it is still not utilized to its full potential despite all the benefits that accompany its usage.

This thesis aimed at exploring the performance of remote sensing in mineral exploration in the North of Cameroon and attempted to derive a prediction model for gold exploration using data from Sentinel-2 and ASTER. Some studies had been done already in this direction in this area before, but the methodology applied was different. While image enhancement used was common (band ratios and principal component analysis), they were applied in different ways. We carried out PCA analysis not on the bands but on the band ratios. Moreover, there has not been any attempt to create a machine learning model that can be applied for gold exploration. Prediction maps were produced from the models, and they show a regional correlation with known gold occurrences in the region. This thesis presents an integrated approach of Sentinel-2 imagery, ASTER, DEM and field data through machine learning models that can be applied to set priority areas for field exploration thereby streamlining cost and effort in a complex geological terrain.

TABLE OF CONTENTS

List of Figures

List of Tables

Integration of Sentinel-2 and ASTER Satellite Data for Gold Mineral Exploration in Northern Cameroon

NGORAN Serge Fondzefe

1 INTRODUCTION

The mining process is split into 5 stages: the exploration & prospecting stage, the discovery stage, the development stage, the production stage, and the reclamation stage. Mineral exploration is the search for evidence of a presence of a given mineral in a location. Almost every location has a certain amount of each given mineral, so what we actually search for in mineral exploration is anomalies. Zones of high concentration of a target mineral. The exploration phase is arguably one of the mining process's riskiest and most resource-consuming phases. Reconnaissance efforts are carried out to find out any potential anomalies. Mineral exploration aims to locate ore deposits in commercially viable quantities through the sequential process of information gathering and interpretation **(Abdolmaleki et al., 2020)**. A lot of human resources, as well as financial and technological efforts, are invested. In this stage, geological, geochemical, and geophysical methods are employed to follow up on the anomalies or targets **(Macheyeki et al., 2020)**. Sometimes, the mineral exploration phase results are not economically worth pursuing in a given location. This makes the exploration phase a strategic part of the mining process and sustainability as a whole.

When mining is mentioned, most minds drift toward the planet's destruction. However, it must be noted that the resources we need for the renewable energy process do have to be mined. The demand for minerals is rising from the copper used to build assets for renewable energy systems to the Lithium used in battery cells of electric vehicles. Despite our efforts to increase the usage of recyclable materials in various domains, the demand for minerals continues to increase. Developing sustainable, economical, and environmentally friendly ways to satisfy our mineral needs is paramount.

Picking areas to explore can be challenging. Other challenges include remoteness and accessibility, exploration costs, and other environmental impacts. Remote sensing has emerged as a powerful tool in geological investigations and mineral exploration. The Sentinel-2 mission, part of the Copernicus program

by the European Space Agency (ESA), offers high-resolution, multispectral imagery that has proven beneficial in various earth observation applications, including geological studies **(Drusch et al., 2012)**. Similarly, ASTER, a collaboration between NASA and Japan's Ministry of Economy, Trade and Industry (METI), provides critical data for mapping rock. The community has duly explored a strong and direct link between reflected radiation and mineralogy through quantum physical processes **(Van Der Meer et al., 2012)**. Various materials have distinct ways of reflecting and emitting electromagnetic radiation, providing a foundation for detecting and mapping materials across areas of interest **(Leverington, 2011)**. The narrow spectral bands of Multi Spectral Image (MSI) ***(Figure 2)*** are highly effective in geological studies for identifying minerals containing iron **(Karimzadeh & H. Tangestani, 2022)**. Unlike the meteorological sector, which employs a model-based approach where data assimilation parameters are calculated using physics principles, the geological remote sensing community largely relies on empirical methods and estimates parameters through correlation. This approach impedes the validation and reproducibility of these parameters and renders the results precise to individual sites, suggesting that there is no universally applicable solution **(Van Der Meer et al., 2012)**. Sentinel-2 has one of the better spatio-spectral resolution and signal-to-noise ratio among freely available multispectral data **(Abdolmaleki et al., 2020)**. The satellite sensor's high spatial, spectral, and temporal resolution, coupled with the accessibility of its data at no cost, underscore its significance for investigating its potential across various applications in remote sensing **(Karimzadeh & H. Tangestani, 2022)**. The Advanced Spaceborne Thermal Emission and Reflection (ASTER) multispectral sensor with six bands in the Short Wave Infra-Red (SWIR) and five bands in the Thermal Infra-Red (TIR) ***(Figure 2)*** was mainly designed for geological purposes **(Bedini, 2011)**.

The most frequently studied systems are hydrothermal, which include abundant spectrally active mineral groups such as hydroxyl-bearing minerals **(Van Der Meer et al., 2012)**. A distinctly lower reflectance in band seven than in band five is exhibited by altered rocks **(Sabins, 1999)**.

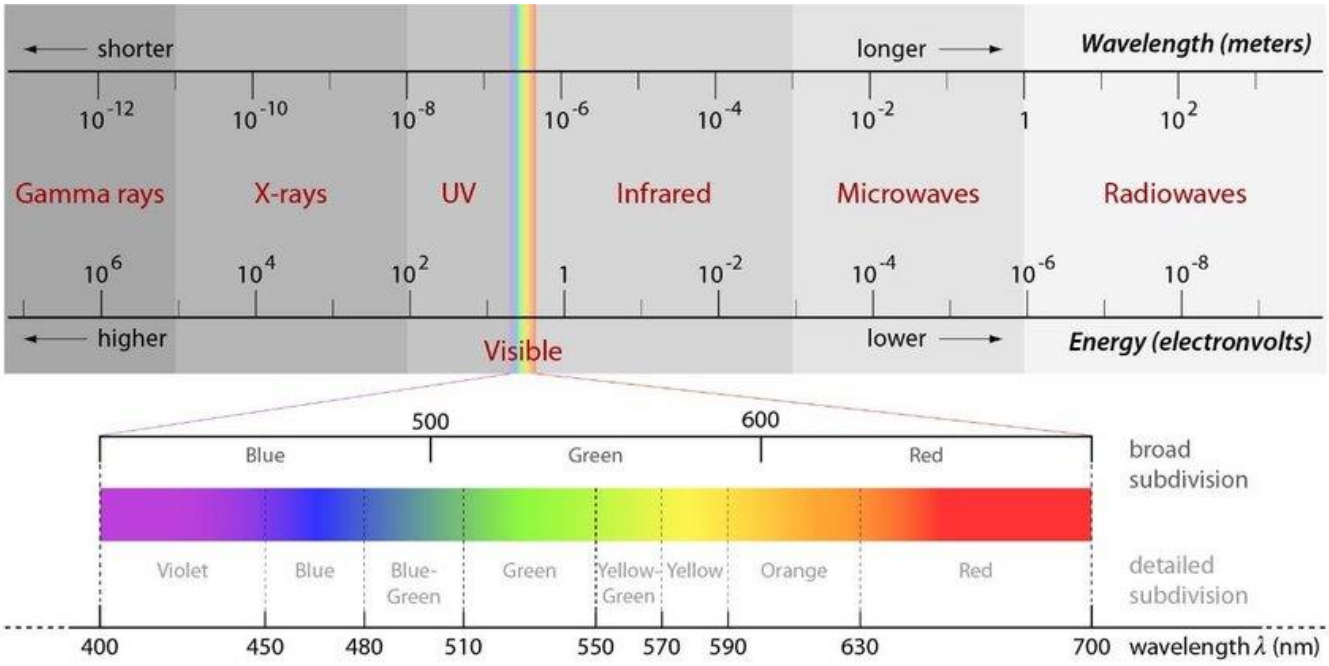


Figure 1 *The complete electromagnetic spectrum with the spectral subdivisions of the visible waveband* ***(Verhoeven, 2017)***

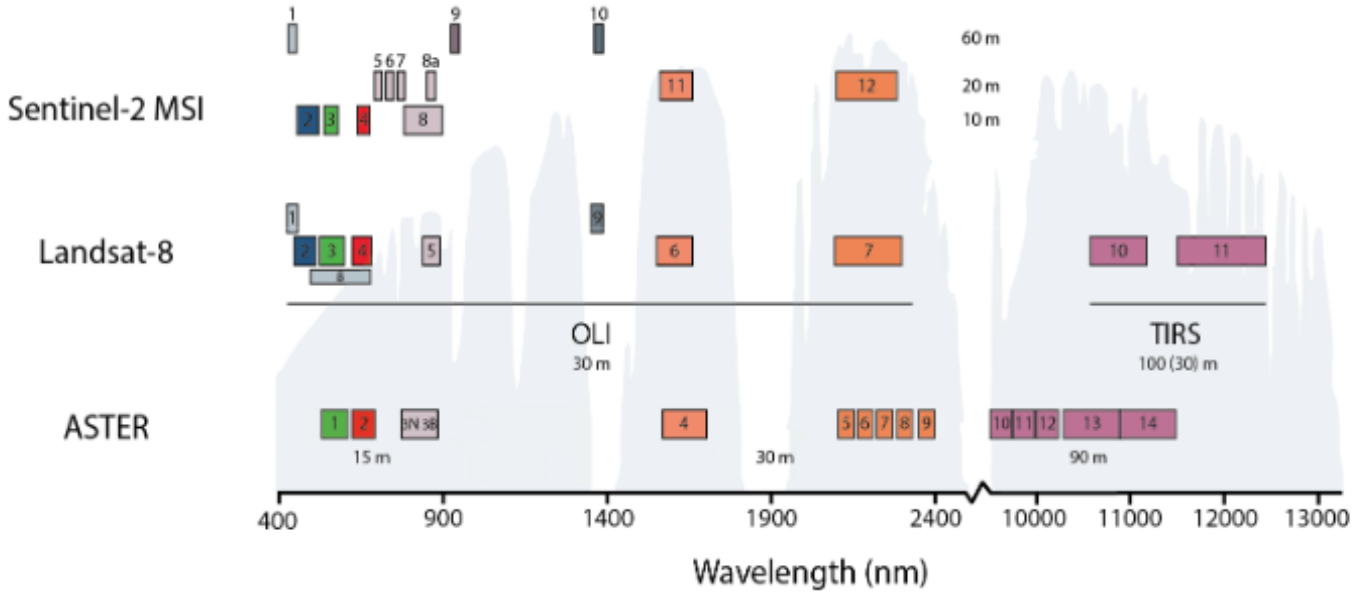


Figure 2 *Comparison between ASTER, Landsat 8, and Sentinel-2 bands. Adapted from USGS Landsat Program, ASTER: 1 green; 2 red; 3N/3B—NIR nadir-looking/NIR backward looking; 4–9 SWIR; 10–14 TIR. Landsat-8: 1 coastal aerosol; 2 blue; 3 green; 4 red; 5 NIR; 6 SWIR 1; 7 SWIR 2; 8 Panchromatic; 9 cirrus; 10 TIRS 1; 11 TIRS 2. Sentinel-2: 1 coastal aerosol; 2 blue; 3 green; 4 red; 5, 6, 7, and 9 vegetation red edge; 8 NIR; 10 cirrus; 11 and 12 SWIR* ***(Cardoso-Fernandes et al., 2020)***

Some of the advantages of using remote sensing in mineral exploration include

Multi-Spectral Analysis: Multispectral images provide information beyond the visible bands of the Electromagnetic Spectrum (EMS) ***(Figure 1)***. Relevant information can be obtained regarding minerals in the Near Infra-Red (NIR), SWIR, and even the TIR of the EMS. This information can be used to map out alteration zones and mineral deposits.

Wide-Area Coverage: Global coverage is advantageous. Coverage of more expansive areas can be done using satellite images. Even difficult-to-access regions can be covered using remote sensing.

There is the advantage of archived information, which can compare and contrast information.

Cost-Effectiveness: Compared to traditional exploration methods, remote sensing can significantly reduce costs by minimizing the need for extensive field trips, drilling, and sampling in the preliminary exploration stages.

Time Efficiency: It enables rapid data acquisition, which is crucial for time-sensitive exploration projects. This efficiency can give companies a competitive edge by allowing quicker decision-making and faster progression to the drilling and development stages.

Environmental Monitoring: Beyond exploration, remote sensing is also used for monitoring environmental changes and impacts associated with mining activities. This can help plan and implement more sustainable mining practices.

Integration with GIS and Other Data: Remote sensing data can be easily integrated with Geographic Information Systems (GIS) and other geological data, enhancing the analysis and interpretation capabilities for more accurate targeting and exploration strategy development.

Advanced Technologies and Algorithms: The continuous advancement in remote sensing technologies, including high-resolution imaging, Light Detection and Ranging (LiDAR), and sophisticated data processing algorithms, allows for ever-improving accuracy and detail in mineral exploration.

Risk Reduction: By providing detailed preliminary data, remote sensing helps to narrow down target areas for more focused exploration, reducing the risk and uncertainty associated with mineral exploration projects.

Some of the disadvantages of using remote sensing in mineral exploration include

Resolution Limitations: The spatial, spectral, and temporal resolution of remote sensing data can sometimes be insufficient for detailed mineralogical analysis, especially when using freely available satellite imagery. Higher-resolution data can be very costly.

Cloud Cover and Atmospheric Interference: Data acquisition from satellite and aerial imagery can be significantly affected by cloud cover, atmospheric conditions, vegetation, and other environmental factors, leading to delays or the need for data correction and filtering. Given that most minerals exhibit spectral characteristics in the IR region of the EMS, zones with vegetation cover are unsuitable for utilizing remote sensing techniques for mineral exploration. LiDAR techniques can still be used but may not always provide detailed mineralogical information.

Complex Data Interpretation: The interpretation of remote sensing data requires specialized knowledge and experience. The spectral signatures of various minerals can be complex and ambiguous, making it challenging to identify specific minerals or alteration zones accurately.

Dependence on Surface Expressions: Remote sensing techniques are mainly limited to identifying mineral signatures exposed at or near the Earth's surface. Deeply buried deposits that lack surface expressions can remain undetected.

Data Processing and Analysis Costs: Although the acquisition of remote sensing data can be cost-effective compared to ground surveys, the processing, analysis, and interpretation of this data can be resource-intensive and require access to advanced software and computational resources.

Integration with Other Data: While integrating remote sensing data with GIS and other geological data can enhance exploration efforts, it can also be complex and time-consuming, requiring a high level of expertise to ensure accurate and meaningful results.

There is an ever-increasing volume of data: Data management represents a challenge. With better sensors being created and better-quality images being produced, such as hyperspectral images, the volume of data is increasing, and robust data management systems need to be implemented.

2 STUDY AREA

Cameroon's geological framework is characterized by a complex interplay of Precambrian and recent formations, including ancient crystalline and metamorphic rocks, volcanic sequences, and sedimentary basins**(Toteu et al., 2004).** This framework reflects a dynamic geological history shaped by tectonic movements, volcanic activity, and sedimentation processes, contributing to the country's rich mineral resources and diverse landscapes **(Bernard T. et al., 2019).**

Tcholire's ***(Figure 4)*** geology is characterized by a complex interplay of Precambrian rocks, which are prevalent in the Adamawa-Yadé domain of the Central African Fold Belt **(Toteu et al., 2006)**. This domain is known for its diverse lithological units, including metamorphic, igneous, and sedimentary rocks. These varied rock types and the region's geological history suggest potential for mineralization, particularly gold. The area's geological diversity is marked by various rock types, such as granites, gneisses, and schists, often associated with gold mineralization **(Toteu et al., 2006; Vicat et al., 1998)**. Additionally, numerous regional shear zones and faults, typical conduits for mineral-bearing fluids, further enhance the prospect of gold ***(Figure 5)***. These structural features are critical targets in mineral exploration, as they often control the localization of mineral deposits. The minerals predominantly found include Quartz, Feldspar, Micas (Biotite and Muscovite), Amphiboles (Mainly hornblende), Pyroxenes and other accessory minerals such as Apatite, Zircon and Magnetite.

In Cameroon, gold mining has a history dating back to 1934, reaching production peaks of 717 kg in 1942 and 2,256 kg in 1951, primarily through artisanal alluvial mining **(Ngon et al., 2015)**. Numerous research efforts have focused on elucidating the origin and characteristics of gold deposits in the northern part of Cameroon, linking gold mineralization to hydrothermal fluid systems hosted by shear zones, where gold is found in quartz veins and altered rocks **(Toteu et al., 2001)**. Previous studies have indicated that gold mineralization results from the deposition or precipitation of high-temperature hydrothermal fluids **(Azeuda Ndonfack et al., 2021; Kpeou et al., 2020; Ngatcha et al., 2019; Takodjou Wambo et al., 2020)**. Gold mining activity is concentrated mainly in the eastern part of Cameroon **(Akwinga, 2010)**. There is significant advancement in research on gold in this region of the country **(Vishiti et al., 2015)**.

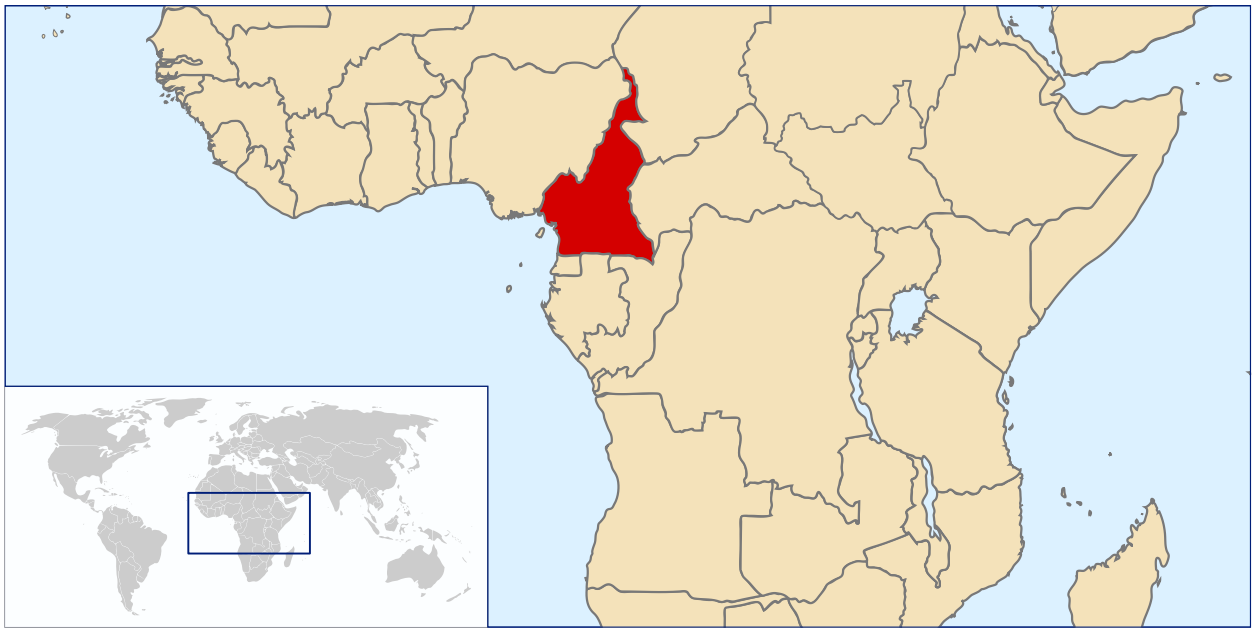


Figure 3 *The location map of Cameroon* ***(Rei-artur, 2006)***

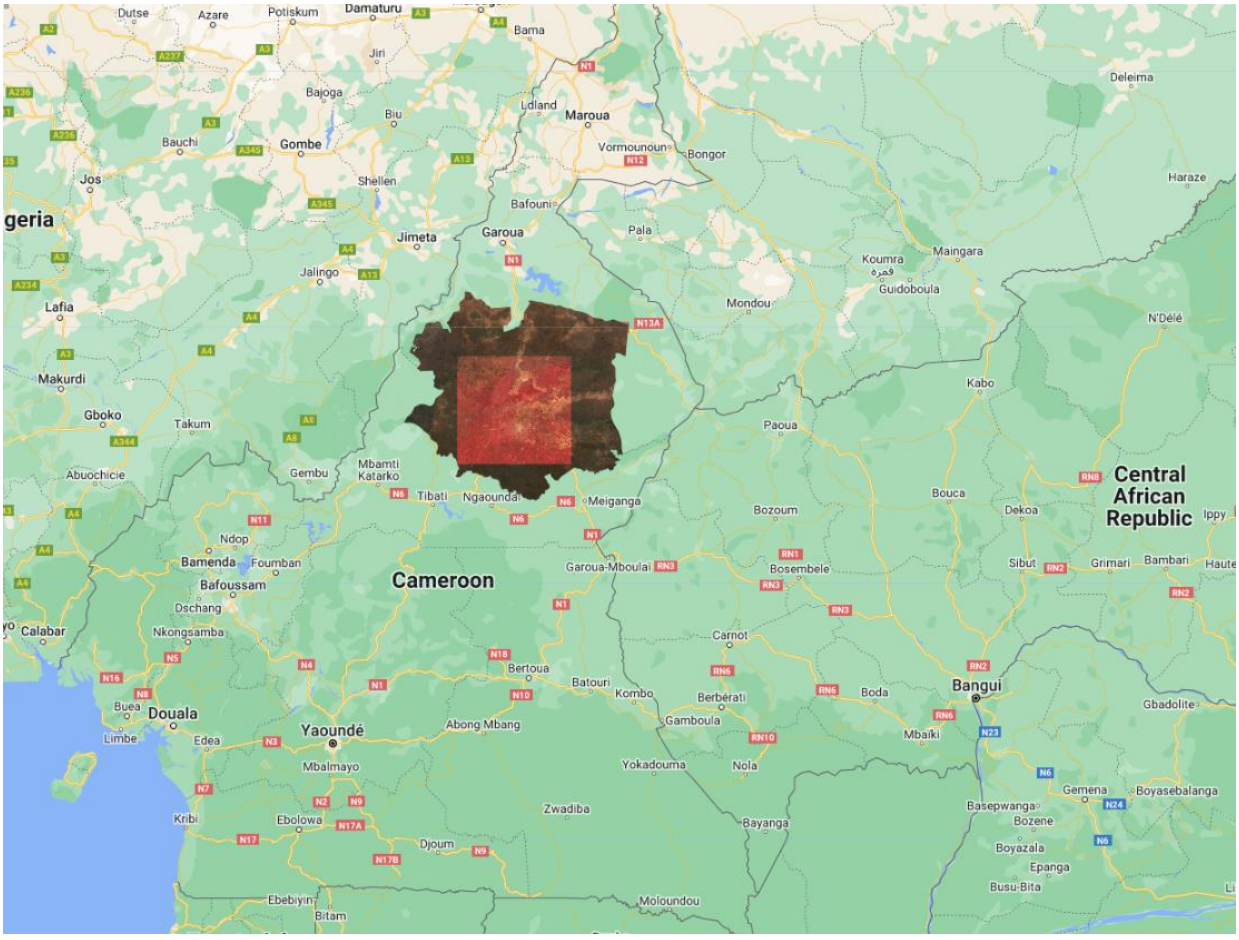


Figure 4 *Google Earth Engine revealing study area in Sentinel-2 RGB image format showing the region covered by samples used in the geochemical data in red shade.*

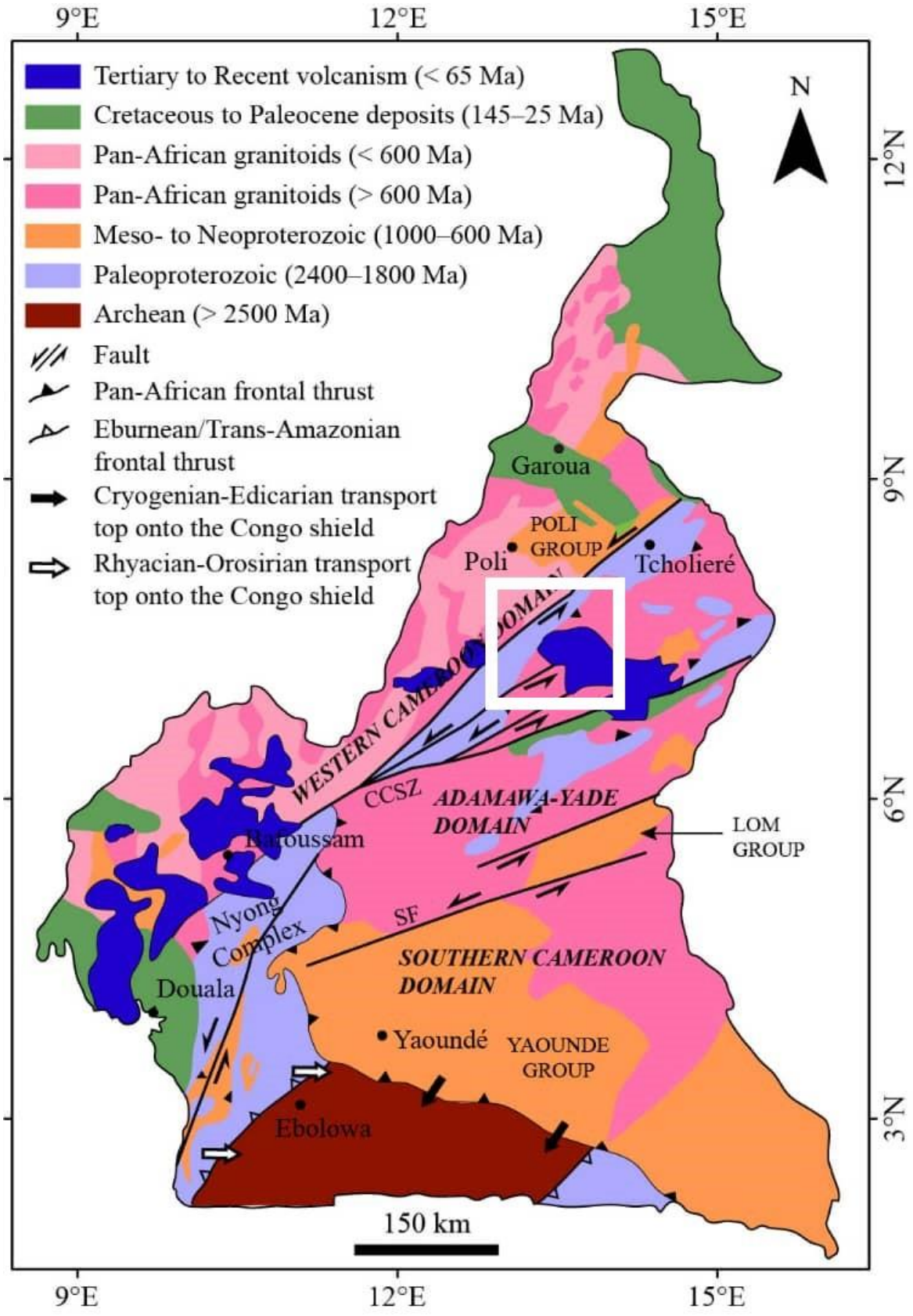


Figure 5 *Geology Map of Cameroon Modified after* ***Dumort et al. 1968; Toteu et al. 2001; Owona et al. 2012*** *showing the position of the study area in white box*

3 RESEARCH OBJECTIVES

Remote sensing for mineral exploration is widely popular in developed countries. Reviews highlighted that numerous image-processing techniques could be utilized with ASTER data to identify hydrothermal alteration mineral zones linked to epithermal gold deposits and the associated host-rock lithology at both regional and local scales **(Pour & Hashim, 2012)**. This, however, is not the case in most of the regions of Africa and some developing countries. Given that sensors on a satellite will perform differently based on solar illumination, cloud coverage, magnetic field, and many other factors, studies in these parts of the globe need to be done independently following methodologies applied in previous works.

This study is set against the backdrop of a growing recognition of the importance of remote sensing in African geosciences. The African Union and various African governments have increasingly emphasized the need for technological advancement in the exploration of mineral resources **(Hilson, 2020)(African Union, 2009)**. Integrating cutting-edge satellite remote sensing into mineral exploration strategies aligns with these objectives and offers a promising avenue for sustainable and efficient resource management. Using satellite imagery for mineral exploration offers a cost-effective, broad-coverage, and non-intrusive method to identify and map mineral deposits over large, often inaccessible areas **(Bedini, 2011; Sabins, 1999)**. This is particularly relevant in African contexts, where logistical challenges and the vastness of unexplored territories pose significant obstacles to traditional exploration methods.

The title of the thesis, "Integration of Sentinel-2 and ASTER Satellite Data for Gold Mineral Exploration in Northern Cameroon," reflects an alternative approach to leveraging advanced satellite remote sensing technologies in African mineral exploration. The work draws inspiration from the limited application of satellite remote sensing in mineral exploration in Africa, specifically Cameroon. The study aims to:

- Investigate the effectiveness of satellite imagery, particularly the integration of Sentinel-2 and ASTER data in exploring gold deposits in Cameroon.
- Develop and evaluate a machine learning models that can be used to predict high-potential mineralization zones.

In conclusion, this thesis aims to bridge the gap between advanced remote sensing technology and African mineral exploration. By focusing on the integration of Sentinel-2 and ASTER data for gold exploration in Northern Cameroon, the study seeks to provide a model or the foundations of a model that could be replicated in other parts of Africa, thereby enhancing the continent's capacity to exploit its mineral resources sustainably and effectively. The best exploration results are obtained by combining geologic and fracture mapping with recognizing hydrothermally altered rocks **(Sabins, 1999)**.

4 MATERIALS AND METHODOLOGY

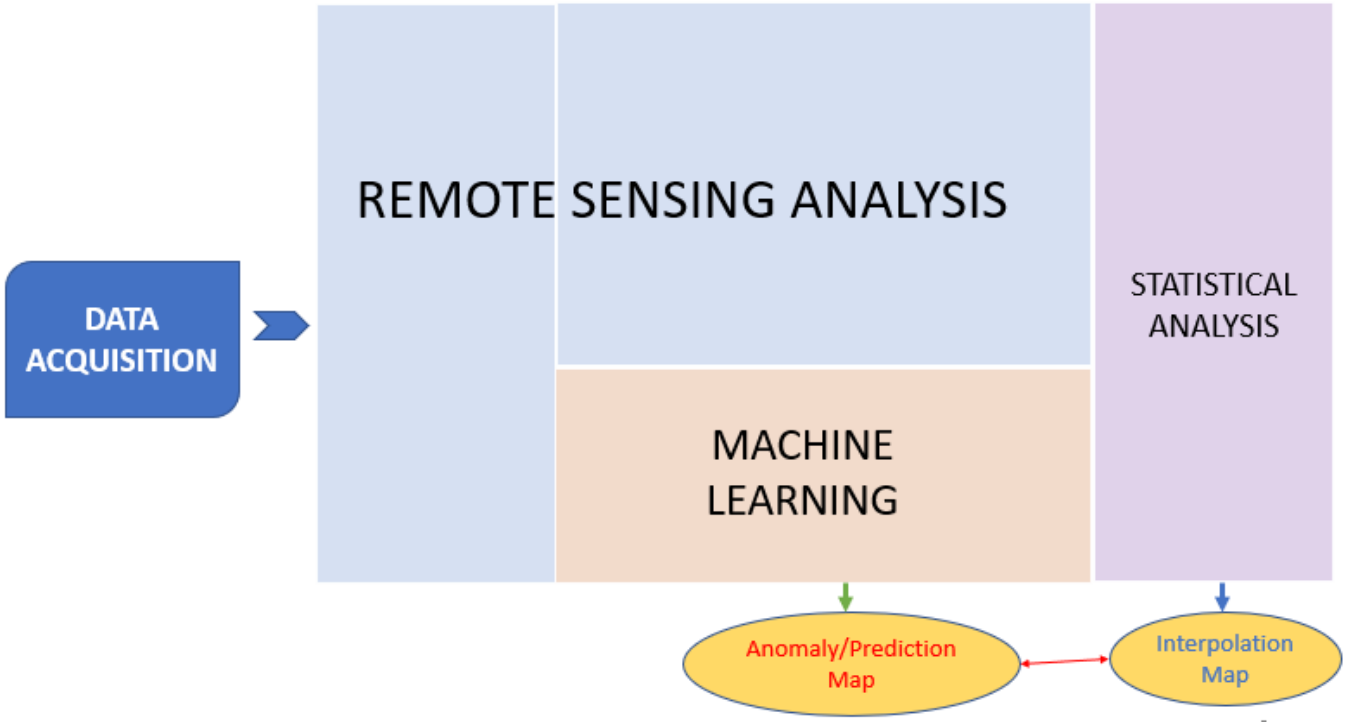


Figure 6 *Overall Methodology split into broad categories along with overall expected outputs*

The broad methodology was applied as represented by ***Figure 6***. This master's thesis, which focuses on the Integration of Sentinel 2 and ASTER satellite images for mineral exploration in Northern Cameroon, a comprehensive workflow divided into four principal tasks was developed and meticulously implemented, each tailored to the specific needs of this investigation. The Data Acquisition phase was the foundation of the research, during which Sentinel 2 and ASTER satellite imagery were collected alongside mineralogical data pertinent to Northern Cameroon. Following data collection, the Remote Sensing Analysis phase was initiated, employing image processing and analysis techniques to classify and interpret the satellite images. This step was critical for identifying potential mineral-rich areas by analyzing the spectral signatures captured in the imagery. In the Statistical Analysis phase, applied statistical methods was used to validate the remote sensing interpretations and machine learning model, ensuring the robustness of the results through spatial statistics. This phase was instrumental in quantifying the confidence of the potential mineral deposits identified. The final phase, Machine Learning, involved developing and training predictive models using the processed satellite data. These models aimed to predict the likelihood of mineral occurrences

across unexplored regions of the study area, leveraging the power of artificial intelligence to uncover new insights and opportunities for mineral exploration. Each of these tasks, underpinned by a series of detailed activities, was designed to optimize the integration of Sentinel 2 and ASTER imagery, ultimately contributing significant advancements to the region's mineral exploration field.

4.1 Data Acquisition

4.1.1 Geochemical Data

The first activity was ensuring availability of geochemical data for the study environment. This was an essential and integral part of the process because of the lack of reliable data from most African countries. The geochemical data was graciously provided by EXPLORERS 33, a geology consultancy firm based in Cameroon. Numerous mines have been found by identifying outcrops displaying altered rocks, followed by assays of rock samples **(Sabins, 1999)**. The technique used by **(H. Tangestani, 2022)**, where the locality of the control points and rock samples were recorded using a Global Positioning System (GPS) device and were matched with their locations on MSI was implemented. ***1232 Rocks*** were sampled during fieldwork in November 2021 and geolocated using handheld GPS. The samples were processed and sent to South Africa for chemical analysis. The geochemical data was obtained in the form of an Excel table. The columns included coordinates of each location from which each rock was sampled, a reference ID number for each rock, ***42 elements*** quantified in ppm, and gold quantified in ppb.

4.1.2 Remote Sensing Data

In areas where bedrock is visible, multispectral remote sensing can aid in identifying altered rocks, as their reflectance spectra exhibit differences from those of the unaltered country rock **(Sabins, 1999)**. For the satellite images, accounts were created in the USGS Global Visualization Viewer and Europe's Copernicus program. The spectral reflectance of minerals is predominantly observed in the NIR and SWIR bands of the EMS. Plants exhibit good reflectance of NIR, so the study area should ideally be void of vegetation coverage to enable the target part of the spectrum to reach the ground surface. It is equally important that the images be free from cloud coverage. Though clouds often affect instead the visible bands of the EMS, specific band ratios are calculated using these bandwidths. Therefore, it is essential to get images with minimal cloud coverage. All remote sensing data was obtained using GEE. Snippets of code for each data are provided in ***Figure 7, Figure 8, and Figure 9***.

4.1.2.1 Sentinel-2 Data

Thirty-six images were downloaded using GEE with a cloud cover of less than 20%. The acquisition period was January 2022, which corresponds to the dry season, which improves the visibility of the acquisition of images. Mean composites were generated to smooth out the images and mitigate the effects of any underlying cloud coverage and shadows. The *maskS2clouds* function was applied to each image in the collection to mask out clouds and cirrus clouds based on the *QA60 band*, which is a quality assurance layer providing cloud cover metadata. The code snippet in Figure 6 shows the process.

```
function maskS2clouds(image) {
  var qa = image.select('QA60');

  // Bits 10 and 11 are clouds and cirrus, respectively.
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  // Both flags should be set to zero, indicating clear conditions.
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
      .and(qa.bitwiseAnd(cirrusBitMask).eq(0));

  return image.updateMask(mask).divide(10000);
}

var dataset = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
                  .filterDate('2022-01-01', '2022-01-30')
                  .filterBounds(ROI)
                  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20))
                  .map(maskS2clouds);

```

Figure 7 *GEE Code snippet for acquisition of Sentinel-2 images, and masking of clouds*

4.1.2.2 ASTER

Fourteen images were downloaded between 07/01/2007 and 18/03/2007. These different images were mosaiced to match the study area. More recent images did not have all 14 bands, especially in the IR region of EMS, which are critical for mineral analysis in remote sensing. However, this time difference does not constitute a problem because geologic events occur on time scales far superior to centuries of years. The study area has not witnessed any disasters in recent times, so acquired data should not pose a problem for our work.

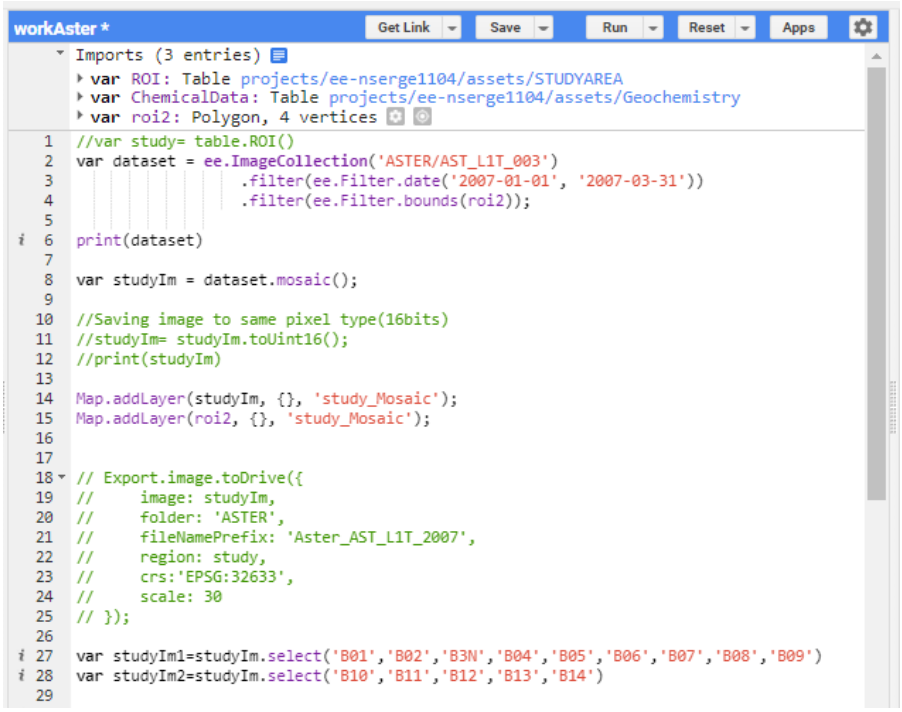


Figure 8 *GEE Code snippet for acquisition and mosaicking of ASTER images. Due to problem with TIR bands, the acquired data was split into 2 sets, the first 9 bands and the last 5 bands featuring in the TIR*

4.1.2.3 DEM

The Digital Elevation Model (DEM) was obtained from the Shuttle Radar Topography Mission (SRTM). The SRTM mission was a one-time data collection event in February 2000, so the dataset does not have varying dates as satellite image collections do.

```
// DEM Download
var dem = ee.Image('USGS/SRTMGL1_003').clip(roi2);
// var elevation = dataset.select('elevationmodel');
// var slope = ee.Terrain.slope(elevation);
Map.addLayer(dem, {min: -1000, max: 2000}, 'dem');
```

Figure 9 *GEE Code snippet for acquisition DEM*

The rest of the work was done in accordance with the workflow described in ***Figure 10.***

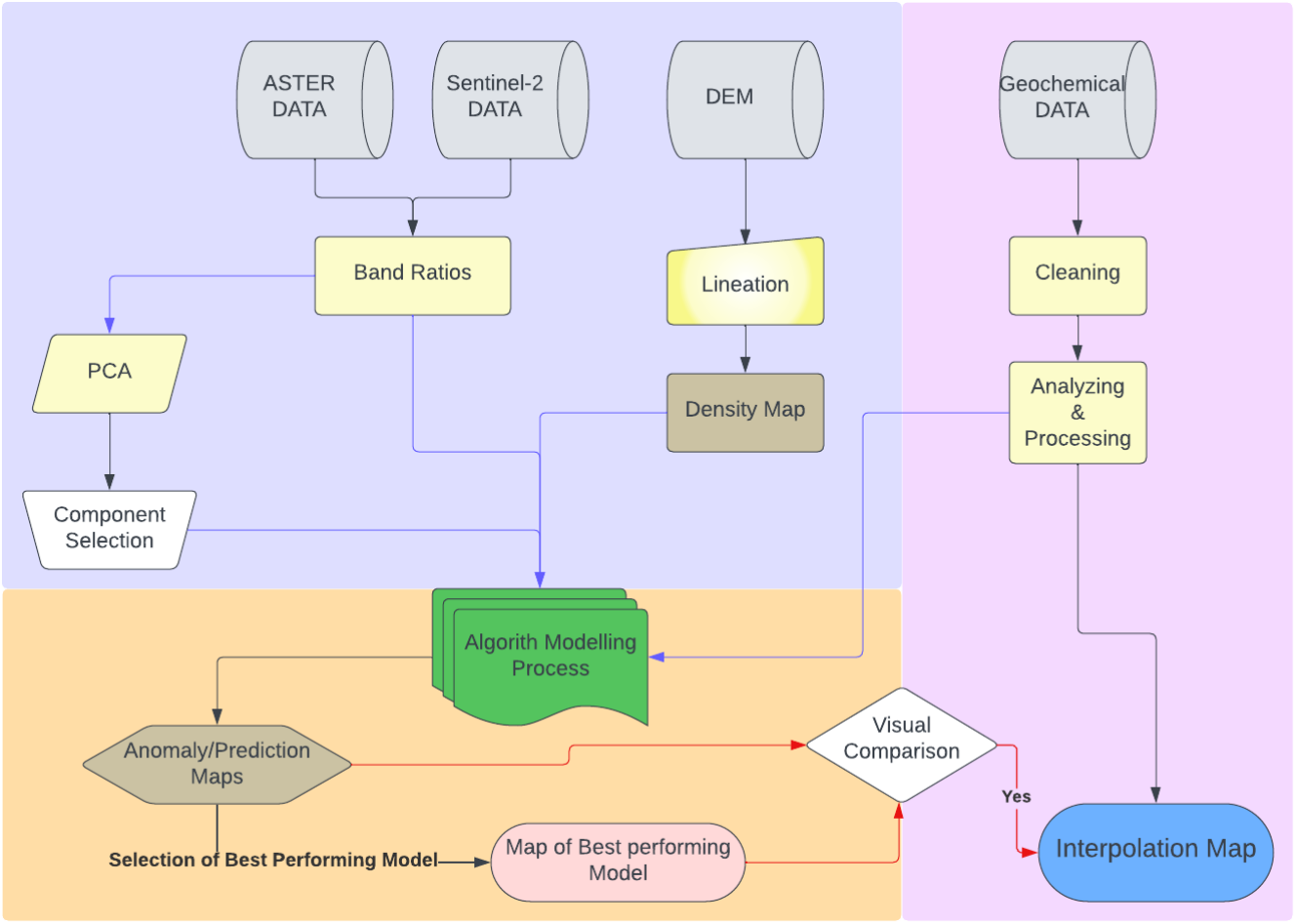


Figure 10 *Detailed breakdown of methodology. Black connecting arrows representing connecting tasks, blue lines is the integration of results from previous tasks, red line represents visual comparison of output*

4.2 Analysis

4.2.1 Remote Sensing Analysis

Geologic remote sensing is based on laboratory spectral analysis of minerals and rocks conducted by Graham Hunt and John Salisbury in the 1970's through 1980's **(Van Der Meer et al., 2012)**. The data from satellite images such as Sentinel-2 and ASTER comprises visible, shortwave infrared, and thermal bands. Through appropriate preprocessing and merging of these bands, it is possible to generate relative distributions of mineral alterations, including iron oxides, siliceous rocks, carbonates, sericite, illite, alunite, and kaolinite **(*ASTER_Processing_for_Mineral_Exploration.Pdf*, n.d.)**. Alteration zones frequently accompany mineral deposits and serve as crucial indicators for mineral exploration. They hold particular importance, especially in hydrothermal sulfide deposits, encompassing metals such as copper, lead, zinc, cobalt, molybdenum, gold, silver, and others. **(Gupta, 2018).**

4.2.1.1 Band Ratios

Band ratio is a method wherein one band's digital numbers (brightness values) are divided by those of another. It is commonly employed to mitigate the impact of solar illumination and topography variability and enhance the spectral information in the images **(Gupta, 2018)**. The selection of ratios depends on the peaks and valleys observed in a reflectance curve. Typically, a band with higher reflectance is divided by one with lower reflectance **(Abdolmaleki et al., 2020)**. Some minerals have perfect reflectance curves with peaks and troughs which when analyzed using band ratios can be able to provide relevant information about the given mineral ***(Figure 11)***.

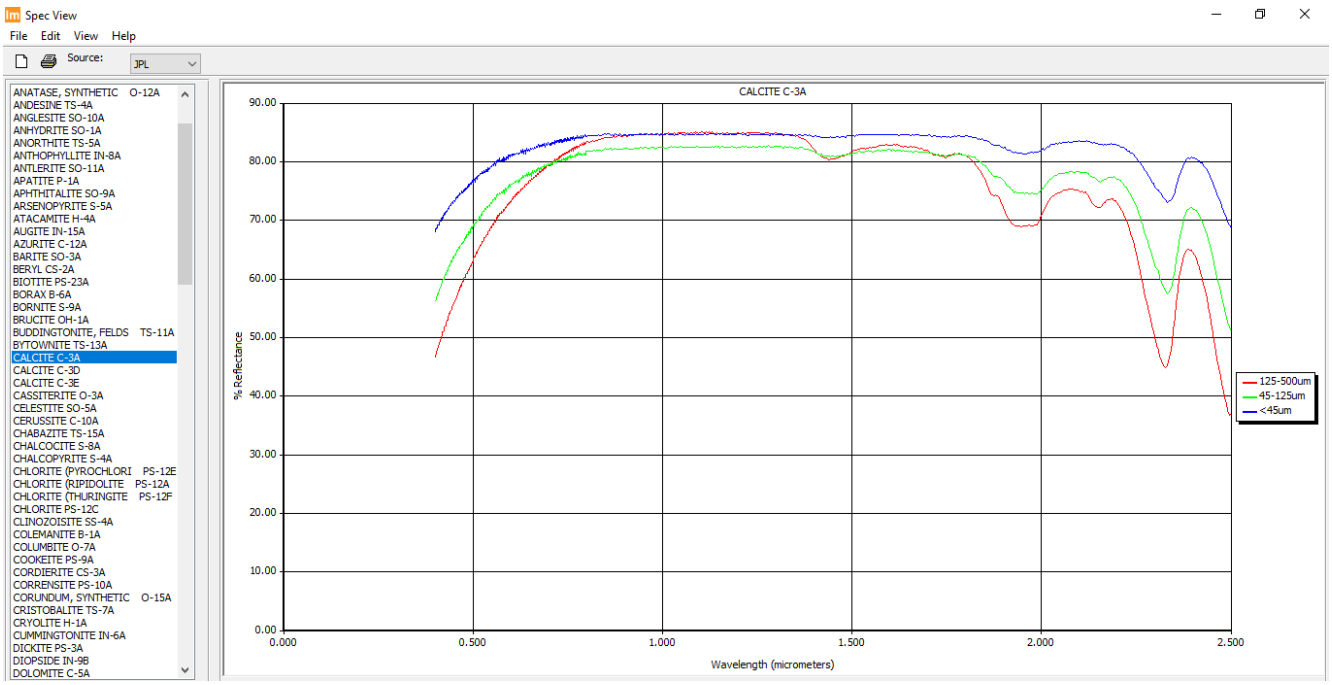


Figure 11 *Erdas Spectral Signature of Calcite, a Carbonate mineral showing peaks and troughs in the SWIR which can be analyzed using band ratios to enhance their mapping in a given study area*

On the flip side, some do no provide so much fluctuation (Figure 11) and analyzing them will require other methods. This however could be handled through the usage of hyperspectral images with finer wavelengths enabling a more detailed analysis.

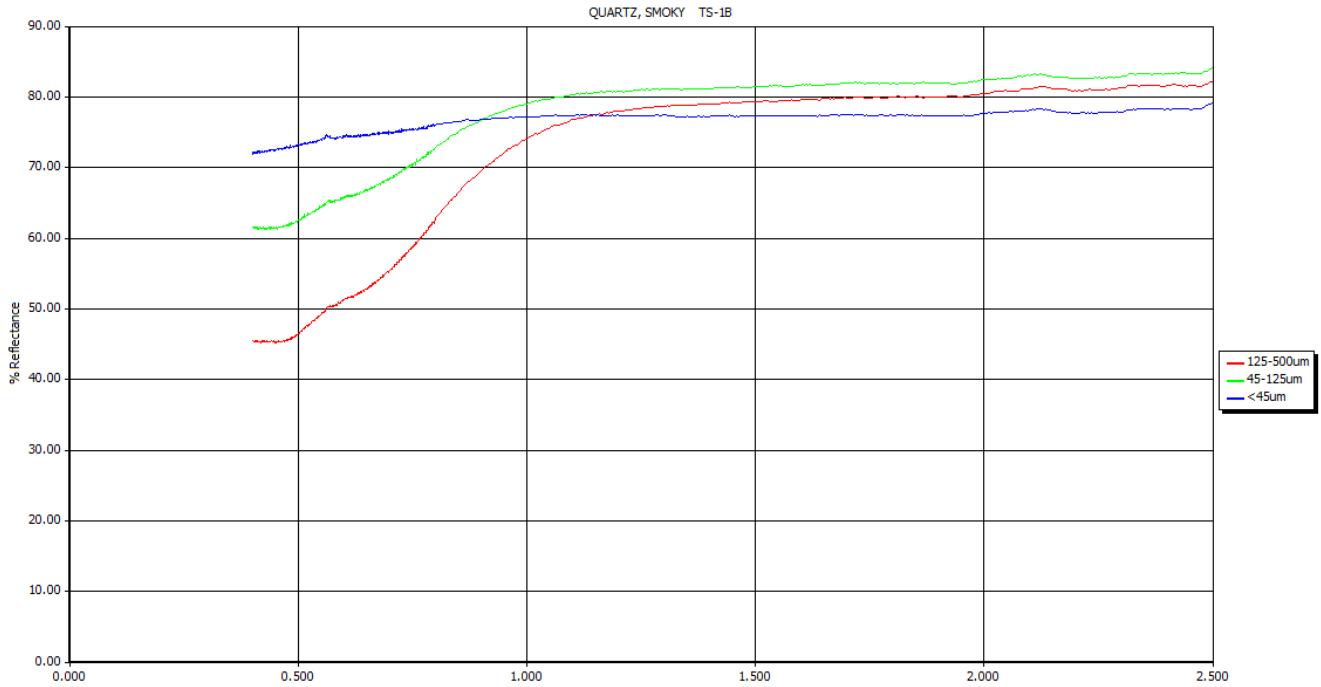


Figure 12 *Erdas Spectral Signature of Quartz, major constituent of host rock in which carries veins in which gold is sometimes mined. An evident lack of clearly distinguishable peaks and troughs*

Figure 12 shows some of the spectral responses for specific mineral categories are capable of varying depending on the rock type. This is especially so because most of these mineral groups are constituted of different minerals that might have slightly differing spectral signatures. These responses are, in turn, used through research to determine band ratios. However, the derivation of these band ratios is not directly given from these spectral responses. Certain spectral bands have been observed to react to the clay and iron absorptions **(*ASTER_Processing_for_Mineral_Exploration.Pdf*, n.d.).** The primary benefit of rationing is its ability to generate an image that is largely unaffected by changes in illumination conditions. If ground materials exhibit the same spectral ratio, the pixels obtain identical DN values in well- or poorly-illuminated areas **(Gupta, 2018)**. ***Table 1*** provides a set of standard ASTER alteration enhancements that can be applied to identify mineral zones or groupings broadly. The selection of band secondary images for which to calculate the band ratios is based on field studies and literature review of with geologic and geographic similarities. **Nzulu (2023)** used X-ray diffraction (XRD) used to identify SiO2 (quartz), Fe3O4 (magnetite), garnet, pyrite (FeS2), periclase (MgO), arsenopyrites, pyrrhotite, biotite, titanium oxide, and Fe2O3 (hematite) as the main gold indicator minerals. Other minerals characteristic of alteration zones were also calculated such as Carbonates, alteration minerals, Chlorites and epidotes.

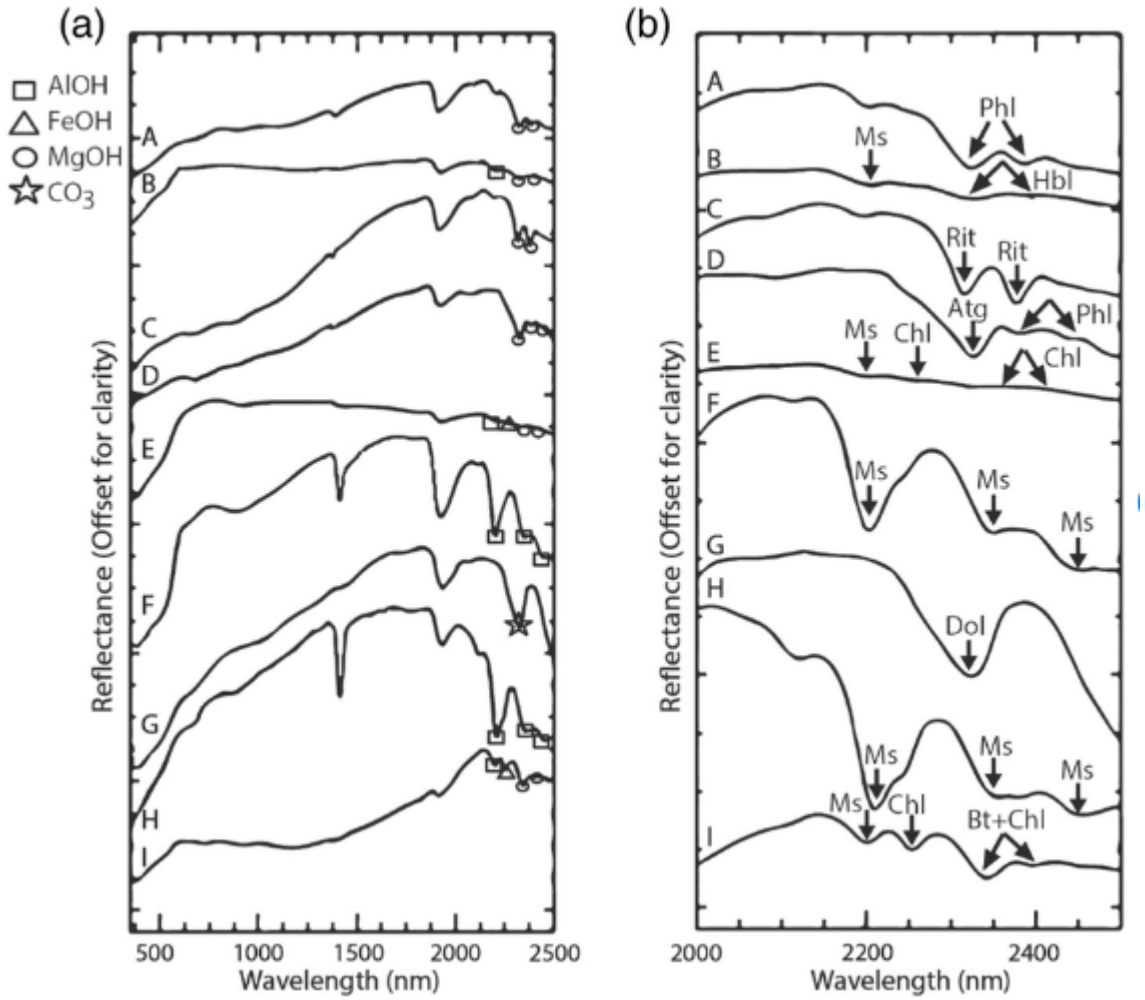


Figure 13 *Averaged spectral reflectance curves for certain rocks. In a) the full spectra range are shown for kersantite (A), monzonite (B), lamproite (C), kimberlite (D), granite (E), fenite (F), carbonatite (G), quartzite (H) and gneiss (I). In b) only the SWIR range is shown. Characteristic features are indicated that are related to the mica group minerals (muscovite: Ms., phlogopite: Phl, biotite: Bt), amphibole group minerals (hornblende: Hbl, richterite: Rit), serpentine group minerals (antigorite: Atg), chlorite group minerals (chlorite: Chl) and carbonate group minerals (dolomite: Dol)* ***(Salehi, 2018)***

Feature	Band or Ratio	Comments	Reference
Iron			
Ferric iron, Fe^{3+}	2/1		Rowan
Ferrous iron, Fe^{2+}	5/3 + 1/2		Rowan
Laterite	4/5		Bierwith
Gossan	4/2		Volesky
Ferrous Silicates (biot, chl, amph)	5/4	Fe oxide Cu-Au alteration	CSIRO
Ferric Oxides	4/3	Can be ambiguous*	CSIRO
Carbonates / Mafic Minerals			
Carbonate / Chlorite / Epidote	(7+9)/8		Rowan
Epiote / chlorite / Amphibole	(6+9)/(7+8)	Endoskarn	CSIRO
Amphibole / MgOH	(6+9)/8	Can be other MgOH or carbonate*	Hewson
Amphibole	6/8		Bierwith
Dolomite	(6+8)/7		Rowan, USGS
Carbonate	13/14	Exoskarn (cal/dolom)	Bierwith, Nimoyima, CSIRO

Table 1 Band Ratios used for enhancing mineral analysis **(after Kaliknowski and Oliver, 2004; *ASTER_Processing_for_Mineral_Exploration.Pdf*, n.d., 2008)**

4.2.1.1.1 Sentinel 2 Band Ratio Calculation

The Copernicus data can enhance exploration efficiency and precision by pinpointing regions of particular interest and more significant mineral potential **(Abdolmaleki et al., 2020)**. ASTER has been widely utilized by the geological remote sensing community since 2000 as a primary tool for surface mineralogy mapping. Numerous ASTER band ratios act as indicators for mineralogy, some of which coincide with the spectral range covered by Sentinel-2 **(Van Der Meer et al., 2014)** (refer to ***Table 2***).

Feature	ASTER	Sentinel-2
Iron		
Ferric Iron, Fe3+	2/1	4/3
Ferrous Iron, Fe2+	5/3 + 1/2	12/8 + 3/4
Laterite	4/5	11/12†
Gossan	4/2	11/4
Ferrous silicates (Biotite, chloride, amphibole)	5/4	12/11†
Ferric oxides	4/3	11/8
Carbonates / Mafic minerals		
Carbonate / Chlorite / Epidote	(7+9)/8	–
Epidote / Chlorite/ Amphibole	(6+9)/(7+8)	–
Amphibole / MgOH	(6+9)/8	–
Amphibole	6/8	–
Dolomite	(6+8)/7	–
Silicates		
Sericite / Muscovite / Illite / Smectite	(5+7)/6	–
Alunite, Kaolinite, Pyrophyllite	(4+6)/5	–
Phengitic	5/6	–
Muscovite	7/6	–
Kaolinite	7/5	–
Clay	$(5\times7)/6^2$	–
Alteration	4/5	11/12†
Host rock	5/6	–
Other		
Vegetation	3/2	8/4
NDVI	(3-2)/(3+2)	(8-4)/(8+4)

Table 2 Senntinel-2 band ratios serve as equivalents to ASTER band ratios, serving as proxies for mineralogy. Adapted from Kalinowski and Oliver, the table is confined to ratios falling within the wavelength spectrum of Sentinel-2 **(Van Der Meer et al., 2014)**

The band ratios for Sentinel-2 data were performed on Google Earth Engine (GEE). Given its volume, code was a better option to traditional GIS software to ease data manipulation. The map function was applied to the dataset, and operations were performed on each image in the collection, as seen in the code snippet in ***Figure 13***. Specific bands were selected for each image using the select function depending on the target band ratio; then, a mean method was applied to the band ratios for the collection. This calculates the mean value for each pixel across all images in the collection. The result is a single composite image where each pixel represents the mean value of the corresponding pixel across all images in the collection. The output was exported as tiff files.

```
// Compute and add Ferric Iron ratio layer (B4/B3)
var ferricIron = dataset
  .map(function(image) {
    var ratio = image.select(['B4']).divide(image.select(['B3']));
    return ratio.rename('FerricIron');
  })
  .mean().clip(roi2);
Map.addLayer(ferricIron, {min: 0.8, max: 1.8, palette: ['black', 'blue', 'white', 'red']}, 'Ferric Iron Ratio', 0);

// Compute and add Ferrous Iron ratio layer ((B12/B8) + (B3/B4))
var ferrousIron = dataset
  .map(function(image) {
    var ratio1 = image.select(['B12']).divide(image.select(['B8']));
    var ratio2 = image.select(['B3']).divide(image.select(['B4']));
    return ratio1.add(ratio2);
  })
  .mean().clip(roi2);
Map.addLayer(ferrousIron, {min: 1.3, max: 2.3, palette: ['black', 'blue', 'white', 'red']}, 'Ferrous Iron Ratio', 0);
```

Fig 10

Figure 14 *GEE Code snippet for calculation of Sentinel-2 Band ratios*

4.2.1.1.2 ASTER band Ratio Calculation

Featuring six bands in the SWIR wavelength range and five bands in the TIR, The ASTER multispectral sensor was primarily developed for geological applications **(Bedini, 2011)**. As a result of the finer separation in the Infra-Red (IR) part of the EMS, more band ratios can be derived from ASTER data. The calculated band ratios relevant for gold exploration in the study area include Alteration, Amphiboles, Amphibole MgOH, Carbonates, Chlorites, Ferrous Silicates, and Host rock. The analysis was done on Erdas using the model builder. These mineral indices were calculated using the model builder function in Erdas. The output was exported as tiff files. ***Figure 15*** shows a sample of the process of modeling a band ratio on Erdas Imagine software platform

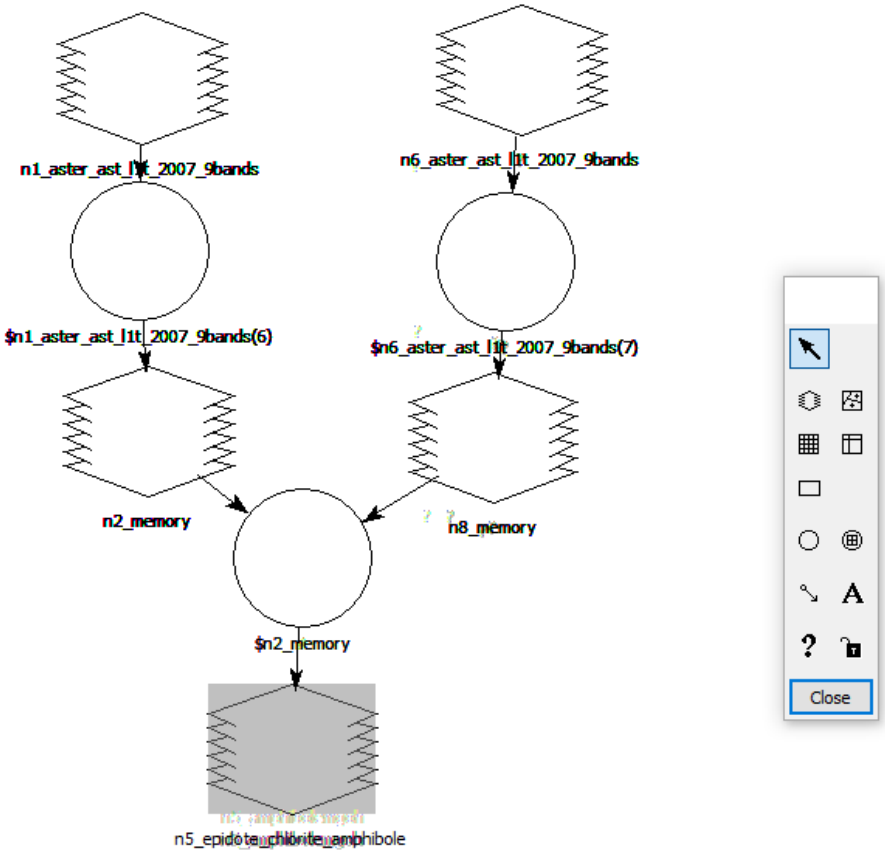


Figure 15 *Sample of model builder used on Erdas to compute the band ratios using ASTER data taking reference from information on Table 1*

4.2.1.2 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method to enhance and isolate specific spectral signatures from background noise **(Abdolmaleki et al., 2020)**. The PCA enhances overall distinctiveness and decreases the number of dimensions, making it valuable in classification tasks **(Gupta, 2018)**. Through a linear transformation, PCA minimizes the redundancy of correlated data and converts a set of interrelated input bands into uncorrelated principal components (PCs) **(Abdolmaleki et al., 2020)**.

The input data was the calculated mineral indices from ASTER data. All indices except the amphibole mineral ratio (Alteration, Amphibole MgOH, Carbonates, Chlorites, Ferrous Silicates, and Quartz) were stacked. The resultant image was used to perform PCA using the built-in PCA function in Erdas, and we set the resultant components to 3. This was in an attempt to reduce the dimensionality of our data to 3 dimensions: PC1, PC2, and PC3. The resultant image was exported in .img format and later converted to .tiff format.

4.2.1.3 Lineaments

Recently, the term "lineament" has been widely utilized, often with varying interpretations. Photo lines, which refer to linear alignments of features observed in photographs and images, are among the most conspicuous features visible in high-altitude aerial and space imagery. Consequently, the term "lineament" has become increasingly common in the literature on remote sensing geology **(Gupta, 2018)**. Shear zones are often associated with the presence of mineralization, especially gold, as they can serve as conduits for hydrothermal fluids that deposit minerals. Understanding the elevation trends is important for planning exploration logistics, as higher terrain may be more difficult to access. It also helps in understanding erosion patterns which can affect the secondary dispersion of minerals.

A DEM is the next-generation digital method for depicting the terrain. It describes the elevations of various points digitally **(Gupta, 2018).** It has various derivatives and multiple applications in geology, including terrain topography, geochemical and geophysical properties of the terrain topography, and our area of interest, shear zones, and faults. Geological surface structures are frequently identified using spatial features in images. The brightness contrast and patterns observed in satellite images indicate lineaments and folds within a study area **(Abdolmaleki et al., 2020)**.

An elevation map was obtained from the DEM. Fault lines were derived from the elevation maps in the form of lineation. Different layers of elevation, as outlined in ***Table 3***, were mapped, and a lineament map was created by deriving lines from each layer map with the details outlined in ***Table 4***. The hillshade is created by simulating sunlight casting shadows at various altitudes and azimuth. This technique accentuates the topographic features and can highlight subtle variations in relief not easily noticeable in a standard DEM.

Significant and minor lineament are consistently noticeable in any given photograph or image. Major lineaments often align with significant shear zones, faults, fractures, and prominent tectonic structures or boundaries. Conversely, minor lineaments typically correspond to comparatively minor faults, joints, fractures, bedding traces, and similar features. **(Gupta, 2018)**. A density map was produced from the lineament map by working on the symbology. All this was done on ArcGIS Pro.

Layer	Azimut	Altitude
Layer 1	315	45
Layer 2	200	50
Layer 3	100	60
Layer 4	75	75

Table 3 Observed hillshade layers from which shear zones were manually mapped through lineaments

Difficulties arise when attempting to incorporate lineament maps into mineral exploration models for the following reasons **(Rowan & Bowers, 1995)**: (a) Some of the features identified as lineaments may not have a structural-geological origin, and (b) distinguishing between post-mineralization and pre-mineralization structures may prove challenging **(Gupta, 2018)**. When comparing lineament patterns with ore occurrences, it was observed that mining districts often align with lineaments and are clustered at their intersections **(Sabins, 1999)**. In his work, **Sabins (1999)** also noted that the best exploration works were obtained by combining fracture mapping with the recognition of hydrothermally altered rocks.

4.2.2 Statistical Analysis

Geochemical methods have traditionally been conducted mainly on the ground. This process typically entails establishing baseline measurements and a grid pattern, followed by sampling rocks, soil, or water and subsequently analyzing their chemical composition **(Gupta, 2018)**

The geochemical data provided contained information on ***1232 samples***. The information on each sample included a reference number, UTM zone, XUTM and YUTM coordinates, the batch number, and the concentration of ***42 mineral elements*** in ppm and gold concentration in ppb. Exploratory Data Analysis (EDA) was carried out on the Geochemical Data. It was cleaned so that any missing data could be addressed. This was done using the filter tool on Microsoft Excel. The relationship between the mineral elements was explored through Pearson's correlation analysis, shown in Table in the Appendix.

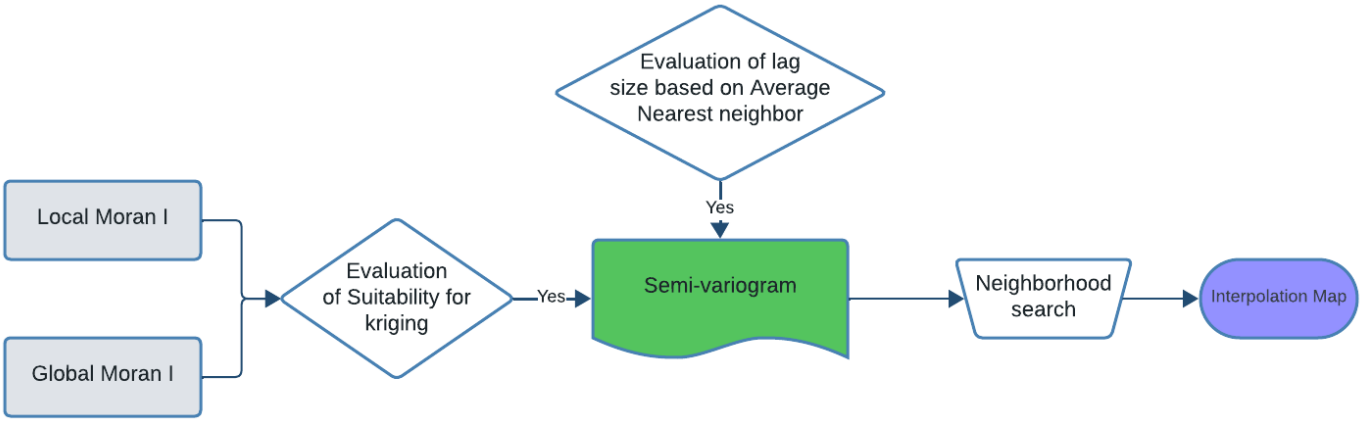


Figure 16 *Steps undertaken for creation of Interpolation map using Kriging*

The existence of local clusters was performed by obtaining the Anselin Local Moran's I on ArcGIS Pro. In spatial analysis, a local Moran's I statistic is used to identify clusters of similar values in a geographical area. It helps understand a variable's spatial autocorrelation, indicating whether similar or dissimilar values tend to cluster in certain areas. The Moran's I value can range from -1 (indicating perfect dispersion) to +1 (indicating perfect clustering), with values near 0 suggesting random spatial patterns. Local Moran's I value of ***0.032653*** is positive and closer to 0, suggesting a weak clustering of similar values in your study area. It indicates a slight tendency for areas with similar values to be located near each other, but the effect is not strong.

A global Moran I was also carried out using ArcGIS Pro. The output favored applying the kriging method using a lag distance of 14 m. The results section discusses the report attached to the appendix in more depth. These 2 reports were enough justification for the implementation of Kriging as a suitable interpolation method. The process of Interpolation is summarized in the workflow in ***Figure 16***.

Ordinary kriging was done using information from the above analysis. Kriging not only estimates values at unsampled locations but also provides an estimation of the uncertainty of those predictions. It considers the spatial correlation between sample points, offering a more statistically reliable interpolation compared to simpler methods like inverse distance weighting. The semi variogram was obtained using values as indicated in ***Figure 17***.

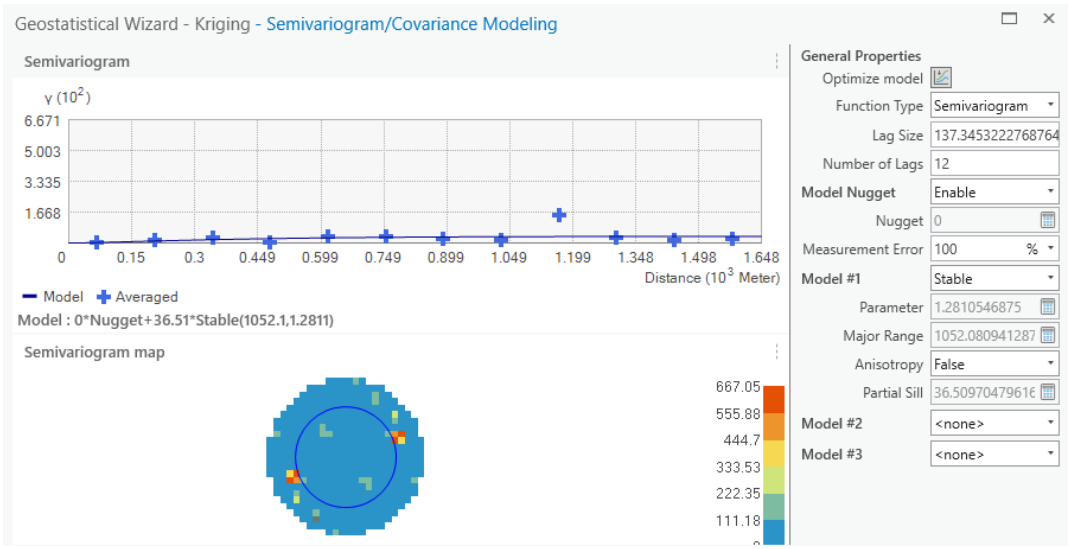


Figure 17 *Semi-variogram model showing the final values that were used*

The Average Nearest Neighbor was evaluated to justify the lag size used in the semi-variogram. The lag size multiplied by the number of lags **should** be less than one-half the most significant distance in your dataset. The Average Nearest Neighbor (featured in the appendix) shows a mean distance of ***966 meters*** and justifies a lag size of ***137***. Number of lags was kept at ***12***.

sector	Max Neig.	Min Neig	Mean Error	RMSE
4 by 45 degree	5	2	0.03	6.0379
4 by 45 degree	8	3	-0.0157	5.913
4	8	3	0.013	5.9864
8	8	3	0.00046	5.7801

Table 4 Samples of different neighborhood searches and the resulting values

Some of the neighborhood parameters used during calibration are shown in ***Table 4***. While the Root Mean Square Error (RMSE) did not change much, I focused on the mean error, and the best value was obtained through 8 sectors, maximum neighbors of 8, and minimum of 3. The prediction map was obtained using the above details.

4.2.3 Machine Learning

Adiri (2020) suggests that the performance of classifications depends on the amount of input data, training samples, and how they represent the study area. They have to be balanced in order to achieve the best performance.

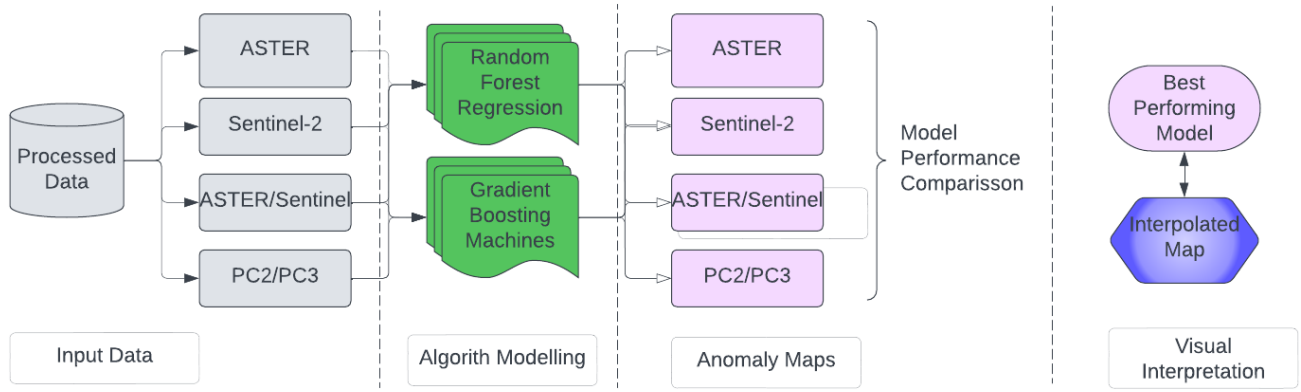


Figure 18 ***Processes involved in the Machine learning part of the project***

The methodology focused on the application and comparative analysis of two machine learning algorithms: Random Forest Regression and Gradient Boosting Machines (via the xgboost package). How they were implemented is outlined in the workflow in ***Figure 18.*** These algorithms were selected for their robustness and effectiveness in handling complex spatial datasets characteristic of remote sensing and geological data. Ensemble-based methods, mainly RF and GBM, outperform other machine-learning classification approaches **(Adiri et al., 2020)**. Random Forest ***(Figure 19)*** offers robustness against overfitting and is relatively straightforward to tune, making it suitable for preliminary analyses.

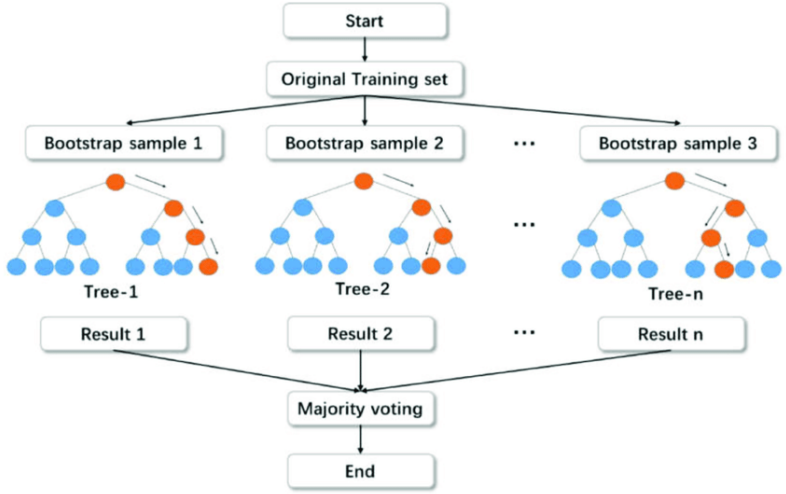


Figure 19 *Random forest algorithm structure* ***(Khan et al., 2022)***

Gradient Boosting Machines, particularly the implementation via xgboost ***(Figure 20)***, are known for their efficiency and performance and can handle nonlinear relationships and interactions between features.

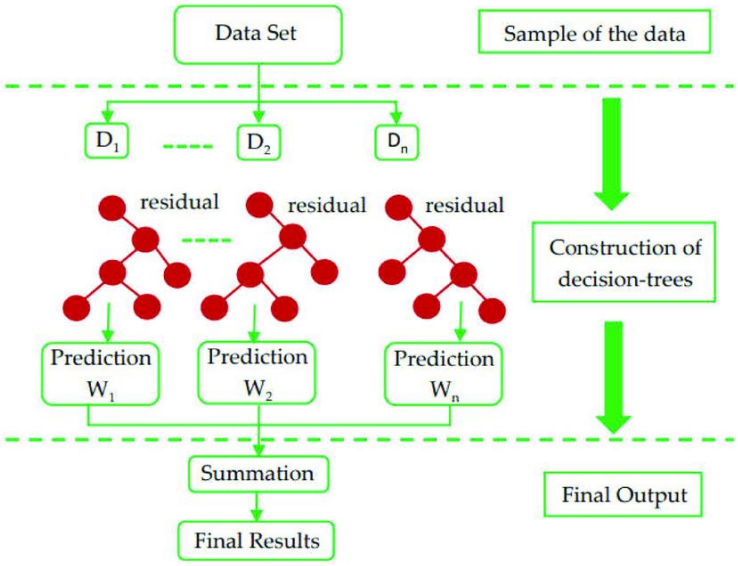


Figure 20 *XGBoost (extreme gradient-boosting) algorithm structure* ***(Khan et al., 2022)***

The modeling process involved the use of R Studio and Google Colab platforms, leveraging several essential R packages for data manipulation (dplyr), spatial data handling (terra, sf), and machine learning (mlr3, xgboost). Four distinct sets of features were tested across both algorithms: (1) three Sentinel 2 derived ratios plus lineament density map, (2) seven ASTER-derived ratios plus lineament density map, (3) all ten ratios from both satellites plus lineament density map, and (4) PCA components (PC2, PC3) plus lineament density map. The primary benefit of this method lies in its ability to resolve potential ambiguities that may arise from relying solely on one dataset by integrating multiple datasets **(Gupta, 2018)**. Each run included remote sensing data and a density map obtained from the lineament map. The best exploration results are obtained by combining geologic and fracture mapping with recognizing hydrothermally altered rocks **(Sabins, 1999)**. The models were trained using geochemical data in the form of shapefiles as training labels, aiming to predict the spatial distribution of gold mineralization zones.

To optimize the model's performance, hyperparameters of the respective algorithm were tuned. In the case of Random Forest, this involved adjusting the number of trees, the depth of the trees, and the number of entries. The model was trained with a ***5-fold*** cross-validation system.

```
###hyperparameter optimization and getting best performing model
random_f= lrn('regr.ranger',mtry=to_tune(1,7),num.trees=to_tune(c(100,300,500)))


instance = ti(
  task = task_gold,
  learner = random_f,
  resampling = rsmp("cv", folds = 5),
  measures = measures_r,
  terminator = trm("none")
)
```

Figure 21 *Sample code used for training using the Random Forest model*

For the Gradient Boosting Machines, the learning rate range was set between ***0.2*** and ***0.5***, tree depth was maintained at the default value of **6**, the sub-sample was between ***0.5*** and ***0.75***, and the number of folds was ***5***.

The output from the predictive models was a geospatial TIFF file representing the predicted high-potential zones for gold mineralization. This file format is compatible with most GIS software, enabling easy integration into further analysis and decision-making processes within the field of mineral exploration.

5 RESULTS

5.1 Remote Sensing Analysis

The output from the remote sensing analysis was maps obtained from Band ratios and density of lineaments. The band ratios were in turn used in the PCA analysis.

5.1.1 Band Ratios

The main purpose of the output from the band ratios was to serve as variables for the machine learning models.

5.1.1.1 Sentinel 2 Band Ratios

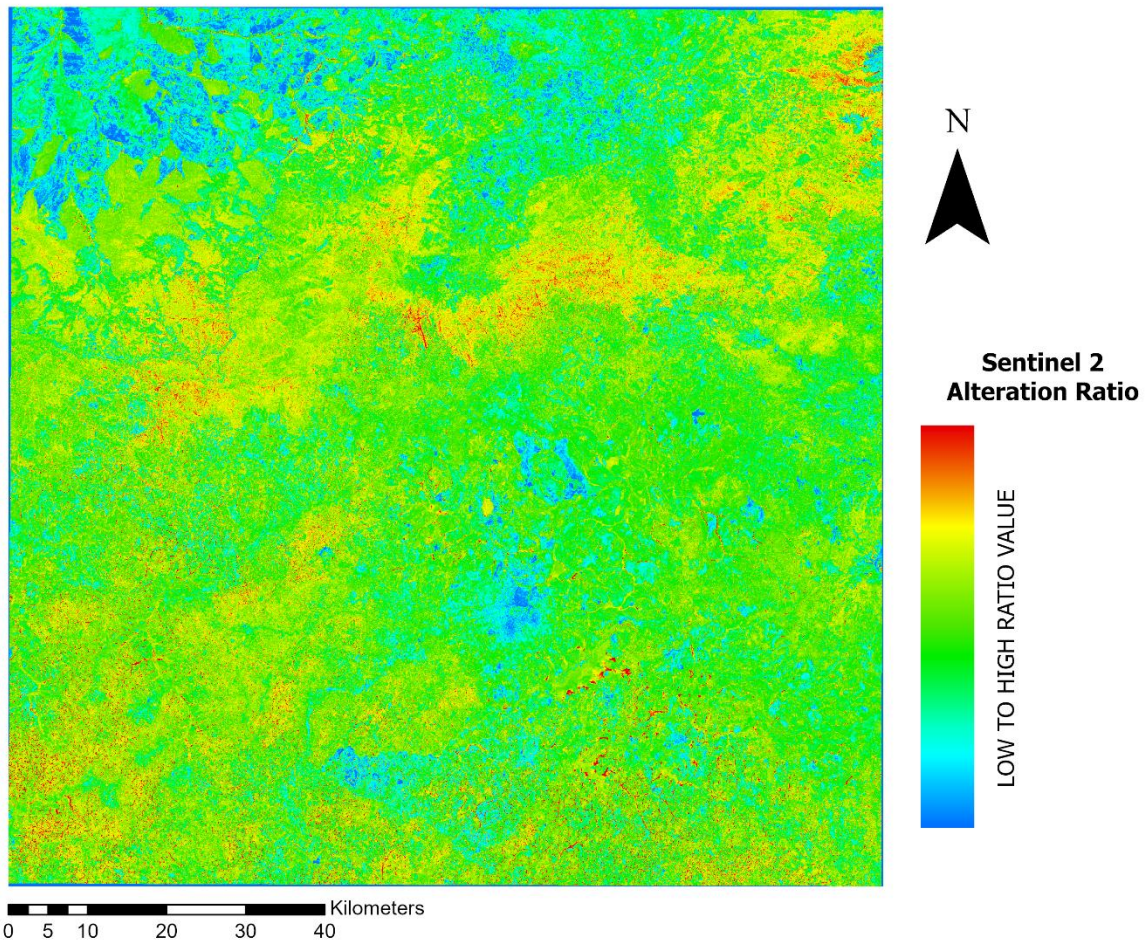


Figure 22 *Alteration Index using Sentinel-2 multispectral data showing a heterogeneous spatial distribution of altered rocks in red color*

Figure 22 showing heterogeneous spatial distribution of high alteration suggesting a complex geological history with localized alteration events. ***Figure 23*** on the other hand indicating the presence and concentration of ferric iron (map a) and ferrous silicate minerals (map b) based on Sentinel-2 satellite data. The areas with high ferric iron (map a) concentration are shown in red, particularly noticeable in the central to southern parts of the map. These could indicate the presence of oxidized iron minerals. The bright red areas signify higher concentrations of ferrous silicates (map b), which could suggest the presence of minerals like olivine or pyroxene, typically found in mafic to ultramafic rocks. The distribution of ferrous silicates is less uniform compared to ferric iron, with more scattered, localized high concentration areas. This might reflect the original distribution of mafic and ultramafic rocks or subsequent geological processes that have redistributed these minerals.

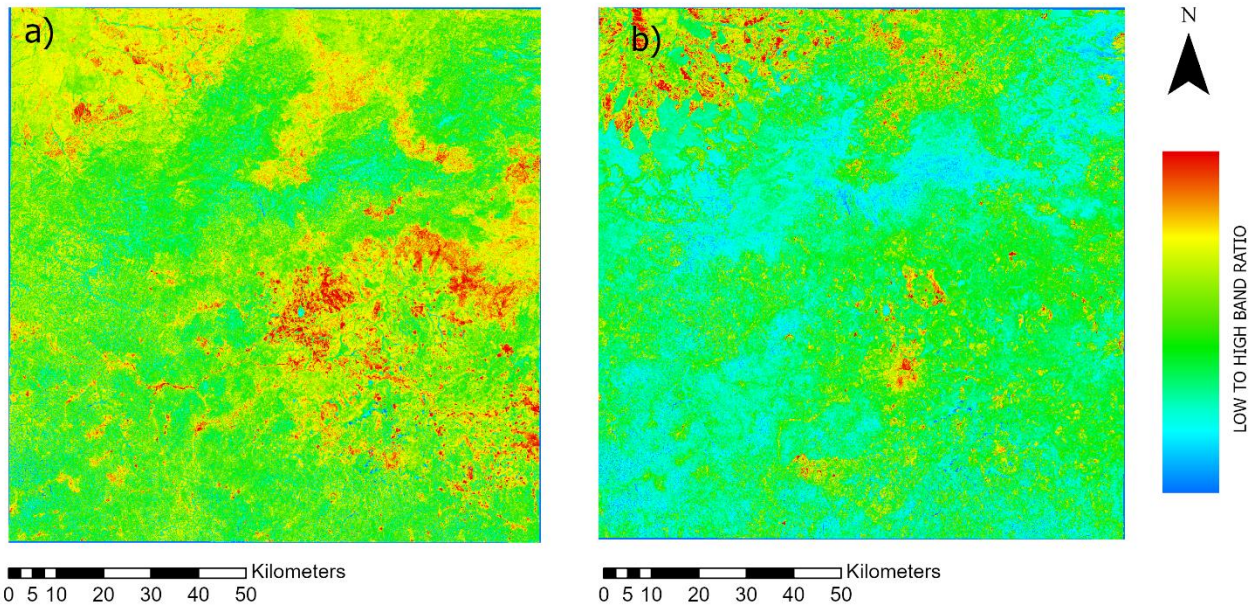


Figure 23 *a) Ferric iron Index and b) Ferrous Silicate Index obtained using Sentinel-2 multispectral data showing the green to blue areas indicating lower concentrations of ferric iron and ferrous silicates, which could correspond to areas with less weathering or the presence of rocks and minerals with low iron content*

5.1.1.2 ASTER Band Ratios

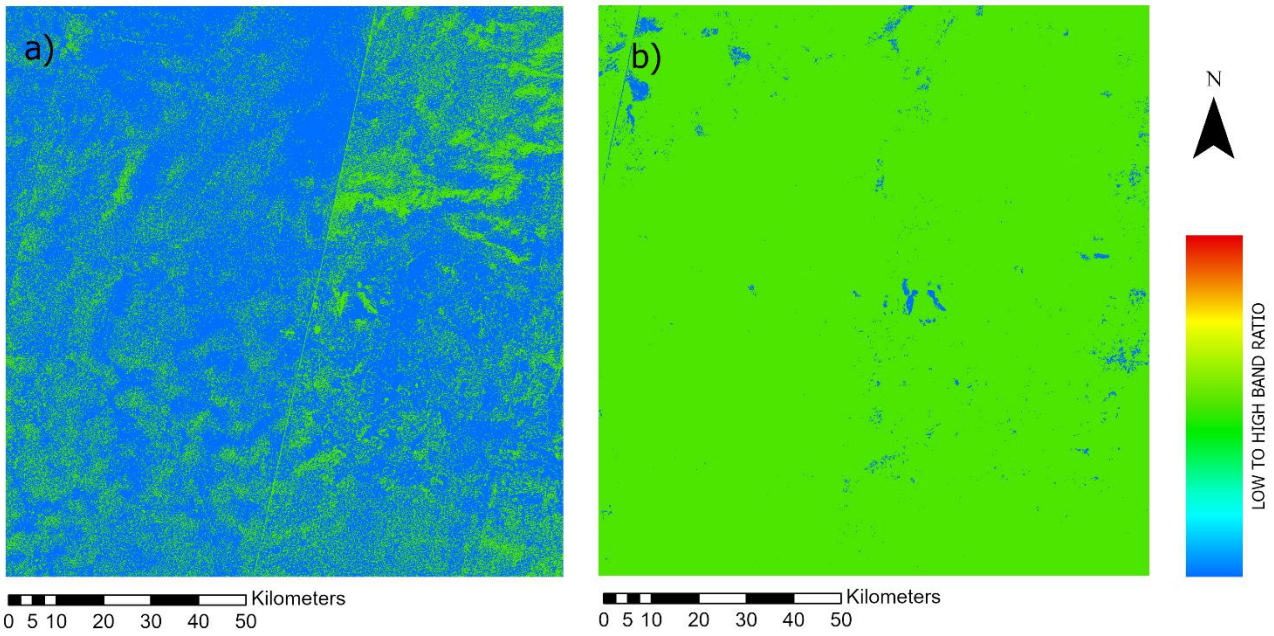
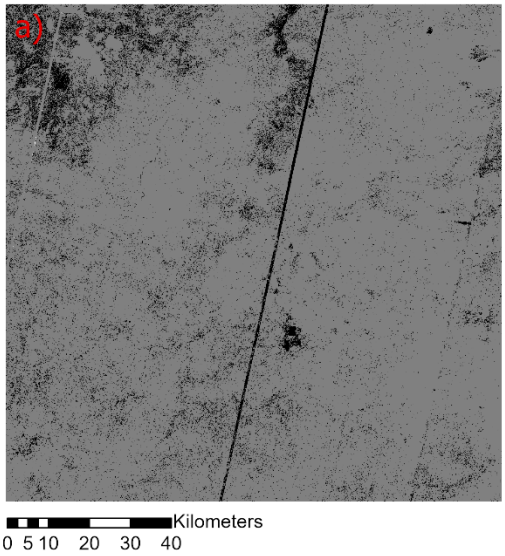
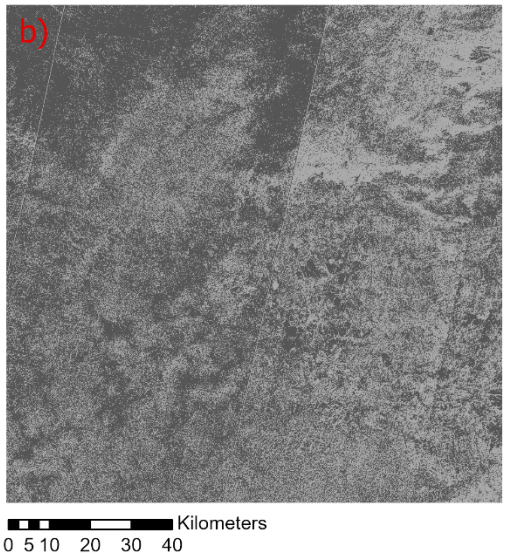
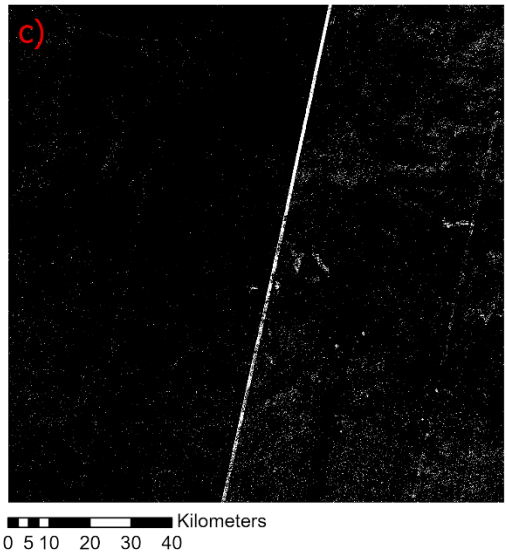


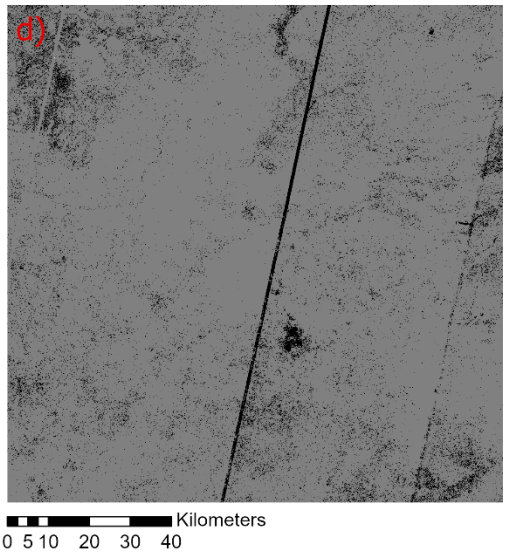
Figure 24 *a) Quartz Index and b) Alteration Index obtained using ASTER multispectral data*

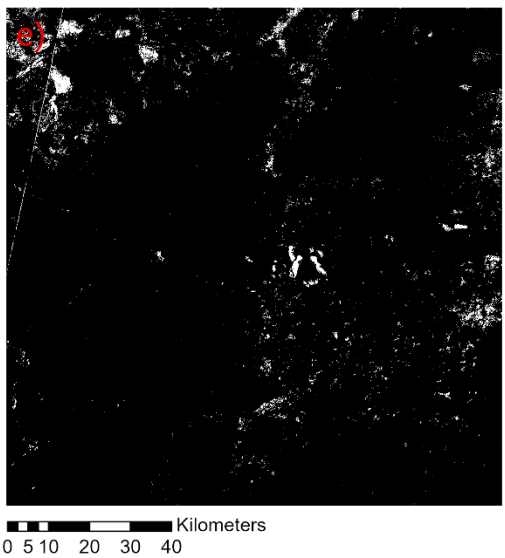
High quartz zones may be of interest for exploration, particularly for gold, as quartz veins can be associated with gold deposits. The dominant blue regions suggest a lower concentration of quartz. These could represent a variety of geological materials, from volcanic rocks without significant quartz content to soil-covered areas. The patchy distribution could also suggest differential weathering or erosion exposing quartz-rich rocks. The alteration Index ***Figure 24*** (map b) shows patches of very low alteration zones in blue and generally intermediate alteration represented by the green color. Indeed, each mineral possesses distinctive absorption characteristics within the electromagnetic spectrum. These specific regions, detectable in the visible and near-infrared (VNIR), shortwave infrared (SWIR), and thermal infrared (TIR) spectral ranges, play a crucial role in identifying minerals **(Adiri et al., 2020)**. ***Figure 25*** show the rest of indices obtained using ASTER data have a fairly homogeneous distribution across the study area.











MAP DESCRIPTION NEXT PAGE

Figure 25 *a) Amphiboles MgOH Index, b) Carbonates Index, c) Clay Index, d) Chlorites Index, and e) Ferrous Silicates Index obtained using ASTER multispectral data*

5.1.2 PCA

Typically, the initial principal component (PC1) image broadly represents albedo or intensity, while the rest of the PC images contain spectral details. The eigenvector matrix provides insight into how input spectral bands contribute to each principal component analysis (PCA) **(Gupta, 2018)**. The eigenvalues from your Principal Component Analysis (PCA) provide insight into the variance captured by each principal component:

*Eigenvalue **0**:* These components have no variance and do not contribute to the data analysis. They can be disregarded in further analysis ***(Figure 26).***

*Largest Eigenvalue **(4.101097774760777**):* This principal component captures the most variance. It's significantly larger than the others ***(Figure 27)*** , indicating that this component alone explains a substantial part of the total variance in the dataset. This component is likely to be the most significant in understanding spatial relationships related to gold.

Other Non-Zero Eigenvalues: The remaining eigenvalues (***0.7851234839676881, 0.06940444337197048, and 0.0004336744158351162***) represent smaller portions of the variance. Each successive component adds to the cumulative understanding of the data, but with diminishing returns.

*Very Small Eigenvalue (**2.317920408624025e-06**):* This value is close to zero, suggesting that the corresponding component contributes very little to the total variance and may not be useful for your analysis.

The meaningful information in my dataset is captured in the first few components. This PCA output can help in focusing the analysis on the most informative aspects of the data, especially in relation to the spatial distribution and potential correlation with gold.

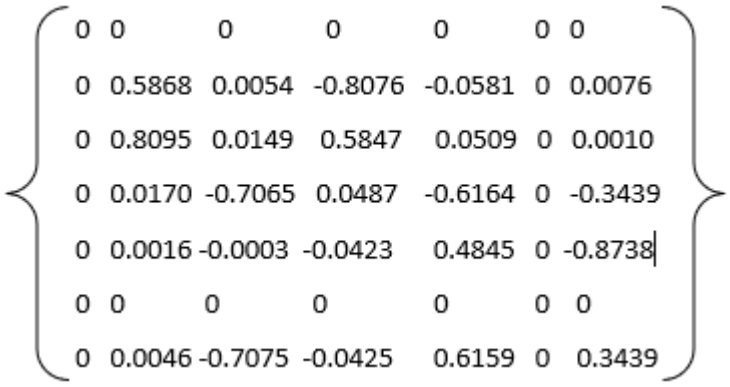


Figure 26 *Eigen matrix obtained from PCA of 7 band ratios from ASTER data namely* Alteration Index, *Amphiboles MgOH Index, Carbonates Index, Clay Index, Chlorites Index, Ferrous Silicates Index and Quartz Index*

```
0
4.101097774760777
0.7851234839676881
0.06940444337197048
0.0004336744158351162
0
2.317920408624025e-06
```

Figure 27 *Eigen values obtained from PCA of 7 band ratios from ASTER data namely* Alteration Index, *Amphiboles MgOH Index, Carbonates Index, Clay Index, Chlorites Index, Ferrous Silicates Index and Quartz Index*

5.1.3 Lineament and Density Map

The map shows the density of lineaments, which are linear or curvilinear features on the Earth's surface, often associated with fractures or faults ***(Figure 28)***. This part of the project was subjective as the lines were drawn based on visual appreciation of the different layers outlined in table 3. High-density areas (shown in red and orange) suggest a high frequency of these structural features, which could be indicative of zones of weakness that have been reactivated multiple times. These areas are often exploration targets for structurally controlled mineral deposits.

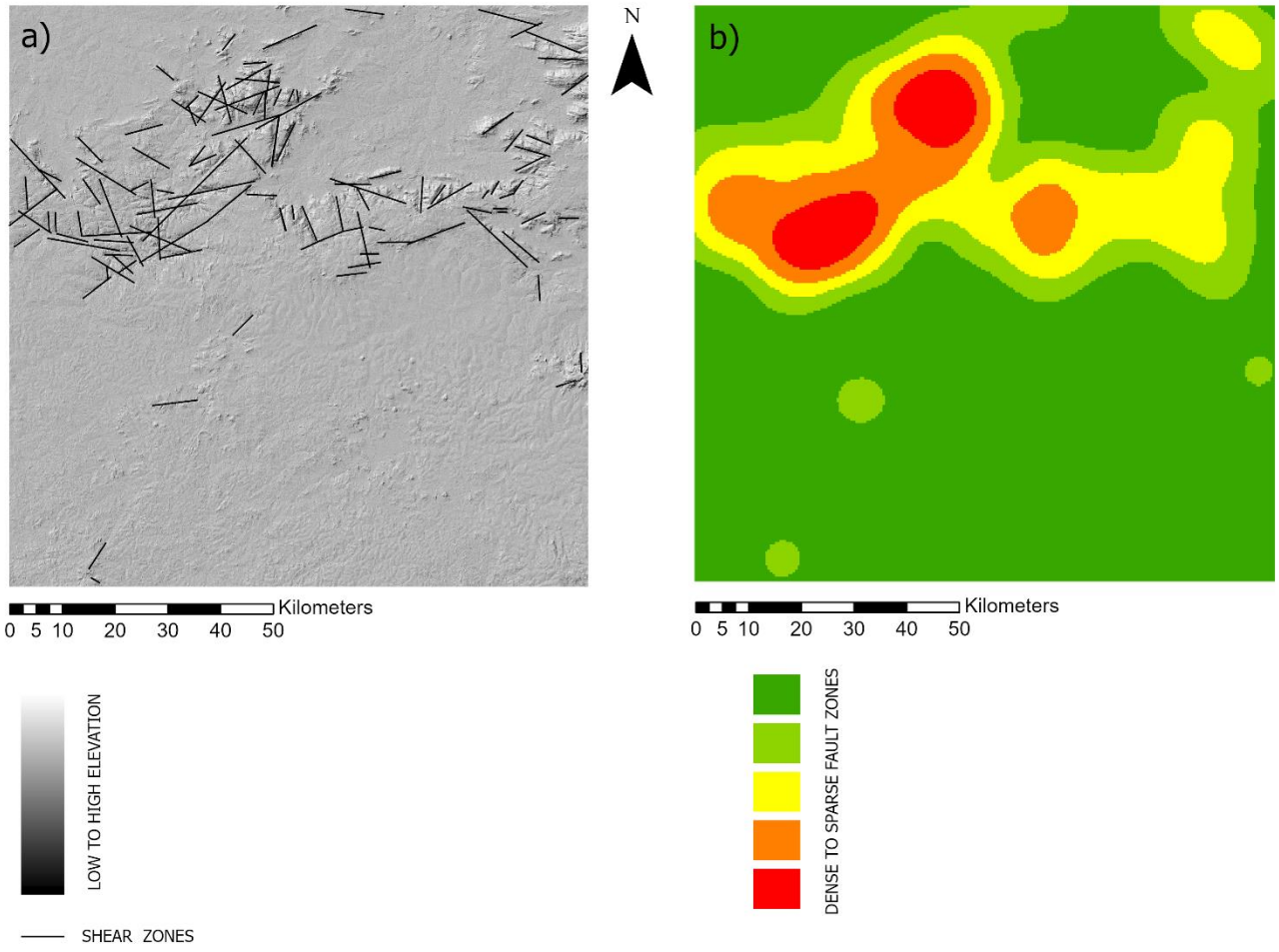


Figure 28 *a) Map of elevation model at an azimuth of 315 and altitude of 45. This also shows the derived lineaments representing shear zones. b) Density map of the lineaments*

5.2 Statistical Analysis

The Average Nearest Neighbor Summary results indicating significant clustering (with a z-score of ***-25.771228***) align with the findings from the Local Moran's I analysis, which also suggested clustering or spatial autocorrelation within the study area. The significant negative z-score from the Average Nearest Neighbor analysis indicates that the spatial pattern of features is much more clustered than would be expected if the features were randomly distributed. If the distribution were random, the observed mean distance between features is significantly less ***(966.4086 meters)*** than the expected mean distance ***(1568.3222 meters)***. This supports the Local Moran's I findings of spatial autocorrelation, where a positive Moran's I value, and a significant p-value indicated non-random clustering of similar values within the area. Both statistics underline a significant departure from spatial randomness, reinforcing the conclusion that the observed patterns are unlikely to be due to chance.

For the Local Moran, I had the following values as per the report attached in the appendix.
P-value of ***0.000626***: This measures the statistical significance of your Moran's I result. A p-value this low (less than ***0.05***) is typically considered statistically significant, suggesting that the observed spatial pattern (the weak clustering indicated by your Moran's I value) is unlikely to have occurred by chance. In other words, there is strong evidence to support the presence of spatial autocorrelation in your data.

Z-score of ***3.420252***: The Z-score indicates how many standard deviations the observed Moran's I value is from the expected value (assuming no spatial autocorrelation). A Z-score of ***3.42***, which is more than three standard deviations away, suggests that the observed clustering is statistically significant. In practice, Z-scores above +1.96 or below -1.96 (corresponding to a 95% confidence level) are typically considered significant.

The analysis indicated a statistically significant, albeit weak, spatial autocorrelation in the study area, with a slight tendency for similar values to cluster together. The low p-value and the relatively high Z-score strengthen the confidence in these results, suggesting that the observed pattern is unlikely to be due to random chance.

The document's Global Moran's I analysis indicates varying degrees of spatial autocorrelation across different distances within the study area. Key insights include:

The Moran's Index shows values ranging from ***0.020197*** to ***0.044103*** across different distances, indicating varying but generally weak spatial autocorrelation strength. These values were all positive, suggesting some degree of clustering rather than dispersion.

The z-scores range significantly, with the maximum peak at a distance of ***13,978.27*** meters and a z-score of ***9.086539***, which is highly significant statistically. This suggested that the spatial autocorrelation is substantial and statistically significant at certain distances.

P-values were very low (many essentially 0), especially at more considerable distances, indicating vital statistical significance of the observed spatial patterns at those scales.

These details suggested that spatial autocorrelation exists and varies with distance, which is crucial for performing kriging. Specifically, the variogram used in kriging, which models the degree and scale of spatial autocorrelation, can be more accurately parameterized with this information, leading to more precise interpolation results.

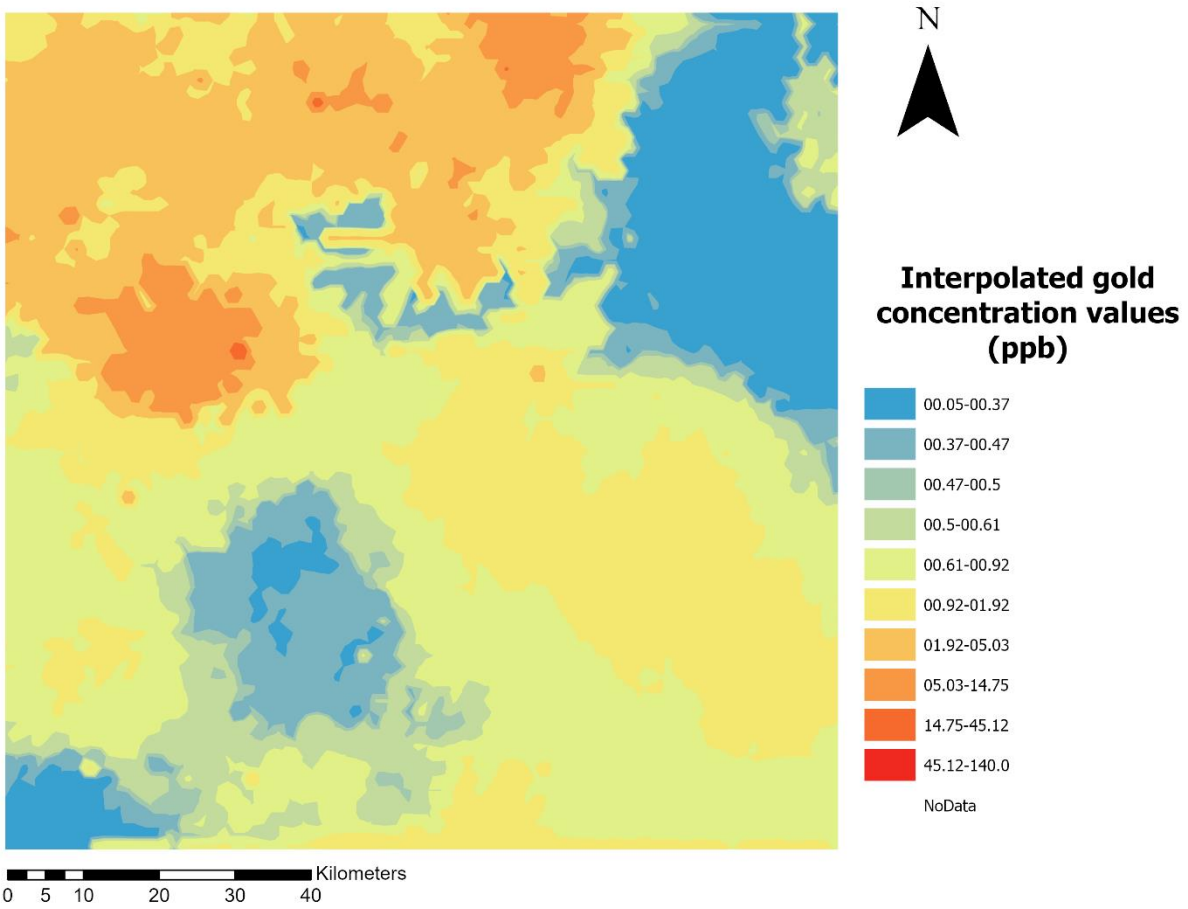


Figure 29 *Map of outcome from interpolation using ordinary Kriging showing interpolated gold concentration values in parts per billion*

This information is crucial for understanding spatial relationships and can guide targeted interventions or further research. The interpolation map obtained ***(Figure 29)*** also indicates that the transition between different concentrations appears smooth, suggesting a gradual change in gold distribution rather than abrupt shifts. This indicates the interpolation method's effectiveness.

5.3 Machine Learning

Visualizing prediction results spatially provides a direct and actionable insight into potential exploration targets. The models were evaluated based on their ability to predict gold mineralization zones accurately. Model performance was assessed using cross-validation techniques, focusing on metrics such as the Root Mean Square Error (RMSE) and the coefficient of determination (R^2). This approach allowed for identifying the most compelling feature set and algorithm combination in predicting potential gold mineralization areas. The different results obtained will be discussed below. In each of the tables, the following parameters will be displayed;

Random Forest Model

The optimum performance model's parameters include:

num.trees parameter = Number of trees

mtry parameter = number of variables tried at each split

Gradient Boosting Machines Model

The optimum performance model's parameters include:

eta = learning rate values (Determines the contribution of each tree on the final outcome and controls how quickly the algorithm proceeds down the gradient descent (learns))

subsample = Subsample (proportion of training data (rows) available for creation of each tree)

colsample_bytree = Subsample (proportion of training data (columns) available for creation of each tree)

Both models' performance is given by:

regr.rmse = RMSE (Root Mean Square Error)

regr.rsq = RSQ (R Squared)

Sentinel-2 derived Indices plus fault density variables: The RMSE is lower for the GBM model, which suggests that it may have a better fit for the data compared to the RF model. However, the RSQ values for both models are low (and even negative for the RF model), which usually indicates a poor fit. In the context of spatial data, traditional RSQ may not always be the best metric due to spatial autocorrelation ***Tables 5 and 6)***.

mtry	num.trees	learner_param_vals	x_domain	regr.rmse	regr.rsq
<int>	<chr>	<list>	<list>	<dbl>	<dbl>
1	300	1, 1, 300	1, 300	5.162218	-0.1111268

Table 5 Parameters and performance of optimum RF for Sentinel-2 data integrated with density map from lineation

eta	subsample	colsample_bytree	learner_param_vals	x_domain	regr.rmse	regr.rsq
<dbl>	<chr>	<dbl>	<list>	<list>	<dbl>	<dbl>
0.4132110	0.5	0.8400521	2 , 4 , 1 , none , 0.413210980966687, 0.5 , 0.840052146650851	0.4132110, 0.5000000, 0.8400521	4.676102	-0.03213545
0.3120135	0.5	0.9353159	2 , 4 , 1 , none , 0.312013527425006, 0.5 , 0.935315885953605	0.3120135, 0.5000000, 0.9353159	4.719725	0.04150138

Table 6 Parameters and performance of optimum GBM for Sentinel-2 data integrated with density map from lineation

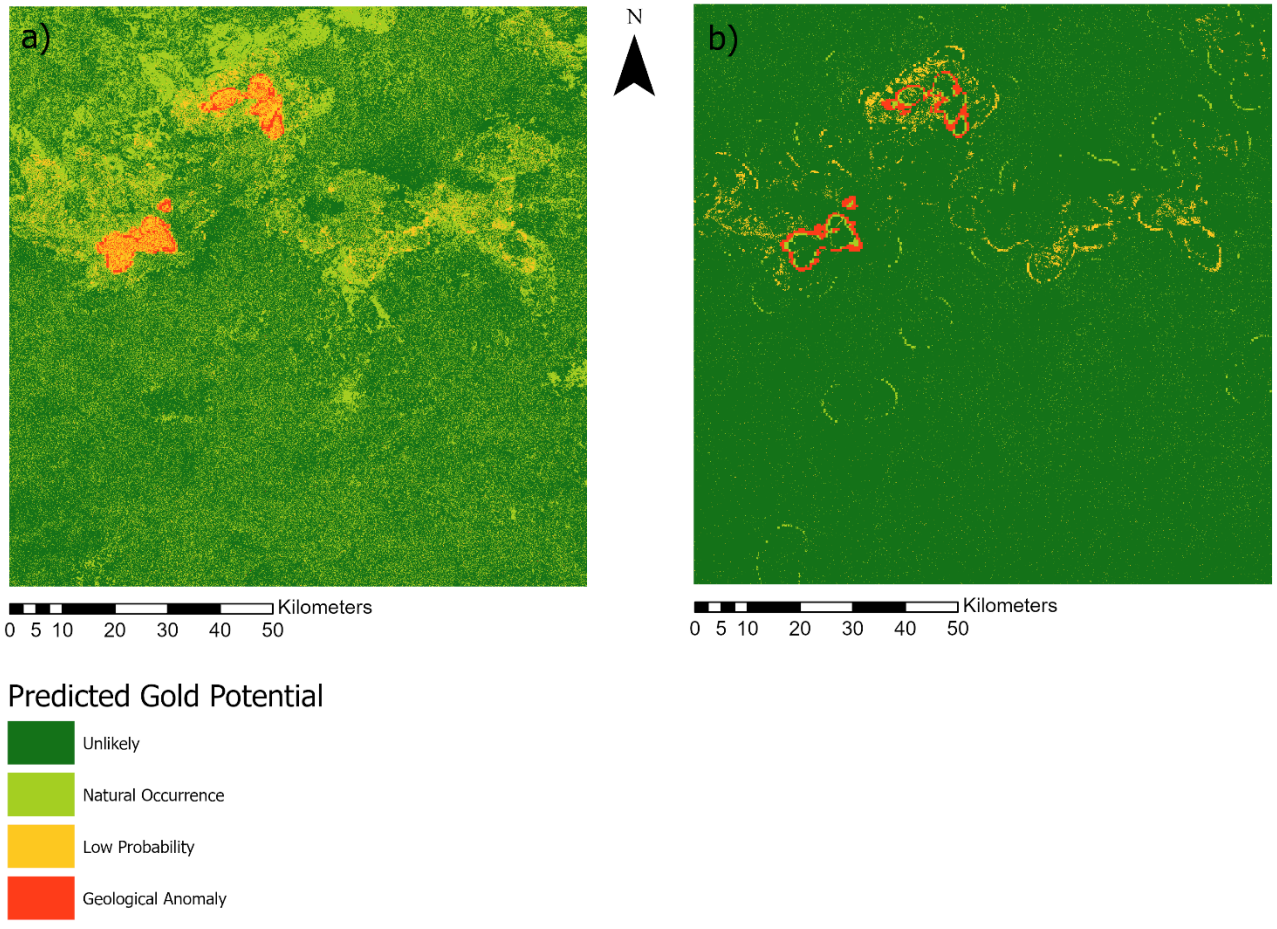


Figure 30 *Maps of gold potential prediction from Sentinel-2 and DEM derived data using a) Random Forest Model and b) Gradient Boosting Machines Model. The natural occurrence value considered is 5 ppb* ***(Weick, 1994)***

The RF model seems to predict a higher number of potential areas, which might be useful for exploratory purposes but could result in higher false positives. The GBM model appears to be more selective, which could reduce the number of false positives but at the risk of missing potential sites ***(Figure 30)***.

ASTER derived Indices plus fault density variables: The RSQ values for both models are low. The GBM model has a better RMSE, suggesting it may be more accurate in its predictions ***(Tables 7 and 8)***.

mtry	num.trees	learner_param_vals	x_domain	regr.rmse	regr.rsq
<int>	<chr>	<list>	<list>	<dbl>	<dbl>
1	200	1, 1, 200	1, 200	4.783133	-0.04741318

Table 7 Parameters and performance of optimum RF for ASTER data integrated with density map from lineation

eta	subsample	colsample_bytree	learner_param_vals	x_domain	regr.rmse	regr.rsq
<dbl>	<chr>	<dbl>	<list>	<list>	<dbl>	<dbl>
0.4392394	0.5	0.6796965	2 , 4 , 1 , none , 0.439239399880171, 0.5 , 0.679696505749598	0.4392394, 0.5000000, 0.6796965	4.530639	0.1258989

Table 8 Parameters and performance of optimum GBM for ASTER data integrated with density map from lineation

The negative RSQ for the RF model suggests that it may not be capturing the complexity of the data well, and further refinement of the model or the inclusion of additional data might be necessary. The positive RSQ for the GBM model, although low, indicates that there is some explanatory power that might be improved with further tuning.

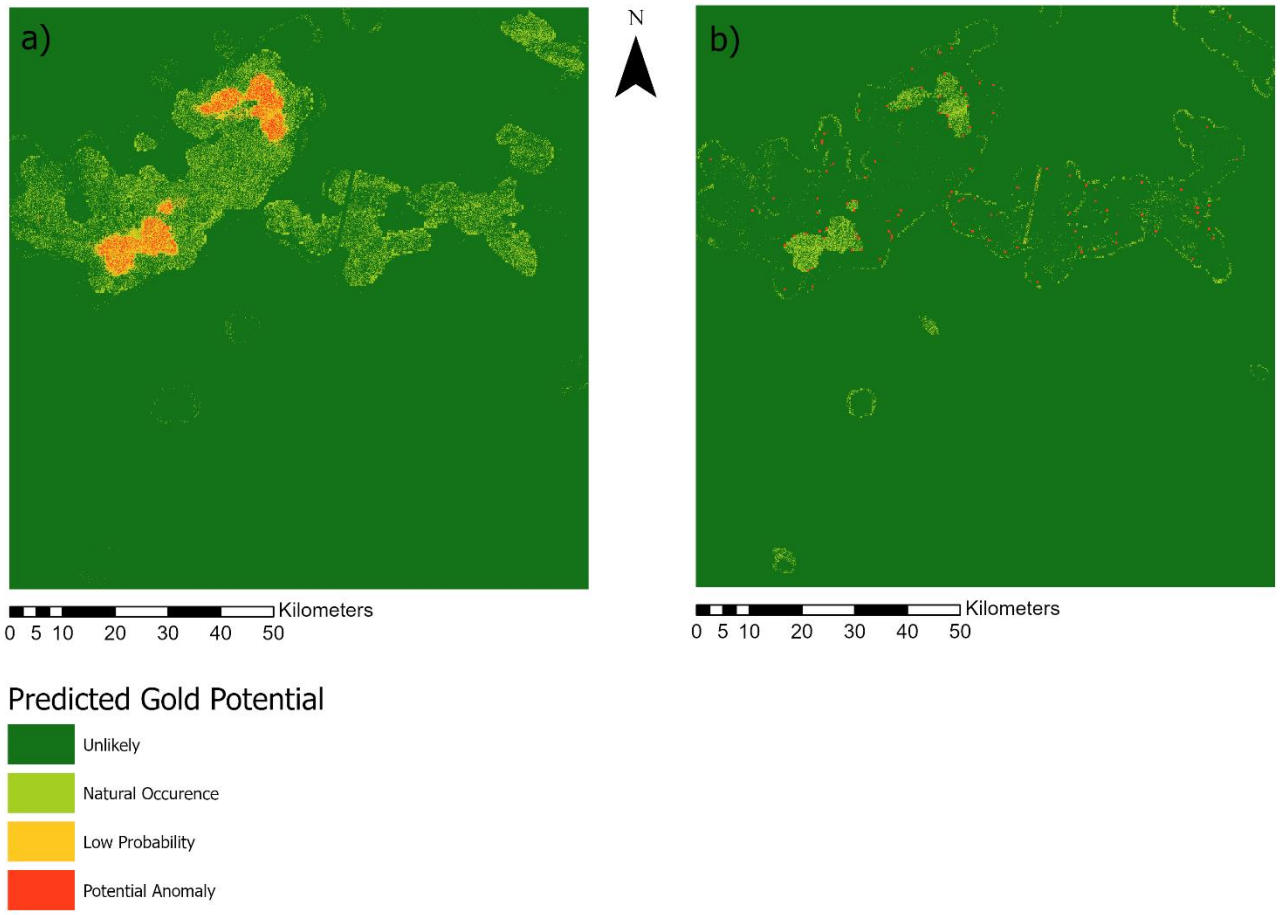


Figure 31 *Maps of gold potential prediction from ASTER and DEM derived data using a) Random Forest Model and b) Gradient Boosting Machines Model. The natural occurrence value considered is 5 ppb* ***(Weick, 1994)***

Sentinel-2 and ASTER derived Indices plus fault density variables: ***Table 9 and 10*** show the performance of data from both satellites.The RMSE is 4.781339, which suggests the model's predictions are on average about 4.78 units away from the actual values. The RSQ is -0.03303724, which is negative, indicating that the model's predictions are not better than a horizontal line fit to the data points. For GBM, the RMSE is 4.578339, which is slightly lower than the RF model, and the RSQ is 0.1101183, which is positive, indicating the model has some predictive power. The GBM model appears to perform better than the RF model in terms of RMSE and RSQ, indicating more accurate and reliable predictions.

mtry	num.trees	learner_param_vals	x_domain	regr.rmse	regr.rsq
<int>	<chr>	<list>	<list>	<dbl>	<dbl>
1	200	1, 1, 200	1, 200	4.781339	-0.03303724

Table 9 Parameters and performance of optimum RF for Sentinel-2 and ASTER data integrated with density map from lineation

eta	subsample	colsample_bytree	learner_param_vals	x_domain	regr.rmse	regr.rsq
<dbl>	<chr>	<dbl>	<list>	<list>	<dbl>	<dbl>
0.4949585	0.5	0.4545542	2 , 4 , 1 , none , 0.494958492694423, 0.5 , 0.454554151277989	0.4949585, 0.5000000, 0.4545542	4.533830	0.0585390
0.3329782	0.75	0.7544629	2 , 4 , 1 , none , 0.332978229271248, 0.75 , 0.754462919011712	0.3329782, 0.7500000, 0.7544629	4.578339	0.1101183

Table 10 Parameters and performance of optimum GBM for Sentinel-2 and ASTER data integrated with density map from lineation

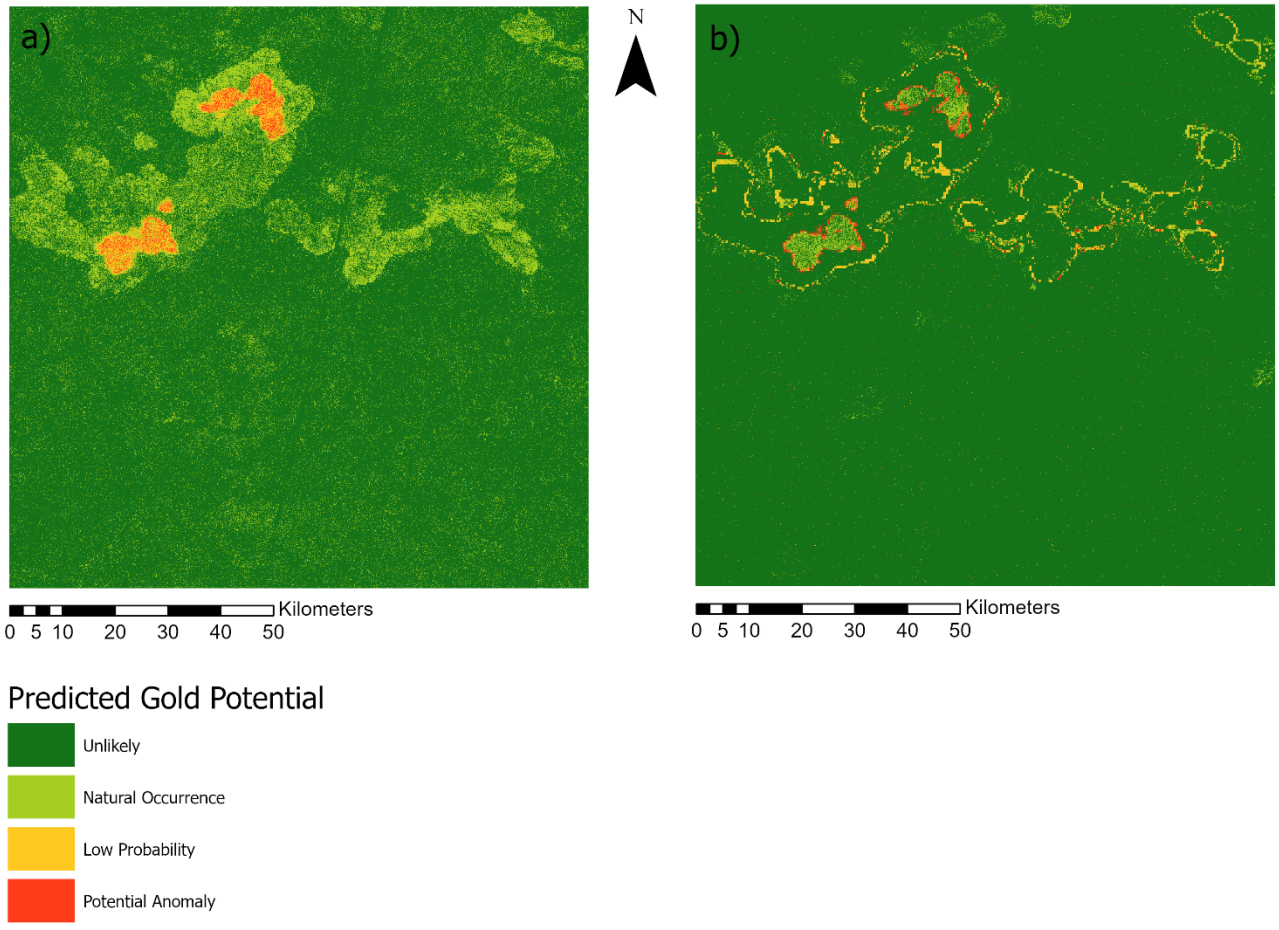


Figure 32 *Maps of gold potential prediction from Sentinel-2 and ASTER and DEM derived data using a) Random Forest Model and b) Gradient Boosting Machines Model. The natural occurrence value considered is 5 ppb* ***(Weick, 1994)***

The inclusion of a larger number of variables (combining Sentinel-2 and ASTER data) may have contributed to the improved performance of the GBM model. This indicates that a multi-sensor approach can enhance model predictions.

Map (a) ***(Figure 32)*** derived from RF Shows extensive areas of Natural Occurrence and Potential Anomaly, suggesting a liberal prediction model that could be useful in exploratory stages to avoid missing potential sites. Map (b) on the other hand from GBM indicates more selective areas with Potential Anomaly, which may represent a more targeted approach, potentially leading to more efficient allocation of exploration resources.

PCA derived data (PC2 and PC3) plus fault density variables: The negative RSQ for the RF model ***(Table 11)*** suggests that it may not be capturing the underlying patterns in the data effectively for this particular problem. In contrast, the GBM model ***(Table 12)***, while having a low RSQ, shows some ability to predict variations in gold potential, which may be attributed to its iterative nature that focuses on correcting previous errors. The use of PCA-derived variables (PC2 and PC3) indicates an attempt to reduce dimensionality and possibly collinearity in the predictor variables. These components should capture the most variance in the mineral indices, but whether they capture the most predictive variance for gold potential is not guaranteed. The map is shown on ***Figure 33***.

mtry	num.trees	learner_param_vals	x_domain	regr.rmse	regr.rsq
<int>	<chr>	<list>	<list>	<dbl>	<dbl>
1	150	1, 1, 150	1, 150	5.157075	-0.1896573

Table 11 Parameters and performance of optimum RF for PC2 and PC3 data integrated with density map from lineation

eta	subsample	colsample_bytree	learner_param_vals	x_domain	regr.rmse	regr.rsq
<dbl>	<chr>	<dbl>	<list>	<list>	<dbl>	<dbl>
0.3811527	0.5	0.9393147	2 , 4 , 1 , none , 0.381152716465294, 0.5 , 0.939314735424705	0.3811527, 0.5000000, 0.9393147	4.574129	0.1043934

Table 12 Parameters and performance of optimum GBM for PC2 and PC3 data integrated with density map from lineation

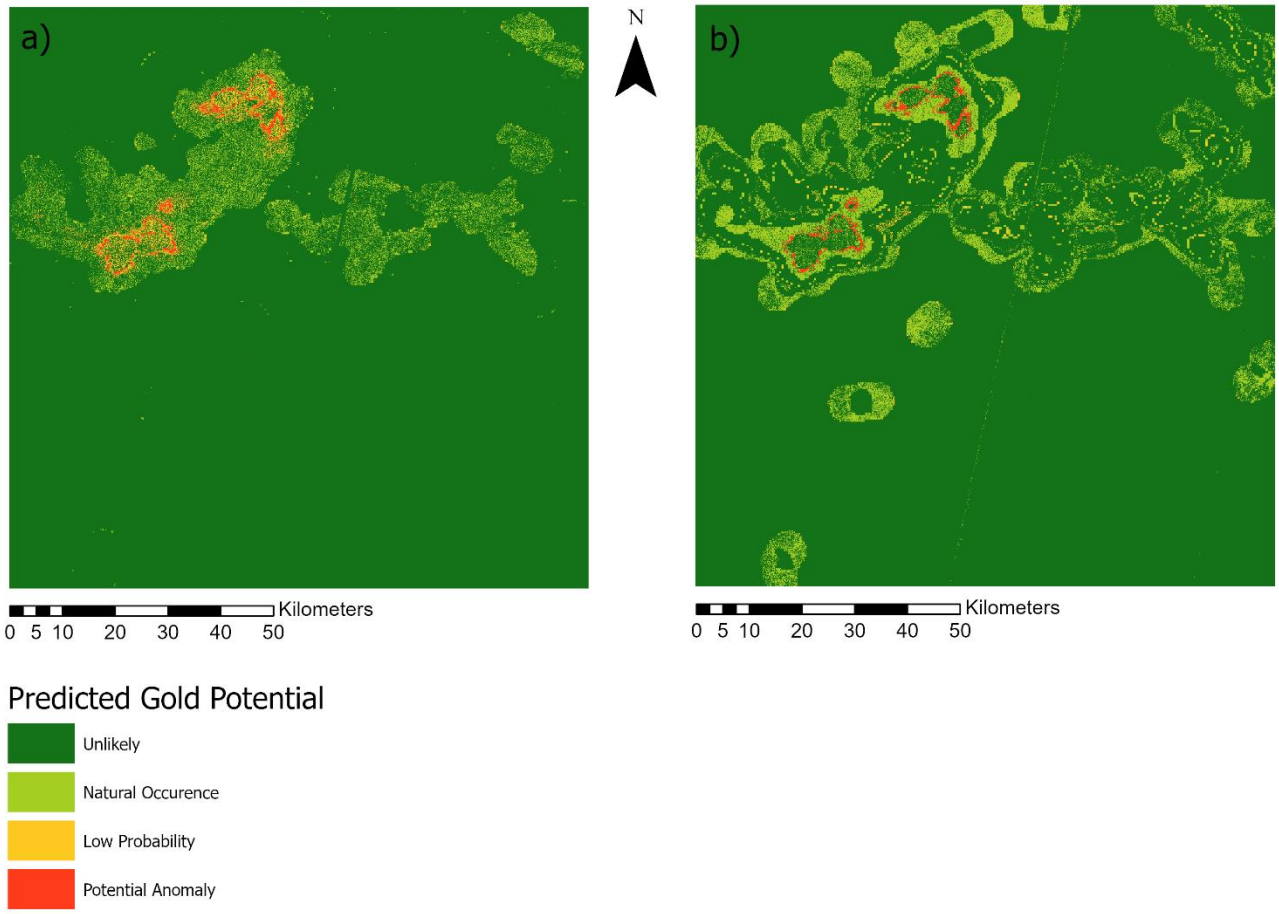


Figure 33 *Maps of gold potential prediction from PC2 and PC3 and DEM derived data using a) Random Forest Model and b) Gradient Boosting Machines Model. The natural occurrence value considered is 5 ppb* ***(Weick, 1994)***

Summary Statistics

Based on these numbers, the Gradient Boosting Machine model using ASTER data, which has an **RMSE of 4.530639** and an **RSQ of 0.1258989**, appears to be the best performing model among those listed. It has the lowest RMSE, indicating the smallest average errors in prediction, and the highest RSQ, indicating the best fit among the models compared. The full summary is displayed on ***Table 13.*** A map of the best performing model is displayed side-by-side with the interpolation map for visual comparison in ***Figure 34***. The outcome is discussed in the discussion and conclusion chapter.

	Model	RMSE	RSQ
GBM Models	GBM with ASTER	4.530639	0.12589890
	GBM with PCA	4.574129	0.10439340
	GBM with Sentinel-2 and ASTER	4.578339	0.11011830
	GBM with Sentinel	4.719725	0.04150138
RF Models	RF with Sentinel-2 and ASTER	4.781339	-0.03303724
	RF with ASTER	4.783133	-0.04741318
	RF with Sentinel-2	5.162218	-0.11126800
	RF with PCA	5.157075	-0.18965730

Table 13 Summary of machine learning models performance. Best performing model is in green shade. The best performing random forest is in sky blue shade

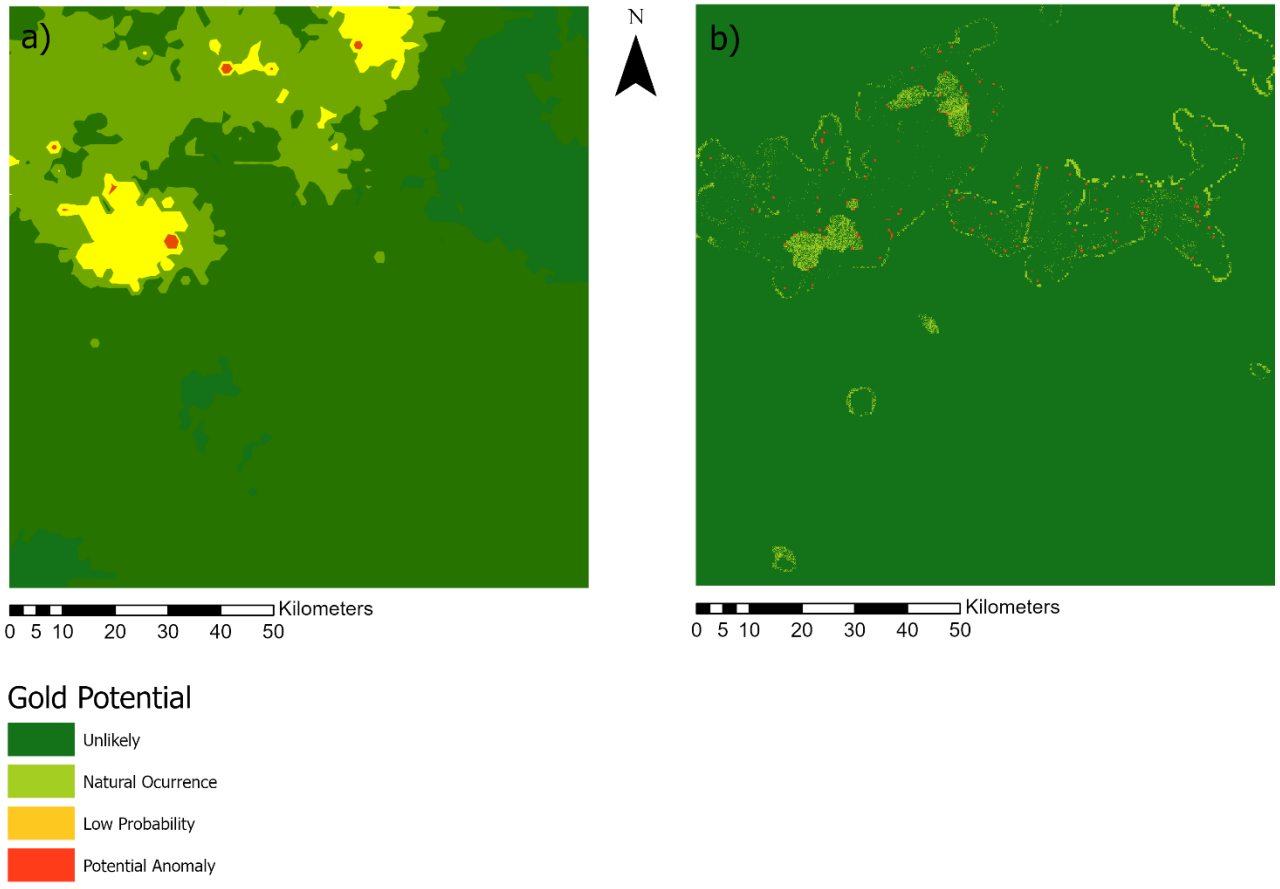


Figure 34 *Maps of a) Interpolated gold with symbology adjusted to adjusted to same symbology as best performing model b)*

6 DISCUSSIONS AND CONCLUSION

Identifying promising areas within a specific region is a critical step in mineral exploration **(Abdolmaleki et al., 2020)**. While not all alteration is linked to ore bodies, and not all ore bodies exhibit alteration, altering rocks is a valuable indicator of potential deposits **(Sabins, 1999)**.
Using satellite images does a great job of addressing these issues, amongst others. Satellites have the advantage of offering a non-destructive and non-invasive form of studying the places of interest. They also reduce risk by prioritizing exploration zones, reducing the chances of unsuccessful drill campaigns.

This research was able to demonstrate a response to the research objectives which were to ascertain if remote sensing can be applied in Africa particularly in the North of Cameroon and the possibility of getting a prediction model that can be used for gold exploration. This is evidenced by visual similarity observed in the different maps produced. Clearly, the Northeastern part of the study area seems to be an interesting zone and a potentially high anomaly zone. While on their own the band ratios ***(Figure 22 to Figure 25)*** were not able to provide some collective clarity on a decent zone of interest, they collectively displayed the Northeast as a good location when applied to the model. Part of the reason could be that alteration process in the study area is not happening in a homogeneous way, exposure of the rock might be playing a role to this. Soil formation processes and vegetation also have an impact.

While the RMSE and RSQ values provide a quantitative measure of model performance, the real value comes from how these models help focus exploration efforts. Both models offer valuable insights, but their performance and the prediction patterns suggest they may be suitable for different phases of the exploration process. It might be beneficial to use both models in a complementary manner. The RF model could be used to establish broad areas of interest, while the GBM model could refine these predictions to prioritize exploration efforts. The delivery of outputs in a geospatial TIFF format further enhances the applicability of the research findings in the practical context of mineral exploration and geological studies.

There are many ways this work can be improved upon, not just in fine tuning of the model. Coupling surface information from remote sensing imagery with subsurface information from geophysical data has been underexploited by the geologic remote sensing community **(Van Der Meer et al., 2012)**. In integrated geo-investigations, geological data serves as a crucial input. This data can consist of descriptions of lithology or structure at specific points or in the form of maps **(Gupta, 2018)**. The absence of depth penetration in

optical remote sensing data underscores the significance of conducting integrated interpretation alongside datasets that offer depth-related information, such as geophysical data **(Abdolmaleki et al., 2020)**.

In summary, remote sensing, when properly employed, is a valuable technical resource for mineral exploration **(Sabins, 1999)**. Integrating cutting-edge satellite remote sensing into mineral exploration strategies aligns with these objectives and offers a promising avenue for sustainable and efficient resource management.

7 BIBLIOGRAPHIES

Abdolmaleki, M., Rasmussen, T. M., & Pal, M. K. (2020). EXPLORATION OF IOCG MINERALIZATIONS USING INTEGRATION OF SPACE-BORNE REMOTE SENSING DATA WITH AIRBORNE GEOPHYSICAL DATA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2020*, 9–16. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-9-2020

Adiri, Z., Lhissou, R., El Harti, A., Jellouli, A., & Chakouri, M. (2020). Recent advances in the use of public domain satellite imagery for mineral exploration: A review of Landsat-8 and Sentinel-2 applications. *Ore Geology Reviews*, *117*, 103332. https://doi.org/10.1016/j.oregeorev.2020.103332

African Union. (2009). Africa Mining Vision. Addis Ababa: African Union.

Akwinga, A. V. (2010). *Lode gold mineralisation in the Neoproterozoic granitoids of Batouri, southeastern Cameroon*. Papierflieger.

ASTER_Processing_for_Mineral_Exploration.pdf. (n.d.).

Azeuda Ndonfack, K. I., Xie, Y., Goldfarb, R., Zhong, R., & Qu, Y. (2021). Genesis and mineralization style of gold occurrences of the Lower Lom Belt, Bétaré Oya district, eastern Cameroon. *Ore Geology Reviews*, *139*, 104586. https://doi.org/10.1016/j.oregeorev.2021.104586

Bedini, E. (2011). Mineral mapping in the Kap Simpson complex, central East Greenland, using HyMap and ASTER remote sensing data. *Advances in Space Research*, *47*(1), 60–73. https://doi.org/10.1016/j.asr.2010.08.021

Bradley, B & Brandon, G (2020, February 01) *Hands-On Machine Learning with R* https://bradleyboehmke.github.io/HOML/

Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., Perrotta, M., & Roda-Robles, E. (2020). Detecting Lithium (Li) Mineralizations from Space: Current Research and Future Perspectives. *Applied Sciences*, *10*(5), 1785. https://doi.org/10.3390/app10051785

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Gupta, R. P. (2018). *Remote Sensing Geology*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-55876-8

Hilson, G. (2020). The Africa Mining Vision: A manifesto for more inclusive extractive industry-led development? *Canadian Journal of Development Studies / Revue Canadienne d'études Du Développement*, *41*(3), 417–431. https://doi.org/10.1080/02255189.2020.1821352

Karimzadeh, S., & H. Tangestani, M. (2022). Potential of Sentinel-2 MSI data in targeting rare earth element (Nd3+) bearing minerals in Esfordi phosphate deposit, Iran. *The Egyptian Journal of Remote Sensing and Space Science*, *25*(3), 697–710. https://doi.org/10.1016/j.ejrs.2022.04.001

Khan, K., Ahmad, W., Amin, M. N., Ahmad, A., Nazar, S., & Alabdullah, A. A. (2022). Compressive Strength Estimation of Steel-Fiber-Reinforced Concrete and Raw Material Interactions Using Advanced Algorithms. *Polymers*, *14*(15), 3065. https://doi.org/10.3390/polym14153065

Kpeou, J., Béziat, D., Salvi, S., Estrade, G., Moloto-A-Kenguemba, G., & Debat, P. (2020). Gold mineralization related to Proterozoic cover in the Congo craton (Central African Republic): A consequence of Panafrican events. *Journal of African Earth Sciences*, *166*, 103825. https://doi.org/10.1016/j.jafrearsci.2020.103825

Leverington, D. (2011). *Remote Sensing for Geological Mapping and Mineral Exploration – Spectral Imaging*. Auracle Geospatial Science, Inc.

Macheyeki, A. S., Li, X., Kafumu, D. P., & Yuan, F. (2020). Elements of exploration geochemistry. In *Applied Geochemistry* (pp. 1–43). Elsevier. https://doi.org/10.1016/B978-0-12-819495-9.00001-3

Ngatcha, R. B., Okunlola, O. A., Suh, C. E., Ateh, K. I., & Hofmann, A. (2019). Petrochemical characterization of Neoproterozoic Colomine granitoids, SE Cameroon: Implications for gold mineralization. *Lithos*, *344–345*, 175–192. https://doi.org/10.1016/j.lithos.2019.06.028

Ngon, G. F. N., Abomo, P. S., Mbog, M. B., Bitchong, A. M., Mbaï, J. S., Minyemeck, T. V. N., & Fouateu, R. Y. (2015). Geological, Mineralogical and Geochemical Studies of Pyrite Deposits in the Eastern Part of Douala Sub-Basin (Cameroon, Central Africa). *International Journal of Geosciences*, *06*(08), 882–893. https://doi.org/10.4236/ijg.2015.68072

Nzulu, Gabriel Kofi (2023). *Characterization of Pathfinders and Indicators of Gold*, https://doi.org/10.3384/9789180750462

Pour, A. B., & Hashim, M. (2012). The application of ASTER remote sensing data to porphyry copper and epithermal gold deposits. *Ore Geology Reviews*, *44*, 1–9. https://doi.org/10.1016/j.oregeorev.2011.09.009

R. James Weick (1994) *Gold Occurrences* https://www.gov.nl.ca/iet/mines/publicoutreach/minerals/gold/

Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore Geology Reviews*, *14*(3–4), 157–183. https://doi.org/10.1016/S0169-1368(99)00007-4

Salehi, S. (2018). *PhD Thesis: Potentials and challenges for hyperspectral mineral mapping in the Arctic Developing innovative strategies for data acquisition and integration*. https://doi.org/10.13140/RG.2.2.24369.63841

Takodjou Wambo, J. D., Pour, A. B., Ganno, S., Asimow, P. D., Zoheir, B., Salles, R. D. R., Nzenti, J. P., Pradhan, B., & Muslim, A. M. (2020). Identifying high potential zones of gold mineralization in a sub-tropical region using Landsat-8 and ASTER remote sensing data: A case study of the Ngoura-Colomines goldfield, eastern Cameroon. *Ore Geology Reviews*, *122*, 103530. https://doi.org/10.1016/j.oregeorev.2020.103530

Toteu, S. F., Penaye, J., & Djomani, Y. P. (2004). Geodynamic evolution of the Pan-African belt in central Africa with special reference to Cameroon. *Canadian Journal of Earth Sciences*, *41*(1), 73–85. https://doi.org/10.1139/e03-079

Toteu, S. F., Van Schmus, W. R., Penaye, J., & Michard, A. (2001). New U–Pb and Sm–Nd data from north-central Cameroon and its bearing on the pre-Pan African history of central Africa. *Precambrian Research*, *108*(1–2), 45–73. https://doi.org/10.1016/S0301-9268(00)00149-2

Van Der Meer, F. D., Van Der Werff, H. M. A., & Van Ruitenbeek, F. J. A. (2014). Potential of ESA's Sentinel-2 for geological applications. *Remote Sensing of Environment*, *148*, 124–133. https://doi.org/10.1016/j.rse.2014.03.022

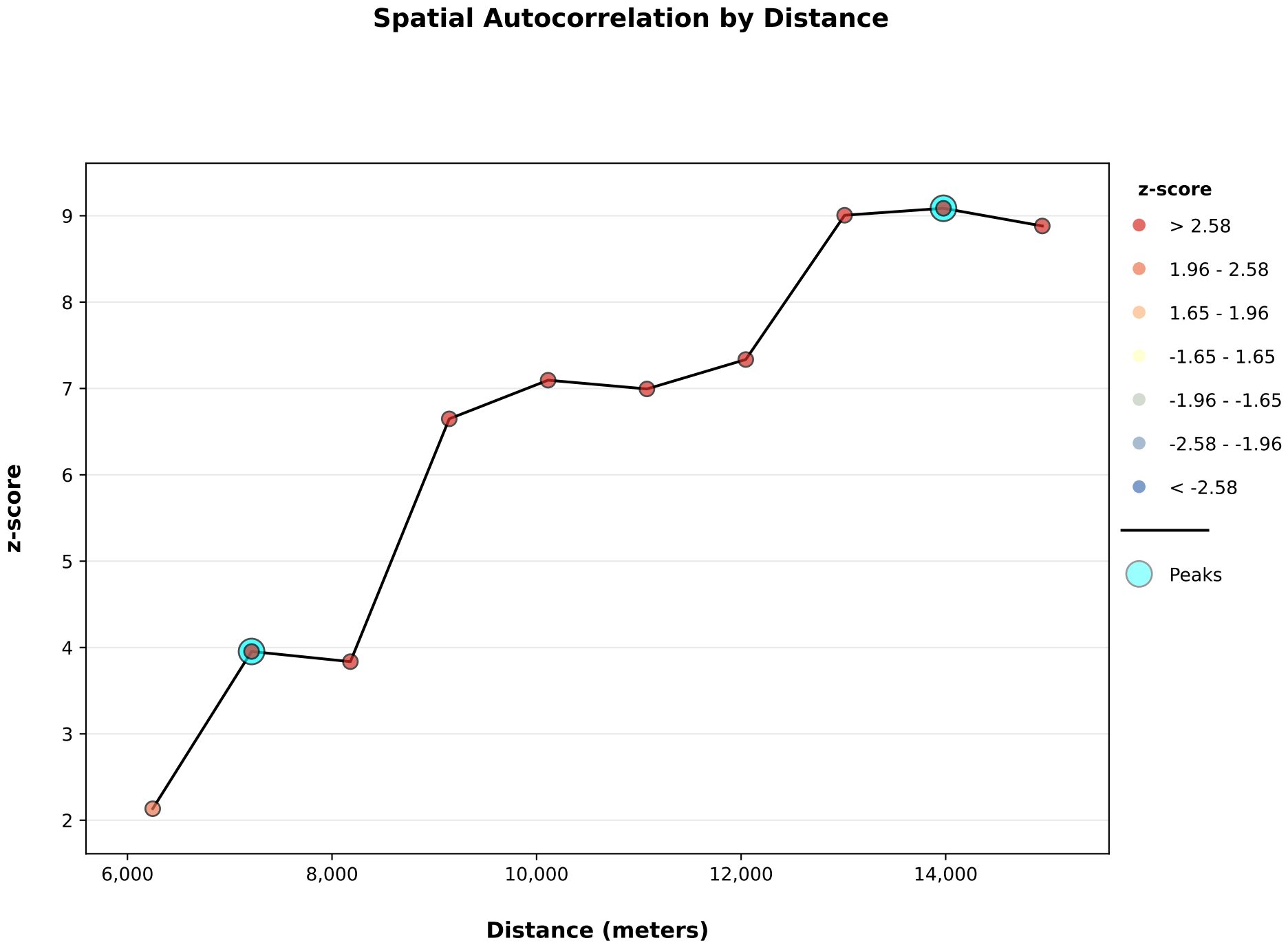
Van Der Meer, F. D., Van Der Werff, H. M. A., Van Ruitenbeek, F. J. A., Hecker, C. A., Bakker, W. H., Noomen, M. F., Van Der Meijde, M., Carranza, E. J. M., Smeth, J. B. D., & Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, *14*(1), 112–128. https://doi.org/10.1016/j.jag.2011.08.002

Verhoeven, G. J. (2017). *The reflection of two fields – Electromagnetic radiation and its role in (aerial) imaging*. https://doi.org/10.5281/ZENODO.3534245

Vicat, J. P., et al. (1998). Notice explicative, Carte géologique de la République du Cameroun. [Geological map of the Republic of Cameroon].

Vishiti, A., Suh, C. E., Lehmann, B., Egbe, J. A., & Shemang, E. M. (2015). Gold grade variation and particle microchemistry in exploration pits of the Batouri gold district, SE Cameroon. *Journal of African Earth Sciences*, *111*, 1–13. https://doi.org/10.1016/j.jafrearsci.2015.07.010

8 APPENDIX


Spatial Autocorrelation by Distance
z-score
9
8
7
6
5
4
3
2
6,000
8,000
10,000
12,000
14,000
Distance (meters)
z-score
> 2.58
1.96 - 2.58
1.65 - 1.96
-1.65 - 1.65
-1.96 - -1.65
-2.58 - -1.96
< -2.58
Peaks

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
6247.00	0.020197	-0.000812	0.000097	2.133492	0.032884
7213.41	0.033087	-0.000812	0.000074	3.954093	0.000077
8179.82	0.028134	-0.000812	0.000057	3.836416	0.000125
9146.23	0.044103	-0.000812	0.000046	6.648170	0.000000
10112.63	0.042719	-0.000812	0.000038	7.095882	0.000000
11079.04	0.038450	-0.000812	0.000032	6.994466	0.000000
12045.45	0.037122	-0.000812	0.000027	7.335176	0.000000
13011.86	0.042332	-0.000812	0.000023	9.005944	0.000000
13978.27	0.039725	-0.000812	0.000020	9.086539	0.000000
14944.68	0.036276	-0.000812	0.000017	8.881666	0.000000

First Peak (Distance; Value): 7213.41; 3.954093

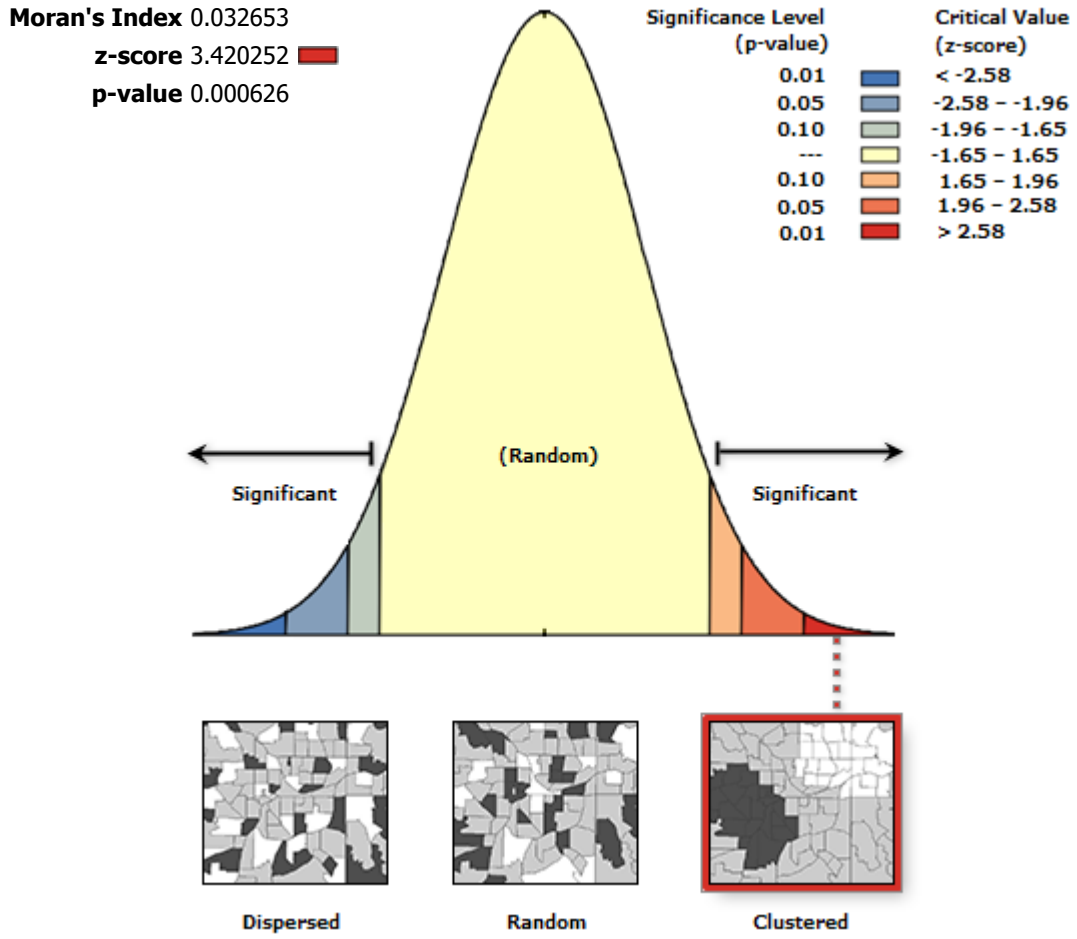
Max Peak (Distance; Value): 13978.27; 9.086539

Distance measured in meters

Incremental Autocorrelation Parameters

Parameter Name	Input Value
Input Features	RESULTATS_ANALYSES_NGAOUNDERE
Input Field	AU__PPB_
Number of Distance Bands	10
Beginning Distance	6247.000000
Distance Increment	966.408573
Distance Method	EUCLIDEAN
Row Standardization	False
Selection Set	False

Spatial Autocorrelation Report



Given the z-score of 3.420252, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

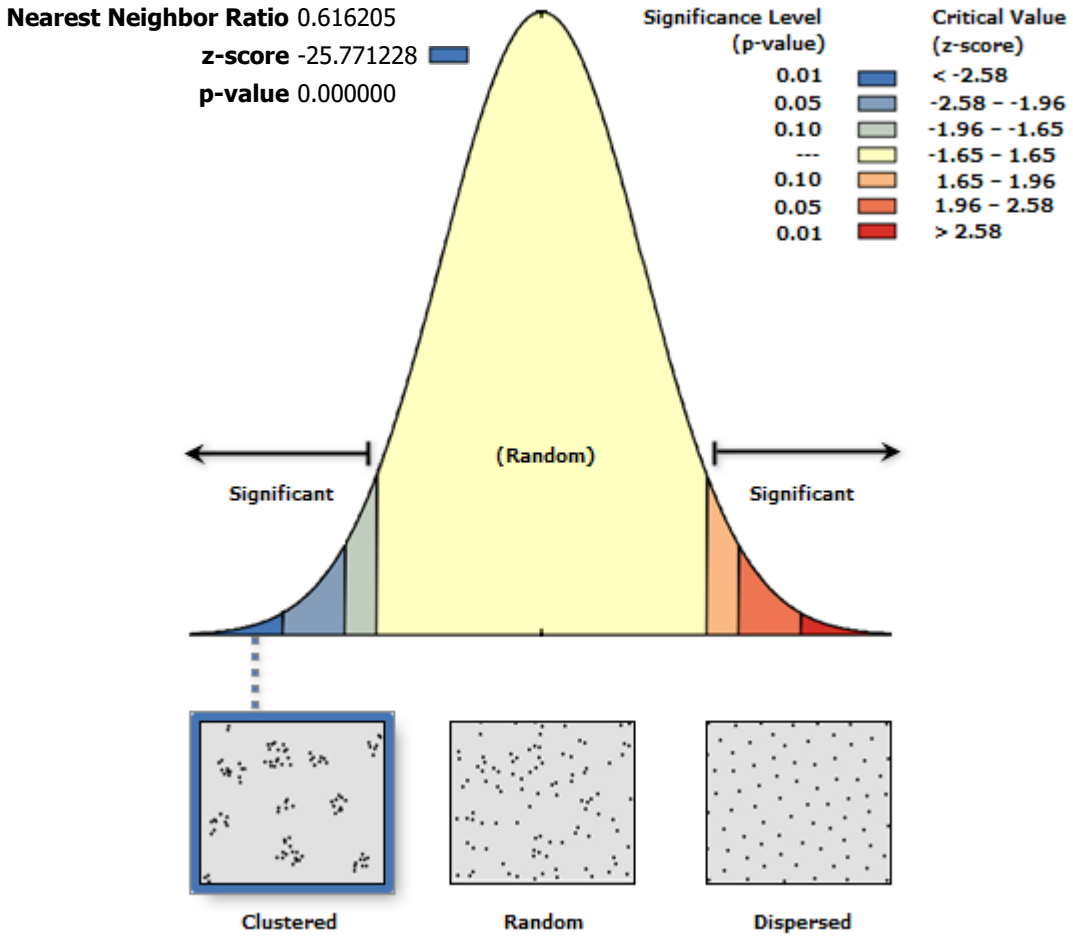
Moran's Index	0.032653
Expected Index	-0.000812
Variance	0.000096
z-score	3.420252
p-value	0.000626

Dataset Information

Input Feature Class:	RESULTATS_ANALYSES_NGAOUNDERE
Input Field:	AU__PPB_
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN

Row Standardization:	True
Distance Threshold:	14000.0000 meters
Weights Matrix File:	None
Selection Set:	False

Average Nearest Neighbor Summary



Given the z-score of -25.771228, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Average Nearest Neighbor Summary

Observed Mean Distance	966.4086 meters
Expected Mean Distance	1568.3222 meters
Nearest Neighbor Ratio	0.616205
z-score	-25.771228
p-value	0.000000

Dataset Information

Input Feature Class:	RESULTATS_ANALYSES_NGAOUNDERE
Distance Method:	EUCLIDEAN
Study Area:	12121078803.926645
Selection Set:	False