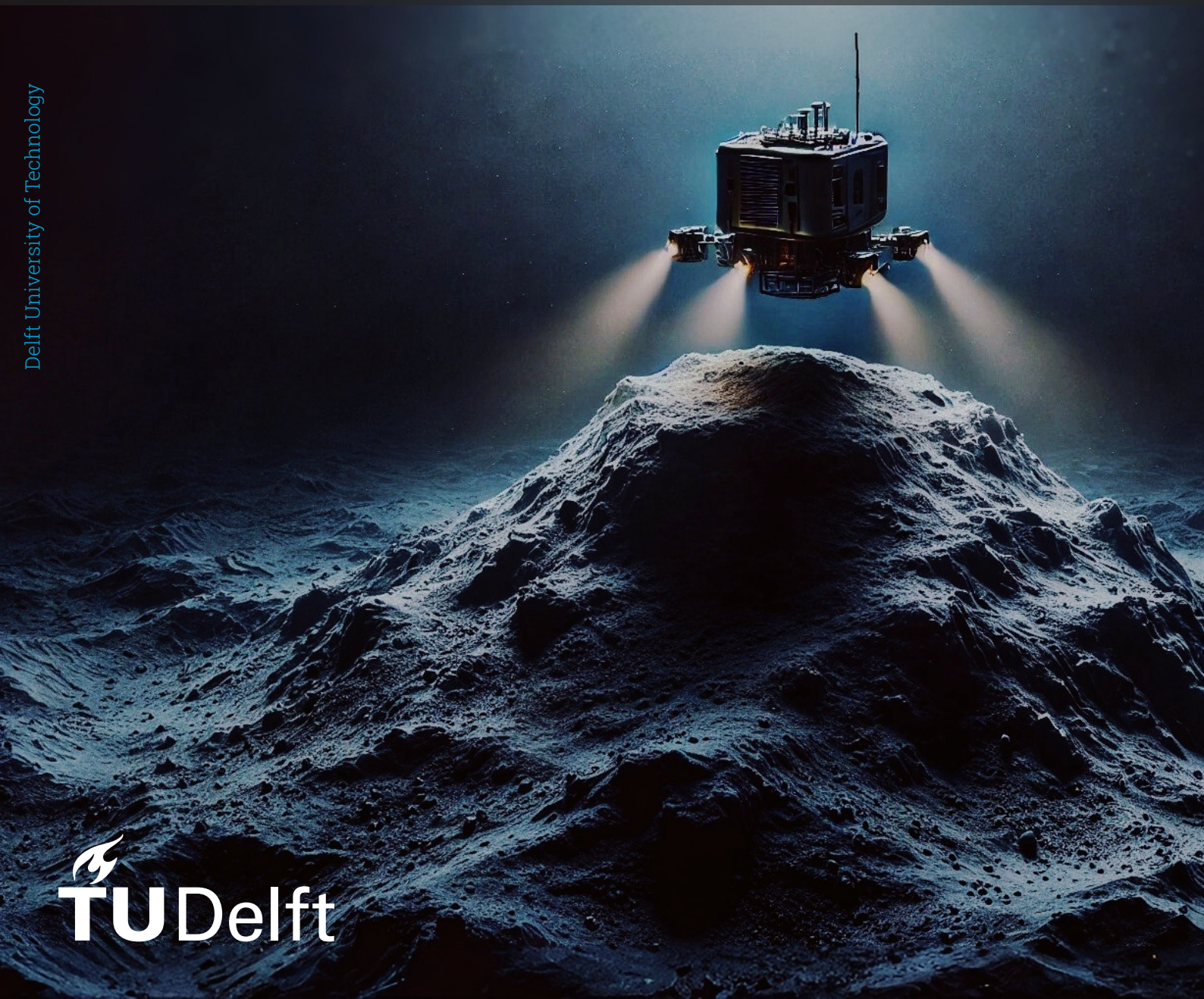
Depths and Deposits

Analyzing relationship between Ferromanganese Crust Composition and Water Depth through Collocated Co-Kriging to Study Potential Improvements in Estimation

AESM7000: Master Thesis

Arjun Prakash Kollamparambil



Delft University of Technology

TU Delft

Depths and Deposits

Analyzing relationship between Ferromanganese Crust Composition and Water Depth through Collocated Co-Kriging to Study Potential Improvements in Estimation

by

Arjun Prakash Kollamparambil

Master Applied Earth Sciences

5911990

Main supervisor:	Dr. Mike Buxton
Supervisor:	Dr. ir. Rudy Helmons
Supervisor:	Dr. Ing. Steinar Løve Ellefmo (NTNU Trondheim)
Supervisor:	Kiarash Pashna
Project Duration:	February, 2025 - August, 2025
Faculty:	Faculty of Civil Engineering and Geosciences, Delft

TU Delft

Preface

This thesis was my first proper dive into academic research, compositional data analysis and geostatistics. And it has helped me understand what I want to do with my career and life in many ways

I would like to thank my supervisors, Mike, Rudy, Steinar, and Kiarash, for their guidance and support throughout this thesis. I am also grateful to Dr. Sarah Atinta Howarth and her supervisor, Prof. Bramley Murton, for granting permission to use data from her doctoral research which is the foundation of this study. Special thanks to Dr. Vera Pawlowsky-Glahn for her valuable advice on compositional data analysis and ILR transformation. Finally, I appreciate the unconditional support and encouragement from my family and friends during this project and my academic journey.

Arjun Prakash Kollamparambil
Delft, August 2025

Abstract

The global transition to green energy has intensified demand for critical raw materials, increasing interest in deep-sea mineral resources such as cobalt-rich ferromanganese (Fe-Mn) crusts on seamounts like Tropic Seamount in the NE Atlantic. This thesis investigates the relationship between water depth and Fe-Mn crust composition and evaluates whether incorporating this relationship through Co-kriging (CK) improves resource estimation compared to Ordinary Kriging (OK). Geochemical and bathymetric data were analyzed using exploratory data analysis. The estimation workflow included block model generation, isometric log-ratio (ILR) transformation, Landmark-ISOMAP embedding for locally varying anisotropy, variogram modeling, and geostatistical estimation with Ordinary Kriging, Simple Co-Kriging (SCK) and Intrinsic Collocated Co-Kriging (ICCK) as well as Inverse Distance Weighted (IDW) Estimation. Model performance was assessed using Quantitative Kriging Neighborhood Analysis (QKNA) metrics and leave-one-out cross-validation (LOOCV), with a critical evaluation of LOOCV's limitations.

Results demonstrate a significant relationship between water depth and Fe-Mn crust composition, with a significant improvement in estimation accuracy and confidence when water depth is used as the secondary variable in Co-Kriging, with SCK and ICCK providing more accurate and confident resource estimates than OK. ICCK also performed better than IDW. For the first time, tonnage calculations for metals in Fe-Mn crusts based on a 3D block model and real geochemical data are presented, highlighting Tropic Seamount's potential as a substantial mineral resource for Europe. The workflow developed in this study, including ILR transformation and L-ISOMAP embedding, proved effective for handling compositional data and spatial anisotropy. The study also identifies methodological limitations, such as the need for improved variography, better sampling distribution, consideration of non-metallic elements, and more advanced cross-validation techniques. The assumption that the top 1 cm of crust represents the entire deposit is noted as a simplification, and future research should address vertical stratigraphy and sampling distribution.

The findings indicate that integrating ILR transformation and Landmark-ISOMAP embedding with Co-Kriging leads to better resource estimation for Fe-Mn crusts, enabling more confident and accurate assessments. This methodology offers significant potential for mineral resource exploration and future research in both marine and terrestrial environments.

Keywords: Deep sea mining, ferromanganese crusts, compositional data, isometric log-ratio, co-kriging, block model, estimation, critical raw materials.

Arjun Prakash Kollamparambil, Master Thesis, Delft University of Technology. *"Analyzing relationship between Ferromanganese Crust Composition and Water Depth through Collocated Co-Kriging to Study Potential Improvements in Estimation"*

Contents

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
ALR	Additive Log Ratio
CFA	Carbonate Fluoroapatite
CCK	Collocated Co-Kriging
CK	Co-Kriging
CLR	Centered Log Ratio
EDS	Energy Dispersive Spectroscopy
Fe-Mn	Ferromanganese
FKA	Factorial Kriging Analysis
Gof	Goodness of Fit
HREE	Heavy Rare Earth Element
ICCK	Intrinsic Collocated Co-Kriging
IDW	Inverse Distance Weighted Estimation
ILR	Isometric Log Ratio
LMC	Linear Model of Coregionalization
LREE	Light Rare Earth Element
LVA	Locally Varying Anisotropy
MAPE	Mean Absolute Percentage Error
MCD	Minimum Covariance Determinant
MDS	Multidimensional Scaling
MAE	Mean Absolute Error
MVN	Multivariate Normal
NRMSE	Normalized Root Mean Squared Error
OK	Ordinary Kriging
OMZ	Oxygen Minimum Zone
QKNA	Quantitative Kriging Neighborhood Analysis
REE	Rare Earth Element
REY	Rare Earth Element plus Yttrium
RMSE	Root Mean Squared Error
SBP	Sequential Binary Partition
SEM-EDS	Scanning Electron Microscopy - Energy Dispersive Spectroscopy
SPD	Shortest Path Distance
UTM	Universal Transverse Mercator
XRCT	X-ray Computed Tomography

Chemical Symbols

Symbol	Element Name
Al	Aluminum
As	Arsenic
Ca	Calcium
Cd	Cadmium
Ce	Cerium
Co	Cobalt
Cs	Cesium
Cu	Copper
Dy	Dysprosium
Er	Erbium
Fe	Iron
Gd	Gadolinium
Hf	Hafnium
Ho	Holmium
K	Potassium
La	Lanthanum
Li	Lithium
Lu	Lutetium
Mg	Magnesium
Mn	Manganese
Mo	Molybdenum
Na	Sodium
Nb	Niobium
Ni	Nickel
P	Phosphorus
Pb	Lead
Pr	Praseodymium
Pt	Platinum
Sb	Antimony
Si	Silicon
Sm	Samarium
Sn	Tin
Tb	Terbium
Te	Tellurium
Ti	Titanium
Tl	Thallium
Tm	Thulium
U	Uranium
V	Vanadium
W	Tungsten
Y	Yttrium
Yb	Ytterbium
Zn	Zinc

1

Introduction

1.1. Introduction and Relevance of the Project

As the world increasingly shifts towards green energy, demand for critical raw materials such as copper, lithium, and rare earth elements (REEs) continues to grow (International Energy Agency, 2025). These materials are essential for manufacturing batteries, renewable energy technologies, and a range of applications vital to the energy transition. However, terrestrial reserves are becoming progressively less viable for exploitation, prompting countries and companies to explore alternative sources. In this context, deep-sea mineral resources, such as deep-sea massive sulfide deposits, polymetallic nodules, and ferromanganese (Fe–Mn) crusts, have attracted renewed interest as potential contributors to future supply (Vysetti, 2023).

Among these, cobalt-rich Fe–Mn crusts have garnered particular attention due to their high concentrations of strategic metals, including cobalt (Co), nickel (Ni), and platinum (Pt) (Usui and Suzuki, 2022). The elevated cobalt content, in particular, exceeds that of nickel and copper in polymetallic nodules, enhancing their economic appeal. Consequently, Co-rich Fe–Mn crusts have emerged as a priority target for mineral exploration.

Fe-Mn crusts often occur in shallower water than manganese nodules, sometimes up to 2000 m closer to the surface, which might lead to a reduced complexity and cost of deep-sea mining compared to other deeper resources. Their denser distribution in some seabed regions increases extractable yield per unit area, and in many cases they occur on seamounts within the jurisdictions of island nations, possibly simplifying regulatory frameworks (Cronan, 2024; Nath, 2008). Taken together, these factors suggest a potentially more favourable operational and permitting environment for crust extraction compared with other deep-sea resources.

However, extracting Fe–Mn crusts presents significant technical challenges. Unlike nodules, which can be collected from the sediment surface, crusts are firmly attached to hard rock substrates and must be mechanically separated (Halbach et al., 2017). The goal is to remove them with minimal inclusion of the underlying non-valuable rock, yet this has no proven, widely-adopted solution. As of today, the areas that can be economically and technically mined remain uncertain (Halbach et al., 2017), directly affecting feasibility assessments.

Further complications arise from local variability. Oxide layer thickness can range from less than 1 cm to over 15–20 cm even on a single crust (Nath, 2008), making it difficult to design mining systems that adapt

to changing scraping depths. Where seafloor relief exceeds 20 cm, loosening and recovering crusts becomes more difficult (Halbach et al., 2017). In other words, fine-scale roughness and heterogeneity need to be accounted for as well.

Despite advances in understanding Fe–Mn crusts, notable gaps remain in knowledge of their small-scale distribution, genetic controls, occurrence modes, and regional patterns (Usui et al., 2016; Usui and Suzuki, 2022). Key aspects such as age, trace element variation, and relationships to geological setting etc remain limited in understanding. Comparing their origins to terrestrial deposits is hindered by undetermined geological parameters, and intra-seamount variations in morphology and composition remain under-characterised. Addressing these gaps is essential for reliable estimation of metal abundance and grade.

Resource estimation is inherently uncertain, even in terrestrial mining where sampling density is higher. Deep-sea conditions amplify these challenges, restricting both sample numbers and accessibility. Traditional estimation methods such as drilling and physical sampling become more costly and limited, highlighting the need for approaches that maximise the value of existing and / or limited datasets and incorporate additional, spatially correlated variables.

In this context, Co-Kriging (CK) offers advantages over Ordinary Kriging (OK) by integrating secondary variables, such as water depth, that are correlated with the primary variable, thereby reducing estimation error and variance (Yalçin, 2005). While another method, adaptive sampling, can also improve model accuracy (Fuhg et al., 2023), it is not applicable in this study due to a fixed sample data set. Adaptive sampling is typically used to focus the sampling campaign further on geologically important areas and improve model accuracy, it is not applicable here due to fixed sampling locations. Accordingly, the focus here is on improving estimation quality through multivariate geostatistics like CK rather than optimising sample placement.

Aim of the Thesis This master's thesis investigates the relationship between water depth and Fe–Mn crust composition and applies Co-Kriging to evaluate potential improvements in resource estimation. By integrating geochemical data with depth as a secondary variable, the study seeks to enhance the precision and confidence of kriging-based models and refine estimation and classification in line with established reporting standards (Howarth, 2022; Usui and Suzuki, 2022; Usui et al., 2016). The research also examines whether incorporating additional datasets, such as ocean current information, can improve knowledge of metal distribution. Together, these components form a coherent framework for reliable resource evaluation under the data-limited, high-cost conditions of offshore exploration. Ultimately, the results could provide inputs for strategic mine planning and support responsible exploitation of deep-sea mineral resources.

Objectives

1. Analyse the relationship between Fe–Mn crust composition and water depth.
2. Apply Co-Kriging to assess potential improvements in estimation accuracy compared to Ordinary Kriging and simpler methods.
3. Evaluate the contribution of secondary datasets (e.g., bathymetry, ocean currents) to understanding metal distribution and enhancing estimation.
4. Consider environmental implications and applicability of the results to other study areas.

Expected outcomes: A quantitative depth–composition analysis, higher-confidence resource estimates, and an economic block model suitable for downstream planning studies.

The next section, *Study area*, introduces the geological and oceanographic setting to which these methods are applied; the subsequent *Research Questions* section elaborates further on the aim and objectives.

1.2. Study Area

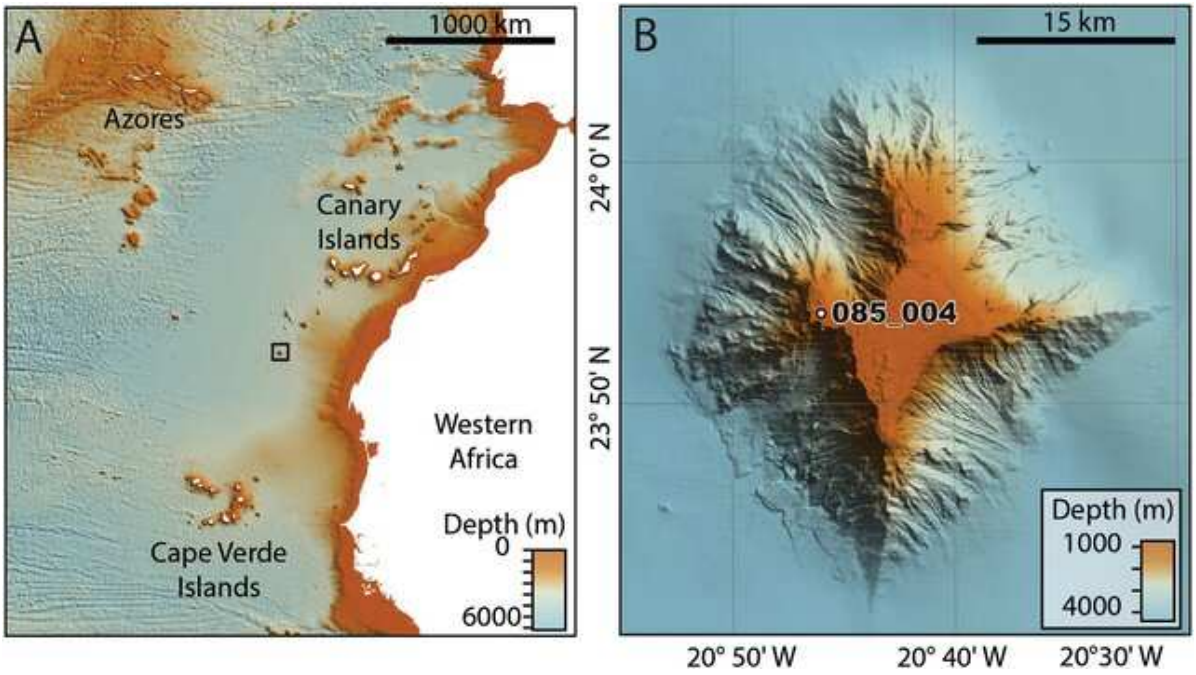


Figure 1.1: (A) Location of Tropic Seamount in the NE Atlantic, situated between the Canary and Cape Verde Islands. (B) Bathymetric map of Tropic Seamount from EM120 multibeam sonar data collected during cruise JC142, gridded at 50 m, showing the position of core 085_004 (Josso et al., 2019).

Tropic Seamount is an isolated volcanic structure located in the northeast Atlantic Ocean, forming part of the Western Saharan Seamount Province (WSSP). It is situated about 400 km off the coast of West Africa, roughly halfway between the Canary Islands and Cape Verde Islands. The seamount began forming in the early Cretaceous, with volcanic activity spanning from approximately 140 million years ago to the present. Most eruptions occurred during the middle Miocene, while Tropic Seamount itself principally developed in the late Aptian (119–114 Ma), with minor volcanic episodes continuing until the middle Paleocene (60 Ma). The seamount is built on oceanic crust that is now around 155 million years old, and its formation started when the oceanic crust was about 35–40 million years old. Tropic Seamount rises from the abyssal plain at 4100 meters below sea level to a summit depth of 950 meters, featuring a diamond-shaped, flat summit. The summit plateau is covered by carbonate platforms encrusted with Fe-Mn oxides, unconsolidated pelagic sediments, and foraminiferal sand. The flanks are marked by gullies, landslide scars, and spurs radiating from the summit, exposing volcanic rocks beneath loose sediments. Erosional processes, such as wave action, have shaped the summit, resulting in terraces and raised beach features (Howarth, 2018; Josso et al., 2019).

Taken together, these morphological and oceanographic characteristics shape the distribution, thickness, and accessibility of Fe–Mn crusts on Tropic Seamount. The research questions will be answered based on the datasets obtained from this seamount.

1.3. Research Questions

Anchored in the Tropic Seamount setting described above, this thesis investigates the depth–composition relationship and its implications for resource estimation.

The primary research question is:

- What is the relationship between water depth and crust composition, and how can Co-Kriging, using depth as a secondary variable, improve resource estimation accuracy and confidence

compared to Ordinary Kriging?

From this, several secondary questions emerge:

1. What factors influence crust composition, particularly the links among composition, bathymetry, and ocean currents?
2. Why use kriging instead of a simpler method such as inverse distance weighting (IDW)? Does kriging offer demonstrable added value in this context?
3. Which benchmarks (e.g., IDW, OK, CK) should be used to assess estimation accuracy and confidence?
4. How will the comparative analysis between estimation methods be conducted?
5. What are the potential benefits and challenges of incorporating ocean current data into the estimation model?
6. What is the effect of the revised resource estimation on the number of drill holes required for a successful exploration campaign?
7. How these results might affect environmental impact?

The research questions outlined are designed to be addressed within the scope of this master's thesis. Related areas of inquiry, including an in-depth Environmental Impact Assessment (EIA), advanced drillhole placement optimisation studies, and expanded oceanographic analyses, are recognised for their potential value but are excluded from the current study due to scope and resource constraints.

1.4. Methods

Literature Review: Existing literature was studied to identify the current state of the art and limitations in understanding regarding the topic and to justify the necessity of the study.

Data Collection: Primary georeferenced geochemical data and secondary bathymetry data regarding ocean floor topography were gathered for this study. These datasets were obtained from a PhD thesis by Sarah Atinta Howarth titled "An Investigation into the Variability of Ferromanganese Crusts in the NE Atlantic" from the University of Southampton (Howarth, 2022). Permission to use this data was granted with the understanding that proper acknowledgment will be provided.

Data Preprocessing: The collected data were transformed using the Isometric Log-Ratio (ILR) Transformation to address the compositional nature of the geochemical data. Compositional data were then collocated with corresponding topography data to ensure spatial alignment.

Exploratory Data Analysis (EDA): The data were visualized through scatter plots and a topography map was created from the topography data. Sample visualizations were overlaid on the topography maps to observe spatial alignment. Extensive statistical analysis, including descriptive statistics (mean, median, mode, standard deviation, histograms, box plots, etc.), correlation analysis (Pearson's Correlation Coefficient, etc.), and clustering, was performed on both the original and transformed data.

Empirical Variograms and Cross-Variograms: Empirical variograms were calculated for both primary (mineral concentrations) and secondary (depth) variables. Cross-variograms were also computed to understand spatial relationships between primary and secondary data.

Fit Variogram Models: Appropriate variogram models were fitted to the empirical variograms and cross-variograms. Parameters such as nugget, sill, and range were estimated, and model selection was optimized based on goodness-of-fit metrics. These models were also used for implementing ordinary kriging (OK).

Linear Model of Coregionalization (LMC): Variogram models were combined into the LMC framework, which integrates individual variograms into a single model that accounts for the spatial relationships between the two variables. Adherence to LMC assumptions was ensured to guarantee that the combined variogram model is valid and leads to meaningful results (e.g., positive definiteness).

Co-Kriging System: The CK system was set up using the fitted variogram models. The co-kriging plan was implemented with geostatistical Python libraries. Ordinary kriging was also implemented, utilizing all samples with the same setup, and the Inverse Distance Squared method was included for later comparison. Mineral concentrations at unsampled locations were predicted, and leave one out cross-validation (LOOCV) as well as Quantitative Kriging Neighbourhood Analysis (QKNA) was performed to assess model accuracy and robustness. Performance metrics were calculated for each method. Comparisons were made for evaluating how CK affects kriging performance.

Resource Estimation: The best performing method based on the LOOCV and/or QKNA metrics was used for back-transformation and resource estimation, identifying areas with high resource potential.

Comparison with OK and IDW: CK results were compared with those of OK, which utilizes all samples, and the simpler IDW method. The study assessed whether CK, using models that link water depth and crust composition data, can affect the number of drillholes required for a successful exploration campaign in terms of accuracy and confidence. Secondary research questions were answered based on these findings.

Tools

Data transformations and CLR variance plot generation was conducted using the Compositional Data Package (CoDaPack) software. All other data analysis and geostatistical analysis were performed using Python in Visual Studio Code.

Data

The bathymetry dataset and the geochemical dataset used in this study was originally obtained from bathymetric surveys and geochemical analyses conducted during a multidisciplinary oceanographic expedition aboard the RRS James Cook (JC142) in late 2016, as part of the MarineE-Tech project under the NERC 'SoS Minerals' research programme. The dataset being used here consists of 86 samples (out of more than 100) collected from various locations on the surface of Tropic Seamount. For each sample, the top 1 cm of the bulk sample was used for elemental composition analyses. The dataset includes concentrations of major, minor, and trace elements, all of them metallic elements except Tellurium which is a metalloid, reported in weight percentage, along with sample names and geographic coordinates. Sample preparation involved air drying and agate-milling with geochemical measurements performed using advanced instrumentation such as the Agilent 8900 ICP-QQQ and Thermo Scientific X-Series III ICP-MS. Analytical procedures included calibration against certified reference materials, correction for instrumental drift, blank values, and dilution, as well as regular use of internal standards to ensure data quality. The resulting dataset provides a geochemical profile of the Fe-Mn crusts across Tropic Seamount, enabling detailed analysis of elemental distributions in relation to water depth and spatial location (Howarth, 2018). A copy of the resulting dataset was provided by Steinar with the permission of Sarah's supervisor Prof. Bramley Murton. One thing to note is that the growth rates of Fe-Mn crusts vary from roughly 0.5-25mm per million years (Usui and Suzuki, 2022, Howarth, 2022) depending on the location on a specific crust and the crust's geological setting and therefore this 1 cm sample could represent anywhere from 20 million years of depositional history and composition to maybe less than a million years. However, this study will consider the 1 cm samples to be representative of the full thickness of the crust due to the constraints of available data.

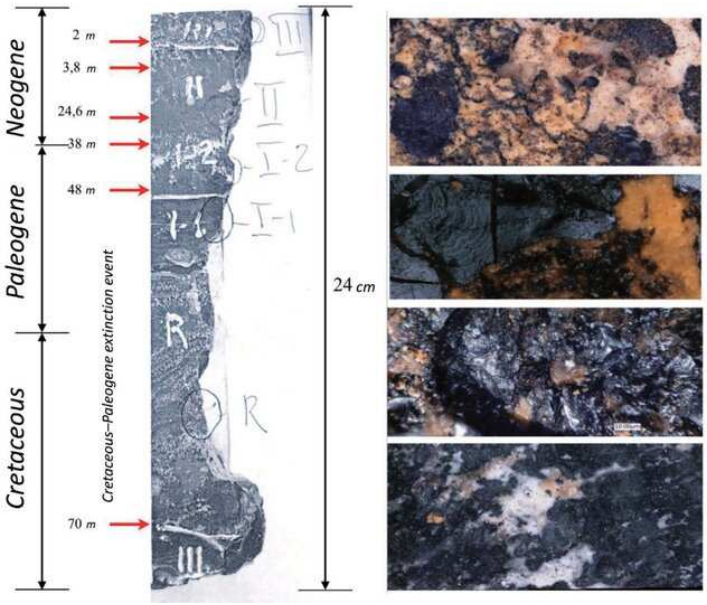


Figure 1.2: An example of the stratigraphy of Fe-Mn crusts, showing vertical layering and corresponding geological time periods. Obtained from *"Study of Deep-Ocean Ferromanganese Crusts Ore Components"*(Novakova and Novikov, 2021).

Figure 1.2 shows the vertical cross section of a Fe-Mn crust that shows the stratigraphy, highlighting the vertical layering and how different layers correspond to distinct geological time periods such as the Neogene, Paleogene, and Cretaceous that a single crust have formed through (Novakova and Novikov, 2021). This stratigraphy is important to consider when interpreting geochemical data, as the geochemical analysis may not fully capture the complexity present throughout the entire crust thickness if only a certain depth of the crust is sample.

Further details about the data can be found in D.

2

Literature Review

This chapter reviews the existing literature on the relationships between water depth and ferromanganese (Fe-Mn) crust composition, the geochemical and oceanographic processes influencing element concentrations, and the mechanisms underlying Fe-Mn crust formation. It provides context for the present study by summarizing key findings and current understanding in the field.

2.1. Formation, Geochemistry, and Resource Context of Fe-Mn Crusts

This section synthesizes current research on the formation, distribution, geochemistry, textural controls, resource assessments, and mining challenges of Fe–Mn crusts, identifying key knowledge gaps and informing the research objectives of this thesis.

Fe–Mn crust formation is described by a colloidal–chemical model of hydrogenetic precipitation involving inorganic chemical and surface-chemical mechanisms, sometimes influenced by microbial activity (Halbach et al., 2017; X.-h. Wang et al., 2009). As illustrated in Figure 2.1 and Figure 2.2, elements in seawater occur as dissolved ions or complexes whose charges vary with pH, which also affects Mn^{2+} oxidation rates (Stumm and Morgan, 2012). Hydrated colloids form and interact with dissolved metal ions: positively charged metal ions such as Co, Ni, Zn, Sn, and Ce are attracted to negatively charged Mn-oxide particles, whereas negatively charged ions and low-charge complexes such as U, As, Pb, Hf, Th, Nb, and REEs are attracted to slightly positively charged Fe-oxyhydroxide (FeOOH) particles (Halbach et al., 2017). Adsorption onto Mn-oxide and Fe-oxyhydroxide particles, together with redox processes, enrich the crusts with trace elements, while oxygen supplied from deeper waters by turbulent movements oxidizes Mn^{2+} and supports layer-by-layer precipitation and aggregation; slow growth rates favor the incorporation of substantial amounts of trace metals into hydrated oxide and oxyhydroxide phases within the water column and on crust surfaces (Halbach et al., 2017).

Recent works highlight both advances and needs in understanding the geochemistry and resource potential of these deposits. J. Hein and Koschinsky (2014) emphasize the potential of Fe–Mn crusts and nodules for REEs and other metals, recent advances in geochemistry, and new data on modern high-tech elements, while underscoring the importance of element speciation, surface chemistry, and physical properties. Although there has been significant progress in deciphering metal accumulation mechanisms, further research is needed in real-time in situ measurements, environmental evaluations, and the analysis of a wide variety of samples, especially thick phosphatized samples, alongside technological challenges in crust mining and the implementation of green technologies (J. Hein and

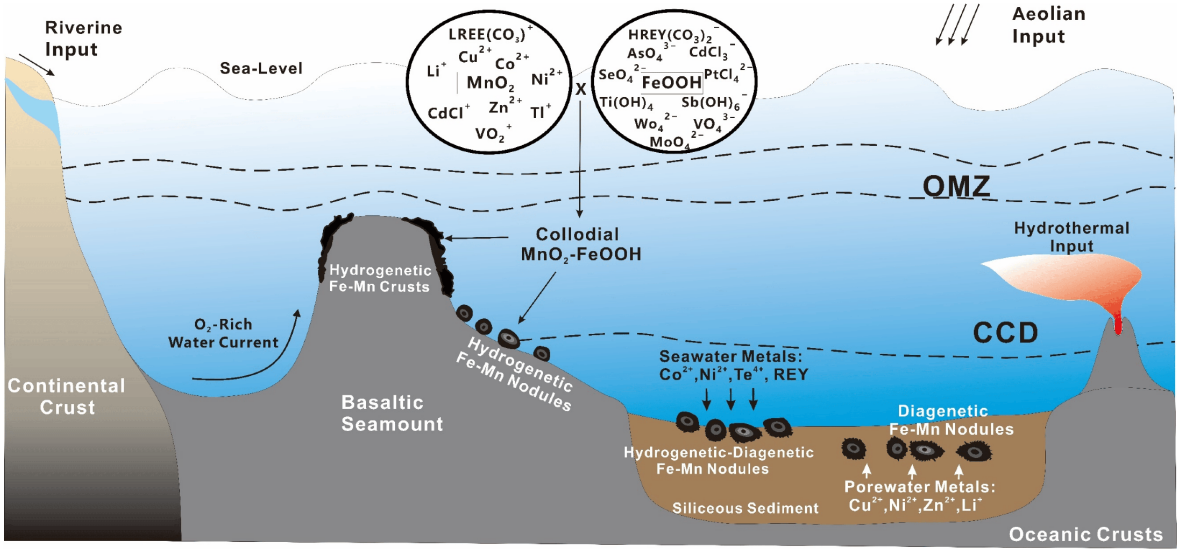


Figure 2.1: Schematic showing the formation, depths of formation, and locations of oceanic Fe-Mn crusts and nodules. (Huang and Fu, 2023).

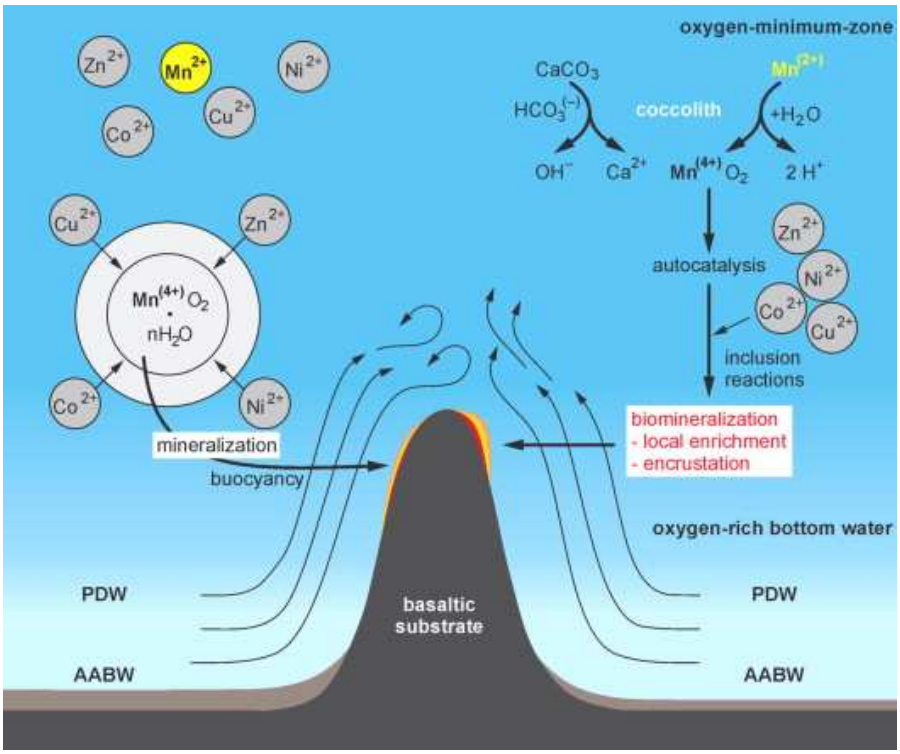


Figure 2.2: Schematic representation of processes controlling the deposition of minerals in hydrogenetic Fe-Mn crusts on basaltic seamounts, highlighting colloid-chemical processes and biomineralization (X.-h. Wang et al., 2009).

Koschinsky, 2014).

Geochemical distinctions between hydrogenous and hydrothermal contributions are central to interpreting crust origin and composition. In the equatorial western Pacific, hydrogenous crusts show a narrow and consistent range of Fe and Mn with a stable Fe/Mn ratio around 0.72, whereas hydrothermal crusts display broad Fe and Mn ranges with Fe/Mn ratios spanning from <0.002 to 3930; hydrogenous crusts are richer in Co, Ni, and Cu (G. Wang et al., 2015). At the same time, mineralogical indicators can be ambiguous: the occurrence of minerals such as todorokite in non-hydrothermal settings cautions against their use as sole hydrothermal indicators, and crusts identified as hydrogenetic may still incorporate hydrothermal material in areas with tectonomagmatic activity (Baturin and Dubinchuk, 2011). More generally, the origin of Fe–Mn crusts is constrained by the contents and ratios of Fe, Mn, ore metals, and REEs (Baturin and Dubinchuk, 2011).

Post-depositional processes further modify trace-element systematics. Phosphatization, driven by interactions with phosphate-rich seawater, remobilizes and reorganizes elements and minerals, enriching certain trace metals such as REY (Zhou et al., 2023). Enhanced sorption of REY(III) and HPO_4^{2-} on Fe oxyhydroxides during phosphatization facilitates REY enrichment (Zhou et al., 2023). Multiple microanalysis methods applied to sample MP2D32A from the Line Islands reveal that phosphatization plays a significant role in REY enrichment and that interactions among carbonate fluorapatite (CFA), Fe–Mn (oxyhydr)oxides, and phosphate complexes are important, while acknowledging additional, not yet fully understood factors contributing to REY enrichment and the limitations in the use of these signatures as proxies for oceanic environmental variability over millions of years (Zhou et al., 2023).

Textural heterogeneity at the microscale also governs metal contents and resource assessments. High-resolution XRCT and SEM–EDS mapping of Atlantic crusts from Tropic Seamount identify dense pillared textures (up to 93% Fe–Mn oxides) and less dense cuspate zones (as low as 57% Fe–Mn oxides) with higher detrital material and encrusting foraminifera; Co is highest in pillared textures, whereas cuspate textures show dilution by detritus and lower metal contents (Yeo et al., 2018). Although metal concentrations can be significant, the thin and variable textures present challenges for economic extraction under current mining models, and accurate resource estimation requires accounting for these textural variations (Yeo et al., 2018).

Regional studies further link composition to depth, age, and resource potential. At Takuyo Daigo Seamount, Fe–Mn crusts are highly enriched in Co (6500 ppm), Ni (4000 ppm), Pt (0.19 ppm), and REEs (1700 ppm), with element concentrations strongly correlating with Fe or Mn; deeper or younger crusts show promise for REE resources, whereas shallower or older crusts are rich in Co, Ni, Ce, and Pt, and the Fe/Mn ratio varies with water depth and age (Nozaki et al., 2016). Assuming a homogeneous distribution, the estimated reserve is roughly 65 million tons of ore, including 420,000 tons of Co and 110,000 tons of REYs, but variability in crust composition, the precision of sampling methods, and geographical constraints limit the accuracy and applicability of such evaluations and the feasibility of commercial mining, suggesting future research should address these limitations (Nozaki et al., 2016).

Global studies reinforce both the scale of the resource and the uncertainties of current estimates. Mizell et al. (2022), in a chapter from *Perspectives on Deep-Sea Mining* (Sharma, 2022), present global tonnage and metal concentration estimates for Fe–Mn crusts and nodules using extensive geochemical data and identify hydrothermal input as a significant control on crust composition, with crusts near spreading centers showing varying hydrothermal inputs relative to those formed farther away. While concluding that these deposits contribute a significant portion of the planet's metal reservoir, the study notes methodological limitations, including a uniform crust-thickness assumption, reliance on two-dimensional footprints rather than full seamount surface areas (ignoring shapes such as cones or guyots), lack of a formal code-based tonnage calculation, and additional uncertainties from sediment

cover, erosion, and rugosity; collectively, these could skew tonnage estimates by a factor of two (Mizell et al., 2022).

Finally, the feasibility of mining remains constrained. Growth interruptions spanning millions of years can lead to diagenetic changes, crusts are firmly attached to substrates, and current methods for detaching crusts from the seafloor without dilution by substrate rock are untested and no proven technology exists for this task (Halbach et al., 2017). Consequently, reported crust tonnages are "identified resources" rather than reserves, given the difficult separation and recovery process and uncertain economic viability. Overall, the realization of Fe–Mn crust mining is more complex than other seabed mineral extraction efforts (Halbach et al., 2017). In parallel, broader geochemical and technological priorities like real-time measurements, environmental evaluations, diverse sampling (including thick phosphatized samples), and green technologies etc frame the path forward (J. Hein and Koschinsky, 2014).

Given that the Fe/Mn ratio varies with water depth and age and that deeper or younger crusts differ compositionally from shallower or older crusts (Nozaki et al., 2016), the following section examines the Relationship between Depth and Crust Composition.

2.2. Relationship between Depth and Crust Composition

The relationship between deep-sea Fe-Mn crust composition and the depth at which they are found has been recently studied, revealing significant depth-dependent variations in elemental concentrations. Understanding this relationship is vital for effective exploration and extraction of these valuable resources.

Fe-Mn oxides precipitate most efficiently in typical seawater conditions, including areas with low oxygen, known as oxygen minimum zones (OMZs). Metals sensitive to redox conditions, such as Co and Ce, tend to concentrate at specific water depths where these OMZs occur (Usui and Suzuki, 2022). Oxygen content profiles of seawater at various depths are invaluable for determining ideal conditions for Co-rich Fe-Mn crust formation (Glasby et al., 2010).

Building on this understanding, Usui and Suzuki (2022) in their chapter *Geological Characterization of Fe-Mn Crust Deposits in the NW Pacific Seamounts for Prudent Deep-Sea Mining* from the book *Perspectives on Deep-Sea Mining* (Sharma, 2022) proposed a hypothetical genetic model based on studies of the Takuyo-Daigo seamount in the Northwest Pacific Ocean. According to this model, Fe-Mn crusts have undergone slow, consistent growth since the Neogene period (23.04 Ma) or earlier, across all water depths. The model highlights that the concentration of most metallic elements is strongly influenced by water depth. For instance, **younger layers at the Takuyo-Daigo seamount show higher Fe/Mn ratios and Co content, while older layers are enriched in total REEs, P, and Pt.** The chapter highlights that the geological characteristics of these deposits are not fully determined at various scales, and the reasons behind mineral diversity are not well understood. Small-scale variations in the occurrence, shape, and composition of the crusts within a seamount are also poorly documented, primarily due to difficulties in vehicle positioning, sampling of bottom materials, and the scarcity of geological information (Usui and Suzuki, 2022).

Further supporting this depth-dependent variation, Usui et al. (2016), in the article *Continuous growth of hydrogenetic Fe-Mn crusts since 17 Myr ago on Takuyo-Daigo Seamount, NW Pacific, at water depths of 800–5500 m* explains how they examined Fe-Mn crusts from the same seamount, focusing on their temporal and spatial variations to assess their resource potential. **Their findings indicate that as depth increases, Fe concentrations decrease while the Mn/Fe ratio rises, suggesting a shift from Fe-oxyhydroxides at shallower depths to Mn-oxides at greater depths**. Co concentrations also exhibit strong depth dependency, with higher levels in shallower regions due to elevated Co concentrations in the OMZ and redox conditions of seawater. Despite continuous growth over 17 million years, regional

variations are controlled by the seamount's evolution, substrate rock, and sedimentary history. Current knowledge is limited by the variability in crust composition and the challenges of precise sampling. The study emphasizes the need for further global- to nano-scale characterization, on-site exploration, and advanced analytical techniques to develop robust models for selective metal concentration and improve resource evaluations.

In the Northeast Atlantic, according to the chapter *Seamount-scale controls on Fe-Mn crust composition, NE Atlantic* of Sarah Atinta Howarths's Phd Thesis titled *An Investigation into the Variability of Fe-Mn Crusts in the NE Atlantic* (Howarth, 2022), the depth of the water seems to be the primary factor influencing the presence of Fe-Mn crusts. This chapter investigates how water depth influences the composition of Fe-Mn crusts at Tropic Seamount in the NE Atlantic. By examining 121 surface scrape samples, the study reveals significant variations in element concentrations related to water depth. It also finds that **as water depth increases, Fe concentrations, along with Co and Te, decrease, while the Mn/Fe ratio increases,** indicating a shift from Fe-oxyhydroxides at shallower depths to Mn-oxides at greater depths. Dissolved oxygen levels, closely linked to water depth, play a crucial role, with the OMZ indirectly promoting Mn-oxide formation through upwelling and mixing processes. The unique conditions in the NE Atlantic, including influences from Saharan dust and biogenic remineralization, result in different depth-related patterns. Understanding these controls is essential for accurately assessing the resource potential and mining feasibility of Fe-Mn crusts in this region. Local environmental factors like slope angle, current velocities, and topography also affect crust composition by influencing the distribution of detrital phases and the dilution of hydrogenetic phases. For example, Al shows negative correlation with maximum current velocity and slope (Howarth, 2022).

In contrast, research conducted along the Ninetyeast Ridge in the Indian Ocean highlights a different depth-related pattern in Fe–Mn crust composition. According to J. R. Hein et al. (2016), **Fe/Mn ratios exhibit a significant increase with water depth,** largely attributed to the rising availability of iron from deep-sourced hydrothermal systems and the dissolution of biogenic carbonates. While dissolved manganese derived from the oxygen minimum zone (OMZ) plays a role in Mn enrichment at intermediate depths, its influence diminishes below the OMZ, especially as oxidation conditions intensify with depth. This process results in an enhanced incorporation of iron and a concurrent decline in Mn/Fe ratios with increasing depth, indicating a transition toward Fe-oxyhydroxide dominance at greater depths. **The study further notes that Fe–Mn crust growth rates correlate positively with water depth,** underscoring the role of deep-ocean chemistry, sediment sources, and geologic features in shaping crust composition. These findings, in tandem with regional environmental variables, refine our understanding of the formation mechanisms and resource viability of crusts in different oceanic provinces.

These studies collectively highlight that water depth, along with associated oceanographic and geological conditions, significantly influences the composition and structure of Fe-Mn crusts. Understanding these depth-dependent variations is essential for accurately assessing resource potential and mining feasibility with confidence. This knowledge has the potential to better inform exploration strategies and enhance resource estimation models, which is a central focus of this thesis.

In light of the critical importance of accurate resource estimation for Fe-Mn crusts and the limitations of current methods under deep-sea conditions, there is a clear need for alternative approaches. Specifically, understanding how water depth influences crust composition could unlock new avenues for improving estimation techniques. Therefore, this study will explore the linkage between depth and crust composition and assess the efficacy of CK in enhancing resource estimation accuracy and confidence compared to traditional methods.

While this section has established the critical influence of water depth and associated oceanographic

factors on the elemental composition and resource potential of Fe-Mn crusts. To further contextualize these findings, it is important to examine the chemical relationships between individual elements within the crusts as described in existing literature. The following section synthesizes published research on geochemical associations, enrichment pathways, and post-depositional modifications concerning Fe-Mn crusts, providing a detailed context for how elemental distributions are governed by both environmental conditions and intrinsic chemical and mineralogical processes.

2.3. Chemical Relationships Between Elements in Fe-Mn Crusts

Sources and Distribution Mechanisms: Marine Fe-Mn crusts and nodules obtain their metal content primarily from seawater, with additional contributions from terrestrial riverine input, groundwater fluxes, hydrothermal activity, aeolian deposition, basaltic substrate weathering, and even cosmogenic sources (Huang and Fu, 2023). In certain regions like the Rio Grande Rise, regeneration of nutrients from local biological productivity provides metals that regenerate rapidly in the water column, particularly Ni, As, V, and Cd. Important contributions also come from various water masses such as the South Atlantic Mode Water and Antarctic Intermediate Water (Benites et al., 2020). In the North Atlantic, Saharan dust (aeolian) input significantly influences Fe-Mn crust composition (Marino et al., 2017).

Element Associations and Phase Preferences: Metal enrichment in Fe-Mn crusts and nodules is largely phase-dependent, with most elements exhibiting preferences for either Mn or Fe phases. This selectivity is governed primarily by electrostatic interactions between the metal species and the oxide surfaces (Huang and Fu, 2023). At typical seawater pH (approximately 8), δ-MnO_2 surfaces carry a negative charge (pH_{zpc} = 2.8) while amorphous FeOOH surfaces maintain a slight positive charge (pH_{zpc} = 8.5). This charge difference creates selective binding environments for different metal species. Analysis of Fe-Mn crusts from the Canary Island Seamount Province reveals that Fe shows a strong positive correlation with Mn (0.87) while exhibiting negative correlations (-0.5 to -0.7) with Mg, Al, K, and Si (Marino et al., 2017). Both Mn and Fe demonstrate positive correlations with REEs, V, Mo, Pb, As, and W, though Fe typically shows stronger correlations with REEs than Mn does (Marino et al., 2017).

Enrichment Pathways and Metal Groupings: The enrichment of metals in Fe-Mn crusts follows several distinct pathways depending on their chemical properties (Huang and Fu, 2023). Phase associations play a key role, with Light Rare Earth Elements (LREEs) tending to be more attracted to negatively charged δ-MnO_2 surfaces due to their positively charged nature in seawater, while Heavy Rare Earth Elements (HREEs) show a preference for slightly positively charged amorphous FeOOH surfaces. Direct substitution on δ-MnO_2 surfaces is observed for Ni, Cu, Zn, and Li, while oxidative substitution on these surfaces involves Co, Ce, and Tl. Some elements, such as REY (except Ce), Cd, Mo, W, V, Te, Pt, As, and Sb, partition between Mn and Fe phases. Dominant sorption by amorphous FeOOH is seen for Ti and Se. These pathways are influenced by the metals' electrochemical properties in seawater. Free cations and weak chloro-complexes (Co^{2+}, Ni^{2+}, $ZnCl^{+}$) and positively charged LREEs are attracted to negatively charged δ-MnO_2, while oxyanion complexes (Mo, Te) and HREEs preferentially bind to the slightly positive amorphous FeOOH (Huang and Fu, 2023).

Coordination Preferences in Oxide Structures: Certain metals exhibit specific coordination preferences within the layered and tunneled manganese oxide structures (Huang and Fu, 2023). Layer or tunnel-wall incorporation is typical for Co, Ni, and Cu. Triple-corner-sharing configurations are observed for Co, Ni, Cu, Zn, and Tl, while double-corner-sharing configurations are preferred by As, Sb, Mo, W, V, and Te. Edge-sharing configurations occur at layer rims for corner-sharing metals when they are less competitive or under less oxidizing conditions. Hydrated interlayer or tunnel-center sorption is common for Ni, Cu, Zn, Cd, Tl, and Li.

Genetic Process Influences: Different formation processes significantly impact metal distribution in

Fe-Mn deposits (Josso et al., 2017). Hydrogenetic precipitation favors enrichment of Co, high field strength elements (HFSE), and REY. Diagenetic processes produce higher Mn, Cu, and Ni concentrations through oxic remobilization in sedimentary columns. Suboxic conditions promote greater Mn and Fe remobilization, which competes with the incorporation of Cu and Ni. Hydrothermal Fe-Mn deposits show strong depletion in HFSE and REY due to rapid formation and high contents of either Fe or Mn oxides.

Post-Depositional Modifications: Post-depositional processes, particularly phosphatization, significantly alter the chemical composition of Fe-Mn crusts. In crusts from the Tropic Seamount, phosphatized samples maintain their total REY content but show redistribution of these elements, with a 22% preferential loss of LREEs, minor gains of HREEs (+7%), and a significant 42% enrichment in Y (Josso et al., 2020).

In Rio Grande Rise crusts, phosphatization occurred during the Miocene (20 to 6.8 Ma) under suboxic conditions, resulting in incorporation of Li from diagenetic reactions (Benites et al., 2020). Sulfate (SO_4) substitutes for phosphate (PO_4) in the apatite structure, while Y^{3+} and REE^{3+} replace Ca^{2+} in the carbonate fluorapatite structure, explaining the enrichment in HREEs (Benites et al., 2020).

Cobalt Relationships: Cobalt enrichment in Fe-Mn crusts occurs through incorporation into the vernadite structure, influenced by substitution and adsorption processes on crust surfaces (Marino et al., 2017). In seawater, dissolved Co^{2+} and Co^{3+} oxidize to their less soluble trivalent state. Crusts with higher cobalt content (0.7 wt%) exhibit slower growth rates (1.2 mm/Ma) compared to those with lower content (0.4 wt%, 2.7 mm/Ma). Co shows positive correlations with both Mn (0.58) and Fe (0.55), as well as strong positive correlations with As (0.7), Ce (0.61), and other metals including Mo, Zn, V, Tl, and LREEs (0.54-0.61) (Marino et al., 2017).

REE Patterns: REY elements in Fe-Mn crusts show strong positive correlations among themselves (>0.7) and with Fe (0.68-0.97), while displaying weaker positive correlations with Mn (0.59-0.85) (Marino et al., 2017). These correlation patterns suggest that REY elements primarily associate with Fe-oxyhydroxides, with a secondary affiliation to Mn oxides. Additionally, REY elements exhibit strong positive correlations with other metals found in Fe-Mn crusts, including As, V, Pb, Mo, and W, all hosted within Fe and Mn phases (Marino et al., 2017).

Geochemical Discrimination Using Element Ratios: Various elemental proportions serve as critical tools for distinguishing between formation mechanisms in oceanic Fe-Mn deposits. Particularly noteworthy is the relationship between Y and Ho; when Y/Ho ratios (normalized to Post-Archean Australian Shale) are below 1, they strongly indicate either hydrogenetic or diagenetic origin. In contrast, deposits formed through hydrothermal processes display considerably more variable Y/Ho values (Josso et al., 2017). Another significant indicator is the Ce anomaly, expressed mathematically as $Ce/Ce^* = Ce_n \,/\, \sqrt{La_n \times Pr_n}$ (subscript n implies normalized values), which functions as an effective redox state marker and typically shows enrichment in hydrogenetically formed crusts (Josso et al., 2020). Further diagnostic ratios include the Gd/Yb ratio, which indicates middle-to-heavy REE fractionation, and La/Sm, which reflects light REE fractionation patterns (Josso et al., 2020, Benites et al., 2020).

Understanding of Phosphatization Effects: Phosphatization processes significantly alter REE distributions while preserving certain diagnostic features. Research indicates that despite overall REE content remaining unaffected, phosphatization typically produces a flatter HREE profiles with $(Gd/Yb)_n$ values approaching 1, alongside slightly negative to positive Y anomalies (Josso et al., 2020). Chemical analyses reveal that phosphatized materials typically experience selective depletion, with LREEs decreasing by approximately 22%, while HREEs show minor enrichment (+7%) and Y concentrations increase substantially (42%). The characteristic positive Ce anomaly commonly intensifies during phosphatization, primarily due to substantial Pr depletion (42%) relative to unaltered samples, while La and Ce

exhibit comparatively moderate reductions (28% and 17%, respectively) (Josso et al., 2020). The Y/Ho relationship can undergo particularly notable increase during phosphatization, potentially reaching superchondritic ratios as high as 50 in extensively phosphatized Pacific Ocean samples, although Atlantic specimens generally display less dramatic alterations (Josso et al., 2020).

Surface Charge Mechanisms and Element Fractionation: The fractionation of elements between manganese and iron phases is largely governed by the contrasting surface charges of these minerals at seawater pH (8). The δ-MnO2 phase, with its pH zero point of charge (pHzpc) of 2.8, carries a negative surface charge that preferentially attracts free cations and weak chloro-complexes such as Co^{2+}, Ni^{2+}, $ZnCl^{+}$, and positively charged LREEs. Conversely, amorphous FeOOH with a pH_{zpc} of 8.5 maintains a slight positive charge, preferentially binding oxyanion complexes like Mo and Te, as well as HREEs (Huang and Fu, 2023). This electrostatic mechanism explains many of the observed element associations in Fe-Mn crusts, though some elements (such as Pt and W) show enrichment patterns that suggest additional mechanisms beyond simple electrostatic attraction (Huang and Fu, 2023).

It needs to be mentioned that none of the literature given in this section (2.3) explicitly mention any compositional data transformation methods which are crucial when dealing with geochemical datasets constrained by the constant-sum rule. Therefore the correlations mentioned here might not reflect real world relationships due to the spurious correlations found in compositional data that sum to a constant.

Summary

Across the reviewed literature, Fe–Mn crusts form by slow hydrogenetic precipitation in which dissolved ions and complexes in seawater, modulated by pH, redox conditions, oxygen supply and other environmental factors, adsorb to Mn-oxide and Fe-oxyhydroxide surfaces and accrete layer by layer, with post-depositional phosphatization redistributing REY. Composition varies systematically with water depth and regional oceanography: OMZs and water-mass structure influence Co and Ce at specific depths, and with depth, Fe and Mn partitioning and Fe/Mn or Mn/Fe ratios shift, and trends differ among oceanic provinces. Microscale textures and detrital inputs further modify metal contents and extraction feasibility. Resource potential appears significant but is uncertain due to variability in composition and texture, sampling precision, simplifying assumptions in tonnage estimation, and unresolved mining challenges related to substrate attachment and economic viability. Viewed through a Source–Fluid–Transport–Trap/Sink lens, metals derive primarily from seawater with additional inputs from terrestrial, groundwater, hydrothermal, aeolian, basalt weathering, and cosmogenic sources (Source); seawater is the carrier (Fluid); ocean circulation, OMZs, upwelling and mixing, and colloid formation and complexation provide the pathways (Transport); adsorption to Mn-oxide and Fe-oxyhydroxide particles, redox-driven oxidation, layer-by-layer precipitation, phosphatization, and textural controls act as the traps and sinks (Trap/Sink).

3

Methodology

This chapter outlines the theoretical background and methodological framework used to analyze the relationship between water depth and ferromanganese crust composition. It details the data structure, preprocessing steps, exploratory analyses, geostatistical modeling approaches and the tools employed to address the research questions and hypotheses of this study.

3.1. Block Modelling of Tropic Seamount FeMn Crust

Generation of crust 3D volume: The creation of a three-dimensional volumetric model of the Fe-Mn crust deposit began with the acquisition and integration of two crucial datasets. Bathymetric data provided the spatial foundation through x, y coordinates (in UTM coordinates) and corresponding z depth values (in metres) of the ocean floor topography. This was complemented by geochemical measurements, which contained spatial coordinates and thickness measurements of the Fe-Mn crust deposits along with metal concentrations. These thickness values, originally recorded in millimetres, were converted to metres for consistency with the bathymetric scale.

To transform the discrete bathymetric points into a continuous thickness model spanning the entire study area, an Ordinary Kriging interpolation approach was implemented. This geostatistical method employed a linear variogram model to characterize the spatial correlation structure of the crust thickness. To manage computational demands efficiently, the interpolation was executed in batches of 1000 points, systematically mapping thickness values across the complete bathymetric grid. This process created a comprehensive spatial model of crust thickness variations throughout the region of interest.

The interpolated thickness data enabled the generation of two spatially correlated surfaces that defined the vertical boundaries of the Fe-Mn crust deposit. The top surface preserved the original bathymetry, representing the interface between seawater and the upper crust boundary. The bottom surface was mathematically derived by subtracting the interpolated crust thickness from the bathymetric elevation at each grid point, effectively modeling the lower boundary of the mineral deposit, as visualized in Figures 3.1a, 3.1b, and 3.1c from different viewing angles (these figures are only for visualization of the volume and hence is not to actual scale or represent units). These surfaces were exported as GOCAD TSurf files, preserving both vertex coordinates and triangulation information necessary for 3D representation.

To complete the volumetric model, the top and bottom surfaces needed to be joined along their edges to create a watertight 3D solid. This process began with reading the vertex and triangulation data from

both TSurf files. The boundary vertices of each surface were systematically identified using a convex hull algorithm in the xy-plane. Corresponding vertices between top and bottom surfaces were precisely matched, enabling the generation of side faces by connecting these boundary points. The resulting quadrilateral faces were triangulated to maintain proper mesh topology and ensure compatibility with standard 3D visualization formats.

The final step involved integrating all mesh components; top surface triangles, bottom surface triangles, and side triangles into a unified data structure with carefully managed vertex indexing. This complete 3D model was written to a VTK file in legacy ASCII format, containing the full available geometric definition of the Fe-Mn crust deposit, as illustrated in Figures 3.1d and 3.1e with exaggerated thickness, and in Figure 3.1f with actual proportions. This volumetric representation provides a comprehensive basis for quantitative resource assessment, visualization in specialized scientific software, and further geospatial analysis of the marine mineral deposit. The resulting model enables accurate volume calculations and spatial characterization critical for both scientific understanding and potential resource evaluation.

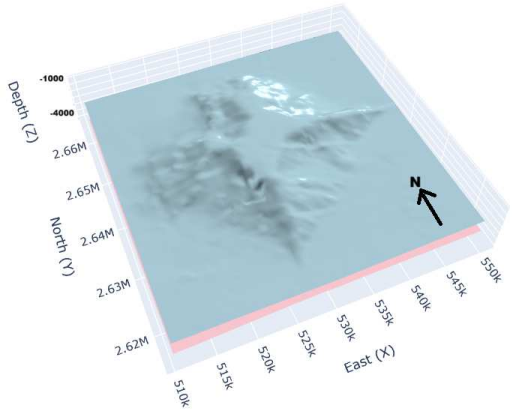
Creation of constrained block model from the 3D crust volume: The creation of a block model represents a critical step in transforming the continuous 3D volume of Fe-Mn crust into a discretized representation suitable for resource estimation and geostatistical analysis. This process begins with the preparation of input data, which includes geochemical measurements, bathymetric data, and the previously generated 3D volumetric model stored in VTK format.

The initial phase involves establishing the spatial domain for block generation. This is accomplished by analyzing the spatial extents of the geochemical data points to determine the natural boundaries of the study area. From these points, minimum and maximum coordinates are extracted along each dimension (X, Y, Z) to define a bounding box. To ensure comprehensive coverage of the region of interest, this bounding box is expanded with a small padding factor (5%) and additional significant points are explicitly included, particularly, four carefully selected bathymetry data points that were manually identified through visual comparison of the bathymetry data and the top surface of the volume. These manually selected points form the vertices of a convex hull that effectively encompasses the area containing meaningful Fe-Mn crust deposits.

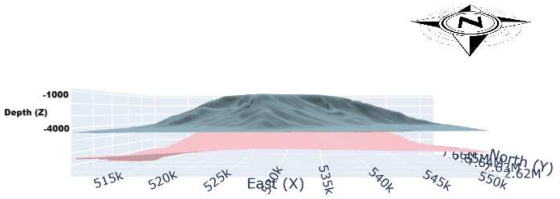
With the spatial domain established, the next critical step involves creating a convex hull constraint to focus the block model on geologically meaningful areas. Rather than filling the entire rectangular bounding box with blocks, a convex hull is defined using the four manually selected vertices that represent the outer boundaries of the meaningful study area. These vertices are projected onto the XY plane to create a two-dimensional polygonal boundary using the Shapely library's Polygon class, with further optimization through Matplotlib's Path functionality for efficient point-in-polygon tests. This approach reduces computational requirements by eliminating blocks in areas outside the area of interest.

After the convex hull is established, the block generation process begins by first creating a single horizontal layer of blocks within the hull boundary. Each block in this initial layer has its centroid at a specific XY location within the hull, with set block dimensions horizontally. Subsequently, this initial layer pattern is replicated vertically to cover the entire Z-extent of the data domain, with each Z-layer spaced at set intervals corresponding to block height. This creates a complete three-dimensional grid of blocks, with finer resolution in the vertical dimension to capture the relatively thin nature of the Fe-Mn crust deposits. To optimize memory usage on high-performance computing systems, the grid generation is implemented using a file-backed approach where blocks are created and stored in batches organized by Z-layers, leveraging Python's multiprocessing capabilities.

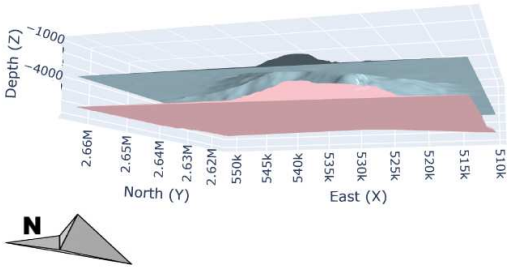
A particularly sophisticated aspect of the block model generation is the filtering of blocks based on



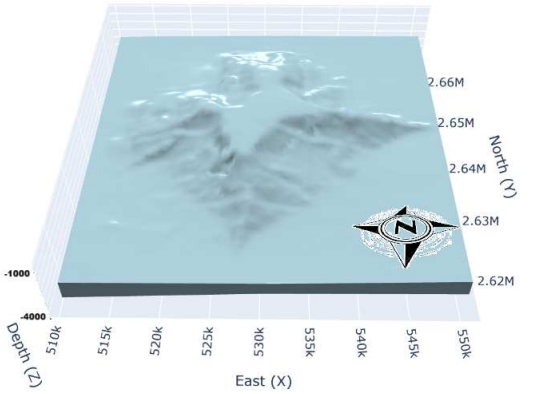
(a) Top and bottom surfaces with thickness exaggerated by a factor of 50,000.



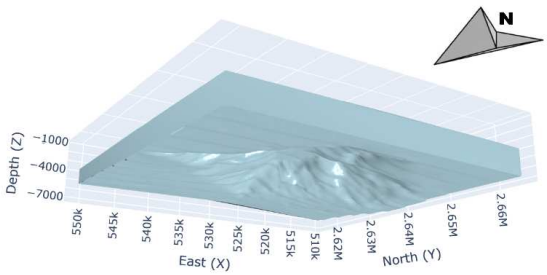
(b) Alternative angle of top and bottom surfaces with exaggerated thickness.



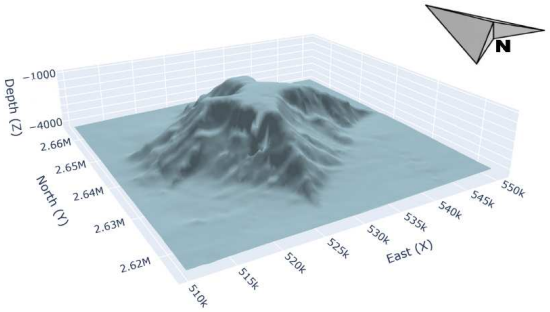
(c) Third perspective of top and bottom surfaces with exaggerated thickness.



(d) Complete 3D volume with exaggerated thickness factor.



(e) Alternative view of 3D volume with exaggerated thickness.



(f) 3D volume model with actual thickness proportions and slightly exaggerated vertical scale for visibility.

Figure 3.1: Visualization of the Fe-Mn crust 3D model generation process, showing both the individual surfaces and the final volumetric representation. All the axes are in metres (m)

their intersection with the 3D Fe-Mn crust volume since only those blocks that encompass the crust is required. This intersection testing employs ray tracing techniques, a computational method borrowed from computer graphics, implemented through the PyVista library's `ray_trace` functionality. For each candidate block, multiple rays are cast from points slightly outside each of its six faces, directed inward toward the block center. These rays from different directions (top, bottom, left, right, front, and back) check whether they intersect with the triangulated surfaces of the crust volume. If any ray intersects with the volume's surface, it indicates that the block potentially contains part of the crust deposit and is therefore retained in the model. This relatively complicated approach is required because the resolution of the bathymetry data is much lower than the resolution of the blocks that simple filtering of the blocks based on coordinates is not feasible.

The ray tracing algorithm functions by calculating intersections between each ray and the triangulated mesh surfaces that define the crust volume. Ray origins are positioned at strategic points outside each block face, and rays are directed perpendicular to these faces toward the block center. When a ray encounters a triangulated face of the mesh, the intersection point is recorded. By analyzing these intersections, the algorithm can determine whether a block is fully outside, fully inside, or partially intersecting the volume. This approach ensures accurate representation of the crust geometry while maintaining the regular block structure needed for resource estimation.

This was also one of the most computationally intensive part of the resource modeling because due to the geographical extent of the Tropic Seamount (approximately 40km in diameter and 3km in height) and the fine dimensions of the blocks, there were, depending on the dimensions of the block, upto billions of blocks that had to be filtered through to find the required ones. For computational efficiency, this intersection testing is preceded by a preliminary filtering step that eliminates blocks clearly outside the volume's bounding box. Only blocks that pass this initial spatial filter undergo the more computationally intensive ray tracing tests. Furthermore, the intersection testing is parallelized and divides the workload into chunks and processes them concurrently across multiple CPU cores, significantly accelerating the computation for such large models.

The final phase involves organizing the blocks that have passed all geometric constraints into a structured model. Each retained block is assigned a unique identifier (block id) and fundamental physical properties such as volume and position information (centroid coordinates and extents). In preparation for more advanced geostatistical modeling, the finalized block model is exported in multiple formats, including standard mining industry formats such as CSV, Parquet, GSLIB, and VTK, ensuring compatibility with a wide range of specialized software tools for visualization and further analysis.

This comprehensive block modeling approach transforms the continuous volumetric representation of the Fe-Mn crust into a structured, quantifiable format that maintains an accurate representation of the deposit's complex geometry while providing the framework necessary for subsequent resource estimation processes.

3.2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to understand the structure and relationships within the dataset. The process began by loading the data into Python. Numeric features were identified and aggregated, while sample names were retained for reference.

The raw and transformed compositional data were analyzed using statistical summaries and visualizations, including histograms, box plots, and cumulative distribution functions (CDFs), for each element and each transformed variable. The code computed key statistics such as mean, median, variance, standard deviation, and skewness to characterize the data distributions.

A correlation matrix was generated to assess relationships between elements, and between ILR coordinates, and the top correlated pairs were visualized using scatter matrix plots. Principal Component Analysis (PCA) was performed to reduce dimensionality and visualize explained variance across principal components.

Clustering analysis was conducted using K-Means on the principal components, with the optimal number of clusters determined via silhouette scores. The resulting clusters were visualized in three-dimensional space, both in PCA and geospatial coordinates. Further analysis included box plots to compare element concentrations and water depth across clusters, providing insights into the underlying patterns in the data.

3.3. Data Transformations

Why Is Data Transformation Necessary for Kriging? Kriging, like many geostatistical methods, performs optimally when certain statistical assumptions are met. Data transformations serve several critical purposes in this context. First, they help improve the normality of data distributions, which is essential since kriging techniques generally work better when the data or residuals approximate a normal distribution (Osborne, 2002). Second, transformations stabilize variance across the data range, addressing heteroscedasticity (the situation where the variability of the error terms in a model is not consistent across all levels of an independent variable (Clar, 2023)) issues that can undermine kriging accuracy. When data exhibits significant skewness, as is common in environmental and geological variables, kriging estimators become problematic due to their linear nature and sensitivity to outliers. This is particularly challenging in contexts with highly skewed distributions, such as ore grades in gold deposits, where a small number of samples can disproportionately influence the overall estimation results (Chilès and Delfiner, 2012).

It has been demonstrated that for data with clearly lognormal distributions, applying appropriate transformations before kriging typically produces estimates with lower experimental estimation variance compared to classical linear kriging estimators applied directly to untransformed data (Journel and Huijbregts, 1976). The fundamental motivation is that the stationary Gaussian case represents an ideal scenario for estimation, making transformations that bring non-Gaussian data closer to this ideal particularly valuable.

3.3.1. Commonly Used Data Transformations

Several transformation approaches are employed in kriging applications, depending on data characteristics and analysis requirements.

Logarithmic Transformations represent a versatile class rather than a single technique. As Osborne (2002) describes, logarithmic transformations involve expressing values as powers of a chosen base. While base 10 is common, natural logarithms (base $e = 2.7182818$) and base 2 logarithms are also frequently employed. The choice of base depends on data characteristics, for example, base 10 works well for extreme ranges, whereas lower bases provide better resolution for less extreme distributions. An important limitation is that logarithms are undefined for values less than or equal to zero, necessitating the addition of constants when such values are present (Osborne, 2002).

Square Root Transformation offers a moderate alternative to logarithmic approaches. This technique involves taking the square root of each data value, which compresses higher values while spreading out lower ones. It proves particularly valuable for data with Poisson distributions (data representing the tally of events in a fixed interval and these events occur randomly but at a steady mean rate (Haight, 1967)) or when variance is proportional to the mean. However, it may not effectively address left-skewed variables and can sometimes increase distortion in such cases (Choueiry, 2025).

For **compositional data**, where variables represent parts of a whole and sum to a constant, specialized log-ratio transformations are necessary due to spurious correlations between components (Egozcue et al., 2003):

- **Additive Log-Ratio (ALR)** expresses compositions by selecting one component as a reference and computing logarithms of the ratios between all other components and this reference. While straightforward, this approach is not isometric and thus inappropriate for analyses involving distances or angles (Hartmann et al., 2023).
- **Centered Log-Ratio (CLR)** divides each component by the geometric mean of all components before taking logarithms. This provides a more symmetric treatment of all components and facilitates interpretation. However, it produces linear dependency among transformed components, sums to zero, and lacks subcompositional coherence (Hartmann et al., 2023).
- **Isometric Log-Ratio (ILR)** transforms compositional data into a Euclidean space with one dimension less than the original data. While the results can be challenging to interpret directly, ILR ensures full-rank covariance matrices, making it compatible with a wide range of statistical methods (Hartmann et al., 2023).

Why Are Log Ratio Transformations Better Than Simple Logarithmic Transformations (Especially for Compositional Data)?: Compositional data presents unique analytical challenges due to its constrained nature, that is, all components must sum to a constant (typically 1 or 100%). This constraint, known as the closure problem, creates dependencies among components that simple logarithmic transformations applied independently to each component fail to address.

According to Egozcue et al. (2003), applying standard logarithmic transformations to compositional data can lead to several statistical issues. First, the closure constraint creates spurious correlations between components that don't reflect genuine relationships. Second, simple logarithmic transformations neglect the fundamental property of compositional data that only the ratios between components carry meaningful information (Egozcue et al., 2003).

Log-ratio transformations specifically address these limitations by:

1. Emphasizing relative differences between components rather than their absolute values
2. Maintaining scale invariance, acknowledging that only the proportions between components matter
3. Ensuring subcompositional coherence, meaning analyses based on a subset of components remain consistent with analyses of the complete set

These properties make log-ratio transformations mathematically rigorous and conceptually appropriate for compositional data analysis in kriging contexts (Egozcue et al., 2003).

Why Is the ILR Transformation Generally Preferred Over ALR and CLR? While all three log-ratio transformations convert compositional data into unconstrained space, they differ significantly in their mathematical properties and suitability for kriging applications.

CLR transformation, while treating all components symmetrically, produces transformed values that sum to zero. As Egozcue et al. (2003) explain, this constraint results in a singular covariance matrix that complicates statistical analyses like kriging, which typically require invertible covariance matrices.

ALR transformation avoids this singularity issue but introduces asymmetry by arbitrarily selecting one component as a reference denominator. More importantly, Egozcue et al. (2003) highlight that alr is not an isometric transformation, meaning it doesn't preserve the geometric properties (distances and angles) when mapping from the simplex with Aitchison metric to real space with Euclidean metric.

Table 3.1: Data Transformation Methods for Kriging Applications

Transformation	Application Context	Key Benefits
Logarithmic	Right-skewed, positive-valued data; particularly effective for variables with extreme ranges	Improves normality; stabilizes variance; different bases (10, e, 2) offer varying degrees of compression for different data ranges (Osborne, 2002)
Square Root	Data with Poisson distribution or when variance is proportional to the mean	Compresses higher values while spreading out lower ones; moderate alternative to logarithmic approaches; maintains more of the original data structure (Choueiry, 2025)
Additive Log-Ratio (ALR)	Compositional data where components sum to a constant	Expresses compositions using one component as reference; computes logarithms of ratios between components and reference; straightforward implementation but not isometric (Hartmann et al., 2023, Egozcue et al., 2003)
Centered Log-Ratio (CLR)	Compositional data requiring symmetric treatment of components	Divides each component by geometric mean before taking logarithms; facilitates interpretation; maintains symmetry among components but produces singular covariance matrices (Hartmann et al., 2023, Egozcue et al., 2003)
Isometric Log-Ratio (ILR)	Compositional data requiring rigorous statistical analysis	Transforms data into Euclidean space with one dimension less than original data; ensures full-rank covariance matrices; preserves geometric properties like distances and angles; compatible with methods requiring invertible covariance matrices (Hartmann et al., 2023, Egozcue et al., 2003)

ILR is often favored over ALR in distance-based analyses like clustering or PCA because it preserves Euclidean geometry and avoids the subjective, context-dependent challenge of selecting a reference component, which can affect interpretation and consistency (Mangnier, 2024). In the absence of explicit recommendations in the literature, the denominator for the ALR transformation was selected based on expert guidance from supervisor Steinar, who has extensive experience in compositional data analysis.

In contrast, ILR transformation provides the advantages of both approaches while avoiding their limitations. It employs an explicitly defined orthonormal basis to represent compositions, ensuring the transformation is both bijective (one-to-one) and isometric. This means Aitchison distances and angles in the simplex accurately transform into standard Euclidean ones in the transformed space (Egozcue et al., 2003).

This mathematical rigor makes ILR particularly compatible with geostatistical methods like kriging that operate under Euclidean assumptions, explaining its general preference among practitioners working with compositional spatial data.

Why Is ILR with the Sequential Binary Partition (SBP) Matrix a Good Approach? The SBP approach significantly enhances the utility of ILR transformations for kriging applications through several key advantages.

As described by Pawlowsky-Glahn et al. (2015), an SBP establishes a hierarchical structure for compositional components. The process involves systematically dividing components into two groups, then further subdividing each group until all groups contain just a single component. This hierarchical partitioning can be conceptualized either as a division process (starting with all components together) or as a fusion process (starting with each component separate) .

This approach provides considerable analytical benefits. First, the SBP guarantees that the resulting ILR coordinates form an orthonormal basis, that is, they are both independent of one another and properly normalized. As Pawlowsky-Glahn and Buccianti (2011) note, this property simplifies subsequent statistical analyses like kriging because the coordinates fully conform to Euclidean assumptions.

Second, SBP creates interpretable hierarchies by organizing components into binary groups following logical or theory-driven structures. Each ILR coordinate derived through this method represents a meaningful balance (or contrast) between component groups, making results more interpretable within the context of underlying geological or environmental processes (Pawlowsky-Glahn and Buccianti, 2011).

Perhaps most importantly, SBP can be tailored to specific research contexts by incorporating prior knowledge. By designing partitions based on geological understanding, environmental characteristics, or other domain expertise, analysts can ensure that the resulting transformed variables reflect meaningful relationships in the data (Pawlowsky-Glahn and Buccianti, 2011).

Having established the advantages of using the SBP matrix for ILR transformation, namely, its ability to create interpretable, orthonormal coordinates tailored to domain knowledge, the next step is to describe how the SBP is constructed in this study. The following section details the systematic procedure for generating an SBP matrix, including both division and fusion perspectives, and explains how these partitions are used to compute ILR coordinates for compositional data analysis.

3.4. Sequential Binary Partition and ILR Transformation

The SBP represents a systematic approach to creating a hierarchical structure for compositional components. As explained by Pawlowsky-Glahn et al. (2015), this process involves methodically dividing components into groups through a series of binary splits. The construction follows a specific

procedure:

First, all compositional parts are organized into two distinct groups, which are typically denoted using +1 and −1 signs in a sign code matrix. This initial division represents the first order of the hierarchy (Pawlowsky-Glahn et al., 2015). In subsequent steps, each previously formed group undergoes further subdivision into two new groups. This hierarchical division continues until all resulting groups contain exactly one component.

An alternative but equivalent perspective describes SBP as a fusion process rather than division. As Pawlowsky-Glahn et al. (2015) note, one can start with D individual groups (each containing a single part) and progressively merge them. The first step combines two of these single-part groups, resulting in $D - 1$ groups. Through successive merging steps, the number of groups continuously decreases until only one group remains, containing all D parts. This fusion process requires exactly $D - 1$ steps, matching the number of steps in the division approach.

Each step in the SBP process, whether viewed as division or fusion, corresponds to the creation of one ILR coordinate, often called a "balance" (Pawlowsky-Glahn and Buccianti, 2011). These balances typically have meaningful interpretations within the context of the data being analyzed. The binary nature of the partitioning ensures that each resulting coordinate represents a contrast between two distinct groups of components.

It is important to note that the ordering of ILR coordinates produced through SBP is inherently flexible. As Pawlowsky-Glahn et al. (2015) point out, the rows of the associated contrast matrix can be rearranged, resulting only in a permutation of the indices of coordinates or balances without altering their fundamental properties.

If the geochemical knowledge between elements presented in Section 2.3 was exhaustive enough that it could be used to create an SBP matrix for the ILR transformation, it would have resulted in a comprehensive SBP that considers all the available relationships between elemtal groups. However, given the overlapping nature of the classifications in 2.3, alternative approaches need to be considered. One such approach suggested by Dr. Vera Pawlowsky-Glahn is grouping based on the principal components of the CLR transformed data. This was accomplished using the software "CoDaPack" (Comas-Cufí and Thió-Henestrosa, 2011). The following SBP generation and ILR transformation processes are adapted from the book chapters by Pawlowsky-Glahn et al. (2015) and Pawlowsky-Glahn and Buccianti (2011).

3.4.1. Hierarchical Partition Creation and Resulting SBP Structure

The partitioning process follows a hierarchical approach, beginning with fundamental principal component divisions and progressively adding smaller and smaller partitions (Thio-Henestrosa and Comas, 2016):

Conditions for a valid SBP

Each partition must:

- Contain both positive and negative elements
- Be linearly independent from all other partitions

The composition was imported into CoDaPack, and CLR biplots were created for the entire dataset. For each biplot, the elements were separated into two mutually exclusive groups based on their positions on either side of the Principal Component 2 (PC2) axis. These groups were then used to create further biplots, as illustrated in Figure 4.3.

Initially, the whole composition was split into two groups, the metallic elements and the "Other" non-metallic elements column. This partition forms the basis for calculating ilr.1. Then the group of metallic elements were then further split according to their separation along PC2 in the first biplot, using new biplots generated for each subgroup. At each step, the largest group resulting from a split was selected for further partitioning. This process continued recursively until all final groups in one branch contained only a single element (see Fig 4.4).

The order in which the groups are split determines the order of the ILR coordinates: ilr.2 follows ilr.1, ilr.3 follows ilr.2, and so on. Since a biplot requires at least three elements, the final splits also considered separation along Principal Component 1 (PC1) when necessary (when there is only 3 elements in the biplot and the last split needs to be created) (see Fig 4.3f). The CLR biplots shown in Figure 4.3 represent different stages of this sequential partitioning, and the PC1 and PC2 axes in each subplot correspond to the specific group being analyzed at that stage and are therefore not the same.

The resulting SBP matrix:

- Encapsulates statistical relationships between elements and preserves ratios.
- Provides interpretable partitions
- Serves as a robust basis for compositional data analysis
- Enhances the interpretability of subsequent statistical analyses

This approach demonstrates how domain-specific knowledge can potentially be systematically incorporated into the statistical framework of compositional data analysis. In the presence of exhaustive data on geochemical relationships, when structuring the SBP according to established geochemical relationships, each resulting ILR coordinate would retain interpretability in terms of known geological processes affecting Fe-Mn crust formation and evolution.

After describing the hierarchical partitioning process and the resulting SBP structure, it is important to clarify how these partitions are used to compute the ILR coordinates for compositional data analysis.

3.4.2. Calculation of ILR Coordinates from an SBP

Given a composition $x = (x_1, x_2, \ldots, x_D)$, where the D parts sum to a constant, the SBP matrix defines a series of $D - 1$ binary splits of the composition into two groups at each step. Each row of the SBP matrix corresponds to one partition, with +1 and −1 indicating membership in the two groups (numerator and denominator, respectively), and 0 indicating exclusion from the current split.

For each partition, an ILR coordinate, also called a *balance*, is calculated (in this study, the terms "ILR coordinate" and "balance" are used interchangeably.). Specifically, for the jth partition, let x_+ denote the group of r parts marked +1 and x_- the group of s parts marked −1. The jth balance is then defined as:

$$b_j = \sqrt{\frac{rs}{r+s}} \ln\left(\frac{gm(x_+)}{gm(x_-)}\right) \tag{3.1}$$

where $gm(x_+)$ and $gm(x_-)$ are the geometric means of the parts in the +1 and −1 groups, respectively.

Each balance b_j thus represents the log-ratio between the geometric means of the two groups defined by the corresponding SBP partition. Collectively, the $D - 1$ balances form an orthonormal coordinate system for the simplex, ensuring that the ILR coordinates are unit-norm, orthogonal log-contrasts. This construction provides a clear interpretation: each ILR coordinate quantifies the relative abundance of two groups of elements as determined by the SBP.

In this study, the ILR transformation was applied to the compositional data using the SBP structure

resulting from the methodology described above. The resulting ILR coordinates are then used for subsequent statistical analyses, including outlier detection and winsorization. Figure 4.4 illustrates how each group is recursively split into subgroups until only single elements remain, visually representing the hierarchical structure underlying the ILR transformation.

3.4.3. Interpretation of ILR coordinates

An SBP divides the components of a composition into two groups at each step, assigning them codes of +1 (numerator) and −1 (denominator). Each ILR coordinate, also called a balance, is constructed as a normalized logarithm of the ratio between the geometric mean of the parts in the numerator group and the geometric mean of the parts in the denominator group. The normalization factor ensures comparability across balances. The sign of the balance indicates which group is relatively more prominent: a positive value for the balance means the numerator group has a greater geometric mean in the composition, while a negative value means the denominator group is more dominant. Thus, balances provide a way to interpret the relative importance of grouped parts in a composition, with the SBP structure directly informing which components are being contrasted in each ILR coordinate (Pawlowsky-Glahn et al., 2015). For example, if an ILR coordinate has a positive correlation with another variable, say water depth in this case, that means that its value increases with depth. That is, based on the definition of the ILR coordinate, the ratio of the numerator group to the denominator increases with increase in depth, or to simplify, the numerator groups increase in the composition when compared to the denominator group as the depth increases.

3.4.4. Multivariate Outlier Detection

Pawlowsky-Glahn and Buccianti (2011) describes a robust approach for multivariate outlier detection in compositional data by first applying the ILR transformation, which properly accounts for the relative nature of compositional datasets. After transforming the data, robust estimators of location and covariance, such as those derived from the Minimum Covariance Determinant (MCD) method, are computed. These estimators are then used to calculate Mahalanobis distances for each observation, quantifying how far each sample deviates from the robust center of the data cloud. Observations with Mahalanobis distances exceeding a threshold, typically set at the 97.5th percentile of the chi-squared distribution, are flagged as multivariate outliers. This method ensures that outlier detection is sensitive to unusual ratios between components, rather than just extreme values in individual variables. The process highlights the importance of using transformations that respect the compositional structure of the data, as inappropriate transformations can lead to misleading conclusions. Identifying and addressing these outliers enhances the reliability and interpretability of subsequent analyses (Pawlowsky-Glahn and Buccianti, 2011).

This was the basis for outlier detection for the data used in this study. The process begins by loading data that has already been transformed using the ILR transformation, ensuring that the compositional structure is properly respected. Principal Component Analysis (PCA) is then applied to the ILR coordinates to facilitate clustering, and k-means clustering is used to identify groups of similar samples.

For each cluster, the Minimum Covariance Determinant (MCD) estimator is applied to robustly estimate the location and covariance matrix of the ILR-transformed data. Mahalanobis distances are computed for each observation within its cluster, quantifying how far each sample lies from the robust center. Observations with squared Mahalanobis distances exceeding the 97.5th percentile of the chi-squared distribution (with appropriate degrees of freedom) are flagged as outliers, following the thresholding approach described in the literature.

To address the influence of outliers, the code performs winsorization: outlier values are adjusted by shrinking them toward the robust cluster center along the direction of their deviation (Statistics How

To, 2024), scaled so that their Mahalanobis distance matches the cutoff. After winsorization, the code re-evaluates outliers using the original robust models and cutoffs, ensuring that the cleaning process is effective and does not introduce new anomalies. Winsorization was chosen rather than removal of outliers because of the already sparse number of data points and also to ensure that all available geochemical ratios are preserved.

This winsorized ILR transformed data was used for kriging.

3.5. Interpolation

There are numerous approaches that can be used for interpolation of geochemical grades onto the empty blocks. The ones focused in this study will be OK, SCK, ICCK and IDW.

3.5.1. Inverse Distance Weighted Estimation

Inverse Distance Weighted Estimation (IDW) interpolation operates under the principle that spatially close points are more similar than those further apart. When estimating a value at an unsampled location, IDW utilizes the measured values from surrounding points. The influence of each measured value on the prediction decreases as its distance from the prediction location increases. Thus, points closer to the prediction site are assigned greater weights, and these weights diminish with increasing distance. This approach ensures that each measured point's local influence is reflected in the estimation, with the weighting scheme directly tied to the inverse of the distance between points (ArcGIS, 2025). Standard IDW interpolation is significantly less computationally intensive when compared to kriging and it assumes isotropy which will be discussed in the following sections. IDW interpolation was conducted on both the untransformed cartesian coordinates as well as a transformed embedded coordinate system (3.5.4) that was used to account for anisotropy. The IDW interpolation in the embedded coordinates were finally chosen for consistency across methods and it will serve as a threshold for comparison of other more complicated methods of interpolation in this study.

Mathematical Formulation and Implementation of IDW

The IDW interpolation method is fundamentally based on the principle that spatial proximity implies greater similarity, reflecting Tobler's first law of geography. In IDW, the similarity between points is modeled as a function that decreases with distance, such that the influence of each known sample on the estimated value diminishes as the distance increases. The method is particularly effective when sample points are evenly distributed across the study area, and it allows for the adjustment of the influence of neighboring points through a power parameter (Emmanuel Romaric et al., 2020).

The estimated value at an unsampled location u, denoted as $Z^*(u)$, is computed as a weighted sum of the values at n surrounding sample points u_i:

$$Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(u_i) \tag{3.2}$$

Here, $Z(u_i)$ represents the measured value at location u_i, and λ_i is the weight assigned to each sample point. The weights are determined by the inverse of the distance between the estimation location and each sample point, raised to the power p:

$$\lambda_i = \frac{1}{d_i^p} \Bigg/ \sum_{j=1}^{n} \frac{1}{d_j^p} \tag{3.3}$$

where d_i is the Euclidean distance between the estimation location u and the sample point u_i, and p is the power parameter that controls the rate at which the influence of a sample point decreases with distance. A higher value of p increases the influence of closer points and reduces the effect of those farther away, while a lower p distributes the weights more evenly among all neighbors. The sum of all weights is constrained to equal one:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{3.4}$$

In this study, the IDW was implemented by calculating the distances between each estimation location and all sample points, applying the chosen power parameter to control the weighting, and normalizing the weights so that their sum equals one. The final estimate at each location was then obtained as the weighted sum of the measured values, following the equations above (Emmanuel Romaric et al., 2020).

3.5.2. Kriging

Kriging is a geostatistical estimation technique designed to provide the best linear unbiased estimate (BLUE) of a regionalized variable, such as the mean grade within a block of a mineral deposit. The primary objective is to estimate in situ resources prior to the application of any cut-off, distinguishing this from the estimation of recoverable reserves, which depends on technical and economic constraints.

The method operates under the assumption of second-order stationarity, where the mean and covariance (or variogram) of the variable are either known or can be reliably inferred. Kriging uses available sample data, which in the context of mineral resources are typically grades measured at discrete locations, and combines them linearly with weights that are determined to ensure two key properties: unbiasedness and minimum estimation variance. The unbiasedness condition requires that the expected value of the estimator equals the unknown mean, while the weights are optimized to minimize the estimation variance.

The estimation process involves solving a system of equations derived from these conditions, resulting in a set of weights that depend on the spatial configuration of the data and the underlying spatial structure as described by the variogram or covariance function. Kriging can be extended to more complex scenarios, such as universal kriging for non-stationary cases and Co-Kriging when multiple correlated variables are available. The technique is widely applicable and forms the foundation for practical resource estimation in mining, enabling the estimation of thousands of blocks efficiently and with quantifiable uncertainty (Journel and Huijbregts, 1976).

While traditional kriging assumes a stationary and isotropic spatial structure, recent advances have enabled the method to accommodate more complex spatial relationships, such as locally varying anisotropy (Dias and Deutsch, 2022).

Boisvert and Deutsch (2011) describe a novel kriging methodology for handling locally varying anisotropy (LVA) through Landmark ISOMAP embedding (L-ISOMAP). Their approach follows these key steps (3.2):

1. Generate an LVA field to characterize spatial relationships;
2. Select strategic landmark points and calculate shortest path distances between sample, between samples, landmarks, ans blocks using Dijkstra's algorithm;
3. Apply L-ISOMAP to transform the grid cells into a higher-dimensional Euclidean space where straight-line distances preserve the original shortest path relationships;

4. Develop an isotropic variogram model in this transformed space, which is valid because the embedding process has effectively incorporated the anisotropy information into the coordinate system itself;
5. For estimation at unsampled locations, determine distances between data points and the estimation location, convert these distances to covariances using the modeled variogram, and solve the kriging system of equations.
6. Repeat this process for all 30 ILR coordinates for OK.

This approach was followed for Ordinary Kriging (OK), Simple Co-Kriging (SCK), and Intrinsic Collocated Co-Kriging (ICCK) in this study. The following section goes into further details of how LVA was tackled using this approach in this study.

3.5.3. Locally Varying Anisotropy in Geostatistical Modeling

Traditional geostatistical methods assume second-order stationarity, where spatial continuity remains constant throughout the modeling domain (Ghosh et al., 2024). However, this assumption proves inadequate when dealing with geological formations exhibiting varying spatial continuity patterns across locations as such geological features often exhibit complex spatial relationships that vary locally, which is visible on the plots of the tropic seamount (4.7), necessitating the incorporation of LVA. LVA addresses this limitation by accommodating location-dependent covariance structures, effectively treating anisotropy as a "trend in the variogram." The need for LVA in kriging arises from natural geological processes that create directional features whose orientation and magnitude change across the modeling domain, including aligned structures and varying geological processes. Implementing LVA in kriging workflows enhances estimation accuracy by properly accounting for local geological features, resulting in models that better reflect real-world spatial relationships and produce more geologically realistic representations of subsurface conditions (Ghosh et al., 2024). According to Ghosh et al. (2024), three primary approaches exist for integrating LVA into geostatistical modeling. The first approach employs hard boundaries between domains with distinct anisotropy characteristics, suitable when variations occur abruptly. The second method involves local reorientation of the variogram, appropriate when anisotropy changes occur at scales larger than sample spacing. The third approach incorporates Shortest Path Distance (SPD) calculations, essential when anisotropy varies significantly at the sample spacing scale. This last method is particularly valuable for modeling complex geological structures such as folds, veins, and channels where traditional Euclidean distances fail to capture true spatial relationships between points. The methodology that utilizes SPD proposed by Ghosh et al. (2024) was utilized for extracting LVA parameters from surface topography and it is as follows:

Methodology

1. **Surface Extraction:** The surface topography is extracted from the Block model containing spatial coordinates (x, y, z) by identifying the maximum elevation (z) at each (x, y) location using Pandas' `groupby` and `max` functions. This creates a regular grid representation of the terrain through NumPy's `meshgrid` function, enabling efficient spatial indexing.
2. **Gradient Calculation:** After applying Gaussian smoothing (`scipy.ndimage.gaussian_filter`) to reduce noise (controlled by the sigma parameter), the first and second derivatives of the surface are calculated using NumPy's `np.gradient` function, which implements finite difference methods for numerical differentiation.
3. **Full 3D Parameterization:** All six LVA parameters are derived from the gradient information using NumPy's array operations:
    - Strike (α): Azimuthal angle from north, calculated as $(\pi/2 - \text{secondary\_angle}) \bmod 2\pi$ using `np.arctan2` for vector direction

- Dip (β): Angle of the bedding plane, derived from the gradient magnitude using `np.arctan` of the gradient magnitude
- Plunge (θ): Angle of linear structure, approximated using surface curvature through second derivatives
- **Anisotropy Ratios:** Two critical ratios are calculated using NumPy's vectorized operations:
  - r_1: Ratio between minor and major axes (0.2–0.5, inversely proportional to gradient magnitude)
  - r_2: Ratio between vertical and major axes (0.1–0.3, inversely proportional to gradient magnitude)

4. **Graph Construction:** A `Networkx` graph is created to represent the spatial structure with anisotropic properties:
   - **Node Creation:** Each point in the LVA field is represented as a node with attributes containing all six LVA parameters. For computational efficiency, a sparse subset of the full LVA field is selected using systematic sampling.
   - **Edge Creation:** Edges are established between neighboring nodes (8-connectivity pattern) with weights computed using anisotropic distance calculations. These calculations transform coordinates through three rotation matrices (for strike, dip, and plunge) before computing the elliptical distance that accounts for directional variability.
   - **Distance Calculation:** Shortest path distances between sample locations (from geochemical sample data with spatial locations) are computed using Dijkstra's algorithm (`Networkx.shortest_path`) to account for the anisotropic nature of the spatial relationships.

This approach follows the gradient-based methodology and extends it to include full 3D parameterization as outlined in Ghosh et al. (2024). The implementation leverages SciPy's spatial data structures (`KDTree`) for efficient nearest neighbor queries and NumPy's vectorized operations for performance. The graph-based representation enables the calculation of geologically meaningful distances that respect anisotropic spatial relationships, critical for subsequent multidimensional scaling and kriging operations. This is ideal for the kriging estimations on the Fe-Mn crusts since the surface has a complex topography and the samples are in roughly 4 varying directions.

3.5.4. Shortest Path Distance Calculation for Complex Geology

When anisotropy varies within local search neighborhoods, as is the case for the Tropic seamount where the samples are spaces hundreds of metres apart, and the surface of the seamount has complex topology in these spaces, traditional distance metrics inadequately represent spatial relationships. Ghosh et al. (2024) describes how the SPD methodology addresses this limitation by calculating distances that follow the geological structure rather than straight-line Euclidean paths. This approach considers the proper relationship between points in folded or curved geological environments, where points that appear distant in Euclidean space may actually be closely related geologically. The implementation utilizes Dijkstra's algorithm on a graph where distances between segments are adjusted according to local anisotropy specifications from the LVA field. For the kriging workflow, exhaustive secondary data from block models provides the foundation for generating the LVA field, as it offers comprehensive spatial coverage necessary for accurate structural representation. While non-Euclidean distances can create mathematical challenges in kriging equations, multidimensional scaling (MDS) technique suggested by Boisvert and Deutsch (2011) effectively embed samples in a high-dimensional space where Euclidean distances appropriately reflect the original shortest path relationships. This is shown in Figure 3.2. The ideal method to embed the blocks in the higher dimensional space is to:

- Calculate the SPD between all sample points using Dijkstra's algorithm.
- Create landmark points based on clustering or a deterministic method. A uniform grid of landmarks were generated based on the x,y extents of the data in this study to reduce computation time and increase repeatability.
- Calculate SPD between sample points and landmarks as well as between each landmarks.
- Identify the location of all blocks using trilateration by calculating the SPD from that block to all land marks.
- Then embed the landmarks, samples, and blocks in the multidimensional space using `sklearn`'s MDS function. However this leads to a time complexity of O(N x L) where N is the number of blocks in the model and L, the number of landmarks in the graph (49), which is far too high for the numerous blocks in the block model. therefore, `Networkx`'s `"nx.single_source_dijkstra_path_length"` functionality was utilized to calculate the SPD between all landmarks and all nodes in the LVA field. Since each block is assigned to its closest node and the nodes are significantly dense in the graph, this allows for an accurate approximation of the SPD between blocks and Landmarks. This reduces the time complexity to O(L) which is orders of magnitude faster than the O(NxL).
- The final result is a block model (4.9) in which the samples and blocks are embedded in a higher dimensional space (24 dimensions in this case) in which a single isotropic variogram can be modelled without worrying about anisotropy.

3.5.5. Variogram Analysis and Modeling

Variogram analysis is performed in the embedded space to quantify spatial autocorrelation and develop appropriate covariance models:

Empirical Variogram Calculation

- Variograms are calculated in the embedded space for each ILR coordinate as described by Boisvert and Deutsch (2011)
- This approach inherently respects the anisotropic nature of the deposit
- Distance bins are optimized to ensure sufficient data pairs for reliable statistics

The empirical semivariogram $\gamma(h)$ is calculated using:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \tag{3.5}$$

where $N(h)$ is the number of data pairs separated by distance h, $z(\mathbf{u}_i)$ is the value at location $\mathbf{u}_i$, and $z(\mathbf{u}_i + \mathbf{h})$ is the value at location $\mathbf{u}_i + \mathbf{h}$.

Variogram Model Fitting

- Multiple parametric theoretical variogram models were tested (exponential, spherical, Gaussian, Matérn) one by one.
- Parameters are optimized using weighted least squares fitting
- The best-fitting model is selected based on goodness-of-fit metrics
- Models are validated through cross-validation techniques

The theoretical variogram models implemented include:

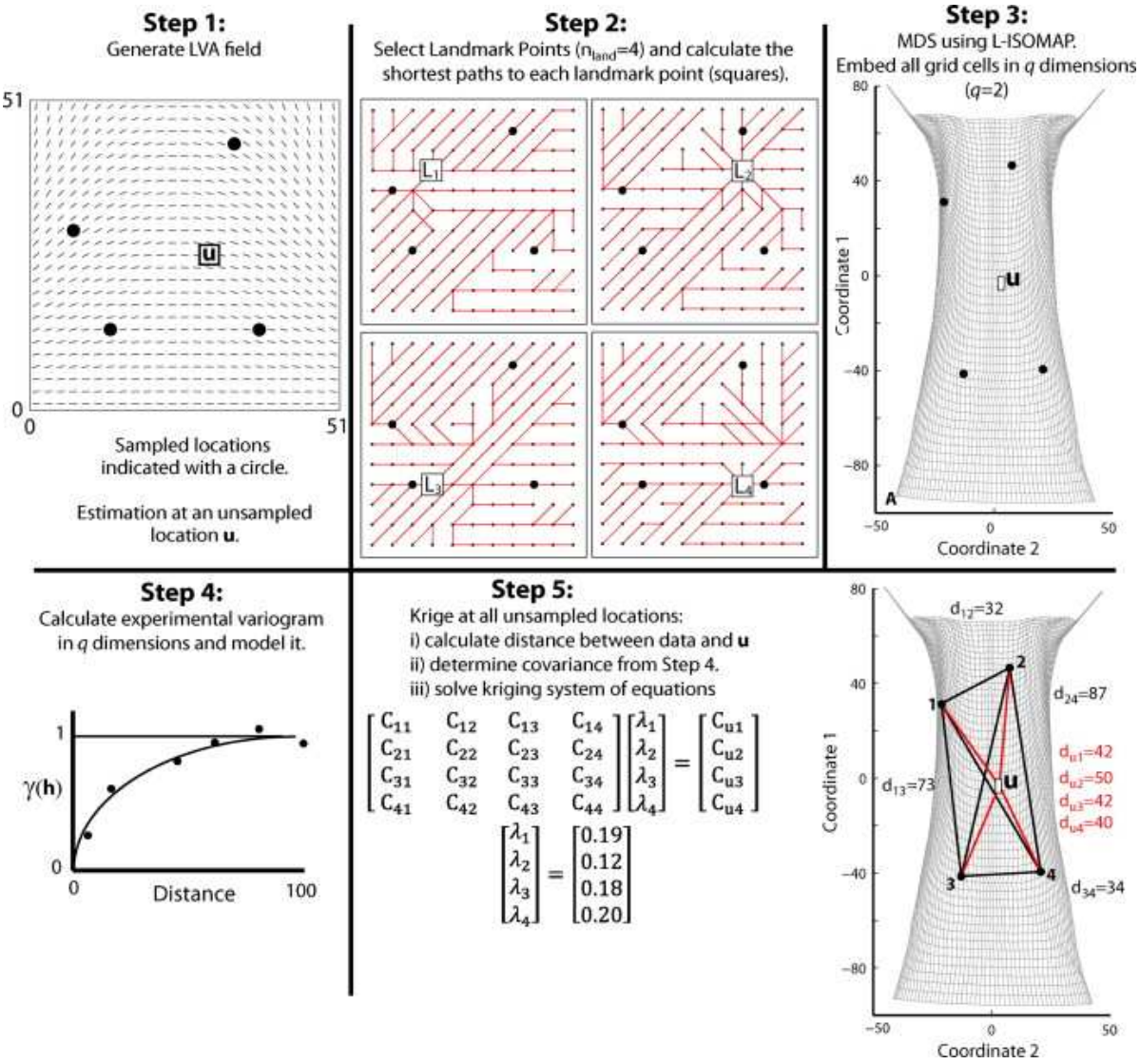


Figure 3.2: Steps for Kriging with LVA proposed by Boisvert and Deutsch, 2011.

Spherical model:(Samson and Deutch, 2025)

$$\gamma(h) = \begin{cases} c_0 + c\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{for } h \leq a \\ c_0 + c & \text{for } h > a \end{cases} \tag{3.6}$$

Exponential model:(Samson and Deutch, 2025)

$$\gamma(h) = c_0 + c\left[1 - \exp\left(-\frac{3h}{a}\right)\right] \tag{3.7}$$

Gaussian model:(Samson and Deutch, 2025)

$$\gamma(h) = c_0 + c\left[1 - \exp\left(-\frac{3h^2}{a^2}\right)\right] \tag{3.8}$$

Matérn model:(Pardo-Iguzquiza and Chica-Olmo, 2008)

$$\gamma(h) = c_0 + c\left[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{a}\right)^{\nu} K_{\nu}\left(\frac{h}{a}\right)\right] \tag{3.9}$$

where c_0 is the nugget effect, c is the sill, a is the range parameter, ν is the smoothness parameter, Γ is the gamma function, and K_ν is the modified Bessel function of the second kind of order ν.

The weighted least squares objective function for model fitting is:

$$\min_{c_0,c,a} \sum_{i=1}^{N} w_i[\gamma_{\text{emp}}(h_i) - \gamma_{\text{model}}(h_i)]^2 \tag{3.10}$$

where $w_i = \frac{1}{h_i}$ are the weights giving higher importance to shorter distances.

Non-Parametric Variogram Modeling

Limitations of Parametric Variogram Models:Traditional parametric variogram models (Gaussian, spherical, exponential etc.) are defined by a fixed mathematical structure that assumes monotonically increasing behavior until reaching a sill. The general form of a parametric variogram can be expressed as:

$$\gamma(h) = \sigma^2\{1 - \exp(-(ah)^m)\} \tag{3.11}$$

where σ^2 represents the sill, a is a range parameter, and m controls the shape (with $1 \leq m \leq 2$). When $m \to 1$, the model approaches an exponential form with more linear behavior at small lags. When $m \to 2$, it approaches the Gaussian form with a more quadratic shape near the origin (Weku et al., 2019).

However, these parametric models impose strict structural constraints that sometimes cannot adequately capture extremely complex spatial correlation patterns, particularly when the spatial correlation exhibits non-monotonic behavior.

Non-Monotonic Spatial Correlation: The "Hole Effect":The geospatial data analyzed in this study exhibits a phenomenon known as the "hole effect," where the spatial correlation does not decrease monotonically as lag distance increases (Black, 2020). Instead, the variogram reaches a maximum, then

decreases to a local minimum before rising again. This pattern indicates a complex spatial structure (Weku et al., 2019).

The traditional parametric models cannot represent this behavior, as they are mathematically constrained to monotonic increases. Using such models could force an inappropriate structure onto the data, potentially leading to incorrect estimations (Weku et al., 2019).

Non-Parametric Bessel Variogram Model:To try to address this limitation, a non-parametric variogram model based on Bessel functions was implemented, which can represent both monotonic and non-monotonic spatial correlation patterns. According to Bochner's theorem, a covariance function $C(h)$ is positive definite if and only if it can be expressed as (Weku et al., 2019):

$$C(h) = \int_0^\infty \Omega_d(th)F(dt) \tag{3.12}$$

where $\Omega_d(x) = (2/x)^{(d-2)/2}\Gamma(d/2)J_{(d-2)/2}(x)$ are basis functions in $\mathbb{R}^d$ (Real coordinate pace in d dimensions).

For two-dimensional data ($d = 2$), this leads to a covariance function $C(h) = J_0(bh)$, where J_0 is the first-kind Bessel function of order zero (Weku et al., 2019). The resulting non-parametric variogram model takes the form:

$$\gamma(h) = \sigma_\epsilon^2 + \sigma^2\{1 - \sum_{k=1}^{p} \alpha_k J_0(kb \cdot h/\omega)\} \tag{3.13}$$

where:

- σ_ϵ^2 is the nugget effect
- σ^2 is the sill (partial sill when nugget is present)
- p is the number of basis functions (typically $p \leq 5$)
- α_k are weights with $\sum_{k=1}^{p} \alpha_k = 1$ and $\alpha_1 > \alpha_2 > ... > \alpha_p$
- b controls the frequency of oscillations
- ω controls the scale of the hole effect

This model provides more flexibility compared to traditional parametric variograms while maintaining mathematical validity as a legitimate variogram function.

As noted by Weku et al. (2019), the non-parametric approach is particularly suitable for capturing complex spatial structures in geological and environmental data. At the same time it sometimes fails to capture the behavior at lag distances close to zero which is very important for geostatistical estimations.

3.5.6. Nested Variogram Models

Both parametric and non-parametric models have thier strenthgs and weaknesses. However,the spatial continuity of geological phenomena sometimes does not follows a single, simple pattern. Instead, these phenomena often represent complex processes operating at multiple scales simultaneously. This complexity could be addressed through nested variogram models (Mundim et al., 1999), which linearly combine different variogram structures to capture multi-scale spatial variability.

This approach was formalized in 1982 through Factorial Kriging Analysis (FKA), which established the theoretical foundation that regionalized phenomena can be decomposed into independent sub-

phenomena operating at different scales. Each sub-phenomenon is characterized by its own variogram structure, and when these structures are combined linearly, they form the comprehensive variogram model of the entire phenomenon (Mundim et al., 1999).

The mathematical formulation of a nested variogram model follows:

$$\gamma(h) = \sum_{i=1}^{n} \gamma_i(h) \tag{3.14}$$

where $\gamma(h)$ is the total variogram, and $\gamma_i(h)$ represents individual variogram structures.

As a concrete example, consider a mineral deposit where the spatial variability might be modeled as:

$$\gamma(h) = c_0 + c_1 \cdot \text{Sph}\left(\frac{h}{a_1}\right) + c_2 \cdot \text{Gau}\left(\frac{h}{a_2}\right) \tag{3.15}$$

This combined model includes:

- A nugget effect (c_0) representing microscale variability (Maptek Pty Ltd, 2024)
- A spherical structure with sill c_1 and range a_1 capturing medium-scale phenomena, showing linear behavior near the origin before steeply increasing and gradually flattening (Maptek Pty Ltd, 2024)
- A Gaussian structure with sill c_2 and range a_2 representing larger-scale regional trends with parabolic behavior near the origin and smooth approach to the sill, ideal for modeling high spatial continuity (Maptek Pty Ltd, 2024)

Mundim et al. (1999) emphasize that understanding the underlying geological processes helps identify appropriate component structures and enhances the model's interpretability, especially when the experimental variogram shows complex patterns where nesting is not immediately obvious.

Thus, nested variogram models might offer not just better statistical fits but also more geologically meaningful representations of spatial variability across multiple scales, leading to more accurate estimation and simulation of spatially distributed phenomena and hence it was decided to use a variogram fitting function that chooses the best fitting model from both parametric variograms, linear combination of a parametric variogram and hole effect, linear combination of multiple parametric variograms and hole effect up to a set number of components. This flexible function would choose the best fit for each ILR variable based on Goodness of Fit (GoF) metrics like Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). This will give the best of both parametric and non parametic spatial correlations.

The ability to accurately model this combined behavior could be critical for subsequent kriging estimates, as it ensures that the spatial correlation structure used for prediction correctly represents the true underlying spatial dependencies in the data set (Weku et al., 2019). This improved modeling of spatial structures could translate into more accurate predictions and more reliable uncertainty quantification in geostatistical analysis.

It must be noted that the goodness-of-fit metrics and visual fit are only an indication of how well the variogram model fits the empirical variogram. However, what it does not ensure is stable kriging and accurate estimates. Therefore, even though the nested variogram approach might give a better fitting variogram, it could also not necessarily lead to a more stable kriging implementation.

In summary, the adoption of nested variogram models was motivated by the need to provide the kriging implementation with sufficient flexibility to capture the complex, non-monotonic spatial correlation structures observed in the data. These patterns are often not adequately represented by traditional single-structure parametric models due to their inherent mathematical limitations (Weku et al., 2019). While the flexible variogram fitting function allowed for both nested and single-structure models, the final selection of the variogram model for each variable was ultimately based not only on goodness-of-fit and visual inspection, but also on the stability and quality of kriging predictions as assessed by QKNA metrics (see 3.5.11) to reach a balance between accuracy and stability.

Cross-Correlation Analysis

- Correlations between primary variables (ILR coordinates) and secondary variable (depth) are quantified
- These correlation coefficients are essential inputs for the Markov Model II (MM II) approach

The Pearson correlation coefficient is calculated as (Samson and Deutch, 2025):

$$\rho_{z,y} = \frac{\mathrm{Cov}(Z,Y)}{\sigma_Z \sigma_Y} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(z_i - \bar{z})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.16}$$

where Z represents the primary variable (ILR coordinate) and Y represents the secondary variable (depth).

3.5.7. Leave-One-Out Cross-Validation

A comprehensive validation procedure is implemented to assess estimation performance:

Sequential Validation Process

- Each data point is temporarily removed from the dataset
- Estimation is performed at the removed location using three methods:OK, SCK and ICCK
- Estimated values are compared with actual values

Error Metrics Calculation

- Root Mean Square Error (RMSE) is calculated for each method and each ILR coordinate
- Mean Absolute Error (MAE) is also computed to assess estimation accuracy
- Normalized Root Mean Square Error (NRMSE) is used to compare errors across different scales defined as RMSE/range (Statistics How To, 2023b).
- Mean Absolute Percentage Error (MAPE) is calculated to express errors as percentages (Statistics How To, 2023a).
- Results are aggregated to evaluate overall performance of each method

The error metrics are defined as (Samson and Deutch, 2025):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{z}_i)^2} \tag{3.17}$$

(Samson and Deutch, 2025)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|z_i - \hat{z}_i| \tag{3.18}$$

(Statistics How To, 2023b)

$$\text{NRMSE} = \frac{\text{RMSE}}{z_{\max} - z_{\min}} \tag{3.19}$$

(Statistics How To, 2023a)

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{z_i - \hat{z}_i}{z_i} \right| \tag{3.20}$$

where z_i is the actual value and $\hat{z}_i$ is the estimated value.

Comparative Analysis

- Performance metrics are compared across methods to determine the most appropriate approach
- Visualizations of error distributions help identify potential biases or limitations

3.5.8. Limitations of LOOCV and Interpretation of Kriging vs. IDW Performance

Certain spatial configurations of sampling data can cause kriging to assign weights in ways that seem unexpected, leading to potential performance issues. These irregularities are rooted in data redundancy and manifest through two main phenomena: String Effect (the linear clustering of samples) and the occurrence of negative weights (Markvoort and Deutsch, 2024).

String Effect: When samples form a narrow linear pattern, either due to geological boundaries or limited search ellipses, the estimation algorithm tends to give more influence to data points at the ends of the sequence. These end points lack neighboring samples on one side, causing kriging to treat them as more informative about the surrounding volume. In contrast, interior points, despite their seemingly central position, are more redundant and thus receive lower weights. This imbalance can skew the estimation output, particularly if the end-point values differ significantly from the average trend.

Negative Weights: A second issue arises when samples located farther from the estimation point are effectively overshadowed by nearer ones. This screening effect can result in negative weights, which, while mathematically valid and reflect redundancy and help kriging maintain an unbiased estimate (Martin, 2020), might amplify subtle or marginal data patterns. Such effects are most prominent when the data are closely spaced and covariance relationships remain positive. Simpler interpolation methods like IDW does not account for how samples are positioned relative to one another, they only factor in their distance to the point being estimated and is not as much affected by this redundancy (Martin, 2020).

In addition to the data-related phenomena described previously, another factor that can influence the perceived performance of kriging versus simpler methods such as IDW is the way performance metrics are computed, specifically through LOOCV. Despite its widespread use, LOOCV has notable limitations when applied to kriging models, which can bias comparisons with methods like IDW.

LOOCV calculates prediction error by successively removing one sample and estimating its value using the remaining data. In cases where the dataset includes clusters of nearby samples, such that each point has a close neighbor, removing one does not significantly impact the quality of the prediction. This scenario introduces a negative bias in LOOCV, as the estimator still performs well without that sample. The resulting error appears lower than it truly is, potentially making distance-based methods like IDW look favorable because they primarily rely on spatial proximity (Pronzato and Rendas, 2024).

In contrast, when data points are more widely spaced, leaving one out results in insufficient similar samples for an accurate prediction. Here, LOOCV tends to overestimate the true prediction error, introducing a positive bias (Pronzato and Rendas, 2024). This can make kriging appear less reliable

than it actually is, especially given that kriging's model structure depends on both sample proximity and their spatial arrangement.

Another technical concern is the inconsistency in hyperparameter settings. In traditional LOOCV implementations for kriging, the model is re-trained independently for each omitted sample. Because the optimal variogram parameters or model settings may differ for each reduced dataset, the estimated prediction errors might not reflect the behavior of the full model trained on the entire data. This discrepancy introduces additional bias and undermines the reliability of LOOCV in evaluating kriging's performance (Pang et al., 2023).

These compounding issues may lead to situations where IDW seems to outperform kriging in terms of cross-validation metrics especially in clustered or irregular spatial configurations despite kriging offering theoretically superior and more robust estimations when applied with full spatial context and appropriately validated model parameters.

3.5.9. Ordinary Kriging Implementation in Embedded Space

In most real-world applications, the mean of the regionalized variable is unknown. While one might consider estimating the mean and subtracting it from the data, which is an approach similar to detrending in time series analysis, this method can be problematic, as it may introduce theoretical uncertainties and obscure the distinction between large-scale trends and local fluctuations.

OK addresses this challenge by assuming that the mean is constant but unknown within the estimation neighborhood. Rather than relying on a separate mean estimation step, OK incorporates the mean as an unknown parameter directly into the kriging system. This approach ensures that the resulting estimator remains unbiased and achieves minimum variance, relying solely on the variogram to model spatial dependence.

OK is particularly well-suited to stationary models and is widely used in practice due to its simplicity and robustness. It serves as the foundational case for more general kriging methods, such as universal kriging, which allow for more complex mean structures. By optimally combining the available data and accounting for spatial correlation, OK provides reliable local estimates even when the mean is not explicitly known (Chilès and Delfiner, 2012). OK is performed for each ILR coordinate in the embedded space, leveraging the L-ISOMAP embedding and variogram models fitted in that space.

OK is the most widely used form of kriging, particularly when the global mean of the regionalized variable is not considered locally reliable. Unlike simple kriging, which assumes a known and constant mean across the entire domain, OK estimates the value at an unsampled location by implicitly determining a local constant mean within a search neighborhood. This approach ensures that the estimator remains unbiased without relying on the global mean, making it more robust in heterogeneous environments.

The OK estimator at an unsampled location $\mathbf{u}_0$ is given by (Ashtiani and Deutsch, 2020):

$$\hat{Z}_{OK}(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i(\mathbf{u}_0) Z(\mathbf{u}_i) + \left[1 - \sum_{i=1}^{n} \lambda_i(\mathbf{u}_0)\right] m \tag{3.21}$$

where $\lambda_i(\mathbf{u}_0)$ are the kriging weights assigned to each data value $Z(\mathbf{u}_i)$, and m is the unknown local mean. The weights are determined such that their sum equals one, ensuring unbiasedness.

The system of equations for OK, derived using the method of Lagrange multipliers to enforce the

unbiasedness constraint, is:

$$\sum_{j=1}^{n} \lambda_j(\mathbf{u}_0) C(\mathbf{u}_j, \mathbf{u}_i) + \mu = C(\mathbf{u}_i, \mathbf{u}_0), \quad i = 1, \ldots, n \tag{3.22}$$

$$\sum_{j=1}^{n} \lambda_j(\mathbf{u}_0) = 1 \tag{3.23}$$

where $C(\mathbf{u}_j, \mathbf{u}_i)$ is the covariance between data locations, $C(\mathbf{u}_i, \mathbf{u}_0)$ is the covariance between the data and the estimation location, and μ is the Lagrange multiplier.

OK minimizes the estimation variance while maintaining the unbiasedness constraint, resulting in weights that differ from those of simple kriging. The method is quasi-stationary, as it assumes the mean is constant only within the local neighborhood rather than across the entire domain. This makes OK particularly suitable for practical applications where local variations in the mean are expected (Ashtiani and Deutsch, 2020).

3.5.10. Co-Kriging Implementation

Co-Kriging is a geostatistical estimation technique that extends OK by incorporating not only the primary variable of interest, which may be sparsely sampled, but also one or more secondary variables that are more densely sampled across the domain. This approach allows for improved estimation of the primary variable at unsampled locations by leveraging the spatial correlation between the primary and secondary variables. According to Monteiro da Rocha et al. (2012), the main advantage of Co-Kriging is its ability to co-estimate poorly sampled variables using auxiliary information, which can lead to reduced estimation error variance and the simultaneous estimation of multiple attributes within the same spatial domain. Furthermore, Co-Kriging remains effective even when primary or secondary data are missing at certain locations, making it a flexible tool for practical applications.

However, the benefits of Co-Kriging is when there is a significant spatial correlation between the primary and secondary variables. If both variables are sampled everywhere or if there is little to no correlation between them, Co-Kriging offers little to no advantage over OK (Monteiro da Rocha et al., 2012).

Collocated Co-Kriging (CCK) is a specialized form of Co-Kriging designed for situations where the primary variable is sparsely sampled, but the secondary variable is available at every estimation location. This method simplifies the Co-Kriging system by only considering the secondary data at the estimation point, which helps to avoid matrix instability issues that can arise in ordinary Co-Kriging due to high autocorrelation among closely spaced secondary data. Additionally, CCK reduces redundancy by screening out the influence of distant secondary data, focusing instead on the most relevant local information. This makes CCK computationally efficient and particularly well-suited for large datasets or cases with heterotopic sampling patterns, where the spatial arrangement of primary and secondary data differs (Monteiro da Rocha et al., 2012).

The estimation process is implemented using OK, SCK and ICCK, following the framework described in Boisvert and Deutsch (2011) for handling locally varying anisotropy and efficient use of secondary information.

Linear Model of Coregionalization (LMC) and Its Role in SCK and ICCK

The Linear Model of Coregionalization (LMC) is a widely adopted approach for modeling the joint spatial continuity of two or more random variables (Babak and Deutsch, 2007). In the LMC framework, the direct and cross-variograms of all variables are represented as a linear combination of a common set of basic nested structures (such as spherical or exponential models). Each structure is associated with a

coregionalization matrix that quantifies the spatial covariance contributions for each variable and their cross-covariances. Mathematically, for two variables Z and Y, the LMC expresses the variograms as:

$$\gamma_{ZZ}(h) = \sum_{i=0}^{n} b_{i,ZZ}\, \gamma_i(h) \tag{3.24}$$

$$\gamma_{YY}(h) = \sum_{i=0}^{n} b_{i,YY}\, \gamma_i(h) \tag{3.25}$$

$$\gamma_{ZY}(h) = \sum_{i=0}^{n} b_{i,ZY}\, \gamma_i(h) \tag{3.26}$$

where $\gamma_i(h)$ are the basic variogram structures and $b_{i,ZZ}$, $b_{i,YY}$, $b_{i,ZY}$ are the sill (contribution) parameters for each structure. The LMC requires that all variables share the same set of structures, but allows the sill parameters to differ, subject to the constraint that each coregionalization matrix is positive semi-definite and all sills are non-negative.

In this study, the LMC was implemented according to these principles. Variogram models were first fitted to both the primary and secondary variables, and a common set of nested structures was established. For each structure, a 2×2 coregionalization matrix was constructed, with diagonal elements representing the direct sills and off-diagonal elements representing the cross-sills, initially set proportional to the observed correlation and then adjusted to ensure positive definiteness. This LMC framework was then used to compute all direct and cross-covariances required for both SCK and CCK.

The use of the LMC in the co-kriging procedures ensures that the spatial relationships between variables are modeled in a consistent and physically meaningful way. By enforcing a shared set of spatial structures and valid covariance matrices, the LMC provides a robust, albeit computationally intensive, foundation for multivariate geostatistical estimation, allowing secondary information to be leveraged to improve predictions of the primary variable.

Limitations of the LMC While the LMC is powerful, it is important to recognize its theoretical limitations. The LMC can only produce symmetric cross-covariance models, meaning that the cross-covariance between two variables at lag h is assumed to be the same regardless of the direction or order of the variables (Marcotte, 2012). Mathematically, this symmetry is expressed as $C_{12}(h) = C_{21}(h)$, where $C_{12}(h)$ is the cross-covariance between variable 1 at location u and variable 2 at location $u + h$, and $C_{21}(h)$ is the cross-covariance in the reverse direction. This leads to the simplified cross-variogram formula:

$$\gamma_{12}(h) = C_{12}(0) - C_{12}(h) \tag{3.27}$$

(Cuba and Deutsch, 2012), where $\gamma_{12}(h)$ is the cross-variogram at lag h, and $C_{12}(0)$ is the collocated cross-covariance.

Within this framework, the cross-correlation is always modeled to be highest at $h = 0$, i.e., when the two variables are measured at the same location. As the lag h increases, the cross-covariance $C_{12}(h)$ typically decreases, resulting in positive cross-variogram values. However, if the empirical cross-covariance at some lag h exceeds the collocated covariance $C_{12}(0)$, the cross-variogram $\gamma_{12}(h)$ becomes negative. This situation is not compatible with the LMC's assumptions, as it implies that the variables are more strongly correlated at a distance than at the same location which is a scenario that LMC cannot represent due to its requirement for maximum cross-correlation at $h = 0$ (Marcotte, 2012).

More generally, asymmetry in the cross-covariance, known as the 'lag effect', can occur when $C_{12}(h) \neq C_{21}(h)$, often due to geological processes that induce a spatial offset or phase shift in the relationship

between variables (Cuba and Deutsch, 2012). The LMC, by construction, cannot model such lag effects, and when negative cross-variogram values are observed at non-zero lags, it may indicate the presence of such effects or the absence of a detectable spatial cross-structure under the constraints of the LMC. In practice, to ensure positive definiteness required for valid Co-Kriging, the fitted cross-variogram model may be set flat at zero, reflecting the lack of meaningful cross-correlation structure within the LMC framework (Marcotte, 2012; Cuba and Deutsch, 2012). The SCK and ICCK implementations closely follow the approach described by Samson and Deutch (2024)

Simple Co-Kriging

Simple Co-Kriging (SCK) is a multivariate extension of kriging that incorporates both primary and secondary variables without requiring equal sampling densities. The estimation at an unsampled location $\mathbf{u}_0$ is given by

$$\hat{Z}(\mathbf{u}_0) = \sum_{\alpha=1}^{n} \lambda_{Z,\alpha} Z(\mathbf{u}_\alpha) + \sum_{\alpha=1}^{n_y} \lambda_{Y,\alpha} Y(\mathbf{u}_\alpha) \tag{3.28}$$

where $Z(\mathbf{u}_\alpha)$ are the primary data, $Y(\mathbf{u}_\alpha)$ are the secondary data, and $\lambda_{Z,\alpha}$, $\lambda_{Y,\alpha}$ are the kriging weights. The weights are determined by solving the following system:

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha} C_z(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{\alpha=1}^{n_y} \lambda_{Y,\alpha} C_{yz}(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_z(\mathbf{u}_\beta - \mathbf{u}_0), \quad \beta = 1, \ldots, n \tag{3.29}$$

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha} C_{yz}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{\alpha=1}^{n_y} \lambda_{Y,\alpha} C_y(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_{yz}(\mathbf{u}_\beta - \mathbf{u}_0), \quad \beta = 1, \ldots, n_y \tag{3.30}$$

where $C_z(\mathbf{h})$ is the covariance of the primary variable, $C_y(\mathbf{h})$ is the covariance of the secondary variable, and $C_{yz}(\mathbf{h})$ is the cross-covariance between primary and secondary variables. In practice, a Linear Model of Coregionalization (LMC) is used to ensure valid covariance models, structuring the covariance matrix in blocks:

$$\mathbf{K} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \tag{3.31}$$

where $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ are the primary and secondary covariance matrices, and $\mathbf{C}_{12} = \mathbf{C}_{21}^T$ are cross-covariance matrices, modeled using the LMC and scaled by the observed correlation coefficient. SCK provides optimal estimates that account for the correlation structure between variables, but requires fitting and interpreting complex cross-covariances, which can be challenging in practice.

Simple Co-Kriging with Markov Model II

SCK with Markov Model II (MM2) leverages the MM2 approach to relate the primary and secondary variables, assuming the secondary variable is more stable and exhaustively sampled. The estimation at an unsampled location $\mathbf{u}_0$ is given by

$$\hat{Z}(\mathbf{u}_0) = \sum_{\alpha=1}^{n} \lambda_{Z,\alpha} Z(\mathbf{u}_\alpha) + \lambda_{Y,0} Y(\mathbf{u}_0) \tag{3.32}$$

where $Z(\mathbf{u}_\alpha)$ are the primary data, $Y(\mathbf{u}_0)$ is the secondary variable at the estimation location, and $\lambda_{Z,\alpha}$, $\lambda_{Y,0}$ are the kriging weights. The weights are determined by solving the following system:

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha}\rho_z(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \lambda_{Y,0}\rho_{zy}(\mathbf{u}_\beta - \mathbf{u}_0) = \rho_z(\mathbf{u}_\beta - \mathbf{u}_0), \quad \beta = 1, \ldots, n \tag{3.33}$$

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha}\rho_{zy}(\mathbf{u}_\alpha - \mathbf{u}_0) + \lambda_{Y,0} = \rho_{zy}(0) \tag{3.34}$$

where $\rho_z(h)$ is the correlogram of the primary variable, and $\rho_{zy}(h)$ is the cross-correlogram between primary and secondary variables. The MM2 model expresses the primary correlogram as

$$\rho_z(h) = \rho_{zy}(0)^2\rho_y(h) + [1 - \rho_{zy}(0)^2]\rho_r(h) \tag{3.35}$$

where $\rho_y(h)$ is the correlogram of the secondary variable, $\rho_{zy}(0)$ is the collocated correlation coefficient, and $\rho_r(h)$ is a residual correlogram. SCK with MM2 is computationally efficient and straightforward, but may suffer from variance inflation due to not being a true intrinsic model.

Intrinsic Collocated Co-Kriging with Markov Model II

Intrinsic Collocated Co-Kriging (ICCK) with MM2 extends SCK by including both the secondary data at the primary locations and the collocated secondary value in the estimation. This approach corrects for variance inflation and more accurately reproduces the variance of the primary variable. The ICCK estimator is

$$\hat{Z}(\mathbf{u}_0) = \sum_{\alpha=1}^{n} \lambda_{Z,\alpha}Z(\mathbf{u}_\alpha) + \sum_{\alpha=1}^{n} \lambda_{Y,\alpha}Y(\mathbf{u}_\alpha) + \lambda_{Y,0}Y(\mathbf{u}_0) \tag{3.36}$$

The weights are obtained by solving the following system:

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha}\rho_z(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{\alpha=1}^{n} \lambda_{Y,\alpha}\rho_{zy}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \lambda_{Y,0}\rho_{zy}(\mathbf{u}_\beta - \mathbf{u}_0) = \rho_z(\mathbf{u}_\beta - \mathbf{u}_0), \quad \beta = 1, \ldots, n \tag{3.37}$$

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha}\rho_{zy}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{\alpha=1}^{n} \lambda_{Y,\alpha}\rho_y(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \lambda_{Y,0}\rho_y(\mathbf{u}_\beta - \mathbf{u}_0) = \rho_{zy}(\mathbf{u}_\beta - \mathbf{u}_0), \quad \beta = 1, \ldots, n \tag{3.38}$$

$$\sum_{\alpha=1}^{n} \lambda_{Z,\alpha}\rho_{zy}(\mathbf{u}_\alpha - \mathbf{u}_0) + \sum_{\alpha=1}^{n} \lambda_{Y,\alpha}\rho_y(\mathbf{u}_\alpha - \mathbf{u}_0) + \lambda_{Y,0} = \rho_{zy}(0) \tag{3.39}$$

ICCK requires solving a larger system but avoids the variance inflation seen in SCK, providing more reliable estimates when secondary data are available at all primary locations.

Both SCK and ICCK approaches are implemented with efficient matrix operations and parallel processing to handle large datasets. Careful numerical handling ensures stable solutions, especially when working with non-Euclidean distances or complex spatial structures, as highlighted in recent studies. These methods allow for robust integration of secondary information, improving the accuracy of spatial predictions (Samson and Deutch, 2024).

3.5.11. Quantitative Kriging Neighbourhood Analysis

To rigorously assess the quality and reliability of kriging estimates, a suite of diagnostic metrics known as Quantitative Kriging Neighbourhood Analysis (QKNA) is employed as explained in Barboza and Deutsch (2024). QKNA provides a comprehensive evaluation of kriging performance at the block level, taking into account the spatial configuration of data, the variogram model, and the geometry of the estimation domain.

QKNA comprises six key metrics:

- **Kriging Variance (KV):** KV quantifies the minimized estimation error for each block, representing the expected squared difference between the true value and the kriging estimate. It is computed using the variogram-derived covariances and the kriging weights assigned to each data point within the search neighborhood. Lower KV values indicate greater estimation precision.
- **Kriging Efficiency (KE):** KE, introduced by Krige, measures the effectiveness of the kriging estimate relative to the block variance. It is defined as one minus the ratio of the kriging variance to the block variance. High KE values correspond to low KV and indicate that the estimate is strongly informed by local data, while low KE suggests greater smoothing or a lack of nearby data.
- **Statistical Efficiency (SE):** SE compares the achieved kriging variance to the theoretical minimum variance obtainable by global simple kriging. Values closer to one indicate that the estimator is operating near its theoretical optimum, while lower values reflect the impact of search restrictions or the use of OK.
- **Slope of Regression (SR):** SR assesses conditional bias by measuring the slope of the regression of the true value on the kriging estimate. An SR value of one indicates local unbiasedness, while values less than one suggest conditional bias, often due to search restrictions or local trends.
- **Negative Weights (NW):** NW quantifies the proportion of negative kriging weights in the estimation. While negative weights are mathematically optimal in some cases and can help capture local trends, excessive negative weights may lead to undesirable effects, such as negative block estimates, particularly in the context of physical quantities like ore grades.
- **Weight to the Mean (WM):** WM measures the influence of the stationary mean in the kriging estimate. In simple kriging, any weight not assigned to local data is allocated to the mean. A high WM indicates that the estimate relies more on the mean, typically due to sparse local data, resulting in smoother predictions.

These QKNA metrics are calculated for each block in the model, providing a detailed, block-by-block assessment of estimation quality. By jointly considering these measures, practitioners can diagnose issues such as excessive smoothing, conditional bias, or instability in the kriging system, and can make informed decisions about parameter selection and model refinement. It is important to note, however, that while QKNA is a powerful diagnostic tool, it does not directly prescribe optimal choices for block size or search radius, but rather helps to evaluate the consequences of such choices in the context of the available data and spatial model (Barboza and Deutsch, 2024).

3.5.12. Optimization of Kriging Parameters

The kriging and Co-Kriging procedures described above require the selection of several key parameters, including the number of neighbors, the number of variogram lags, the number of nested variogram components, and a regularization factor for matrix inversion (a very mall number added to the kriging matrix to ensure stability of the kriging process). To ensure robust and unbiased spatial predictions, these parameters were systematically optimized for each variable and kriging method.

A grid search approach was implemented, where all combinations of the tunable parameters were

evaluated. For each parameter set, the following workflow was executed:

- Empirical variograms (and cross-variograms for Co-Kriging) were computed in the embedded (ISOMAP) space for the ilr-transformed variables.
- Nested variogram models were fitted based on the value of `max_components`, which controls the maximum number of nested structures allowed in the model. For Co-Kriging, a Linear Model of Coregionalization (LMC) was established to ensure valid covariance structures.
- Kriging or Co-Kriging was performed for all blocks in the model, with parallelization used to efficiently process large datasets.
- For each block, kriging estimates, variances, and weights were saved, along with QKNA metrics (such as kriging variance, efficiency, and weight metrics) to assess the quality of the spatial predictions.
- Leave-one-out cross-validation (LOOCV) was performed for each method and parameter set, providing additional validation metrics such as RMSE, NRMSE, and MAPE.

The `max_components` parameter was specifically optimized by allowing the variogram fitting function to test models with up to the specified number of nested structures. The best-fitting model, as determined by goodness-of-fit metrics (such as RMSE and MAE) and subsequent kriging performance (QKNA metrics), was selected for each parameter set.

After evaluating all parameter combinations, the results were summarized and ranked. For each method and variable, the top five parameter sets (based on the lowest mean kriging variance, KV and the highest mean kriging efficiency (KE)) were identified and saved. This systematic optimization ensured that the final block model estimates were based on parameter choices that provided the best balance between prediction accuracy and model stability.

All kriging results, QKNA metrics, LOOCV metrics, and kriging weights were archived for further analysis and visualization. The optimized estimates and variances in ILR space were subsequently backtransformed to the raw variable space using a distribution-based approach (see 3.5.13), allowing for meaningful interpretation of both the mean and uncertainty of the predicted metal grades.

3.5.13. Backtransformation of ILR Coordinates into Original Compositions

The process of converting kriging results from isometric log-ratio (ILR) space to the original compositional (raw) space involves two main steps: (1) backtransformation of ILR means to raw compositions, and (2) propagation of variance through the nonlinear transformation. The approach implemented here closely follows the mathematical framework of the `ilr_inv` function in the `scikit-bio` library (scikit-bio development team, 2024).

1. Backtransformation of ILR Means to Raw Compositions

Reading the Sequential Binary Partition (SBP) Matrix: The SBP matrix is read from a CSV file into a NumPy array of shape $(30, 31)$, where 31 is the number of compositional parts (elements). Each row of the SBP matrix defines a binary partition of the 31 parts for one ILR coordinate, with entries of +1 (positive group), −1 (negative group), or 0 (excluded).

Constructing the ILR Orthonormal Basis: The function `sbp_to_basis` constructs an orthonormal basis matrix B of shape $(31, 30)$ from the SBP matrix. For each row k of the SBP matrix, the indices of the positive group ($\text{SBP}_{k,i} = 1$) and negative group ($\text{SBP}_{k,i} = -1$) are identified. Let n_{pos} and n_{neg} be the number of parts in the positive and negative groups, respectively. For each part i (column of the SBP matrix) in the k^{th} row, the following assignment is made:

$$v_i^{(k)} = \begin{cases} \frac{1}{n_{pos}} \sqrt{\frac{n_{pos} n_{neg}}{n_{pos}+n_{neg}}} & \text{if SBP}_{k,i} = 1 \\ -\frac{1}{n_{neg}} \sqrt{\frac{n_{pos} n_{neg}}{n_{pos}+n_{neg}}} & \text{if SBP}_{k,i} = -1 \\ 0 & \text{otherwise} \end{cases} \tag{3.40}$$

The resulting vector $v^{(k)}$ is the k-th basis vector, and all such vectors are stacked as columns to form B.

Loading ILR-Transformed Data: The ILR-transformed estimates for each block are read from a CSV file into a NumPy array of shape $(N, 30)$, where N is the number of samples (blocks). Each row contains the ILR coordinates for one sample.

Inverse ILR Transformation: The function `ilr_inv_custom` performs the backtransformation for each sample (row vector of ILR coordinates $\mathbf{z}$) by first computing the log-composition:

$$\log(\mathbf{c}) = \mathbf{z} \cdot B^\top \tag{3.41}$$

where $B^\top$ is the transpose of the basis matrix. Each entry is then exponentiated to obtain the unnormalized composition:

$$\mathbf{c}' = \exp(\log(\mathbf{c})) \tag{3.42}$$

Finally, the composition is normalized so that the sum of all parts equals one:

$$\mathbf{c}_{\text{final}} = \frac{\mathbf{c}'}{\sum_{i=1}^{31} c'_i} \tag{3.43}$$

This process is vectorized and applied to all N samples at once. The values are then scaled to 100 to obtain weight percent values consistent with the original raw compositions.

Assigning Element Names and Saving Results: The resulting matrix of compositions is converted to a DataFrame with columns named for each element (e.g., “La (wt%)”, “Ce (wt%)”, etc.). The "block_id" and spatial coordinates (“x”, “y”, “z”) are inserted as the first columns. The final DataFrame is saved as a CSV file for further analysis.

2. Propagation of Variance through the Nonlinear Transformation

It is important to first clarify why the mean of ILR-transformed coordinates can be backtransformed to compositional space, while the variance cannot be treated in the same way. The inverse ILR transformation is nonlinear, and variance does not propagate linearly through such transformations (d’Ache C., 2013, Mert et al., 2016). As a result, directly transforming the kriging-estimated variances in ILR space into raw compositional variances is not mathematically valid or possible.

There exists a gap in the current literature regarding variance propagation from ILR space to compositional space. In particular, no analytical framework has been established that accurately describes how multivariate uncertainty transforms through the inverse ILR mapping. While there exists literature on error propagation into the ILR space (Mert et al., 2016), the same is not the case for back transformation. To approximate this propagation, a simulation-based strategy inspired by multivariate normal distribution modeling, as outlined in *”Parameter Estimation for Multivariate Probability Distributions”* by Texas_A&M_University (2017) was employed. This approach is not formally defined with respect to compositional data and ILR coordinates and may introduce inaccuracies due to geometric constraints

and transformation nonlinearity; however, it offers a possible solution where full co-kriging covariance estimation is unavailable such as in the case of this study.

The procedure is implemented as follows:

Covariance Matrix Construction: For each block, a covariance matrix is constructed in ILR space. The diagonal entries are defined by the kriging-estimated variances for each coordinate, while the off-diagonal entries are scaled using empirical correlation coefficients (ρ_{ij}^{emp}) calculated from the original sample data. The covariance between coordinates i and j is given by

$$\Sigma_{ij} = \rho_{ij}^{\text{emp}} \cdot \sqrt{\sigma_i^2 \cdot \sigma_j^2} \tag{3.44}$$

where i and j index the different ILR coordinates.

Multivariate Normal Distribution: This covariance matrix, together with the kriging mean vector, defines a multivariate normal distribution (MVN) over the $D-1$ ILR coordinates.

Monte Carlo Simulation: A Monte Carlo simulation is performed to generate multiple joint realizations from this MVN distribution, thereby preserving both marginal uncertainty and inter-coordinate dependencies.

Backtransformation to Raw Compositions: Each simulated ILR vector is backtransformed to the original D-part composition using the inverse ILR transformation.

Empirical Variance Calculation: The ensemble of backtransformed samples is then used to empirically compute block-level variance in raw compositional space.

This strategy should allow for a statistically coherent and geometrically valid approximation of variance in the raw compositional space, using available kriging outputs and empirical relationships from sample data (Texas_A&M_University, 2017). Another approach to tackle this could have been to do co-kriging of all the ILR variables together but that is beyond the scope of this study.

3.6. Current Region Analysis

To investigate the influence of ocean currents on the spatial distribution of block model compositions, a current region analysis was performed that integrates hydrodynamic data with the block model grid. The ocean current data used in this analysis was obtained from the Copernicus Marine Service, specifically the global physical analysis and forecast product (Copernicus Marine Service, 2024). The horizontal and vertical current data at a resolution of 24 hours, from 2015 to 2025 was used for the analysis. The process begins by assigning each block a representative current vector, derived from nearby oceanographic model outputs. The study area is then divided into four spatial quadrants, with the quadrant boundaries rotated by 45 degrees to better capture the orientation of the data and prevailing currents.

Within each quadrant, distribution of current directions were analyzed. If the current directions exhibit a single dominant trend, this is used as the reference direction for further classification. In cases where two distinct dominant directions are present (bimodality), both are identified using clustering techniques, and each block is associated with the closest dominant direction.

Blocks are then classified into three regions relative to their quadrant's dominant current direction(s): upstream, downstream, and wake. The downstream region includes all blocks located in the direction of the current, while the wake region is defined as a cone-shaped subset of the downstream area, aligned

closely with the current vector. The upstream region comprises blocks situated opposite to the current direction.

The estimated compositions of each element obtained using ICCK is then clustered using 1D K-means clustering to identify any clusters in the compositions. The current region boundaries are then overlaid onto the cluster plots to further identify spatial correlation between clusters and current regions.

This classification enables a rough spatial analysis of how current regimes may influence the distribution and clustering of block compositions across the study. However, without going into computational fluid dynamics simulations, it cannot be said that this is the exact situation of the seamount in the current. It is also worth mentioning that this analysis considers the current data of the past ten years to be the constant across the $\tilde{1}50$ million years of the sea mount's existence.

3.7. Tonnage calculations

Tonnage calculations are done based on back-transformed estimate values. The total tonnage of each element within the study area was estimated using a block model approach. Elemental weight percentages for each block are taken from the back-transformed ICCK estimate dataset, and the x and y dimensions (100 m) of the blocks were taken (block dimensions are 100 m $\times$ 100 m $\times$ 10 m). The thickness of crust at each block was estimated using ordinary kriging interpolation of measured crust thickness values from sample locations, providing a spatially continuous thickness model across the entire block grid.

The mass of each block was calculated using the estimated crust volume (area $\times$ thickness) and a mean dry bulk density of 1.3 g/cm^3 (J. R. Hein et al., 1999). This mean dry density value is a generalization of Ferromanganese crust densities and not specific to the Tropic seamount and could be considered as an appropriate, yet approximate placeholder in the absence of seamount specific values as the weight percentages were also calculated based on the mass of oven-dried samples (Howarth, 2022). The total tonnage for each element was then determined by summing the block-wise mass of each element (calculated from the block mass and its corresponding weight percentage) across all blocks.

4

Results & Discussion

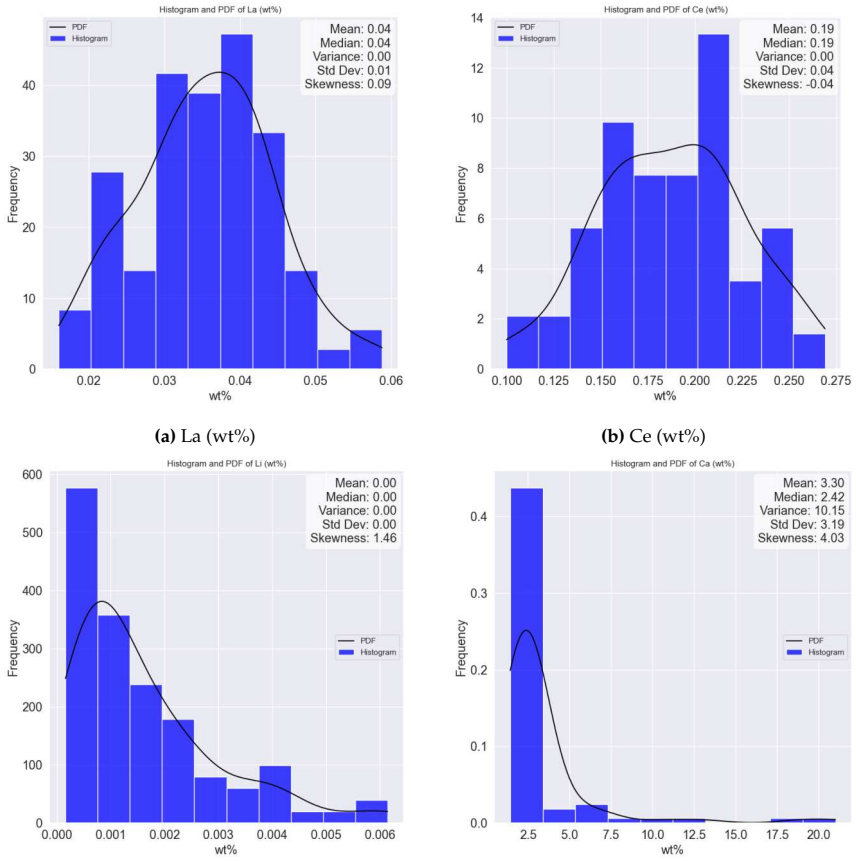
This chapter analyzes the results obtained by following the methodology described in Chapter 3 and also discusses the implications of those results in the context of geostatistics, resource estimation, and CRMs.

4.1. Exploratory Data Analysis of Raw metal grades

The exploratory data analysis plots (Figure 4.1) of the untransformed variables indicate that certain elements, such as La, Ce, and most other REEs, exhibit approximately normal distributions with minimal skewness. In contrast, other variables, including Li and Ca, display highly skewed distributions. These observations underscore the necessity for appropriate preprocessing and transformation steps to ensure the validity of subsequent statistical analyses and the subsequent ILR transformations tried to address this.

The correlation between the absolute value of water depth (where higher values indicate greater depth) and elemental percentages in the raw data is presented in Figure 4.2. Notably, the figure shows that **Fe exhibits a slightly positive correlation with water depth, whereas Mn displays a negative correlation**. This finding contrasts with observations reported in the literature (Section 2.2), where Mn/Fe ratios are generally found to increase with depth. A possible explanation for this discrepancy (in relation to the study by Howarth (2022)) could be that the dataset used in this study is a subset of a much larger analytical dataset with more samples and more diverse analytical methodologies that might exhibit different correlation values that is explained by the data not present in this study. This difference in available data should be considered when comparing results to previous studies

The correlation matrix for the elements in the untransformed data is provided in the appendix. However, these correlations are not emphasized in this report, as correlations in raw compositional data can be spurious due to the closure property inherent to compositional datasets (Section 3.3). Additionally, k-means clustering and principal component analysis were performed but did not yield results relevant to the subsequent geostatistical modeling and are therefore not discussed further.



(a) La (wt%)

(b) Ce (wt%)

(c) Li (wt%)

(d) Ca (wt%)

Figure 4.1: Histograms and PDFs of La, Ce, Li, and Ca (wt%).

Table 4.1: Summary of statistical parameters for all untransformed elements.

Element	Mean	Median	Min	Max	Range	Skewness
La (wt%)	0.04	0.04	0.02	0.06	0.04	0.09
Ce (wt%)	0.19	0.19	0.10	0.27	0.17	-0.05
Pr (wt%)	0.01	0.01	0.00	0.01	0.01	0.47
Nd (wt%)	0.03	0.03	0.01	0.05	0.04	0.08
Sm (wt%)	0.01	0.01	0.00	0.01	0.01	0.13
Eu (wt%)	0.00	0.00	0.00	0.00	0.00	0.13
Gd (wt%)	0.01	0.01	0.00	0.01	0.01	0.09
Tb (wt%)	0.00	0.00	0.00	0.00	0.00	0.11
Dy (wt%)	0.01	0.01	0.00	0.01	0.01	0.08
Ho (wt%)	0.00	0.00	0.00	0.00	0.00	0.02
Er (wt%)	0.00	0.00	0.00	0.00	0.00	-0.04
Tm (wt%)	0.00	0.00	0.00	0.00	0.00	-0.04
Yb (wt%)	0.00	0.00	0.00	0.00	0.00	-0.04
Lu (wt%)	0.00	0.00	0.00	0.00	0.00	0.03
Y (wt%)	0.02	0.02	0.01	0.04	0.03	0.38
Li (wt%)	0.00	0.00	0.00	0.01	0.01	1.43
Co (wt%)	0.56	0.56	0.19	1.07	0.88	0.24
Ni (wt%)	0.33	0.30	0.18	0.67	0.49	1.26
Cu (wt%)	0.08	0.06	0.02	0.22	0.20	1.38
Zn (wt%)	0.07	0.06	0.03	0.11	0.08	0.67
Mo (wt%)	0.05	0.05	0.03	0.09	0.06	0.39
Te (wt%)	0.01	0.01	0.00	0.01	0.01	0.42
Na (wt%)	1.13	1.14	0.84	1.61	0.77	0.51
Mg (wt%)	1.40	1.38	0.71	2.58	1.87	0.78
Al (wt%)	1.58	1.48	0.49	2.92	2.43	0.40
K (wt%)	0.39	0.37	0.23	0.67	0.44	0.84
Ca (wt%)	3.28	2.40	1.42	21.01	19.59	4.08
Ti (wt%)	0.74	0.72	0.28	1.39	1.11	0.36
Mn (wt%)	17.01	16.66	9.12	22.21	13.09	0.01
Fe (wt%)	23.55	23.44	9.33	32.70	23.37	-0.49
Others (wt%)	49.52	49.58	35.91	60.08	24.17	-0.20

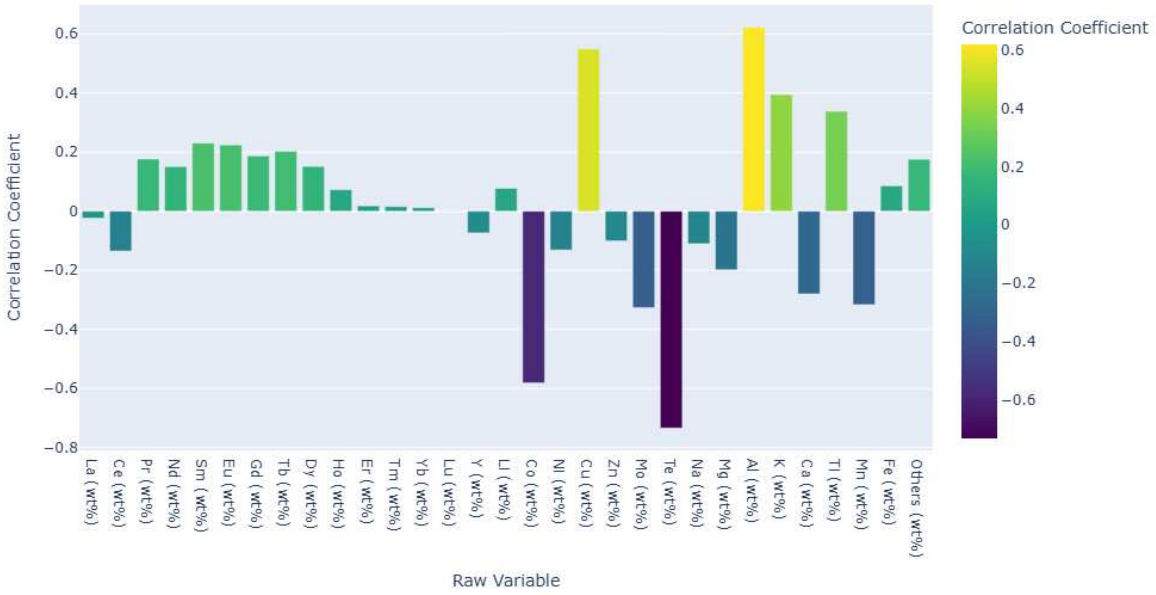


Figure 4.2: Correlation between water depth and weight percentage values of all elements

4.2. Additive Log-Ratio (ALR) Transformation

The ALR transformation was performed using the raw-ALR transformation functionality available in the CoDaPack software suite. Lithium (Li) was selected as the denominator for the ALR transformation based on compositional exploratory data analysis (EDA) using the variation matrix (Table A.2); Li accounted for 26% of the total centered log-ratio (CLR) variance, making it a suitable choice for the denominator. The subjective nature of choosing the denominator for ALR transformation can affect the transformation significantly. Therefore, ILR transformation, which is stated to preserve the compositional relationships better than the other log-ratio transformations was chosen. Also, preliminary kriging models indicated that ALR-transformed variables consistently yielded poorer results in all error metrics compared to ILR variables across all kriging plans. Therefore, further analysis in this report focuses exclusively on ILR variables.

4.3. Sequential Binary partition and ILR transformation

The following is an example of the partitions generated according to the approach outlined in subsection 3.4.1. Again, it must be reiterated that for the ILR transformation, the way this partitioning is done can be arbitrary as long as the partitions obey the conditions given (3.4.1).

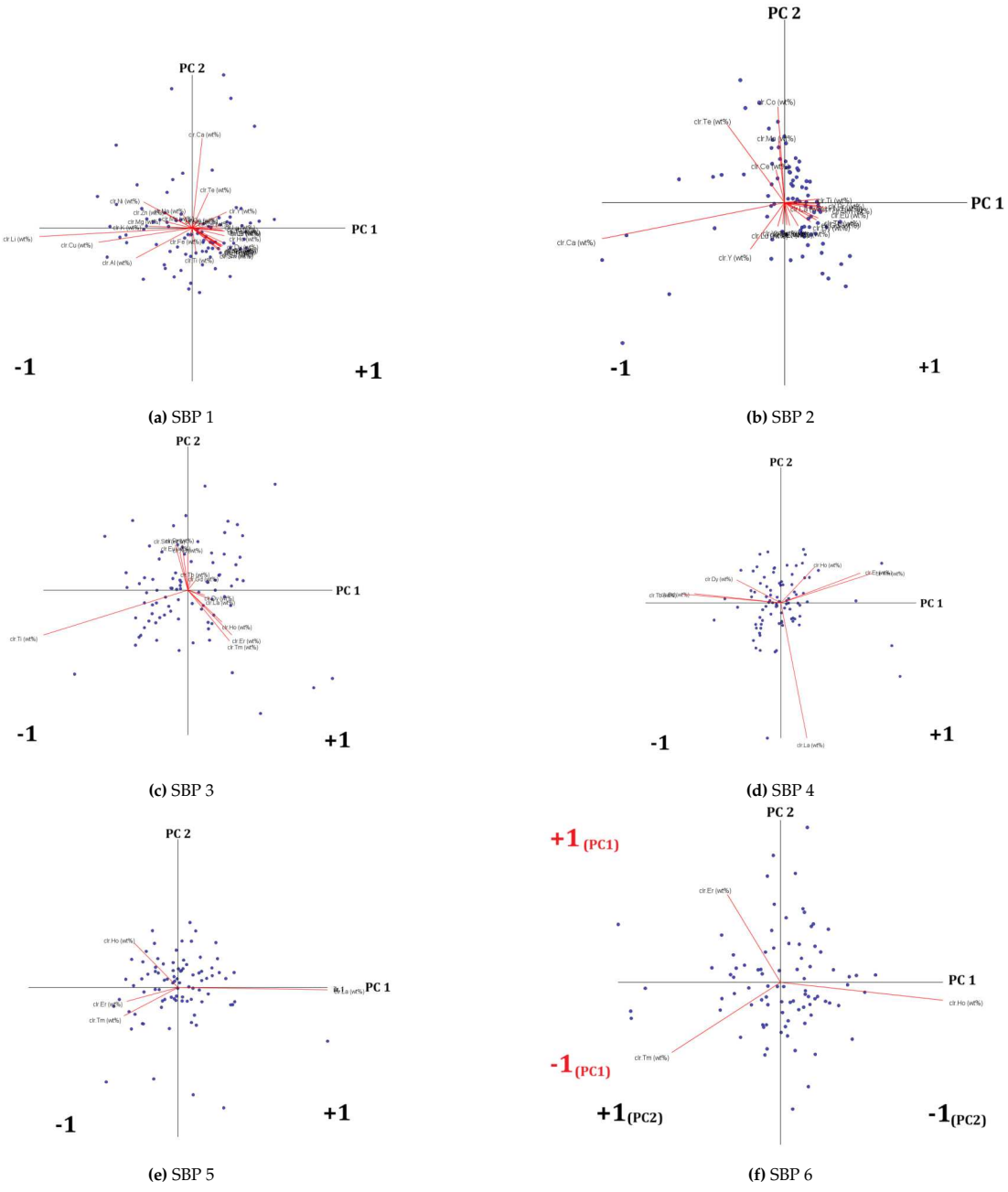
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3a)
  - Right (+1): La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Y, Co, Mo, Te, Ca, Ti
  - Left (-1): Li, Ni, Cu, Zn, Na, Mg, Al, K, Mn, Fe
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3b)
  - Right (+1): La, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Ti
  - Left (-1): Ce, Yb, Lu, Y, Co, Mo, Te, Ca
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3c)
  - Right (+1): La, Gd, Tb, Dy, Ho, Er, Tm

  - Left (-1): Ti, Pr, Nd, Sm, Eu
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3d)
  - Right (+1): La, Ho, Er, Tm
  - Left (-1): Gd, Tb, Dy
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3e)
  - Right (+1): La
  - Left (-1): Ho, Er, Tm
- **Left and right side of PC 2 axis in the CLR Biplot:** (see Fig. 4.3f)
  - Left (+1): ER, TM
  - Right (-1): Ho
- **Top and bottom side of PC 1 axis in the CLR Biplot:** (see Fig. 4.3f)
  - Top (+1): Er
  - Bottom (-1): Tm

Figure 4.4 illustrates all the partitions obtained by following this approach and table 4.2 is the resulting SBP created. This SBP matrix (Table 4.2) represents a comprehensive encoding of relationships between the elements into a mathematically valid partitioning structure. Each row in the matrix corresponds to a binary partition (Figure 4.4) that divides elements based on their associations. This partitioning scheme formed the foundation for the subsequent ILR transformation, which mapped the compositional data onto a Euclidean space where standard statistical methods can be applied while preserving the compositional nature of the data. It must be stated that while the calculation of ILR coordinates from an SBP is explained in the methodology, the actual transformation was done by importing both the data and the SBP into CoDaPack and using the inbuilt `Raw-ILR Transformation` functionality.

It is important to note that, in the context of ILR transformations, while ILR coordinates are geometrically independant, individual ILR variables should not be interpolated independently for estimation purposes if they are correlated to each other as in the case of certain ILR coordinates in this study (example: Table 4.4). Instead, ideally, all ILR coordinates resulting from the transformation that are correlated with each other must be interpolated together using co-kriging. However time and computational constraints of this study meant they can only be interpolated individually.

Finally, only after the complete set of interpolated ILR coordinates is obtained can the data be back-transformed to the original compositional space. This ensures that the compositional relationships are preserved throughout the estimation process.



(a) SBP 1

(b) SBP 2

(c) SBP 3

(d) SBP 4

(e) SBP 5

(f) SBP 6

Figure 4.3: Example of 6 Sequential Binary Partition (SBP) splits made using the CLR Biplot.

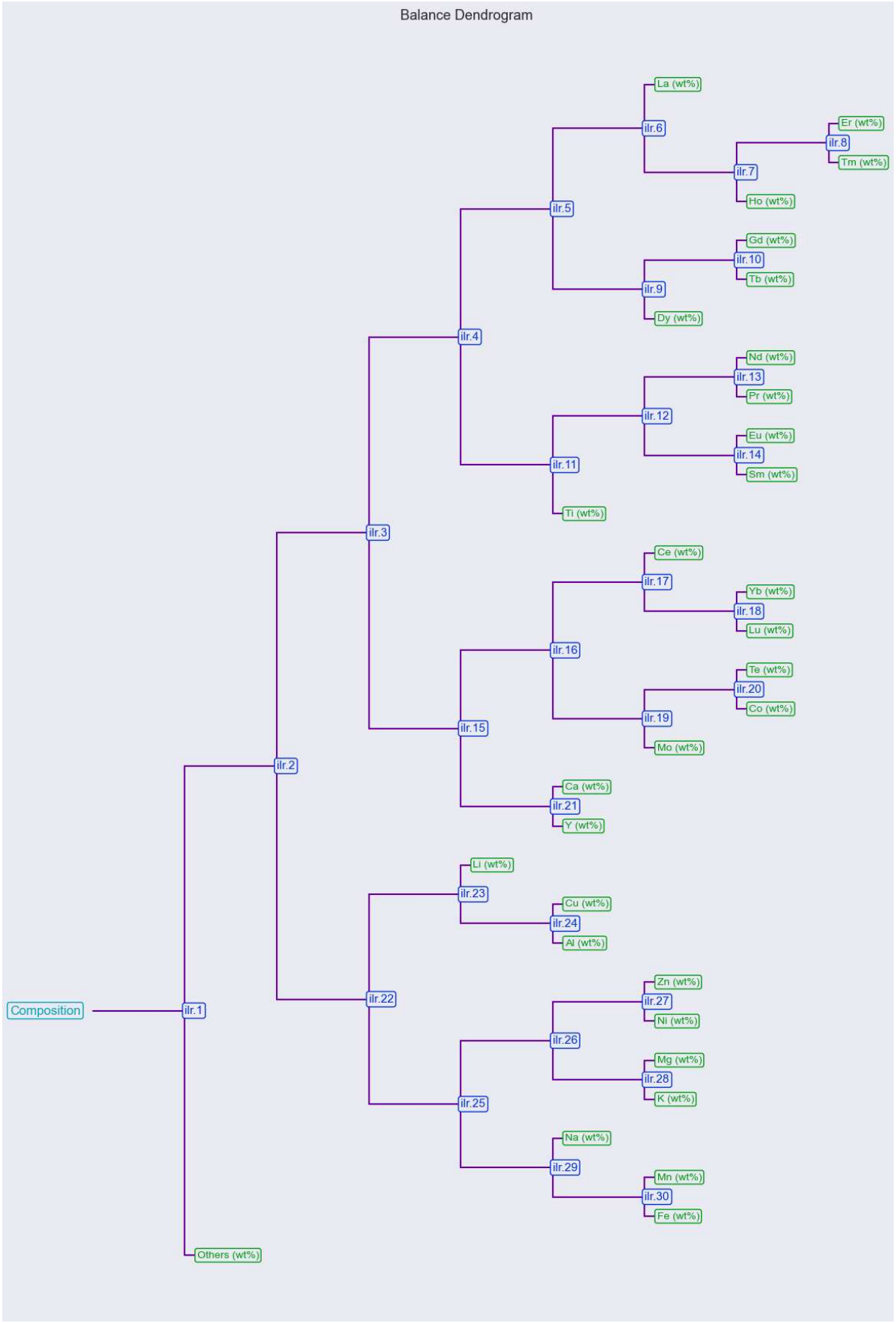
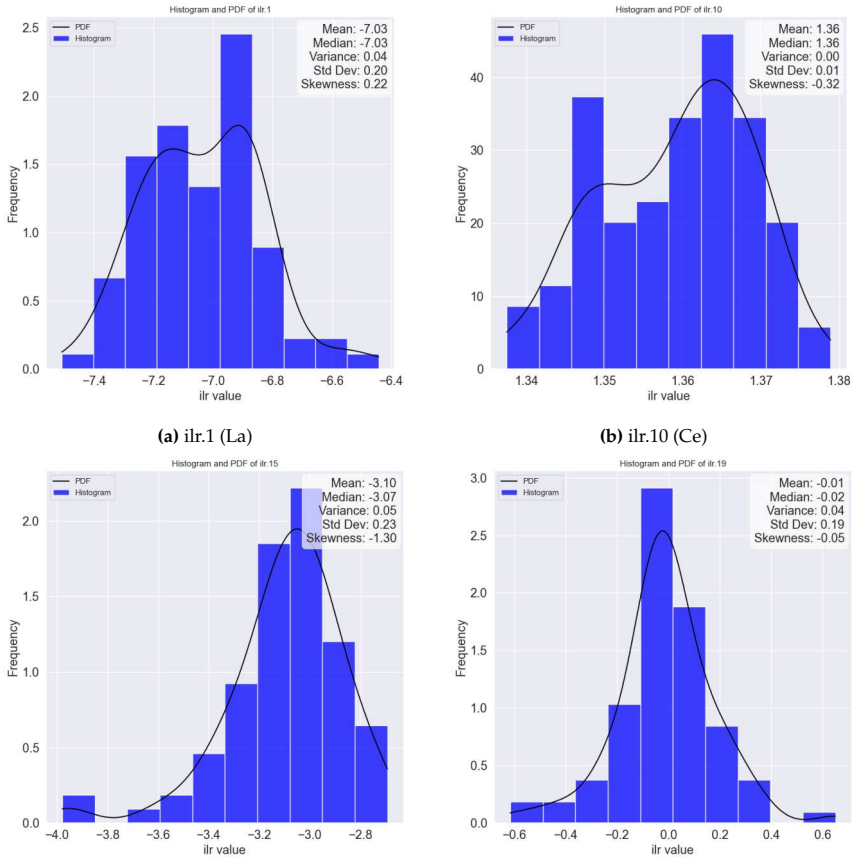


Figure 4.4: Balance dendrogram of the numerator and denominator groups of each ILR coordinate

Table 4.2: Sequential Binary Partition Matrix

Partition	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Y	Li	Co	Ni	Cu	Zn	Mo	Te	Na	Mg	Al	K	Ca	Ti	Mn	Fe	Others
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	0
3	1	-1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	0	-1	0	0	0	-1	-1	0	0	0	0	-1	1	0	0	0
4	1	0	-1	-1	-1	-1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
5	1	0	0	0	0	0	-1	-1	-1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	-1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	1	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
12	0	0	1	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	0	0	0	0	0	0	0	0	0	0	1	1	-1	0	1	0	0	0	1	1	0	0	0	0	-1	0	0	0	0
16	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	-1	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	1	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	1	-1	0	0	-1	-1	1	-1	0	0	-1	-1	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	-1	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	-1	1	0	1	0	0	-1	-1	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	-1	0	-1	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	-1	-1	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0



(a) ilr.1 (La)

(b) ilr.10 (Ce)

(c) ilr.15 (Li)

(d) ilr.19 (Ca)

Figure 4.5: Histograms and PDFs of ilr.1, ilr.10, ilr.15, and ilr.19.

Table 4.3: Summary of statistical parameters for all ILR transformed elements.

Element	Mean	Median	Min	Max	Range	Skewness
ilr.1	-7.03	-7.03	-7.51	-6.44	1.07	0.22
ilr.2	-9.66	-9.52	-12.05	-7.91	4.14	-0.69
ilr.3	-3.77	-3.75	-4.84	-2.95	1.89	-0.43
ilr.4	-3.12	-3.13	-3.37	-2.91	0.46	-0.15
ilr.5	-0.25	-0.25	-0.40	-0.13	0.27	-0.25
ilr.6	3.06	3.06	2.93	3.21	0.27	0.20
ilr.7	-0.04	-0.04	-0.08	0.00	0.08	-0.27
ilr.8	-1.39	-1.38	-1.43	-1.36	0.07	-0.61
ilr.9	-0.69	-0.69	-0.72	-0.64	0.08	0.55
ilr.10	1.36	1.36	1.34	1.38	0.04	-0.29
ilr.11	-4.23	-4.24	-4.58	-3.88	0.70	0.09
ilr.12	1.63	1.64	1.58	1.68	0.10	-0.52
ilr.13	1.03	1.03	1.01	1.05	0.04	-0.37
ilr.14	-1.08	-1.08	-1.11	-1.05	0.06	-0.44
ilr.15	-3.10	-3.07	-3.98	-2.69	1.29	-1.27
ilr.16	-2.78	-2.77	-3.37	-2.08	1.29	0.08
ilr.17	4.31	4.29	4.06	4.64	0.58	0.46
ilr.18	1.36	1.36	1.32	1.41	0.10	0.28
ilr.19	-0.01	-0.02	-0.62	0.65	1.27	-0.07
ilr.20	-3.34	-3.34	-3.49	-3.17	0.32	0.12
ilr.21	3.53	3.50	3.14	4.26	1.13	0.94
ilr.22	-4.66	-4.65	-6.44	-3.46	2.98	-0.39
ilr.23	-4.53	-4.55	-5.45	-3.37	2.08	0.55
ilr.24	-2.22	-2.22	-2.70	-1.66	1.04	0.03
ilr.25	-4.14	-4.17	-4.50	-3.44	1.05	0.96
ilr.26	-1.63	-1.64	-2.05	-1.23	0.82	-0.18
ilr.27	-1.15	-1.14	-1.49	-0.96	0.53	-0.61
ilr.28	0.92	0.93	0.54	1.17	0.63	-0.65
ilr.29	-2.35	-2.37	-2.54	-2.13	0.40	0.40
ilr.30	-0.23	-0.24	-0.43	0.00	0.43	0.30

Another important point is that the original dataset included only major, minor, and trace metallic element concentrations. To normalize the compositional data to 100%, all known metallic concentrations were summed, subtracted from 100, and the remainder was assigned to a new column labeled "Others." This column represents all elements present in the samples whose concentrations are unknown, primarily non-metallic elements. This approach is a compromise, as previous studies have shown that non-metallic elements such as oxygen significantly influence the composition of ferromanganese crusts. However, these relationships are not captured in the SBP and ILR transformations. The "Others" column in general only account for 2.62% of the total CLR variance and therefore might not affect the transformation as much mathematically but it is still worth considering further studies into the non-metallic elements.

4.4. Exploratory Data Analysis of ILR-Transformed Variables

After applying ILR transformations (see Section (3.3) and managing outliers as described in Subsection (3.4.4), most variables exhibit distributions closer to normality, although some skewness remains. The prevalence of outliers has also been reduced. Notably, certain variables, such as ilr.1 and ilr.10 (Figure (4.5), display bimodal distributions. For variables like ilr.15, and ilr.19 (Figure (4.5), a few outliers persist; however, these were retained in the analysis due to the limited sample size and the fact that removing an outlier would necessitate excluding the entire sample. This is particularly important because a sample may be an outlier for one ILR coordinate but not for others, and further reduction in sample size would negatively impact subsequent kriging.

The presence of bimodal distributions in certain ILR coordinates may indicate that the relative balance between element groups is governed by two distinct regimes. This could reflect two types of deposition events, each imparting a unique elemental signature. For example, elements associated with both Fe and Mn may be distributed differently depending on the deposition process. Alternatively, bimodality could result from a combination of deposition followed by post-depositional alteration, such as phosphatization or leaching. In this context, one mode might represent unaltered material deposited directly onto the crust surface, while the other corresponds to zones affected by post-depositional modification. Since ILR coordinates capture relative contrasts rather than absolute concentrations, a bimodal signal may strongly indicate the presence of dual-process behaviour within the crust.

Table 4.4: Top 10 ILR pairs with highest positive and lowest negative correlations.

Top 10 Highest (+ve) Correlations			Top 10 Lowest (-ve) Correlations		
ILR	**ILR**	**Correlation**	**ILR**	**ILR**	**Correlation**
ilr.22	ilr.25	0.839	ilr.2	ilr.25	-0.926
ilr.3	ilr.16	0.827	ilr.2	ilr.22	-0.869
ilr.3	ilr.7	0.769	ilr.12	ilr.16	-0.825
ilr.23	ilr.25	0.731	ilr.8	ilr.18	-0.810
ilr.6	ilr.12	0.682	ilr.5	ilr.7	-0.795
ilr.4	ilr.13	0.674	ilr.3	ilr.5	-0.792
ilr.2	ilr.27	0.662	ilr.4	ilr.9	-0.766
ilr.22	ilr.24	0.661	ilr.2	ilr.24	-0.737
ilr.22	ilr.23	0.636	ilr.25	ilr.27	-0.730
ilr.24	ilr.25	0.626	ilr.5	ilr.9	-0.718

Table 4.4 presents the correlation values between ILR coordinates with the highest and lowest correlations. The presence of significant correlations between ILR coordinates indicates that, in accordance with the interpretability of ILR coordinates (see Subsection 3.4.3), specific groups of elements tend to vary together within the compositional dataset. For instance, strong positive correlations such as between ilr.22 and ilr.25 ($r = 0.84$), ilr.23 and ilr.25 ($r = 0.73$), ilr.7 and ilr.3 ($r = 0.77$), and ilr.3 and ilr.16 ($r = 0.83$) suggest that these groups of elements exhibit similar behavior in their deposition and alteration processes within ferromanganese crusts. Conversely, pronounced negative correlations, such as between ilr.16 and

ilr.12 ($r = -0.83$), ilr.27 and ilr.25 ($r = -0.73$), ilr.25 and ilr.2 ($r = -0.93$), and ilr.22 and ilr.2 ($r = -0.87$), imply that these groups of elements tend to exhibit opposing trends in compositional dominance; as one group increases in relative abundance, the other decreases.

Importantly, the correlations observed between ILR variables are not influenced by the closure problem inherent to compositional data. Therefore, these relationships can be interpreted as genuine associations between elemental groups, rather than artifacts arising from the mathematical constraints of compositional datasets. The full correlation matrix is given in Appendix A

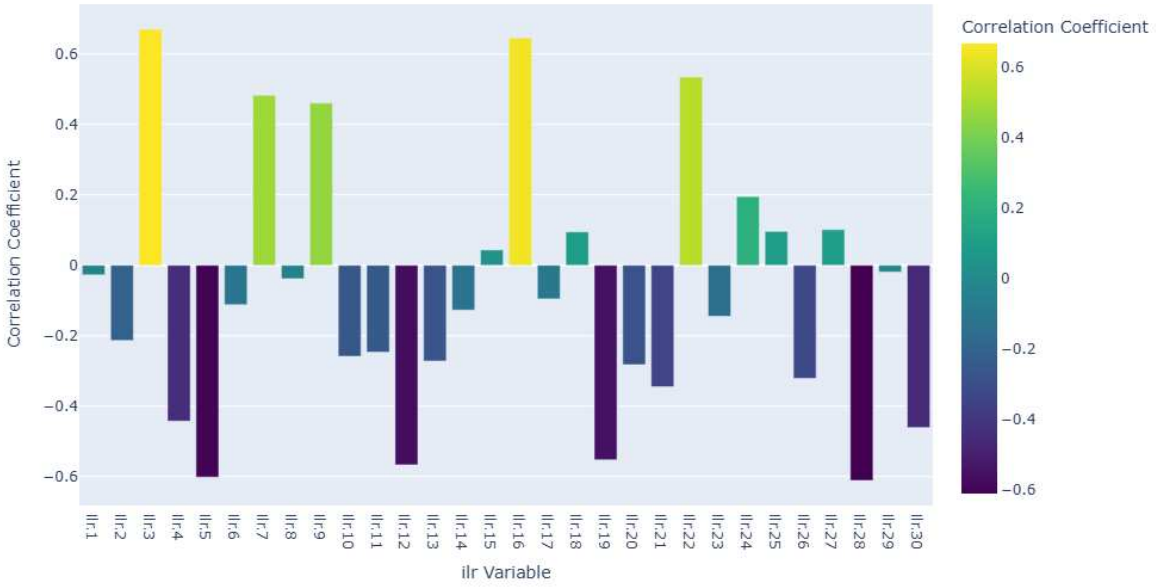


Figure 4.6: Correlation of ILR coordinates with absolute value of water depth

One important aspect of the analysis involves examining the correlation between each ILR coordinate and water depth, as shown in Figure 4.6. These correlations provide insight into how the ratios between groups of elements vary with depth. For example, ilr.4 exhibits a negative correlation with water depth. ilr.4 represents the ratio of HREEs to LREEs (excluding La). This observation aligns with geochemical knowledge (Section 2.3), which suggests that LREEs are preferentially associated with Mn, while HREEs are more closely associated with Fe. A decrease in ilr.4 values with increasing depth implies that the relative abundance of LREEs (associated with Mn) increases compared to HREEs (associated with Fe) at greater depths.

However, this trend is contrary to what is observed in both the raw dataset and the ILR-transformed data. Specifically, ilr.30, which represents the ratio of Mn (numerator) to Fe (denominator), also shows a negative correlation (almost identical to ilr.4) with depth (Figure 4.6), **indicating that the Mn/Fe ratio decreases as depth increases.** This finding is consistent with the results from the raw data (Figure 4.2), where Mn concentrations decrease and Fe concentrations increase with depth.

One possible explanation for this discrepancy is that **the geological setting of the Tropic seamount may promote an atypical association, whereby LREEs are more closely linked to Fe and HREEs to Mn, contrary to the geochemical expectations in the literature review.** This highlights the importance of considering local geological context when interpreting compositional trends. This could be similar to the case of Ninetyeast Ridge in the Indian Ocean (Section 2.2) where Fe/Mn ratios increase with water depth because Mn enrichment diminishes below the OMZ compared to Fe.

Another important aspect of the analysis involves cross-checking the relationship between water depth

and ILR coordinates with the correlations observed between their constituent elemental groups and water depth. For instance, ilr.16, which exhibits a positive correlation with water depth, represents the ratio Ce, Yb, Lu to Co, Mo, Te. As shown in Figure 4.2, Co, Mo, and Te display notably more negative correlation coefficients with water depth compared to Ce, Yb, and Lu. This indicates that, with increasing depth, the relative abundance of Ce, Yb, and Lu increases compared to Co, Mo, and Te, consistent with the interpretation of ilr.16.

Similarly, ilr.19 demonstrates a negative correlation with water depth and is defined by the ratio Co, Te to Mo. The constituent elements Co and Te both exhibit more negative correlations with depth than Mo, supporting the trend indicated by ilr.19. Another example is ilr.22, which has a positive correlation with water depth and corresponds to the ratio Li, Cu, Al to Ni, Zn, Na, Mg, K, Mn, Fe. In Figure 4.2, Li, Cu, and Al show positive correlations with depth, while all denominator elements of ilr.22, except Fe, display negative correlations, further validating the compositional interpretation.

Comparable patterns are observed for ilr.16 and ilr.28 (see Figures 4.6 and 4.2), illustrating the consistency between ILR coordinate correlations and the behavior of their respective elemental groups. **This cross-validation reinforces the interpretability of ILR coordinates in the context of geochemical processes and their relationship with water depth.**

Table 4.5: Numerator and denominator elements for ILR coordinates mentioned in this chapter.

ILR coordinate	Numerator elements	Denominator elements
2	La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Y, Co, Mo, Te, Ti	Li, Ni, Cu, Zn, Na, Mg, Al, K, Mn, Fe
3	La, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm	Ce, Yb, Lu, Y
4	La, Gd, Tb, Dy, Ho, Er, Tm	Pr, Nd, Sm, Eu
7	Ho, Er, Tm	Ho
16	Ce, Yb, Lu	Co, Mo, Te
19	Co	Mo
22	Li, Ni, Cu, Zn, Na, Mg, Al, K, Mn, Fe	Na, Mg, Al, K, Mn, Fe
25	Ni, Zn	Mg, K, Mn, Fe
27	Zn	Mn
28	Mg	Mn
30	Mn	Fe

The complete set of exploratory data analysis plots and the full ILR table is available in the Appendix A.

4.5. Block Model

Two blockmodels were generated by following the methodology outlined in section 3.1. one with block dimension 10m x 10m x 1m and one with block dimensions 100m x 100m x 10m. The lower resolution blockmodel (Fig 4.7) with 90000 blocks was chosen for further estimation and analysis due to computational constraints as the higher resolution block model had 14 million blocks. The initial plan was to generate a low resolution block model for testing and finalization of the kriging plan and then substitute that with the high resolution model for the final resource estimation. However, computational requirements for processing a model with millions of blocks was too intense and therefore the lower resolution model was chosen for all purposes.

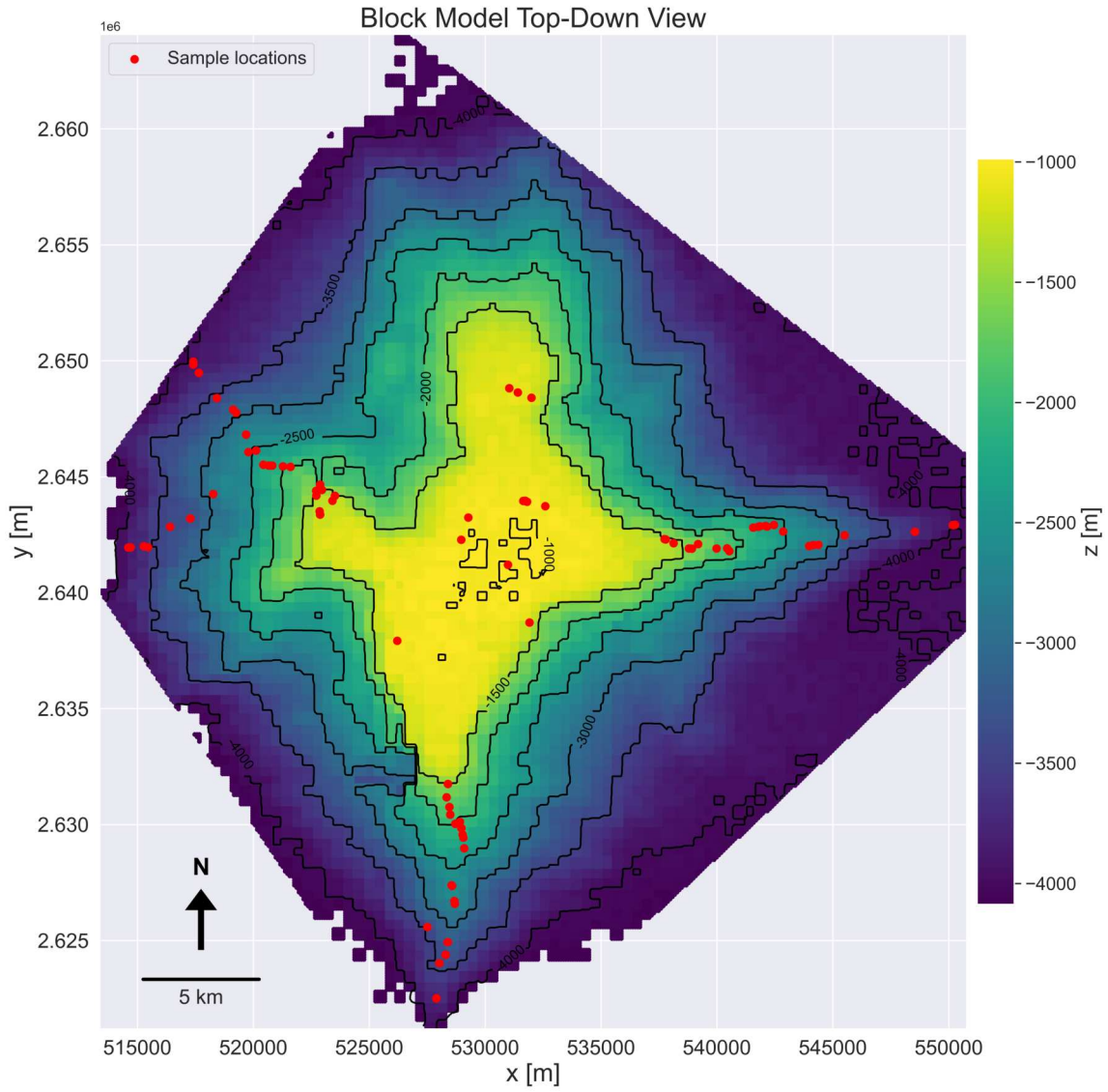


Figure 4.7: Block model visualization with sample locations in red

4.6. Landmark-ISOMAP Embedded Coordinates

The LVA field was generated as described in the Subsection 3.5.3 and visualized in the seventh subplot in Figure 4.8. Figure 4.8 provide a comprehensive visualization of the spatial structure and connectivity within the study area. The first set of plots illustrates: (1) the smoothed surface elevation of the block

model, (2) the spatial distribution of strike angles (azimuth from north), (3) the dip angles indicating surface steepness, (4) the plunge angles representing the direction of maximum curvature, (5) the ratio of minor to major axes of anisotropy (r1), (6) the ratio of vertical to major axes (r2), (7) the primary and secondary anisotropy directions as vector fields, (8) ellipses visualizing the orientation and strength of local anisotropy, and (9) the magnitude of the surface gradient.

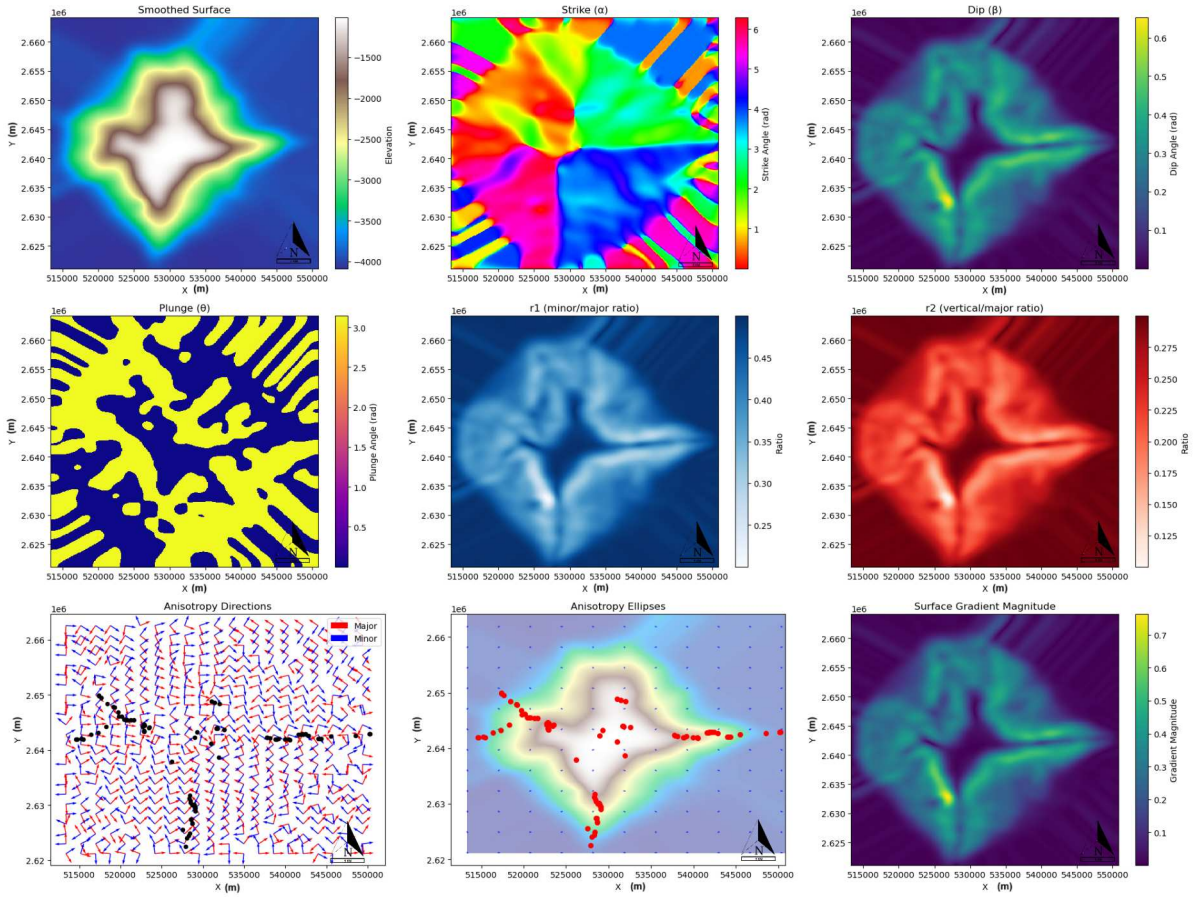


Figure 4.8: LVA Filed Visualizations (x and y axes in metres.)

Using this field and shortest path distances between samples (see Subsection 3.5.4), both samples and blocks were embedded into a 19-dimensional space where Euclidean distances reflect the original shortest path relationships.

This embedded space forms the basis for subsequent interpolation. Its main advantage is that the embedding accounts for complex anisotropic spatial structure, enabling standard geostatistical tools to be applied without explicit anisotropy modeling in the variogram fitting step. This approach should allow efficient and geologically meaningful interpolation in domains with locally varying anisotropy. Another advantage is a space that is effectively flat in terms of spatial relationships and this simplifies how complex geometries can be interpolated since there is no need for countless Search Ellipsoids or complex domianing to account for the anisotropy. Another factor is that MDS and Isomap embeddings are well-established mathematical and geometric methods used in manifold learning and dimensionality reduction. Python libraries such as `scikit-learn`, `SciPy`, `NetworkX` etc offer plug-and-play implementations that make it easy to apply these techniques to spatial and high-dimensional datasets.

The unit of distance in a landmark Isomap embedded space is arbitrary and unitless. It represents a measure of dissimilarity or geodesic distance between points as preserved by the embedding, but it

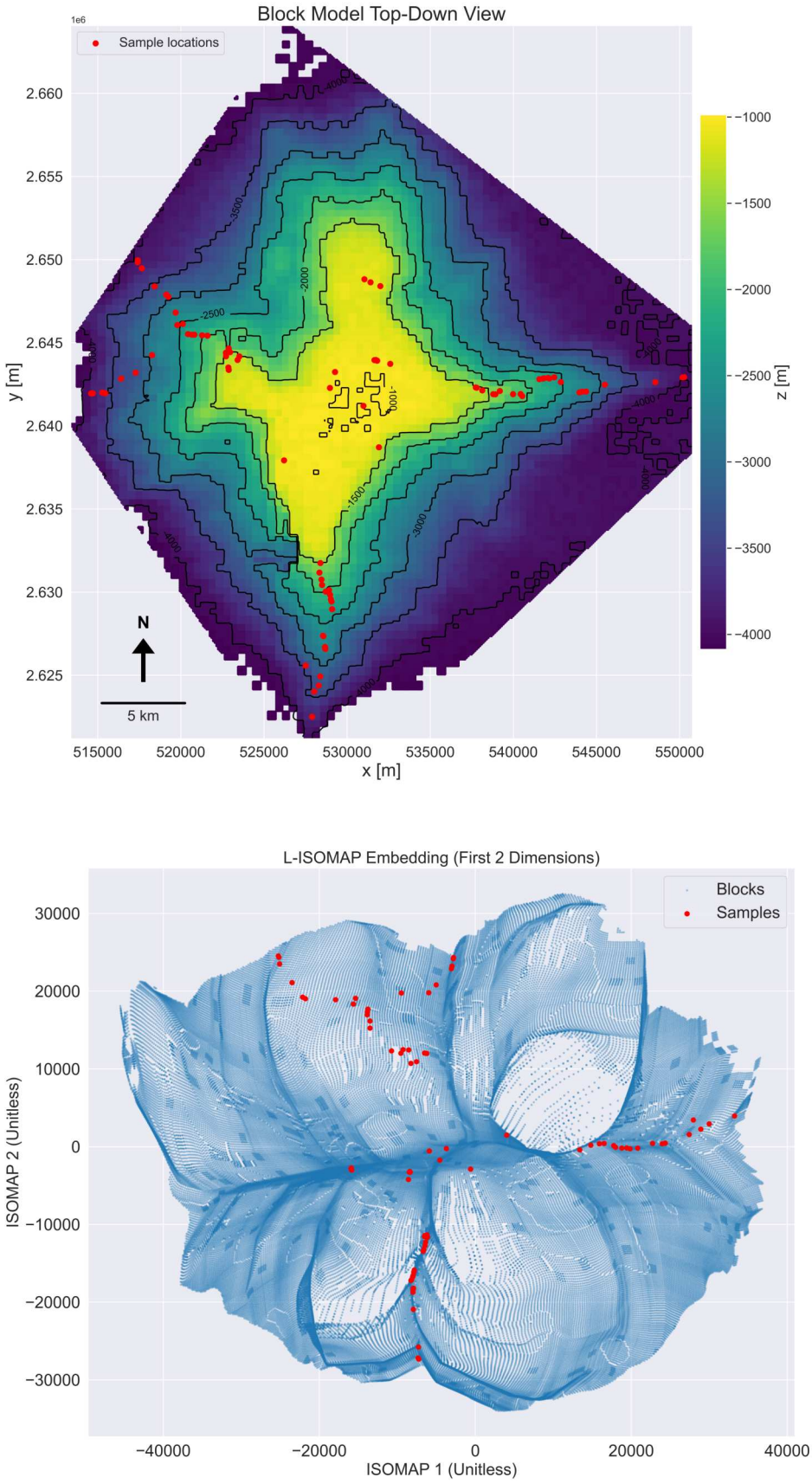


Figure 4.9: Block model representation before (top) and after (bottom) multidimensional embedding using L-ISOMAP. The top images show the original block model, while the bottom image displays the transformed block model after embedding, revealing the intrinsic spatial structure captured by L-ISOMAP. The embedding preserves the shortest path distances between the points but this is relative and therefore the distances in this space is unitless or abstract.

does not correspond to any physical unit (such as meters or kilometers) or cardinal directions from the original data. Any subsequent analysis in the embedded space will therefore be in unit-less distances and relative directions (no northing or easting) including variography.

Figure 4.9 shows the block model and sample locations within the first two dimensions of the embedded space, illustrating how the embedding transforms the coordinate system.

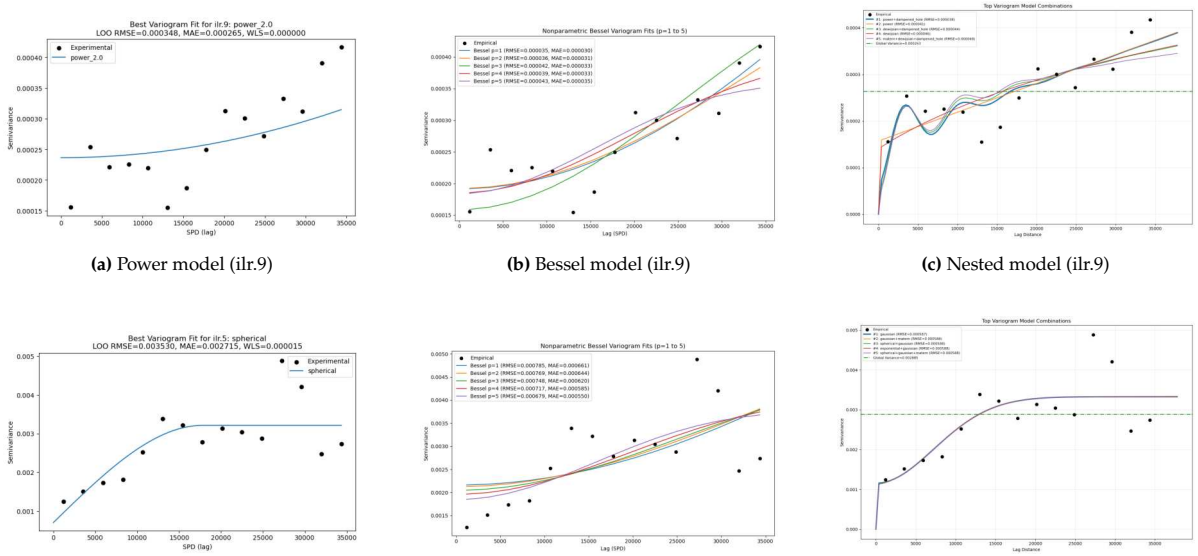
The embedded space and fitted variogram will be used for interpolation and kriging, leveraging the advantages of this method for modeling complex geological structures. However, there are multiple methods:

1. domaining into regions with constant variogram parameters,
2. locally reorienting the variogram at each point based on the LVA field, and
3. using shortest path distances (SPD) between points in the LVA field (used here),

that were suggested by Ghosh et al. (2024) to address local variations in anisotropy. Different results could be obtained with these alternative approaches. The specific approach chosen here was due to the fact that it would utilize all the available sample points and enable estimation with a single isoptropic omnidirectional variogram for each variable.

4.7. Variogram Analysis and Model Fitting

Following the ISOMAP embedding and ILR transformation, isotropic omnidirectional empirical variograms were computed, for each variable and theoretical models for modelling were fitted using three approaches: strictly parametric models, strictly non-parametric models, and flexible nested models that combine multiple functions. The methodology for variogram calculation and model fitting is described in detail in Section 3.5.5. Figure 4.10 illustrates the comparative performance of these approaches for two representative ILR coordinates in regards to their fit to the empirical variogram.



(a) Power model (ilr.9) **(b)** Bessel model (ilr.9) **(c)** Nested model (ilr.9)

(d) Spherical model (ilr.5) **(e)** Bessel model (ilr.5) **(f)** Nested model (ilr.5)

Figure 4.10: Comparison between strictly parametric variogram fitting function, strictly non-parametric variogram fitting function and nested variogram fitting function that finds the ideal linear combination of multiple functions. It can be observed that the nested function is flexible enough to find a better fit when the strict functions cannot in the 3rd sub-figure.

Figure 4.10 demonstrates how the flexible nested variogram model can provide superior fits in cases where strictly parametric or non-parametric models are insufficient. For ilr.5, the basic spherical model offers the best fit, while for ilr.9, a linear combination of power and dampened hole models better

captures the experimental variogram, including the critical "hole effect" structure. This highlights the importance of both goodness-of-fit metrics and visual inspection in variogram model selection.

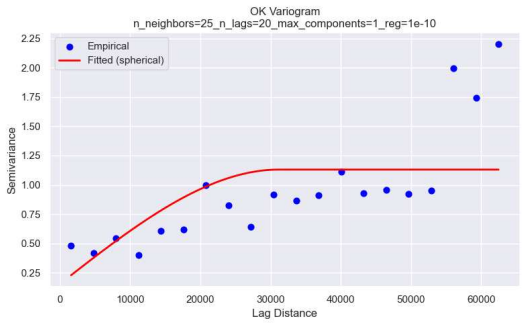
While nested variogram models generally provided greater flexibility and improved fit to the complex spatial correlation structures observed in the data, this did not always translate to more stable kriging or improved estimation accuracy. In practice, the kriging implementation was found to be more stable and reliable when restricted to a single parametric variogram model, as determined by further optimization using QKNA metrics (see Section 3.5.11). This will be elaborated further in the Section 4.8. These metrics directly assess the quality and stability of kriging predictions. It is important to note that this outcome is specific to the present dataset and modeling context, and should not be generalized without further validation.

To illustrate the final parametric variogram models selected for kriging, Figures 4.11 present the fitted parametric variograms for ilr.5 and ilr.9, for both Ordinary Kriging (OK) and Linear Model of Coregionalization (LMC). These plots correspond to the final optimized models used in subsequent interpolation and estimation steps.

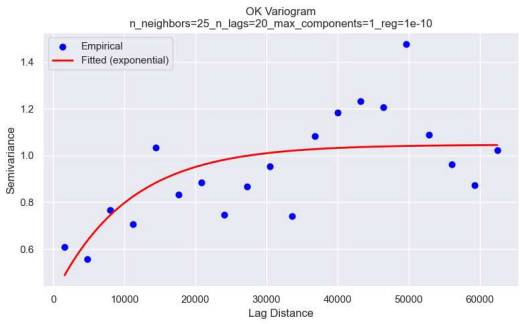
For ilr.9, the empirical cross-variogram (Figure 4.11d) displays negative values at all non-zero lags. As discussed in the methodology (Section 3.5.10), this behavior highlights a key limitation of the LMC: it cannot accommodate situations where the cross-correlation between variables is stronger at a distance than at the same location, nor can it model asymmetric cross-covariances (lag effects). In such cases, to maintain the positive definiteness required for co-kriging, the fitted cross-variogram model is set flat at zero, indicating the absence of a meaningful cross-correlation structure between ilr.9 and the primary variable within the LMC framework.

The negative cross-variogram values at non-zero lags may indicate the presence of a lag effect, where the correlation between variables is maximized at a spatial separation rather than at collocated points. One plausible explanation for this phenomenon is the influence of post-depositional processes such as mass wasting, material flow, or diagenetic alterations (e.g., phosphatization), which could introduce a spatial offset or phase shift in the relationship between water depth and crust composition. These geological processes may result in the observed asymmetry in cross-covariance, further limiting the applicability of the LMC in accurately capturing the spatial relationships between these variables.

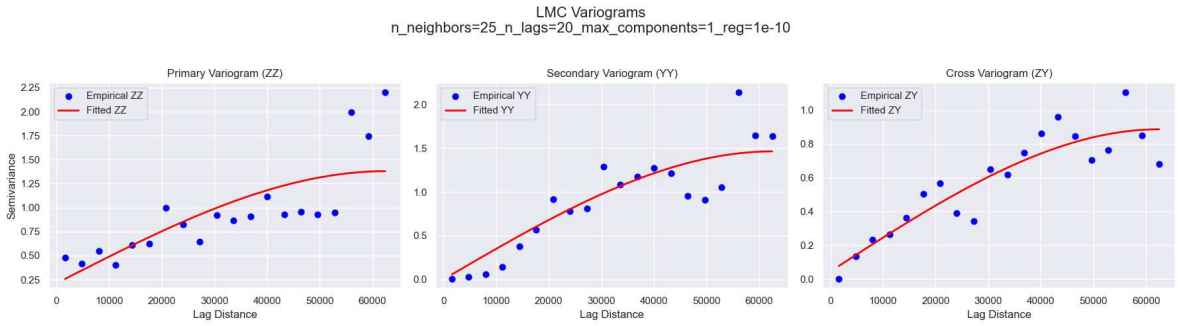
The full set of empirical and fitted variograms for all variables and both methods (OK and LMC) is provided in Appendix B.



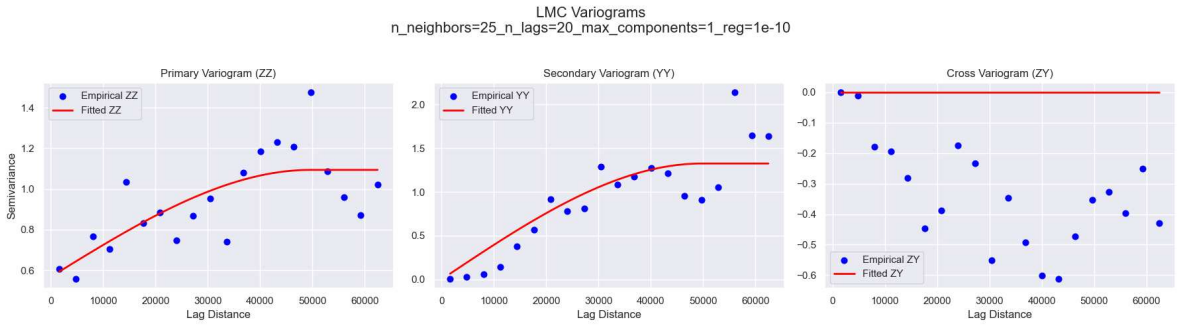
(a) OK parametric variogram fit for ilr.5



(b) OK parametric variogram fit for ilr.9



(c) LMC parametric variogram fit for ilr.5



(d) LMC parametric variogram fit for ilr.9

Figure 4.11: Parametric variogram fits for ilr.5 and ilr.9. Top: Ordinary Kriging (OK) fits for ilr.5 and ilr.9. Bottom: Linear Model of Coregionalization (LMC) fits for ilr.5 and ilr.9.

4.8. Estimation Results

The following subsections present the results of the 2 different variogram fitting approaches, (i) nested variogram functions that used both parametric and non parametric variograms, and (ii) singular parametric variogram functions.

4.8.1. Summary of Kriging using the nested variogram function

Significant improvements in accuracy were observed across nearly all ILR coordinates when employing co-kriging, as compared to ordinary kriging, using the flexible nested variogram fitting function that incorporates linear combinations of parametric and non-parametric variograms. This is illustrated in Figure 4.12. Kriging accuracy seems to significantly improved compared to OK when using depth as a secondary variable for co-kriging. However, the Quantitative Kriging Neighborhood Analysis (QKNA) metrics for this implementation summarized in Table 4.6 revealed invalid values for Kriging Variance (KV) and Kriging Efficiency (KE) in the case of ordinary kriging, with both metrics expected to lie between 0 and 1. For the two co-kriging implementations, while the metric values were mostly valid, they were consistently poor. Ideally, KV should be as close to zero as possible, and KE should approach one (Barboza and Deutsch, 2024). These results indicate that the kriging plans were either unstable (OK) or inefficient (SCK and ICCK), despite the promising gains indicated by leave one out cross-validation (LOOCV).

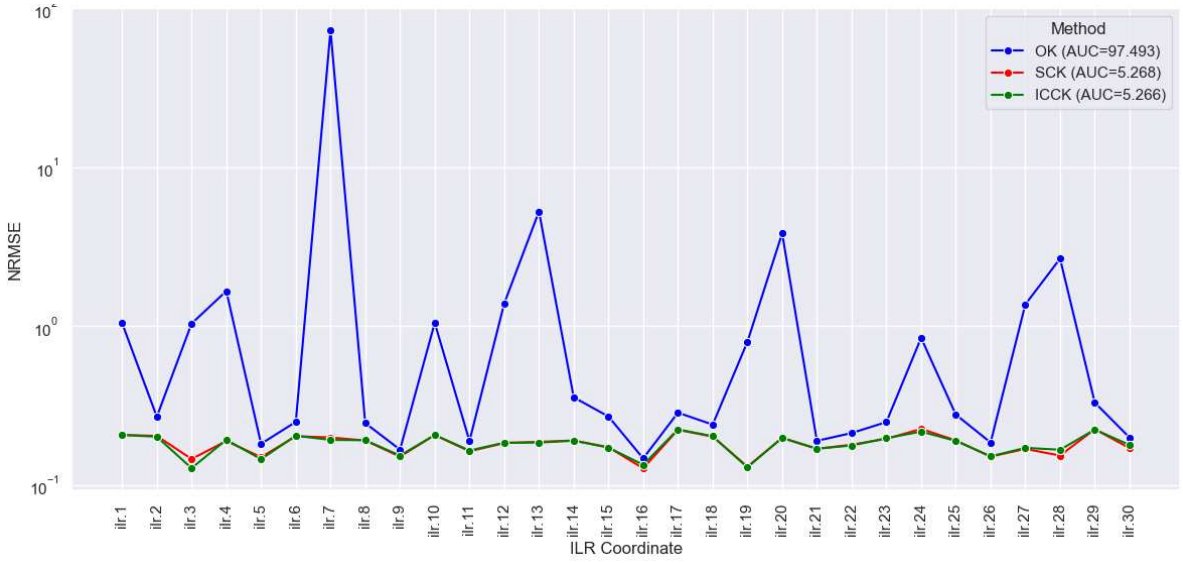


Figure 4.12: co-kriging gain across all 30 ILR coordinates in mean NRMSE values along with Area Under the Curve for all 3 methods for ease of comparison

Figure 4.13: Comparison of NRMSE between Inverse Distance Weighted (IDW) and Intrinsic Collocated Co-Kriging (ICCK) estimates across all ILR coordinates

Table 4.6: Summary of mean Kriging Variance (KV) and mean Kriging Efficiency (KE) values for OK, SCK and ICCK using nested variograms across all ILR coordinates.

ILR coordinate	mean ok KV	mean ok KE	mean sck KV	mean sck KE	mean icck KV	mean icck KE
ilr.1	5018.98	-4959.62	1.06	-0.05	0.77	0.24
ilr.2	1.66	-0.64	1.06	-0.05	1.01	0.01
ilr.3	19125.61	-18902.22	552.89	-545.46	1.04	-0.03
ilr.4	18087.37	-17876.06	0.98	0.03	0.72	0.28
ilr.5	1.26	-0.25	2.20	-1.17	1.07	-0.06
ilr.6	831.09	-820.43	0.83	0.18	0.82	0.19
ilr.7	62735480.00	-62005990.00	4.58	-3.53	0.95	0.06
ilr.8	1.26	-0.24	0.99	0.02	0.99	0.03
ilr.9	0.87	0.14	0.84	0.17	0.75	0.26
ilr.10	1575.48	-1556.16	13.93	-12.77	0.90	0.12
ilr.11	1.06	-0.05	0.96	0.05	0.92	0.09
ilr.12	184.13	-180.99	171.83	-168.83	1.21	-0.20
ilr.13	310036100.00	-306431000.00	0.94	0.07	0.92	0.09
ilr.14	198.69	-195.38	0.75	0.26	0.73	0.27
ilr.15	1.97	-0.95	1.01	0.00	1.01	0.00
ilr.16	0.83	0.18	1.66	-0.64	1.02	-0.01
ilr.17	1.44	-0.42	1.87	-0.85	0.96	0.05
ilr.18	1.18	-0.16	0.98	0.03	0.98	0.03
ilr.19	932403.50	-921560.60	4.66	-3.60	3.64	-2.60
ilr.20	1207346.00	-1193306.00	2.59	-1.56	0.87	0.14
ilr.21	1.76	-0.73	1.15	-0.14	0.96	0.05
ilr.22	1.17	-0.15	1.03	-0.02	0.96	0.05
ilr.23	1.25	-0.23	1.01	0.00	1.01	0.00
ilr.24	217.41	-213.88	1.07	-0.06	0.97	0.04
ilr.25	1.99	-0.97	1.01	0.00	1.01	0.00
ilr.26	1.09	-0.08	0.99	0.03	0.97	0.05
ilr.27	805.67	-795.30	1.08	-0.06	1.01	0.01
ilr.28	162315300.00	-160427900.00	1.78	-0.76	0.49	0.51
ilr.29	575.25	-567.56	0.69	0.32	0.68	0.75
ilr.30	1.13	-0.11	0.99	0.03	0.94	0.07

Consequently, it cannot be concluded with confidence that these results represent real life improvements in estimation accuracy or reliability. The observed instabilities and inefficiencies may be attributed to the incompatibility of nested variograms with the conventional kriging plan which usally uses parametric variograms, or to the nested variograms being unsuitable for the LMC in co-kriging, which is also typically performed using parametric variograms. This also highlights the limitation of LOOCV as a comparative metric across multiple estimation methods.

Figure 4.13 presents a comparison of ICCK, which achieved the lowest NRMSE in Figure 4.12, with IDW interpolation in terms of NRMSE values across the ILR coordinates. Based solely on NRMSE, co-kriging of ILR coordinates with water depth as a secondary variable appears to yield more accurate results than IDW. However, these results are derived from an unconventional approach employing nested variograms that combine linear combinations of parametric and, more importantly, non-parametric models. As indicated by the generally poor or invalid QKNA metrics (Table 4.6), this approach leads to unstable and potentially unreliable kriging solutions.

The presence of invalid QKNA metrics, in the case of Ok, or generally poor metrics for SCK and ICCK, suggests that the kriging solution lacks stability and efficiency, and the observed improvements in NRMSE may not reflect genuine gains in estimation accuracy. These findings could highlight the importance of using conventional, well-validated variogram models in geostatistical interpolation which could ensure more stability. In light of these considerations, the more conventional approach using parametric variograms is subsequently explored and discussed in the following section. However, the better LOOCV values for CK and ICCK suggest that the model could indeed capture the spatial distribution of the variables even with unstable QKNA matrices which could be a result of the previously stated advantages of nested variograms being able to capture complex spatial distribution patterns.

Due to the poor QKNA metrics, parametric variogram functions were used for the estimation next and the results of that approach is as follows.

4.8.2. Summary of Kriging using Parametric Variogram Functions

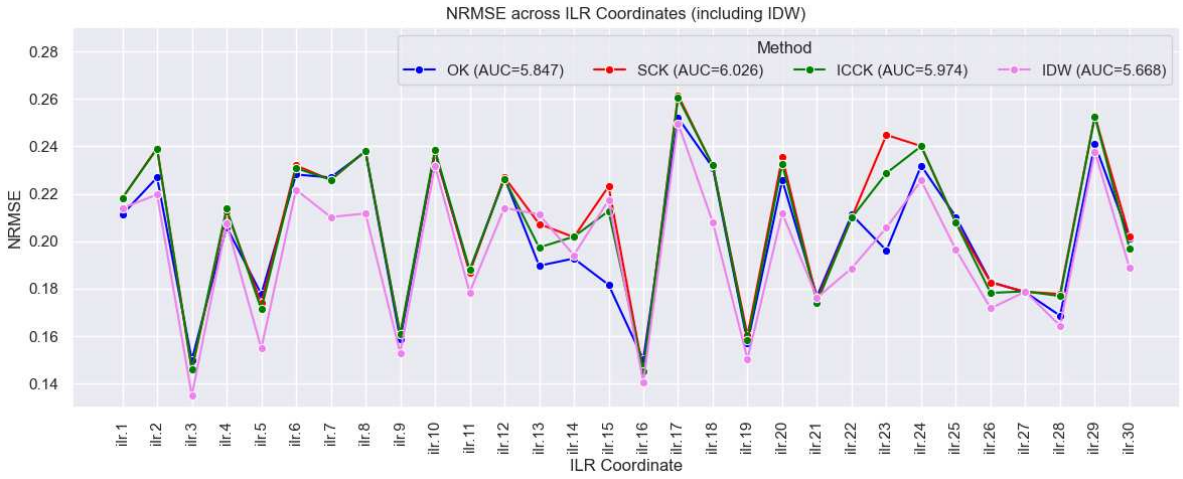


Figure 4.14: Comparison of NRMSE values of IDW and the three kriging methods (OK, SCK, and ICCK) with parametric variograms

Kriging conducted with conventional parametric variogram models did not yield any notable improvement in NRMSE values from cross-validation for co-kriging; in fact, it performed marginally worse than both Ordinary Kriging and Inverse Distance Weighting (IDW), as illustrated in Figure 4.14. This outcome further exemplifies the limitations of relying solely on LOOCV-based metrics for method comparison, as such metrics may not fully capture the underlying performance or reliability of the

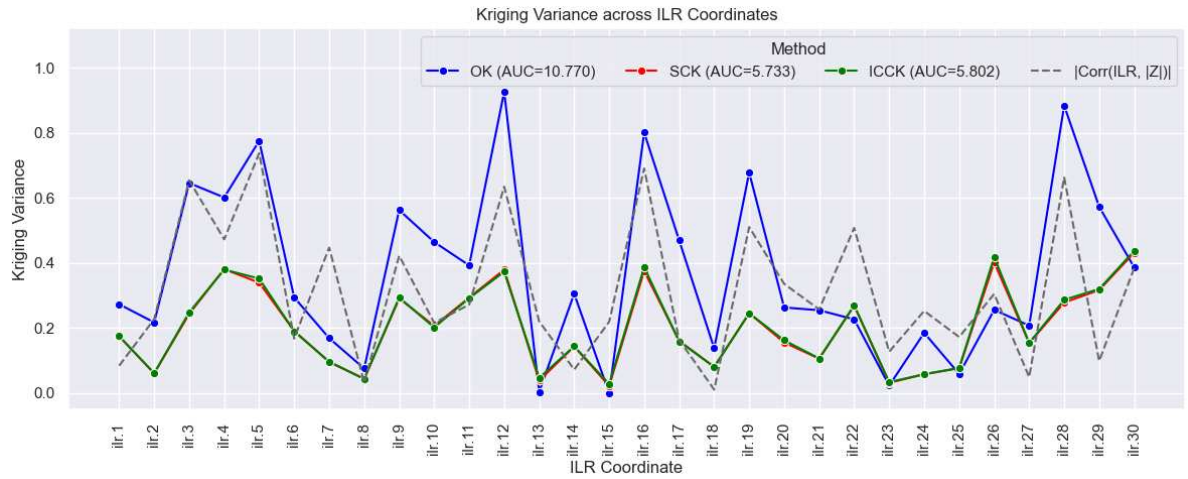


Figure 4.15: Comparison of Kriging Variances of the three methods, OK, SCK, and ICCK along with the absolute value of correlation of ILR coordinates with water depth

interpolation methods. In contrast, the mean kriging variance and kriging efficiency values across all blocks, as derived from QKNA metrics, which are consistently close to ideal, (Figure 4.7), present a different perspective.

Specifically, the analysis revealed a correlation coefficient of 0.624 (Figure 4.15) between the gain in co-kriging efficiency (fall in Kriging Variance) for ICCK and the absolute value of the correlation between water depth and ILR coordinates. The plot shows how the co-kriging gains mostly coincide with the higher correlation values. This finding indicates that improvements in both kriging variance and kriging efficiency through co-kriging are closely linked to the correlation with water depth, thereby quantitatively demonstrating the relationship between water depth and the composition of ferromanganese crusts on the Tropic Seamount. While it is generally said that correlation does not imply causation, in this context **the correlation is observed between co-kriging gain and the secondary variable (water depth) which was specifically chosen for co-kriging. Therefore, it can be reasonably assumed that this correlation does reflect a causal relationship. That is, utilizing water depth as secondary variable in co-kriging leads to better estimation of Fe-Mn crust composition.**

It is important to note, however, that Table 4.7 shows a slightly negative Ordinary kriging variance for ilr.15 (–0.0011), which is theoretically invalid and suggests instability in the kriging solution for this coordinate. This issue may arise from the automated kriging workflow failing to identify optimal variogram parameters for ilr.15, and could potentially be mitigated through more tailored variographic analysis. But it still exhibited valid values for SCK and ICCK. Since ICCK performed consistently well across all the approaches, Further resource estimation and tonnage calculation using back-transformed variables and the block model will utilize the ICCK estimates obtained using kriging using parametric variograms.

These observations also underscore the importance of considering multiple evaluation metrics and geostatistical diagnostics, rather than relying exclusively on LOOCV, when assessing the performance and reliability of interpolation methods.

The limitations of using LOOCV for comparing kriging implementations sated in 3.5.8 can be observed in this study as well. Figure 4.7 shows exactly how samples are oriented as linear clusters and how there is sparse data in areas away from them. The samples were mostly taken from the arms of the Tropic Seamount. Negative weights were observed in the kriging estimation. Both the string effect and the appearance of negative weights reflect the influence of redundancy on kriging's variance minimization. They can lead to a misrepresentation of spatial patterns: by overvaluing data with limited contextual

support and undervaluing denser, high-quality clusters. Consequently, kriging may underestimate or overestimate values in ways that hinder accurate spatial prediction (Markvoort and Deutsch, 2024). This combined with the other limitations of LOOCV like positive and negative biases and inconsistencies in hyperparameter settings make LOOCV a poor choice for comparison of different interpolation methods. This further reinforces the choice of using QKNA metrics for the comparison.

In addition to the limitations of LOOCV, the quality of kriging results in this study is also affected by challenges and limitations in variogram modeling and sample distribution. The automated approach for fitting variograms led to an invalid variance value for ilr.15 as stated earlier, suggesting that a more careful, manual variogram analysis and tailored variography is needed to achieve more reliable results. Then there is the fact that the samples are mostly concentrated along a few linear paths, with large gaps and sparse coverage elsewhere (Figure 4.7). This uneven and sparse sampling makes it difficult for kriging to accurately capture spatial patterns, and can result in negative weights and misleading estimates. Furthermore, these gaps in sampling corresponds to areas on the seamount which are affected by landslides and other mass wasting events and this means that it is possible that those areas could have very thin crusts or even exposed bedrock and no crusts at all. A scenario which is not considered in this study. Improving both the density and distribution of sampling, as well as refining variogram modeling, would help produce more robust and reliable kriging outcomes.

Table 4.7: Summary of kriging variance (KV) and kriging efficiency (KE) QKNA metrics for OK, SCK and ICCK using parametric variogram models.

ilr_coordinate	mean_KV_ok	mean_KE_ok	mean_KV_sck	mean_KE_sck	mean_KV_icck	mean_KE_icck
ilr.1	0.27	0.73	0.17	0.83	0.17	0.83
ilr.2	0.22	0.79	0.06	0.94	0.06	0.94
ilr.3	0.65	0.36	0.24	0.76	0.25	0.76
ilr.4	0.60	0.41	0.38	0.62	0.38	0.63
ilr.5	0.78	0.23	0.34	0.66	0.35	0.65
ilr.6	0.29	0.71	0.19	0.81	0.19	0.81
ilr.7	0.17	0.83	0.10	0.91	0.10	0.91
ilr.8	0.08	0.93	0.04	0.96	0.04	0.96
ilr.9	0.56	0.44	0.29	0.71	0.29	0.71
ilr.10	0.46	0.54	0.21	0.80	0.20	0.80
ilr.11	0.39	0.61	0.29	0.71	0.29	0.71
ilr.12	0.93	0.08	0.38	0.63	0.37	0.63
ilr.13	0.00	1.00	0.04	0.96	0.05	0.95
ilr.14	0.31	0.70	0.14	0.86	0.14	0.86
ilr.15	0.00	1.00	0.02	0.98	0.03	0.97
ilr.16	0.80	0.21	0.38	0.63	0.39	0.62
ilr.17	0.47	0.54	0.16	0.84	0.16	0.84
ilr.18	0.14	0.86	0.08	0.92	0.08	0.92
ilr.19	0.68	0.33	0.25	0.76	0.25	0.76
ilr.20	0.26	0.74	0.16	0.85	0.16	0.84
ilr.21	0.25	0.75	0.11	0.90	0.11	0.90
ilr.22	0.23	0.78	0.27	0.73	0.27	0.73
ilr.23	0.02	0.98	0.03	0.97	0.03	0.97
ilr.24	0.18	0.82	0.06	0.94	0.06	0.94
ilr.25	0.06	0.94	0.08	0.92	0.08	0.92
ilr.26	0.26	0.75	0.40	0.60	0.42	0.59
ilr.27	0.21	0.80	0.15	0.85	0.15	0.85
ilr.28	0.88	0.13	0.28	0.73	0.29	0.72
ilr.29	0.57	0.43	0.32	0.69	0.32	0.68
ilr.30	0.39	0.62	0.43	0.57	0.44	0.57

Impact of Kriging Parameter Optimization

The results of the kriging parameter optimization for the parametric variogram-kriging plan revealed that there was only a negligible difference in kriging variance (KV) and kriging efficiency (KE) among the top five parameter sets. As shown in Table 4.8, all selected parameter sets produced very similar metrics for ilr.6 taken as an example, indicating that any of them could have been chosen without significantly affecting the quality of the spatial predictions. For consistency and ease of interpretation, the parameter set with the lowest mean kriging variance and highest mean kriging efficiency was selected for the final block model estimates. However, the other top parameter sets would also have yielded satisfactory results, as reflected by their comparable performance metrics. This demonstrates the robustness of the kriging procedure to small changes in parameter selection within the optimized range. This also demonstrates how much more stable the parametric variograms are in comparison with the nested variogram kriging (Table 4.6) which produced a KV of 831 and KE of -820 for the same ilr.6. This is the case across all ILR coordinates.

Table 4.8: Comparison of top kriging parameter sets based on mean kriging variance (KV) and mean kriging efficiency (KE).

Parameter set	Mean KV	Mean KE
n neighbors=25, n lags=15, max components=1, reg=1e-10	0.189	0.813
n neighbors=25, n lags=25, max components=1, reg=1e-10	0.192	0.811
n neighbors=20, n lags=15, max components=1, reg=1e-10	0.193	0.810
n neighbors=20, n lags=25, max components=1, reg=1e-10	0.195	0.808
n neighbors=15, n lags=15, max components=1, reg=1e-10	0.198	0.804
n neighbors=15, n lags=25, max components=1, reg=1e-10	0.200	0.802
n neighbors=25, n lags=20, max components=1, reg=1e-10	0.210	0.793
n neighbors=20, n lags=20, max components=1, reg=1e-10	0.213	0.790
n neighbors=15, n lags=20, max components=1, reg=1e-10	0.219	0.784

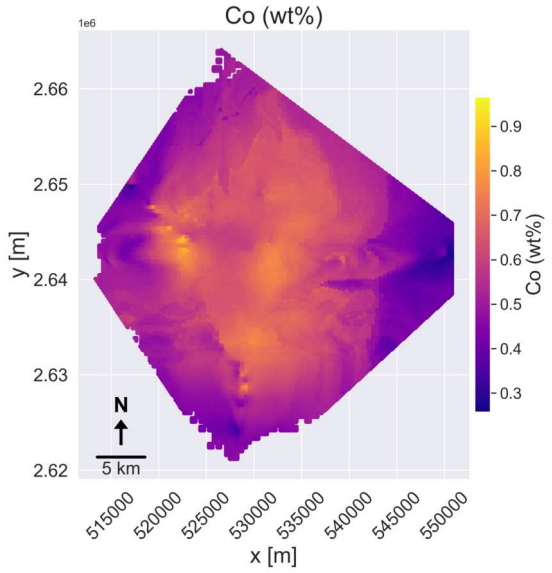
4.8.3. Back transformation to raw variables and visualization of block wise estimates and variances

The block-ids and original x,y and z coordinates were always saved along with the isomap coordinates. Therefore, there was no need to do a back transformation from the embedded space to the real cartesian space and subsequent plots will utilize the cartesian coordinates. To interpret these results of ICCK in the original compositional space, a back transformation from isometric log-ratio (ILR) coordinates to raw metal grades was performed using the methodology described in Section 3.5.13. This approach enables the direct comparison and visualization of estimated metal concentrations in their natural units.

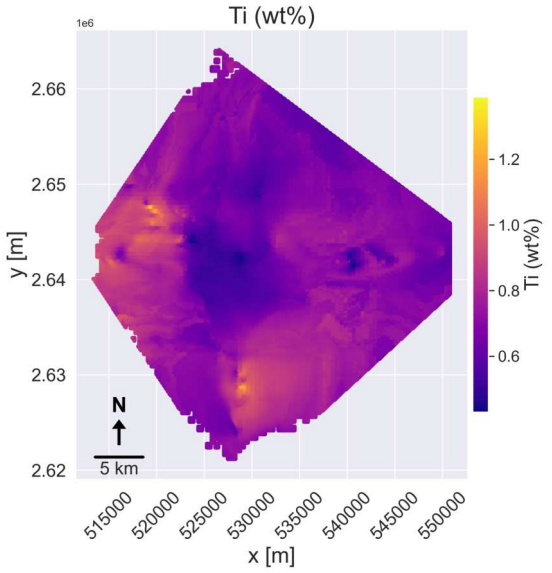
The block-wise estimates and variances of cobalt, titanium, lithium, magnesium, and manganese are given from both a top-down view Figures 4.16 and 4.17, where each subplot corresponds to one of the selected metals, allowing for a spatial assessment of both the estimated grades and their associated uncertainties across the study area.

The choice to focus on these metals were guided by the European Union's 2023 classification of critical raw materials (CRMs), as listed in the *Fifth list 2023 of critical raw materials for the EU* (Commission, 2023). Among the 34 materials deemed critical for strategic and economic resilience, cobalt, titanium, lithium, magnesium, and manganese were present within the geochemical dataset used in this study. Rare earth elements, both light (LREEs) and heavy (HREEs) are also classified as CRMs, but their plots are given in the Appendix C along with the rest of the non-CRM elements due to space constraints.

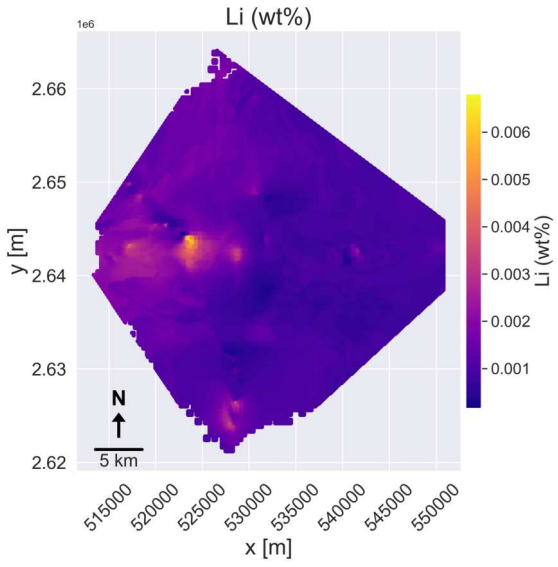
The variance plots (Figure 4.17) demonstrate that regions closer to sampled locations (Figure 4.7) exhibit predictably lower simulated variances, indicating greater confidence in the estimated values. And areas distant from sample points consistently display higher variance, reflecting increased uncertainty in those estimates. Notably, lithium deviates from this spatial pattern, presenting mostly uniformly low



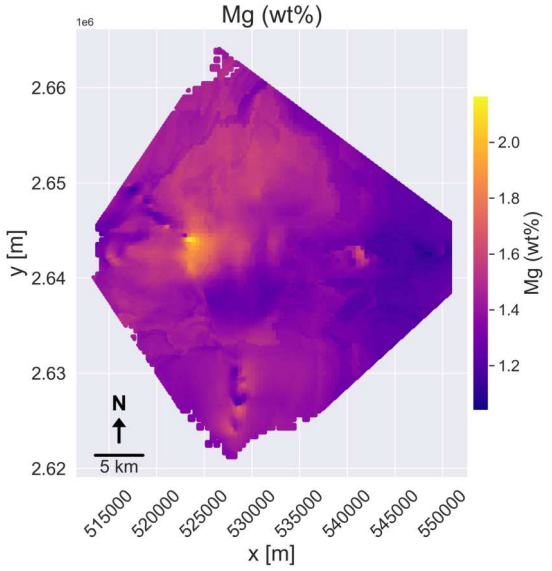
(a) Cobalt



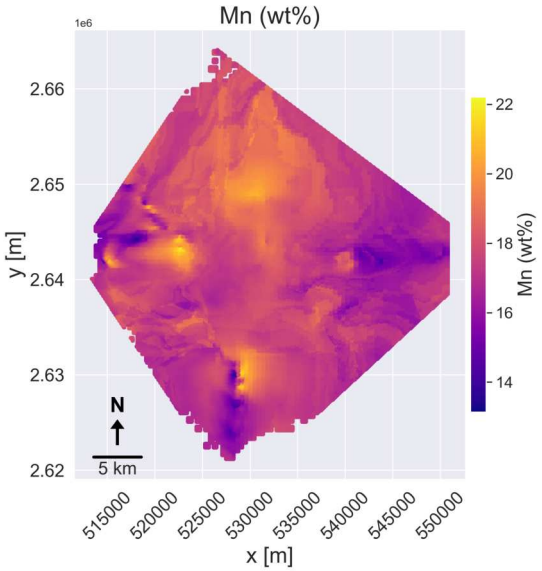
(b) Titanium



(c) Lithium

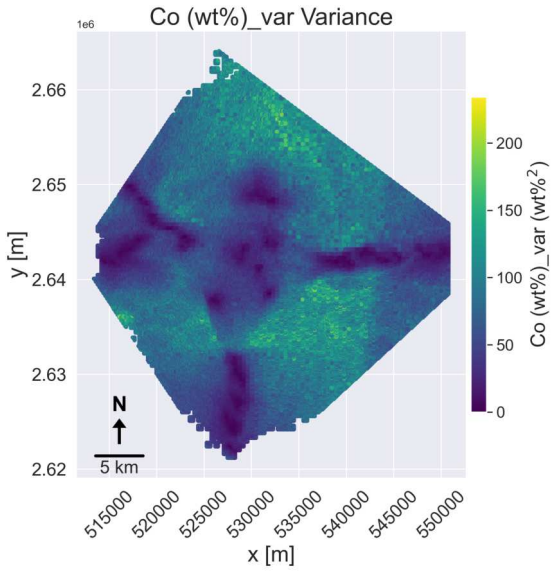


(d) Magnesium

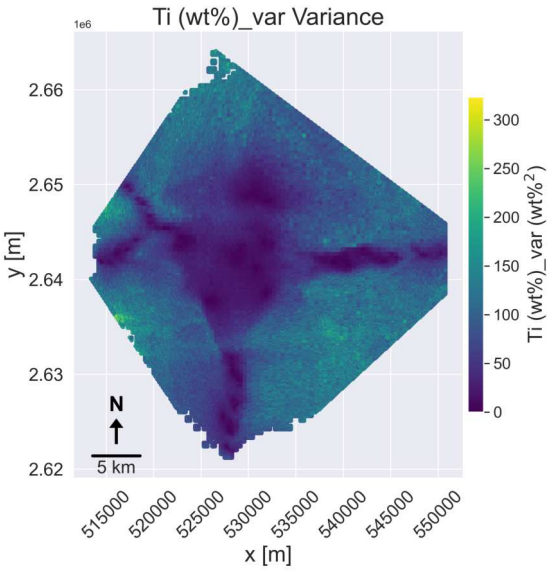


(e) Manganese

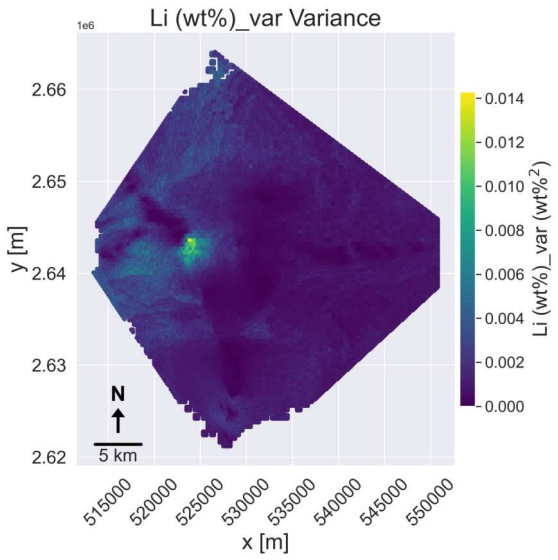
Figure 4.16: Block-wise estimates of selected critical raw materials (CRMs): Cobalt, Titanium, Lithium, Magnesium, and Manganese. Each plot shows spatial grade estimates across the study area, with selected metals drawn from those present in the dataset that are classified as CRMs by the European Union Commission, 2023.



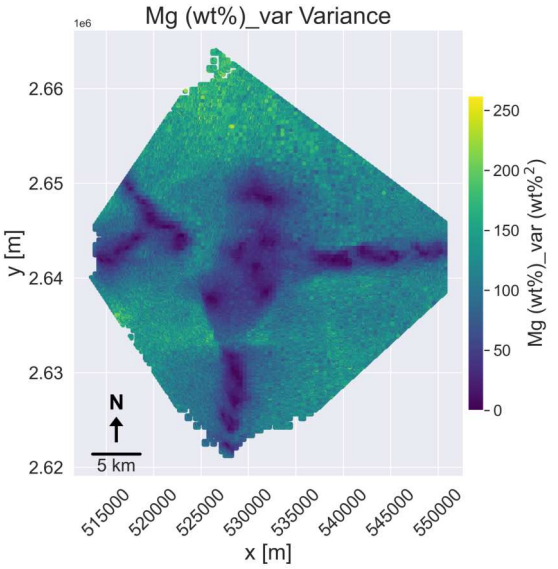
(a) Cobalt



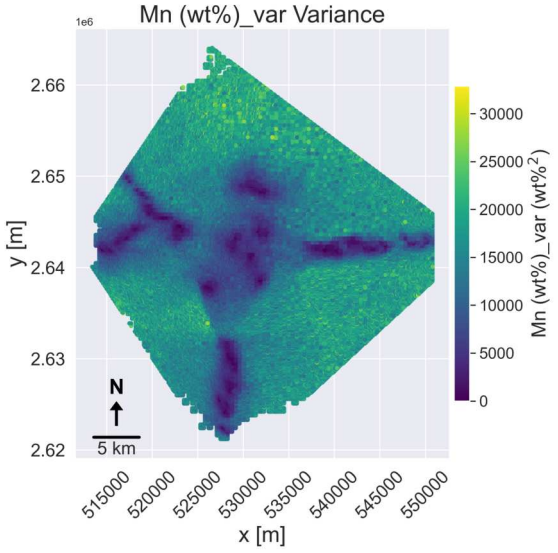
(b) Titanium



(c) Lithium



(d) Magnesium



(e) Manganese

Figure 4.17: Block-wise simulated kriging variances of selected critical raw materials (CRMs): Cobalt, Titanium, Lithium, Magnesium, and Manganese. These plots illustrate the spatial distribution of uncertainty in grade predictions, based on simulated variance of estimates. Lower values of variances means lower uncertainty and higher confidence in the estimation.

variance values across the block model. However, given that these variances are actually simulated values rather than directly obtained from kriging, it would be premature to conclude that Lithium concentrations are estimated with uniformly high confidence throughout the study area.

Following up on that, it is important to reiterate that, as discussed in Section 3.5.13 (see also Paragraph 3.5.13), the variances shown in Figure 4.17 are not direct back transformations of kriging variances. Due to the nonlinear nature of the ILR back transformation, there is currently no established analytical method for propagating variances from ILR space to compositional space. Instead, the variances were approximated through a simulation-based approach, which provides a pragmatic solution in the absence of a formal variance propagation method. Alternative strategies, such as using centered log-ratio (CLR) or log transformations, could allow for direct back transformation of variances due to their linearity, but these approaches may not fully respect the geometry of compositional data. Another theoretically robust, yet computationally intensive, solution would be to perform co-kriging on all ILR coordinates simultaneously, enabling variance simulation as outlined by Texas_A&M_University (2017); however, this was beyond the scope of the present study.

4.9. Tonnage Calculation Results

The tonnage calculations on the back-transformed estimation results indicate a total of 27,122,579.81 tonnes of metallic elements (which is excluding the "Others" elements). The overall crust tonnage is calculated at 52,952,367.28 tonnes (in dry weight), corresponding to a metal content of 51.22% by dry weight. Table 4.9 summarizes the calculated total tonnages for the 31 elements considered in this study based on the ICCK estimates. The total market valuation of each estimated metal present in the Fe–Mn crust on the Tropic Seamount was calculated by multiplying its estimated weight by the corresponding average market price, as reported by reputable metal market and exchange platforms. These figures reflect pricing data from August 2025 and exclude costs associated with extraction, processing, and refinement. It should be noted that market prices are subject to fluctuation and will vary over time. The total market worth estimate of the major, minor and trace metallic elements present in the crust is calculated at 59.41 billion euros.

Cut-off grade vs Tonnage plots of all elements are given in Appendix C.

4.10. Spatial Patterns and Environmental Correlations

4.10.1. Wake Zone Analysis

Comparing the upstream, downstream, and wake zones simulated using ocean current data as described in Section 3.6 with the clustering results of elemental grade estimates (see Figure 4.18) did not reveal any strong spatial correspondence. However, the clustering analysis itself yielded several notable insights.

For instance, aluminium (Al) displays a distinct zonation pattern on and around the flat summit area (see Figure 4.18). This may be related to its known negative correlation with both slope and current velocity, as discussed in Section 2.2.

A particularly striking pattern emerges in the clustering of REEs (see Figure 4.19). Almost all REEs, except Ce, exhibit remarkably similar spatial patterns, with nearly all, except Dy and Tb, showing almost identical spatial clustering. This likely reflects the natural tendency of REEs to occur together in geological environments, a relationship that appears to persist even within ferromanganese crusts.

Despite the similar shapes of the clusters, the cluster assignments themselves are not always consistent. For example, while La and Pr in the light REE group among LREEs show similar clustering with cluster 1 being the larger group, Nd and Sm display clusters that appear to be the inverse, with cluster 2 being the larger group (see Figure 4.19). It is unclear whether this is due to Python randomly assigning cluster

Table 4.9: Estimated total tonnage of elements in the study area and their total estimated market worth (in million euros).

Element	Total Tonnage (tonnes)	Market Worth (million euros)
La	18,787.16	47.09
Ce	101,697.68	343.84
Pr	3,895.50	356.59
Nd	16,745.02	1513.28
Sm	3,403.01	28.97
Eu	740.59	180.41
Gd	3,248.94	1446.43
Tb	476.99	486.42
Dy	2,912.65	699.74
Ho	576.15	36.43
Er	207.53	479.39
Tm	1,469.73	2345.69
Yb	1,360.63	200.01
Lu	200.04	579.72
Y	10,020.55	252.52
Li	616.77	6.10
Co	305,066.88	8542.30
Ni	165,978.83	2112.39
Cu	32,323.98	261.37
Zn	33,917.98	79.21
Mo	32,006.05	1209.83
Te	2,715.86	880.59
Na	600,114.69	1235.04
Mg	741,628.94	1396.69
Al	758,375.64	1622.53
K	207,282.33	3492.62
Ca	1,379,411.95	4985.91
Ti	382,093.50	2109.34
Mn	9,320,003.06	15892.47
Fe	12,995,301.21	6582.38
Others	25,805,266.81	—

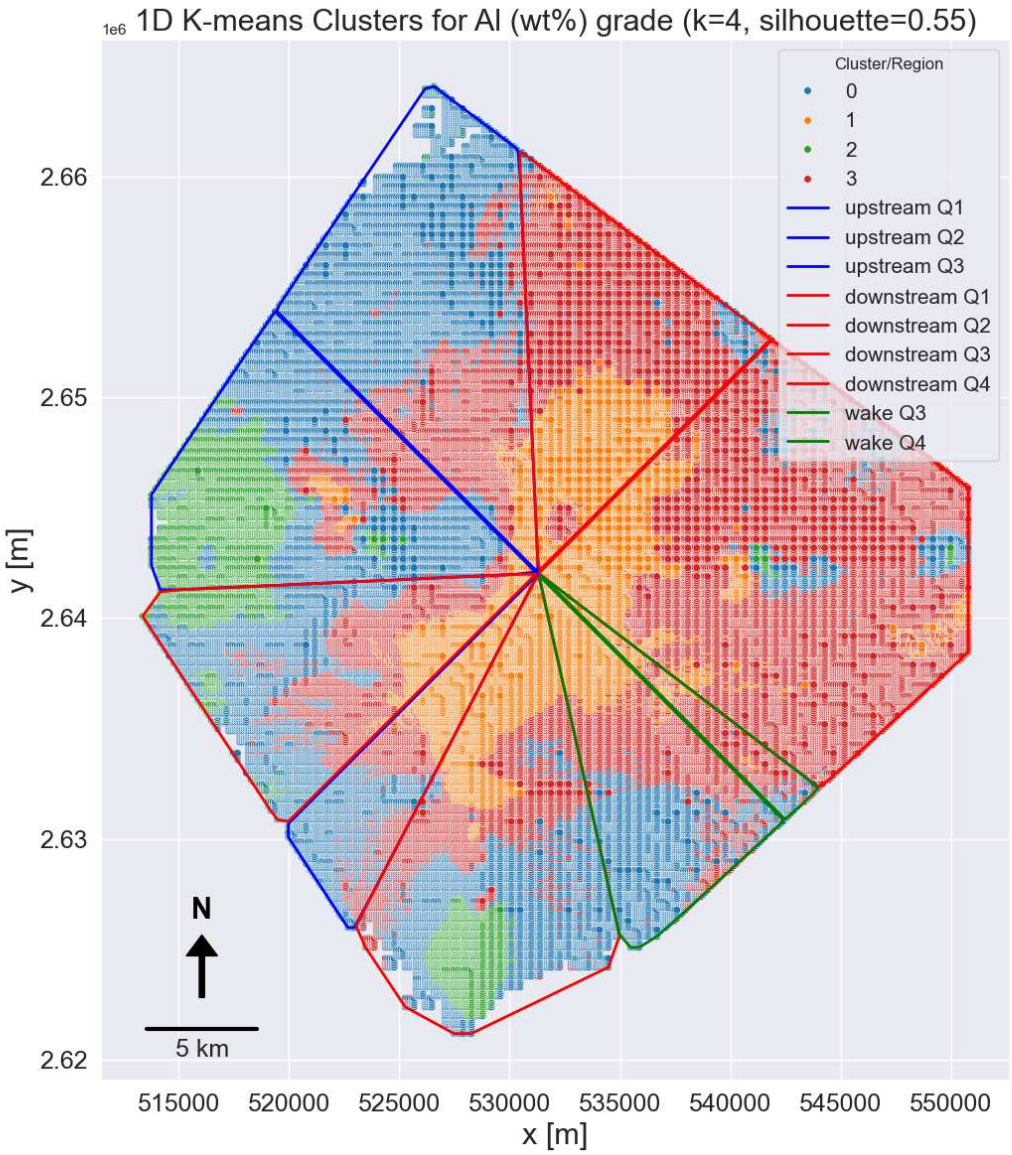


Figure 4.18: Upstream, downstream, and wake zone analysis with clustered estimate grades: example

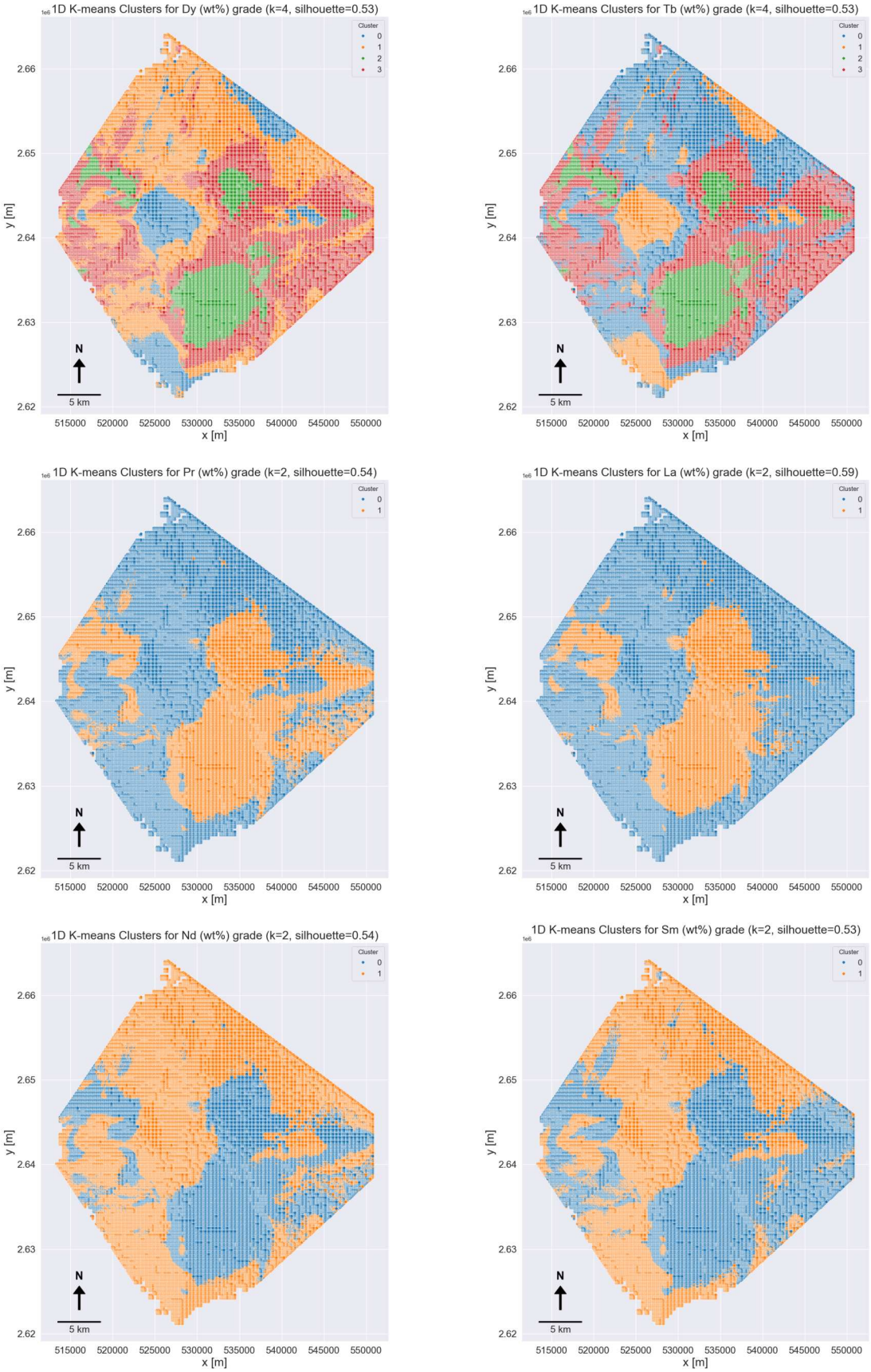


Figure 4.19: 1D K-means cluster plots for Dy, Tb, Pr, La, Nd, and Sm.

labels, or if these elements actually exhibit positive and negative spatial correlations as groups. This effect could also be a result of ILR transformations, which preserves compositional relationships which could have become exaggerated spatially during interpolation.

Both Dy and Tb also exhibit very similar clusters to each other, and some of these clusters are spatially similar to those of the other REEs, though there are even smaller clusters within them (see Figure 4.19). It is unclear why these two show additional number of clusters compared to the others as they all show quite similar distributions in the EDA plots. The full cluster plots and EDA plots for estimates are given in the Appendix A and C.

4.11. Implications for the Environmental Impact of Fe-Mn Crust Exploration and Mining

Although a comprehensive environmental impact assessment (EIA) is beyond the scope of this study, the improved accuracy and confidence provided by ICCK estimation, particularly when incorporating water depth as a secondary variable, can contribute to reducing the environmental footprint of both exploration and mining activities. More precise resource estimates mean that fewer drillholes and samples may be required to achieve the level of confidence mandated by international mineral resource reporting standards, such as the JORC Code. Additionally, enhanced spatial accuracy in grade distribution within the block model enables more selective mining, minimizing unnecessary disturbance of the seafloor. This targeted approach could help reduce the overall area affected by mining operations, thereby lowering the environmental impact associated with uncertainty and inaccuracy in resource estimation and extraction.

5

Conclusion

This study clearly demonstrates a significant relationship between water depth and Fe-Mn crust composition, and shows that incorporating this relationship through co-kriging (CK) improves the confidence and accuracy of resource estimation compared to ordinary kriging (OK). By applying these methods, this work provides, for the first time, tonnage calculations for metals in a Fe-Mn crust based on a block model that accurately captures the three-dimensional structure of the deposit and also using real geochemical data. These results highlight the Tropic Seamount as a substantial mineral resource with the potential to significantly contribute to Europe's critical raw materials supply chain.

Throughout the research process, a comprehensive workflow was developed, including data transformation using the isometric log-ratio (ILR), advanced geostatistical modeling, and the integration of locally varying anisotropy (LVA) via L-ISOMAP embedding. The study also critically evaluated the limitations of leave-one-out cross-validation (LOOCV) for comparing interpolation methods, recommending that future research explore more advanced cross-validation techniques. Additionally, the effectiveness of QKNA metrics as a robust framework for assessing and comparing kriging model performance was demonstrated. The nested variogram approach showed promising results by CK and ICCK being able to outperform IDW and OK in LOOCV metrics even with invalid QKNA metrics. Therefore, future research should look into how a stable nested variogram oriented kriging plan might perform in comparison to the standard parametric variogram oriented plans.

While the potential for CK to reduce the number of required drillholes was clearly identified, this effect was not quantitatively assessed due to the time and computational demands of performing multiple kriging analyses with varying sample sizes. Future work should systematically investigate how changes in sampling density and distribution influence resource estimation outcomes and drillhole optimization. This study considers the landslide scars and gullies to be also definitively covered by Fe-Mn crusts which might not be the ground truth. The sampling distribution in future studies should consider these locations as well since they can affect tonnages significantly. Building on the topic of sampling, this study relied on compositional data from geochemical analysis of only the top 1 cm of the crust samples. However, previous research has shown that Fe-Mn crusts are vertically layered, with compositions varying across different bands within a single crust. Therefore, the assumption that the top 1 cm is representative of the entire crust thickness is a simplification. For more accurate estimations, future studies should aim to analyze samples from the full thickness of the crust, ideally by extracting drill-cores and conducting detailed compositional analyses across all layers.

The advantages of using ILR transformation for compositional data were clear, as ILR coordinates captured meaningful geochemical relationships and enabled correlation studies without the closure problem. However, while ILR coordinates are geometrically independent, some correlations still appear in exploratory data analysis. It is generally not recommended to interpolate correlated variables separately, yet this study did not account for these interdependencies during kriging. Interpolating correlated variables independently can lead to under- or overestimation in the results. Future research should explore cokriging approaches that jointly interpolate all correlated ILR coordinates, thereby incorporating their interdependencies and potentially improving estimation accuracy and confidence. Another limitation of this study was the lack of compositional data for non-metallic elements. Future studies should consider including major non-metallic elements such as oxygen, carbon, and silicon in the SBP, and examine how this affects the ILR transformations, as these elements also contribute to Fe-Mn crust composition.

Automated variogram fitting was employed due to the large number of ILR coordinates and the limitations of available Python tools. However, a more hands-on approach to variography, incorporating expert knowledge and manual model fitting, could further enhance kriging results. Moreover, the assumption of stationarity was not explicitly tested, and instances of negative cross-variogram values suggested possible violations of this spatial autocorrelation. Future studies should incorporate more rigorous experimental variography and investigate the validity of stationarity and ergodicity in similar settings.

The integration of LVA through L-ISOMAP embedding proved highly effective, simplifying the variography and resource estimation process by enabling the use of a single isotropic variogram for each variable while accurately capturing anisotropy and spatial correlations. This approach holds promise for application to other complex marine and terrestrial mineral deposits.

Finally, this study assumed constant sea level and ocean current conditions throughout the history of the Tropic Seamount, despite known changes over geological timescales. Future research should incorporate global paleo sea level models and paleo-current data to better understand how these dynamic oceanographic factors have influenced Fe-Mn crust formation and composition.

In summary, this thesis advances the understanding of the factors controlling Fe-Mn crust composition, most importantly, the water depth-crust composition relationship, and presents an optimized, step-by-step methodology for ILR transformation and kriging of compositional geochemical data, including the handling of locally varying anisotropy using L-ISOMAP embedding. This approach enables more confident and accurate estimation of mineral resources in both marine and terrestrial environments, with significant potential for both the mineral resources industry and future research in this area.

References

ArcGIS. (2025). How inverse distance weighted interpolation works [Retrieved from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm].

Ashtiani, Z., & Deutsch, C. (2020). Kriging with constraints. In J. L. Deutsch (Ed.), *Geostatistics lessons*. Geostatistics Lessons. https://geostatisticslessons.com/lessons/krigingconstraints

Babak, O., & Deutsch, C. V. (2007). *Comparison of cokriging with an lmc versus the intrinsic model* (CCG Annual Report No. 2007-116). Centre for Computational Geostatistics, Department of Civil & Environmental Engineering, University of Alberta. https://www.ccgalberta.com/ccgresources/report09/2007-116_cokriging_comparison.pdf

Barboza, I., & Deutsch, C. V. (2024). Quantitative kriging neighborhood analysis (qkna) [Retrieved from Geostatistics Lessons]. In J. L. Deutsch (Ed.), *Geostatistics lessons*. http://www.geostatisticslessons.com/lessons/qkna

Baturin, G. N., & Dubinchuk, V. T. (2011). Mineralogy and chemistry of ferromanganese crusts from the atlantic ocean. *Geochemistry International*, *49*(6), 563–575. https://doi.org/10.1134/S0016702911060024

Benites, M., Hein, J. R., Mizell, K., Blackburn, T., & Jovane, L. (2020). Genesis and evolution of ferromanganese crusts from the summit of rio grande rise, southwest atlantic ocean. *Minerals*, *10*(4). https://doi.org/10.3390/min10040349

Black, W. (2020, July). The experimental variogram: Can't live with it, can't live without it... [Accessed: June 2, 2025]. https://lazymodellingcrew.com/post/post_09_expvariointro_wb/

Boisvert, J. B., & Deutsch, C. V. (2011). Programs for kriging and sequential gaussian simulation with locally varying anisotropy using non-euclidean distances. *Computers & Geosciences*, *37*(4), 495–510. https://doi.org/https://doi.org/10.1016/j.cageo.2010.03.021

Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. Wiley. https://doi.org/10.1002/9781118136188

Choueiry, G. (2025). Square root transformation: A beginner's guide. https://quantifyinghealth.com/square-root-transformation/

Clar, M. (2023). Heteroscedasticity. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3117–3118). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_3434

Comas-Cufí, M., & Thió-Henestrosa, S. (2011). CoDaPack 2.0: a stand-alone, multi-platform compositional software. In J. Egozcue, R. Tolosana-Delgado, & M. Ortego (Eds.), *Codawork'11: 4th international workshop on compositional data analysis*. https://www.compositionaldata.com/codapack.php

Commission, E. (2023). Critical raw materials [Accessed July 2025].

Copernicus Marine Service. (2024). Global ocean physics analysis and forecast (global_analysisforecast_phy_001_024). https://data.marine.copernicus.eu/product/GLOBAL_ANALYSISFORECAST_PHY_001_024/description

Cronan, D. S. (2024). *Deep-sea minerals developments in the 20th century*. Springer Cham. https://doi.org/10.1007/978-3-031-52342-7

Cuba, M. A., & Deutsch, C. V. (2012). *Cross variography calculation of exploratory, infill, and blasthole drilling* (CCG Annual Report No. Paper 311). Centre for Computational Geostatistics, University of

Alberta. https://www.ccgalberta.com/ccgresources/report14/2012-311_cross-variography.pdf

d'Ache C. (2013). Variance of transformed random variable [Math Stack Exchange. Retrieved from Math Stack Exchange]. https://math.stackexchange.com/questions/341876/variance-of-transformed-random-variable

Dias, P. M., & Deutsch, C. V. (2022). The decision of stationarity (J. L. Deutsch, Ed.) [In J. L. Deutsch (Ed.), Geostatistics Lessons].

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*(3), 279–300. https://doi.org/10.1023/A:1023818214614

Emmanuel Romaric, O., Sangodoyin, A., & Ogundiran, M. (2020). Assessment of ordinary kriging and inverse distance weighting methods for modeling chromium and cadmium soil pollution in e-waste sites in douala, cameroon. *Journal of Health and Pollution*, *10*, 200605. https://doi.org/10.5696/2156-9614-10.26.200605

Fuhg, J. N., Fau, A., & Nackenhorst, U. (2023). State-of-the-art and comparative review of adaptive sampling methods for kriging. *Archives of Computational Methods in Engineering*. https://repo.uni-hannover.de/items/d3c66b2a-3d16-489c-a6a8-44754ca3c8d7

Ghosh, D., Zhang, B., & Boisvert, J. (2024). Locally varying anisotropy [Retrieved from http://www.geostatisticslessons.com In J. L. Deutsch (Ed.), *Geostatistics lessons*. Geostatistics Lessons. http://www.geostatisticslessons.com/lessons/lva

Glasby, G., Mountain, B., tc, V., Banakar, V., Ramesh, R. P., & Ren, X. (2010). Role of hydrology in the formation of co-rich mn crusts from the equatorial n pacific, equatorial s indian ocean and the ne atlantic ocean. *Resource Geology*, *60*(2). https://doi.org/10.1111/j.1751-3928.2010.00123.x

Haight, F. A. (1967). *Handbook of the poisson distribution* [Accessed via National Library of Australia on 31 July 2025]. Wiley. https://nla.gov.au/nla.cat-vn529287

Halbach, P. E., Jahn, A., & Cherkashov, G. (2017). *Marine co-rich ferromanganese crust deposits: Description and formation, occurrences and distribution, estimated world-wide resources*. Springer. https://doi.org/10.1007/978-3-319-52557-0_3

Hartmann, K., Krois, J., & Rudolph, A. (2023). Statistics and geodata analysis using r (soga-r). https://www.geo.fu-berlin.de/en/v/soga-r/index.html

Hein, J., & Koschinsky, A. (2014). Deep-ocean ferromanganese crusts and nodules. *Geochemistry of Mineral Deposits: Treatise of Geochemistry, 2nd Edition*, *13*, 273–291.

Hein, J. R., Conrad, T., Mizell, K., Banakar, V. K., Frey, F. A., & Sager, W. W. (2016). Controls on ferromanganese crust composition and reconnaissance resource potential, ninetyeast ridge, indian ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *110*, 1–19. https://doi.org/https://doi.org/10.1016/j.dsr.2015.11.006

Hein, J. R., Koschinsky, A., Bau, M., Manheim, F. T., Kang, J.-K., & Roberts, L. (1999). Cobalt-rich ferromanganese crusts in the pacific. In D. Cronan (Ed.), *Handbook of marine mineral deposits* (pp. 239–280). CRC Press. https://pubs.usgs.gov/publication/70127614

Howarth, S. A. (2018). *An investigation into the variability of ferromanganese crusts in the ne atlantic* [Doctoral dissertation, National Oceanography Centre Southampton]. https://doi.org/10.3390/min8080327

Howarth, S. A. (2022). *An investigation into the variability of ferromanganese crusts in the ne atlantic* [Doctoral dissertation, University of Southampton]. http://eprints.soton.ac.uk/id/eprint/469116

Huang, S., & Fu, Y. (2023). Enrichment characteristics and mechanisms of critical metals in marine fe-mn crusts and nodules: A review. *Minerals*, *13*(12). https://doi.org/10.3390/min13121532

International Energy Agency. (2025). Global critical minerals outlook 2025 [Licence: CC BY 4.0]. https://www.iea.org/reports/global-critical-minerals-outlook-2025

Josso, P., Pelleter, E., Pourret, O., Fouquet, Y., Etoubleau, J., Cheron, S., & Bollinger, C. (2017). A new discrimination scheme for oceanic ferromanganese deposits using high field strength and rare earth elements [SI:Marine mineral deposits: New resources for base, precious, and critical metals]. *Ore Geology Reviews*, *87*, 3–15. https://doi.org/https://doi.org/10.1016/j.oregeorev.2016.09.003

Josso, P., Parkinson, I., Horstwood, M., Lusty, P., Chenery, S., & Murton, B. (2019). Improving confidence in ferromanganese crust age models: A composite geochemical approach. *Chemical Geology*, *513*. https://doi.org/10.1016/j.chemgeo.2019.03.003

Josso, P., Rushton, J., Lusty, P., Matthews, A., Chenery, S., Holwell, D., Kemp, S. J., & Murton, B. (2020). Late cretaceous and cenozoic paleoceanography from north-east atlantic ferromanganese crust microstratigraphy. *Marine Geology*, *422*, 106122. https://doi.org/https://doi.org/10.1016/j.margeo.2020.106122

Journel, A. G., & Huijbregts, C. J. (1976). *Mining geostatistics*. Academic Press.

Mangnier, L. (2024, February). When are the clr and alr transformations applicable? [Answer posted on Cross Validated (Stack Exchange)]. https://stats.stackexchange.com/questions/660944/when-are-the-clr-and-alr-transformations-applicable

Maptek Pty Ltd. (2024). Variogram models in block modelling [Accessed on [insert access date here]]. https://help.maptek.com/vulcan/2024.0.0/Content/topics/Block/Variography/Block_ModelTypes.htm

Marcotte, D. (2012). Revisiting the linear model of coregionalization. In P. Abrahamsen, R. Hauge, & O. Kolbjørnsen (Eds.), *Geostatistics oslo 2012* (pp. 67–78). Springer Netherlands. https://doi.org/10.1007/978-94-007-4153-9_6

Marino, E., González, F., Somoza, L., Lunar, R., Ortega, L., Vázquez, J., Reyes, J., & Bellido, E. (2017). Strategic and rare elements in cretaceous-cenozoic cobalt-rich ferromanganese crusts from seamounts in the canary island seamount province (northeastern tropical atlantic) [SI:Marine mineral deposits: New resources for base, precious, and critical metals]. *Ore Geology Reviews*, *87*, 41–61. https://doi.org/https://doi.org/10.1016/j.oregeorev.2016.10.005

Markvoort, H. J. B., & Deutsch, C. V. (2024). Kriging weights in the presence of redundant data (J. L. Deutsch, Ed.). http://www.geostatisticslessons.com/lessons/krigingweights

Martin, R. (2020, October). These kriging weights are really, really ridiculously good looking... [Accessed 20 July 2025]. https://lazymodellingcrew.com/post/post_22_weights-from-different-estimators/

Mert, M. C., Filzmoser, P., & Hron, K. (2016). Error propagation in isometric log-ratio coordinates for compositional data: Theoretical and practical considerations. *Mathematical Geosciences*, *48*(8), 941–961. https://doi.org/10.1007/s11004-016-9646-x

Mizell, K., Hein, J. R., Au, M., & Gartman, A. (2022). Estimates of metals contained in abyssal manganese nodules and ferromanganese crusts in the global ocean based on regional variations and genetic types of nodules. In R. Sharma (Ed.), *Perspectives on deep-sea mining: Sustainability, technology, environmental policy and management* (pp. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-030-87982-2_3

Monteiro da Rocha, M., Yamamoto, J., Watanabe, J., & Fonseca, P. (2012). Studying the influence of a secondary variable in collocated cokriging estimates. *Anais da Academia Brasileira de Ciências*, *84*, 335–46. https://doi.org/10.1590/S0001-37652012005000017

Mundim, E., Johann, P., & Remacre, A. (1999). Factorial kriging analysis: Geostatistical filtering applied to reservoir characterization. *The Leading Edge*, *18*, 787–788. https://doi.org/10.1190/1.1438381

Nath, B. N. (2008). Processes of formation of ferromanganese manganese nodules and crusts. *National Institute of Oceanography*. https://drs.nio.res.in/drs/bitstream/handle/2264/741/Refresher_Course_Mar_Geol_Geophys_2007_Lecture_Notes_63.pdf?sequence=2

Novakova, A., & Novikov, D. (2021, May). Study of deep-ocean ferromanganese crusts ore components. https://doi.org/10.5772/intechopen.98200

Nozaki, T., Tokumaru, A., Takaya, Y., Kato, Y., Suzuki, K., & Urabe, T. (2016). Major and trace element compositions and resource potential of ferromanganese crust at takuyo daigo seamount, northwestern pacific ocean. *GEOCHEMICAL JOURNAL*, *50*(6), 527–537. https://doi.org/10.2343/geochemj.2.0430

Osborne, J. W. (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, *8*(6).

Pang, Y., Wang, Y., Lai, X., Zhang, S., Liang, P., & Song, X. (2023). Enhanced kriging leave-one-out cross-validation in improving model estimation and optimization. *Computer Methods in Applied Mechanics and Engineering*, *414*, 116194. https://doi.org/https://doi.org/10.1016/j.cma.2023.116194

Pardo-Iguzquiza, E., & Chica-Olmo, M. (2008). Geostatistics with the matern semivariogram model: A library of computer programs for inference, kriging and simulation. *Computers & Geosciences*, *34*(9), 1073–1079. https://doi.org/https://doi.org/10.1016/j.cageo.2007.09.020

Pawlowsky-Glahn, V., & Buccianti, A. (2011). *Compositional data analysis: Theory and applications*. Wiley. https://doi.org/10.1002/9781119976462

Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. Wiley. https://doi.org/10.1002/9781119003144

Pronzato, L., & Rendas, M.-J. (2024). Weighted leave-one-out cross validation. *SIAM/ASA Journal on Uncertainty Quantification*, *12*(4), 1213–1239. https://doi.org/10.1137/23M1615917

Samson, M., & Deutch, C. (2024). Collocated cokriging. In J. L. Deutsch (Ed.), *Geostatistics lessons*. Geostatistics Lessons. https://geostatisticslessons.com/lessons/collocatedcokriging

Samson, M., & Deutch, C. (2025). *Geostatisticslessonsnotebooks* [Chapter: Collocated Cokriging by Samson, M., & Deutsch, C., 2020]. GitHub. https://github.com/GeostatisticsLessons/GeostatisticsLessonsNotebooks

scikit-bio development team. (2024). Skbio.stats.composition.ilr [Accessed: 2024-07-12]. https://scikit.bio/docs/dev/generated/skbio.stats.composition.ilr.html

Sharma, R. (Ed.). (2022). *Perspectives on deep-sea mining: Sustainability, technology, environmental policy and management*. Springer Cham. https://doi.org/10.1007/978-3-030-87982-2

Statistics How To. (2023a). *Mean absolute percentage error (mape)* [Retrieved June 11, 2025]. https://www.statisticshowto.com/mean-absolute-percentage-error-mape/

Statistics How To. (2023b). *Nrmse: Definition, examples* [Retrieved June 11, 2025]. https://www.statisticshowto.com/nrmse/

Statistics How To. (2024). Winsorize: Definition, examples, when to use [Accessed: 2025-05-23]. https://www.statisticshowto.com/winsorize/

Stumm, W., & Morgan, J. (2012). *Aquatic chemistry: Chemical equilibria and rates in natural waters*. Wiley. https://books.google.nl/books?id=NLV_yfulgkQC

Texas_A&M_University. (2017). Chapter 07: Simulation modeling [Accessed July 2025].

Thio-Henestrosa, S., & Comas, M. (2016). *Codapack v2 user's guide* [April 20, 2016]. Universitat de Girona. Girona, Spain. https://ima.udg.edu/codapack/assets/codapack-manual.pdf

Usui, A., & Suzuki, K. (2022). Geological characterization of ferromanganese crust deposits in the nw pacific seamounts for prudent deep-sea mining. In R. Sharma (Ed.), *Perspectives on deep-sea mining*. Springer, Cham. https://doi.org/10.1007/978-3-030-87982-2_4

Usui, A., Nishi, K., Sato, H., Nakasato, Y., Thornton, B., Kashiwabara, T., Tokumaru, A., Sakaguchi, A., Yamaoka, K., Kato, S., Nitahara, S., Suzuki, K., Iijima, K., & Urabe, T. (2016). Continuous growth of hydrogenetic ferromanganese crusts since 17 myr ago on takuyo-daigo seamount, nw pacific, at water depths of 800–5500 m. *Ore Geology Reviews*, *80*, 47–61. https://doi.org/10.1016/j.oregeorev.2016.09.032

Vysetti, B. (2023). Deep-sea mineral deposits as a future source of critical metals, and environmental issues – a brief review. *Minerals and Materials*, *2*(5). https://doi.org/10.20517/mmm.2022.12

Wang, G., Jansa, L., Chu, F., Zou, C., & Sun, G. (2015). Composition and origin of ferromanganese crusts from equatorial western pacific seamounts. *Journal of Ocean University of China*, *14*(2), 210–219. https://doi.org/10.1007/s11802-015-2391-9

Wang, X.-h., Schloßmacher, U., Natalio, F., Schröder, H. C., Wolf, S. E., Tremel, W., & Müller, W. E. (2009). Evidence for biogenic processes during formation of ferromanganese crusts from the pacific ocean: Implications of biologically induced mineralization. *Micron*, *40*(5), 526–535. https://doi.org/https://doi.org/10.1016/j.micron.2009.04.005

Weku, W., Pramoedyo, H., Widodo, A., & Fitriani, R. (2019). Non-monoton nonparametric variogram to model the land price of manado city with hole effect periodicity structure. *IOP Conference Series: Materials Science and Engineering*, *546*, 052083. https://doi.org/10.1088/1757-899X/546/5/052083

Yalçin, E. (2005). Cokriging and its effect on the estimation precision. *SAIMM Journal*. https://www.saimm.co.za/Journal/v105n04p223.pdf

Yeo, I. A., Dobson, K., Josso, P., Pearce, R. B., Howarth, S. A., Lusty, P. A. J., Le Bas, T. P., & Murton, B. J. (2018). Assessment of the mineral resource potential of atlantic ferromanganese crusts based on their growth history, microstructure, and texture. *Minerals*, *8*(8). https://doi.org/10.3390/min8080327

Zhou, J., Yang, S., Cao, J., Deng, Y., Wei, Z., Li, Y., Tian, D., & Hu, G. (2023). Influence of phosphatization in rey geochemistry in ferromanganese crusts in line islands, central pacific. *Minerals*, *13*(5). https://doi.org/10.3390/min13050647

A

Exploratory Data Analysis

A.1. EDA of Raw Variables

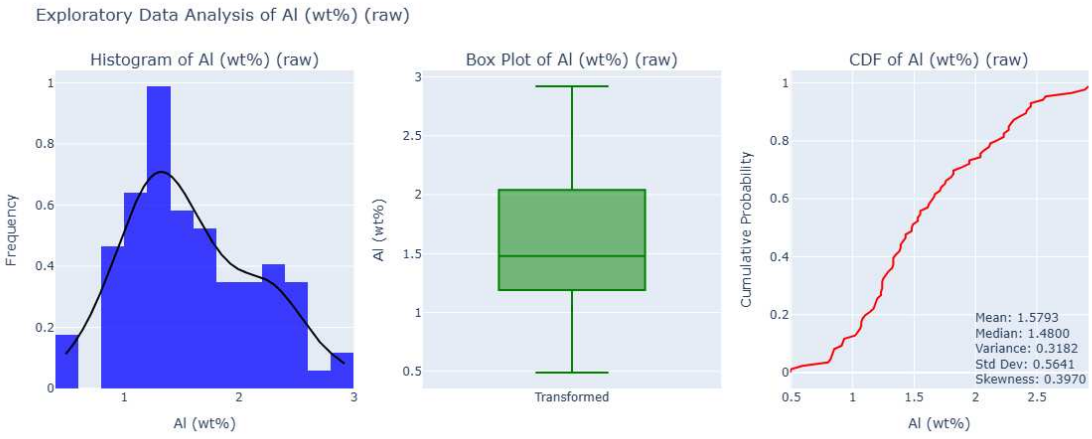


Figure A.1: EDA plot for Al

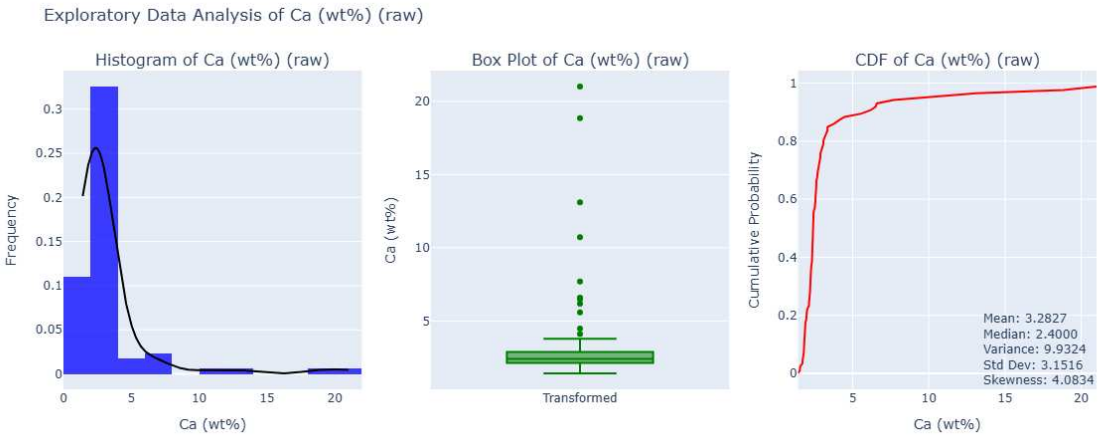


Figure A.2: EDA plot for Ca

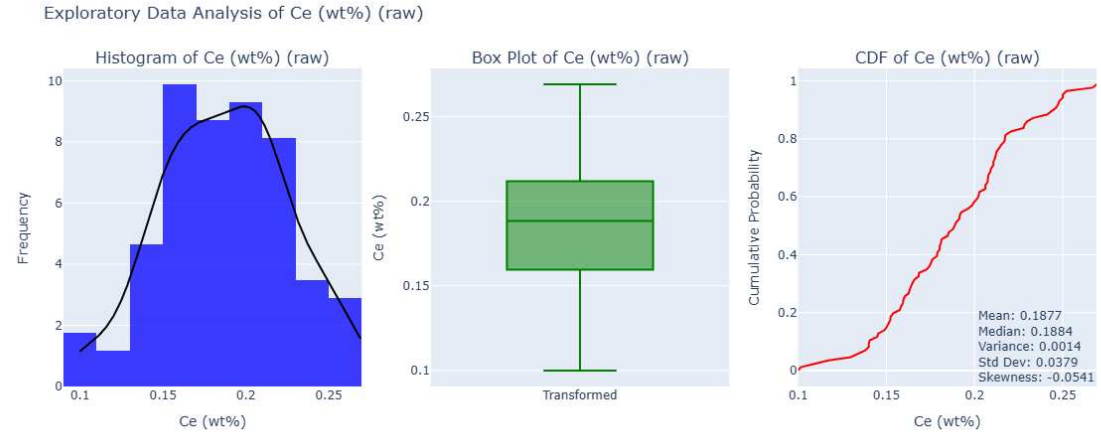


Figure A.3: EDA plot for Ce

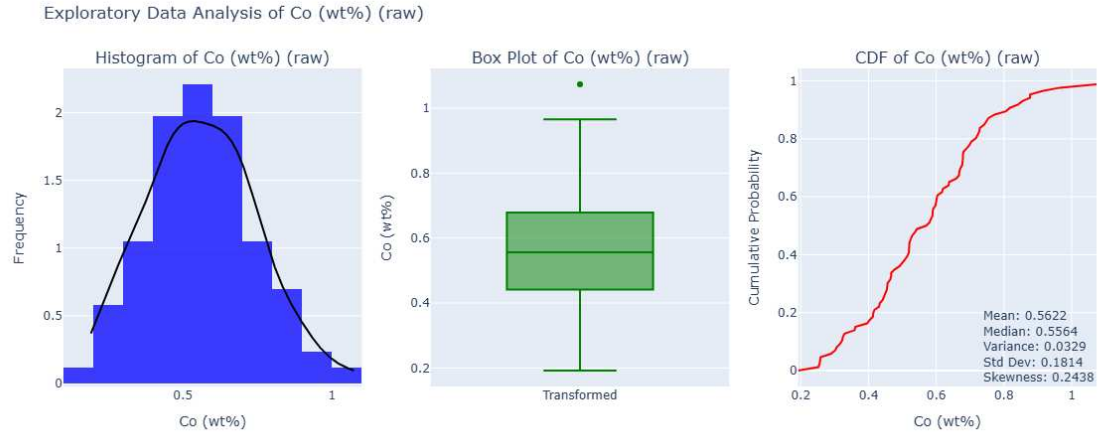


Figure A.4: EDA plot for Co

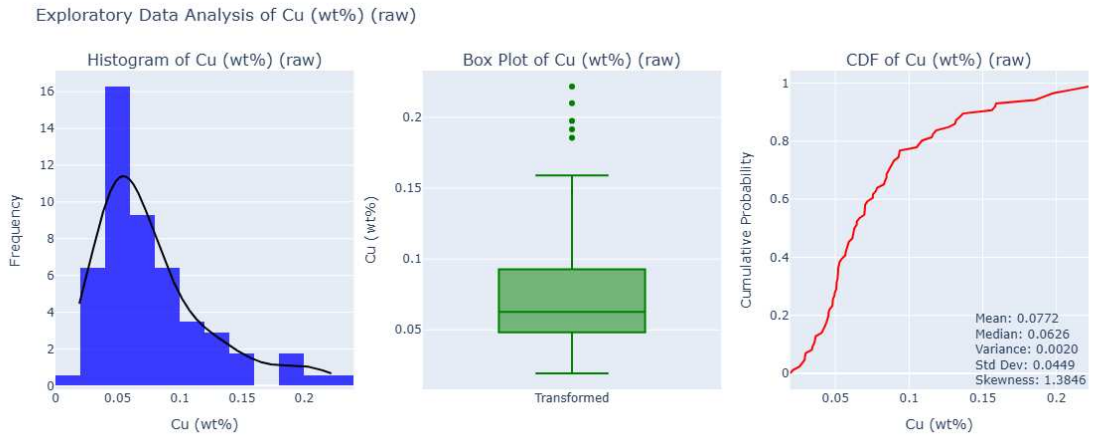


Figure A.5: EDA plot for Cu

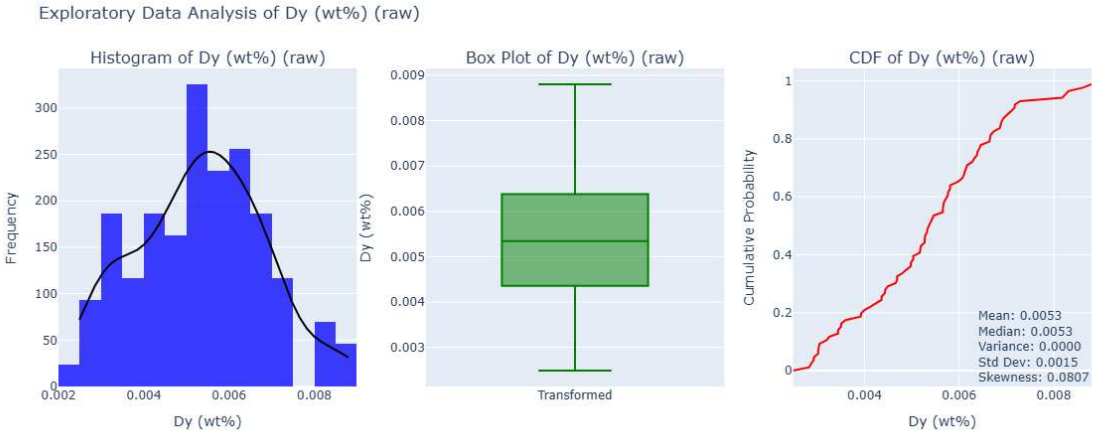


Figure A.6: EDA plot for Dy

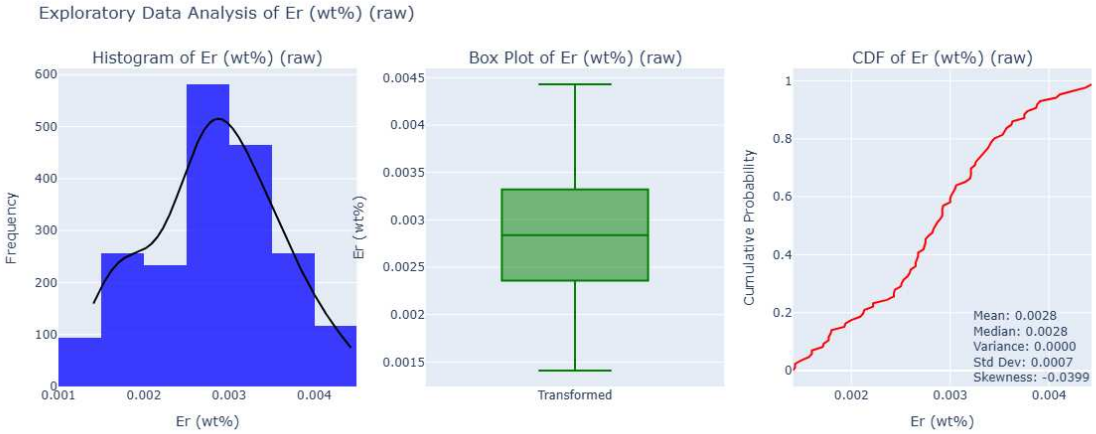


Figure A.7: EDA plot for Er

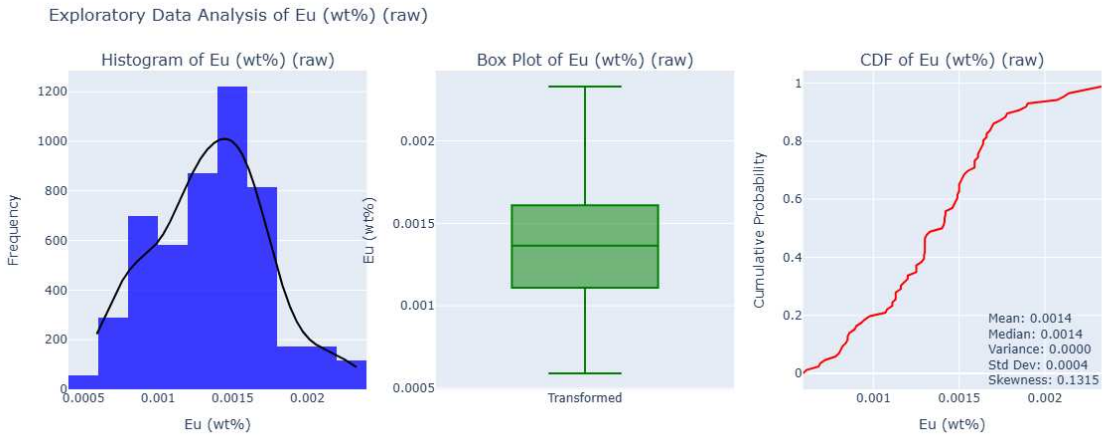


Figure A.8: EDA plot for Eu

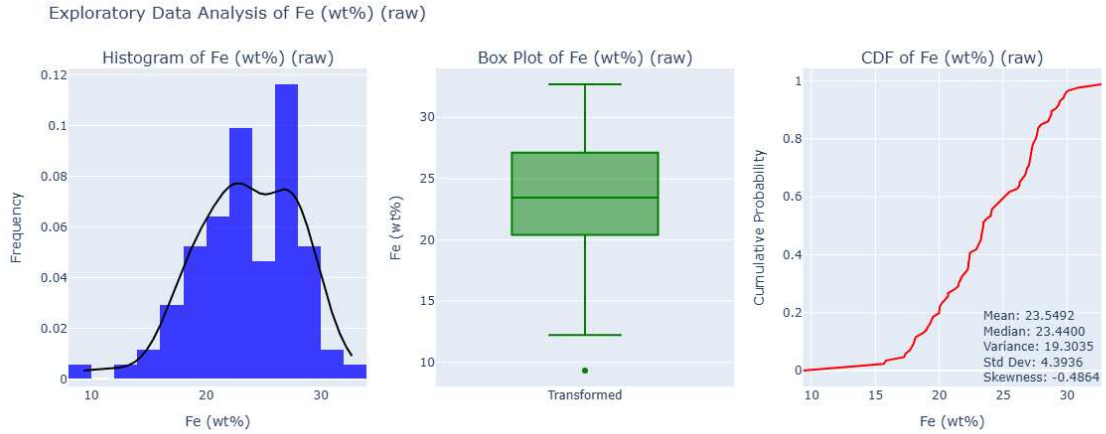


Figure A.9: EDA plot for Fe

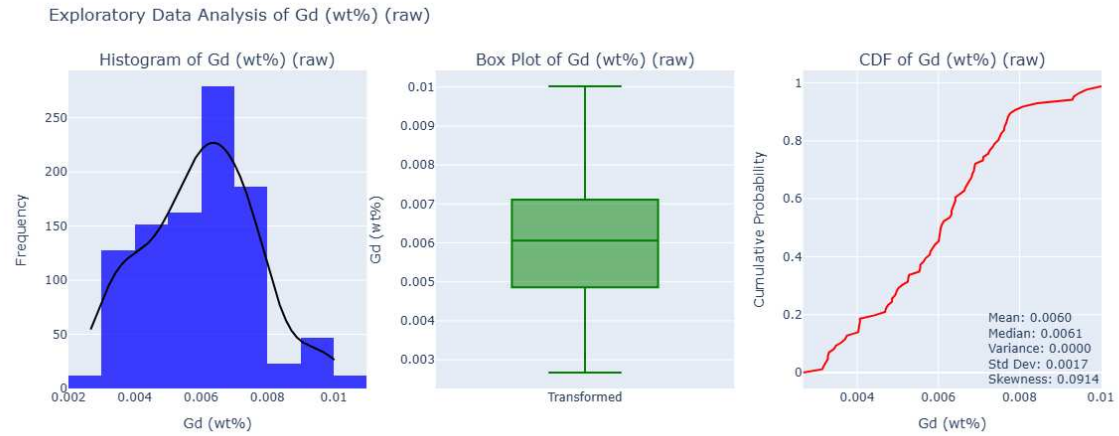


Figure A.10: EDA plot for Gd

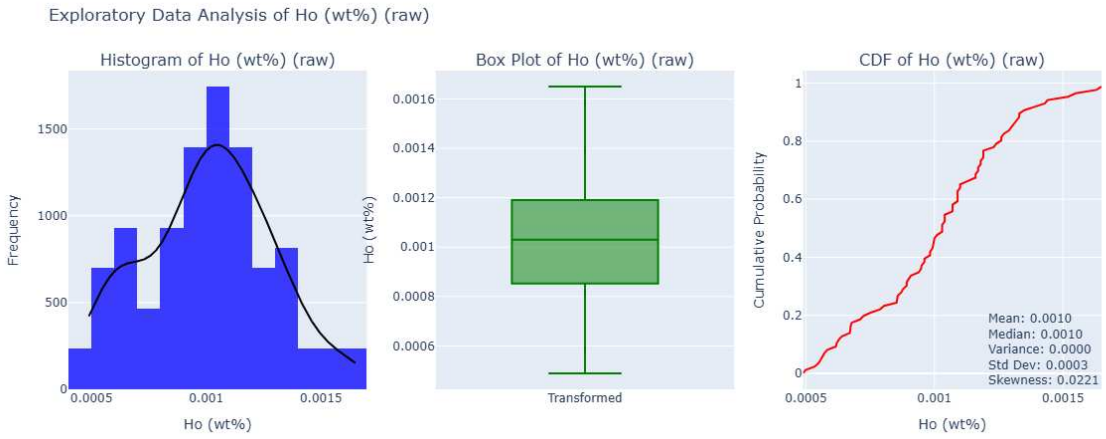


Figure A.11: EDA plot for Ho

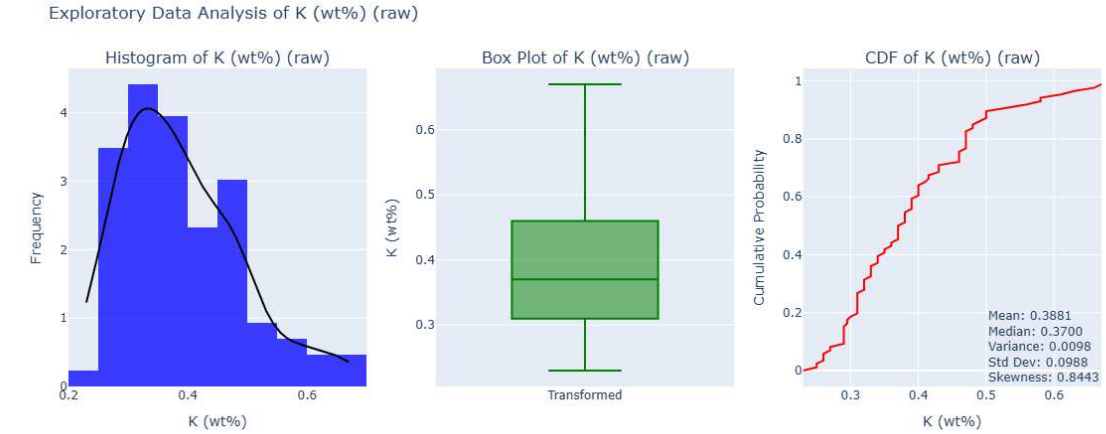


Figure A.12: EDA plot for K

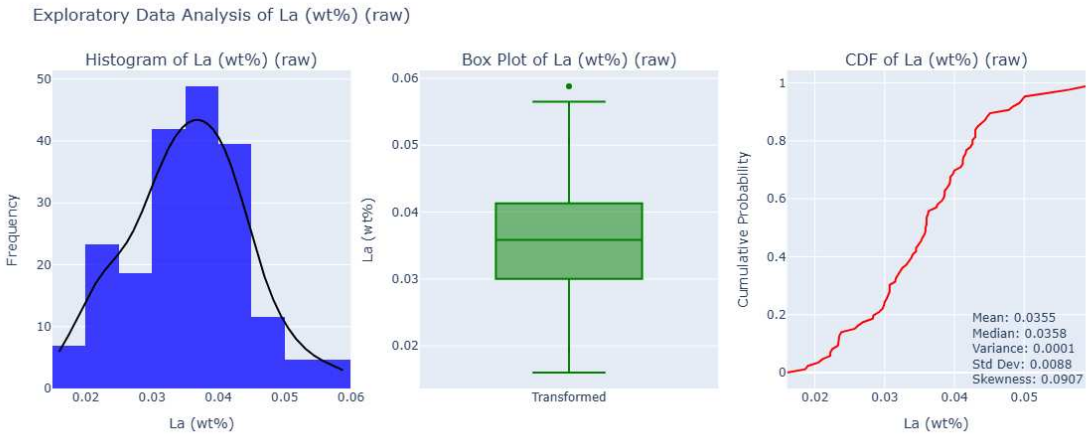


Figure A.13: EDA plot for La

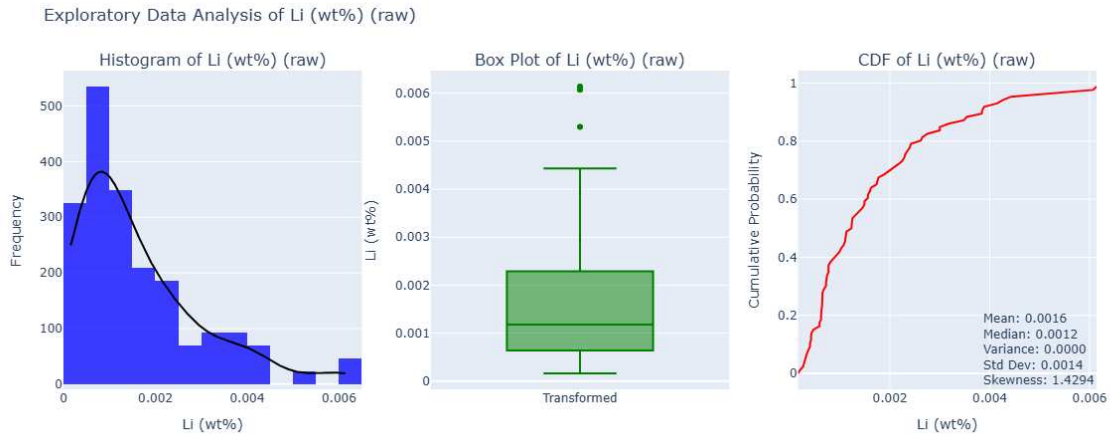


Figure A.14: EDA plot for Li

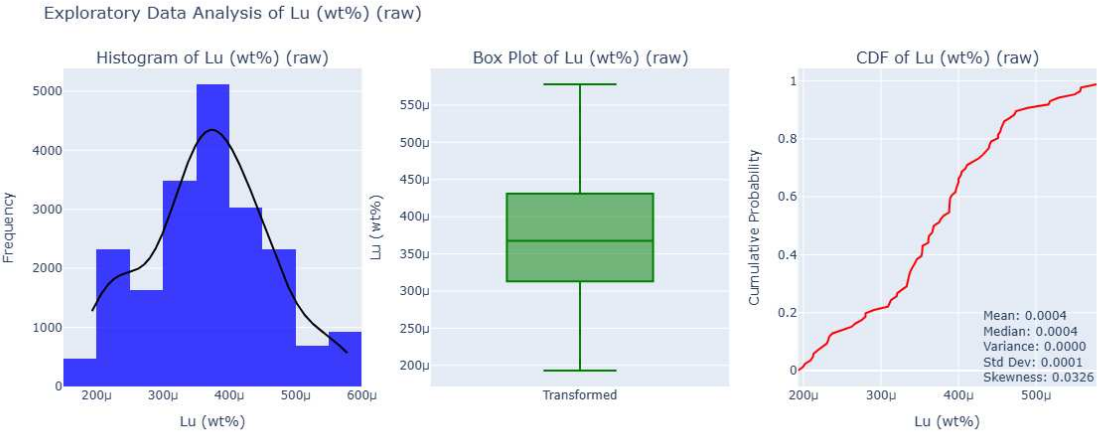


Figure A.15: EDA plot for Lu

Figure A.16: EDA plot for Mg

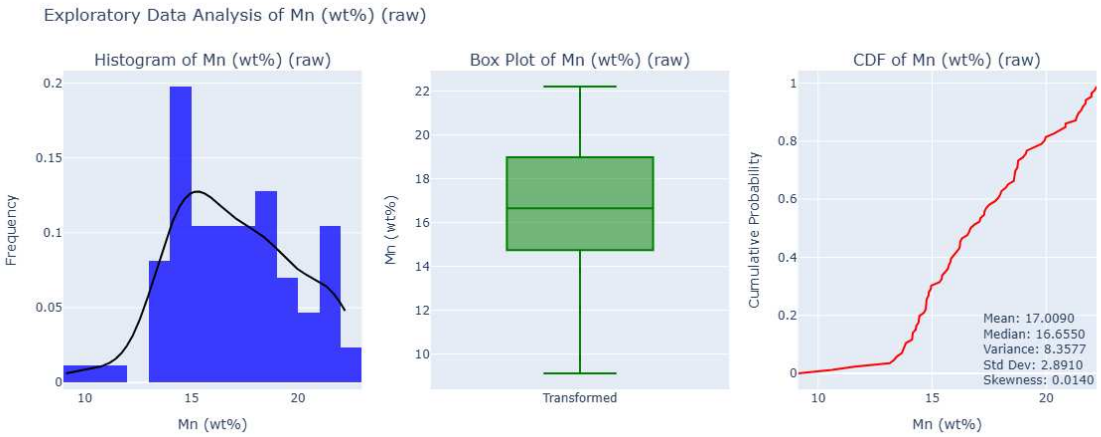


Figure A.17: EDA plot for Mn

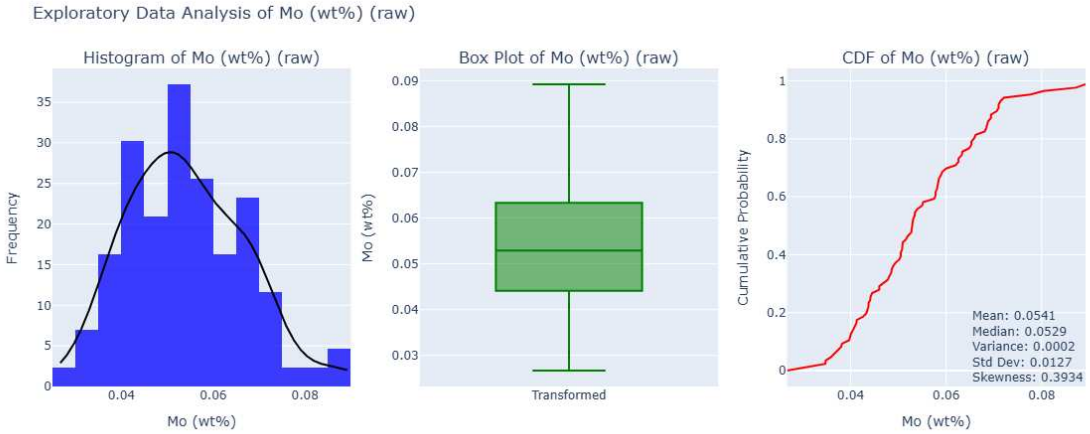


Figure A.18: EDA plot for Mo

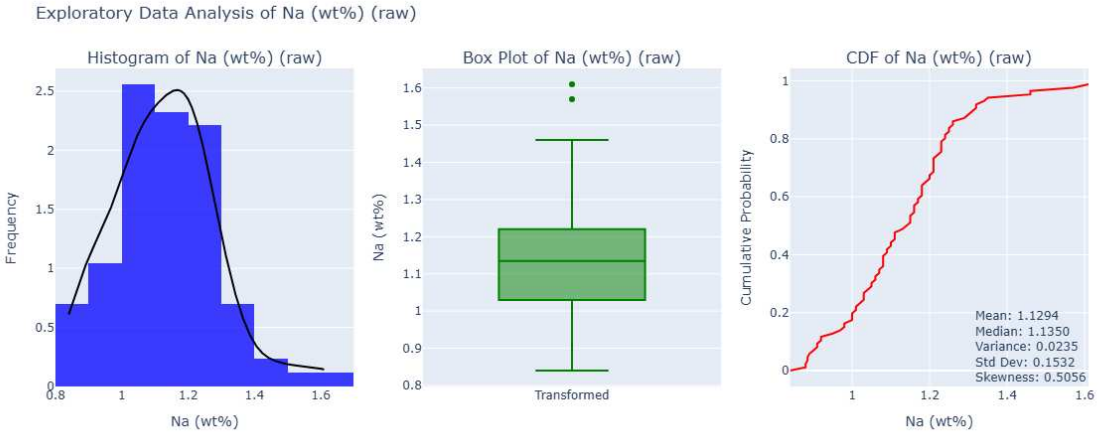


Figure A.19: EDA plot for Na

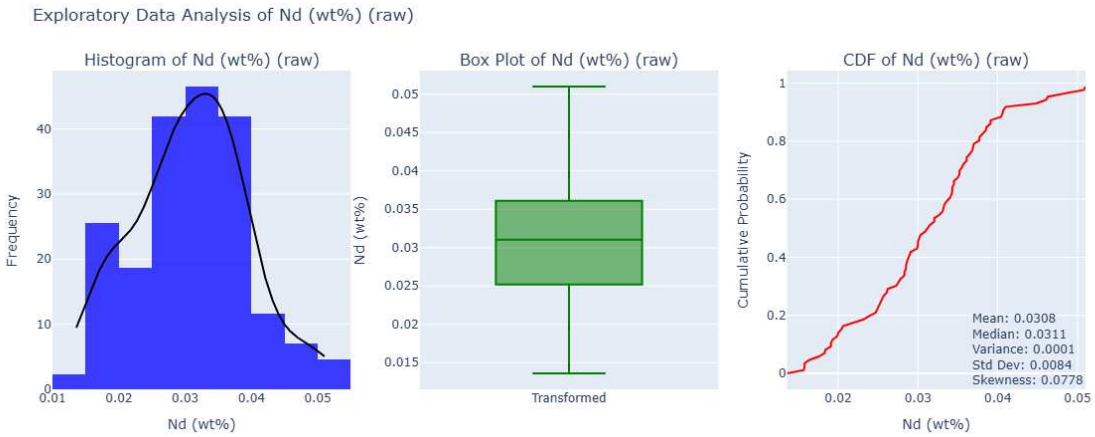


Figure A.20: EDA plot for Nd

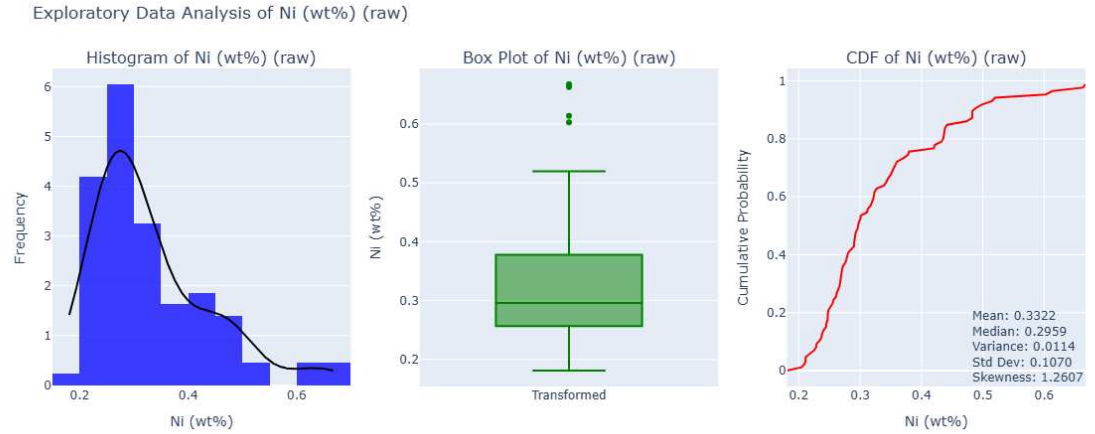


Figure A.21: EDA plot for Ni

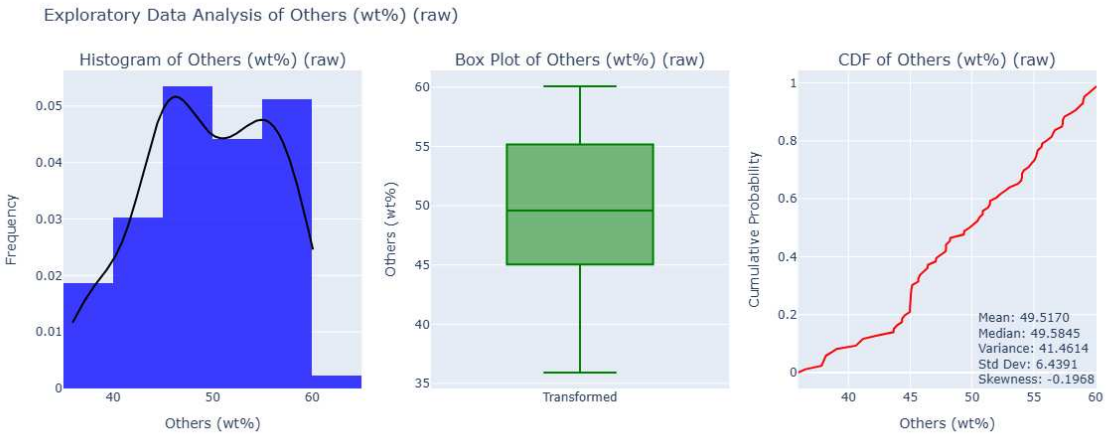


Figure A.22: EDA plot for Others

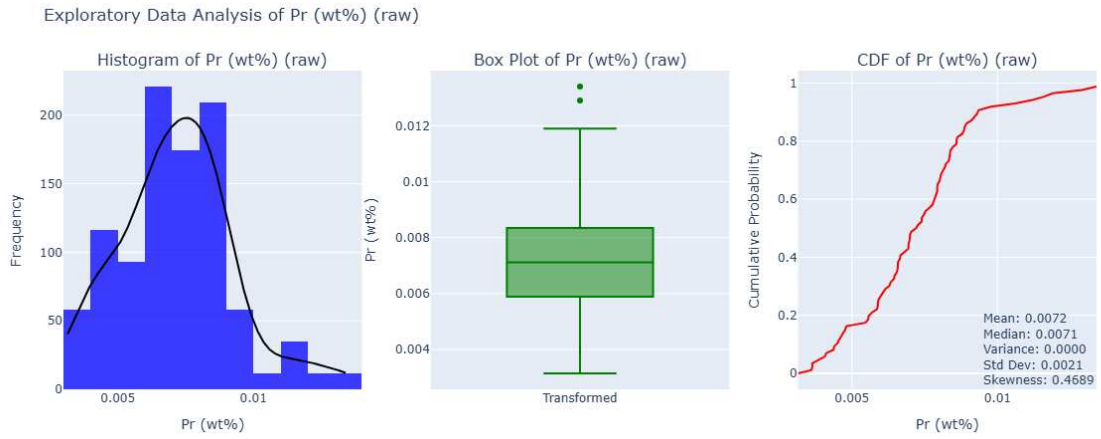


Figure A.23: EDA plot for Pr

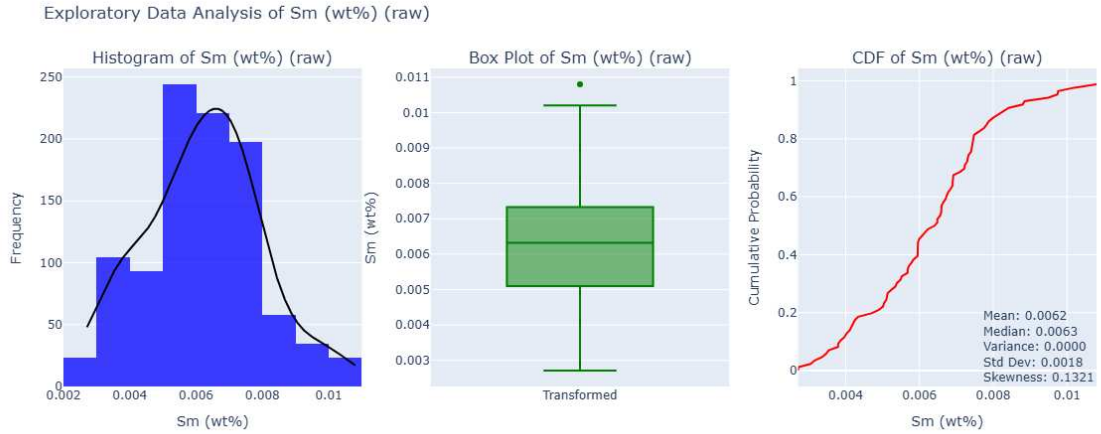


Figure A.24: EDA plot for Sm

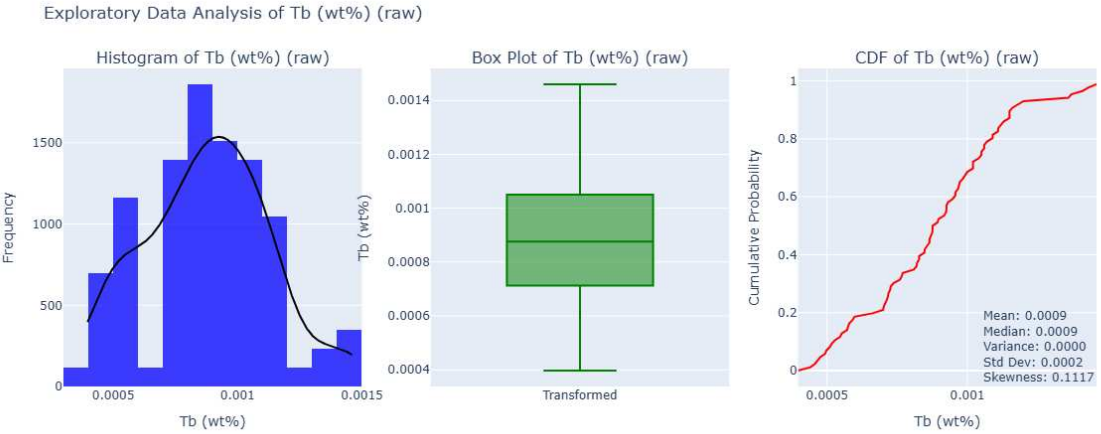


Figure A.25: EDA plot for Tb

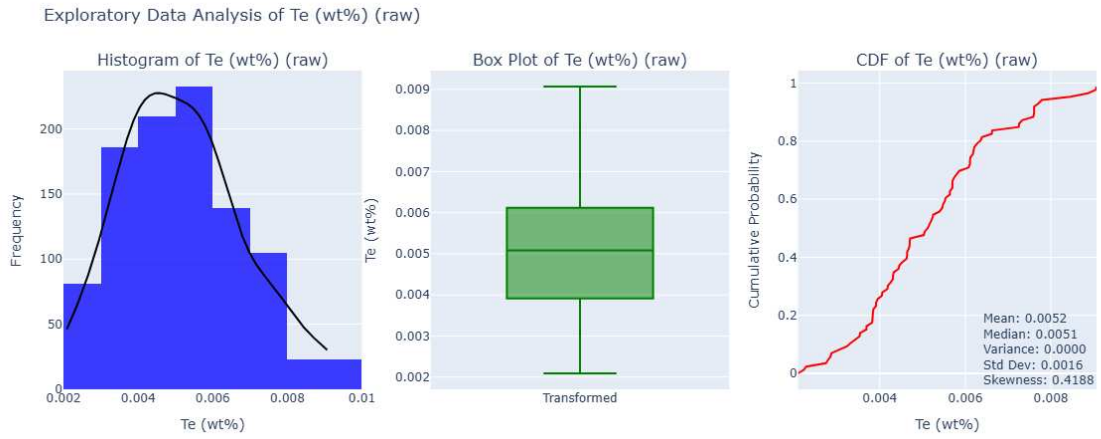


Figure A.26: EDA plot for Te

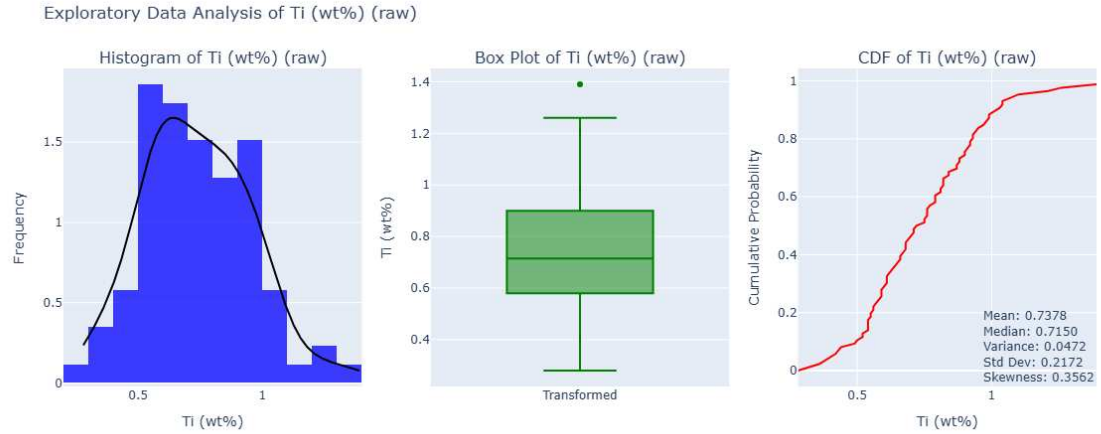


Figure A.27: EDA plot for Ti

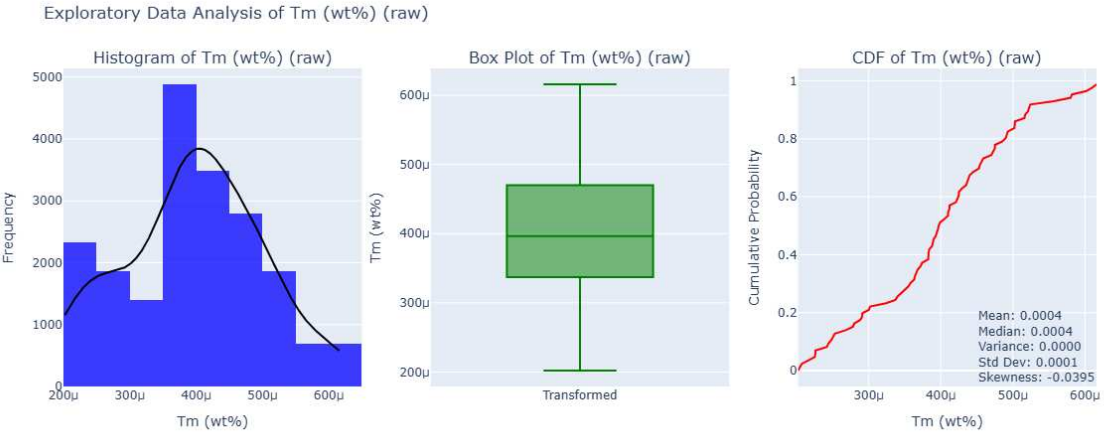


Figure A.28: EDA plot for Tm

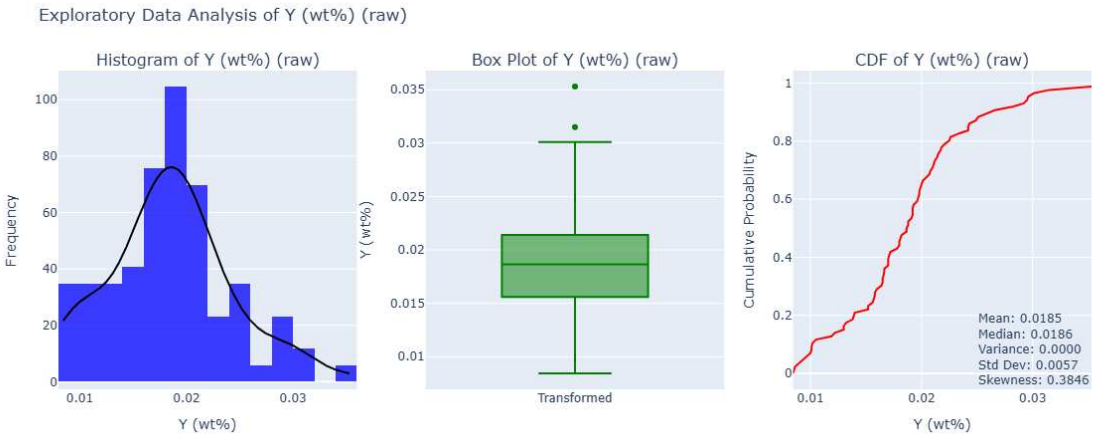


Figure A.29: EDA plot for Y

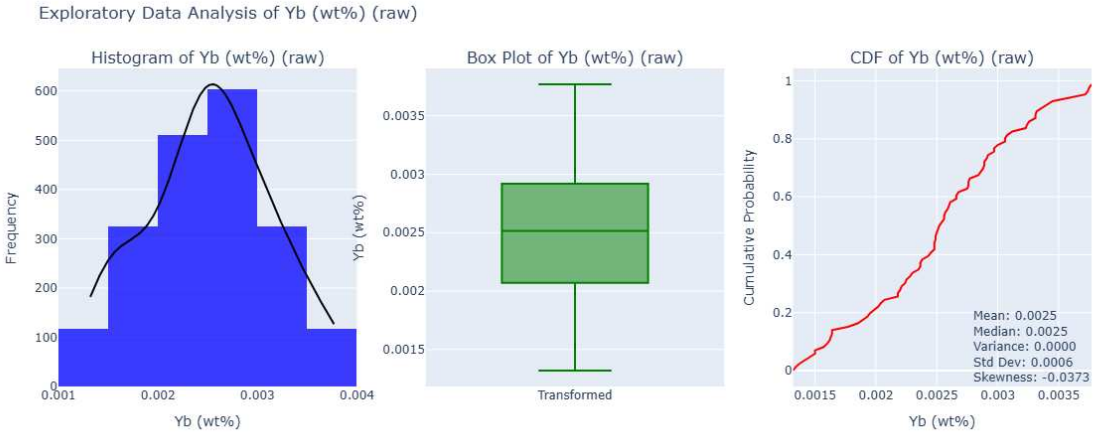


Figure A.30: EDA plot for Yb

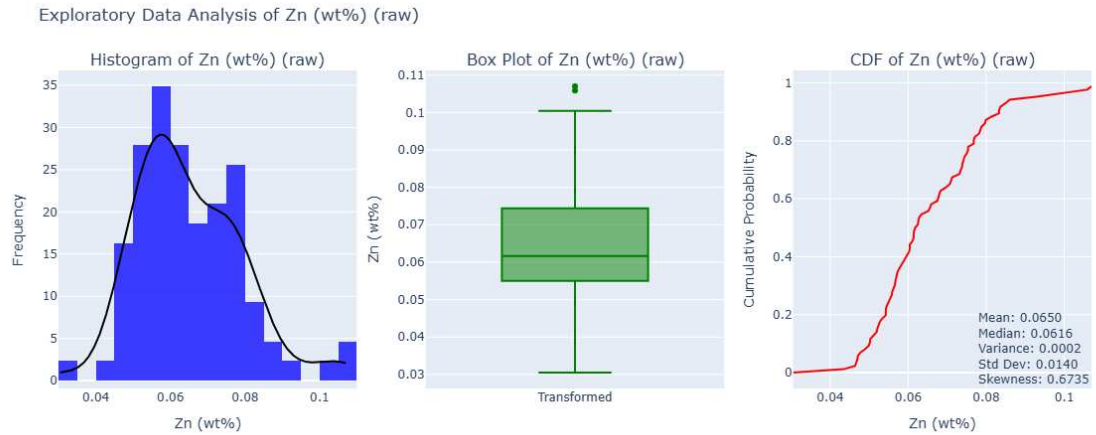


Figure A.31: EDA plot for Zn

A.2. EDA of ILR Variables

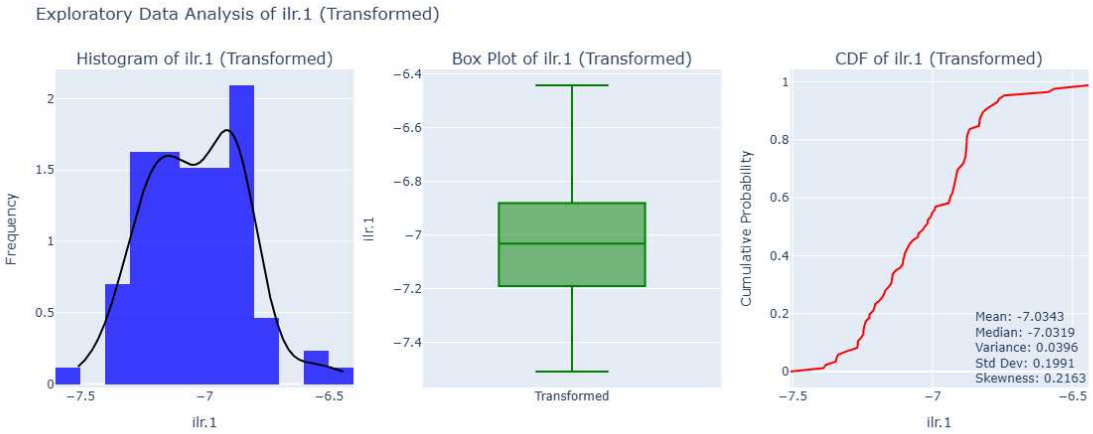


Figure A.32: EDA plot for ilr.1

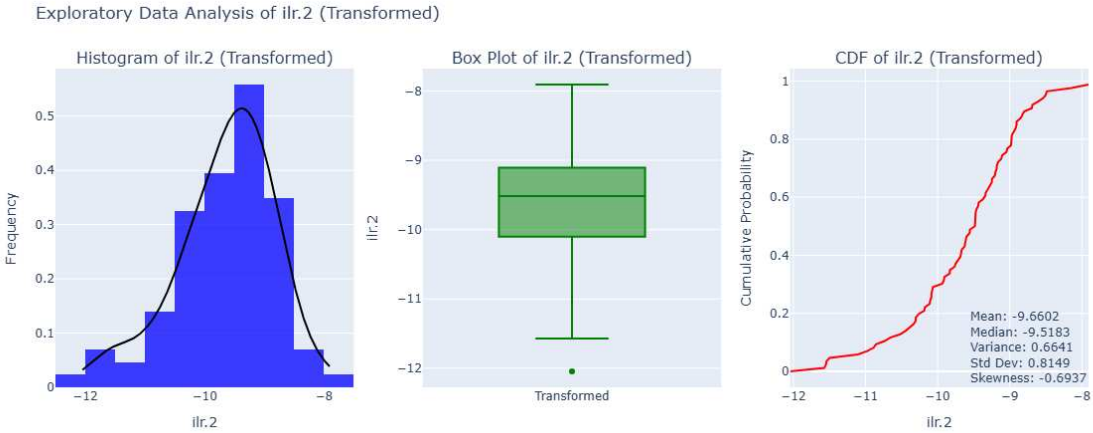


Figure A.33: EDA plot for ilr.2

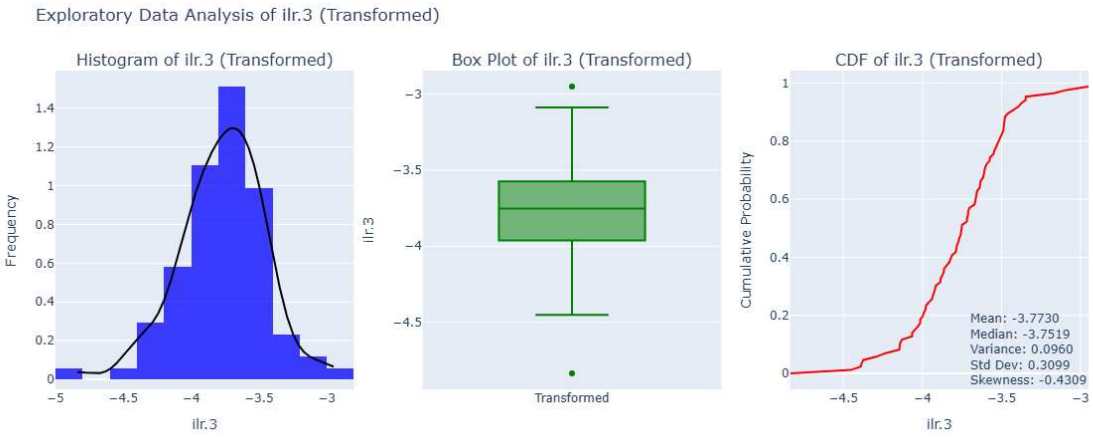


Figure A.34: EDA plot for ilr.3

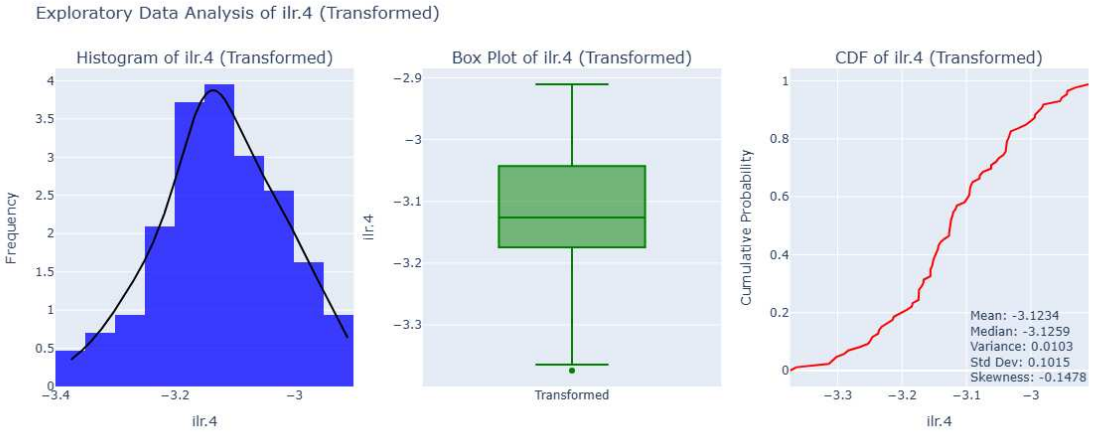


Figure A.35: EDA plot for ilr.4

Figure A.36: EDA plot for ilr.5

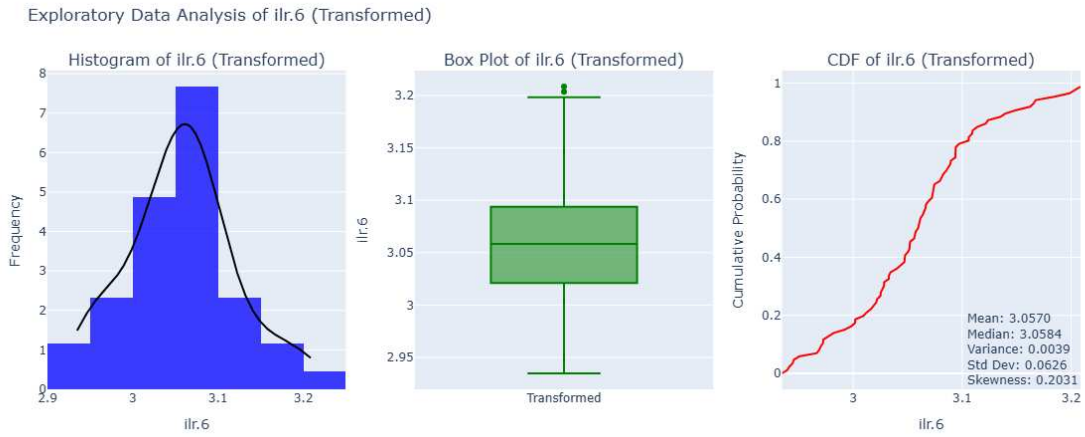


Figure A.37: EDA plot for ilr.6

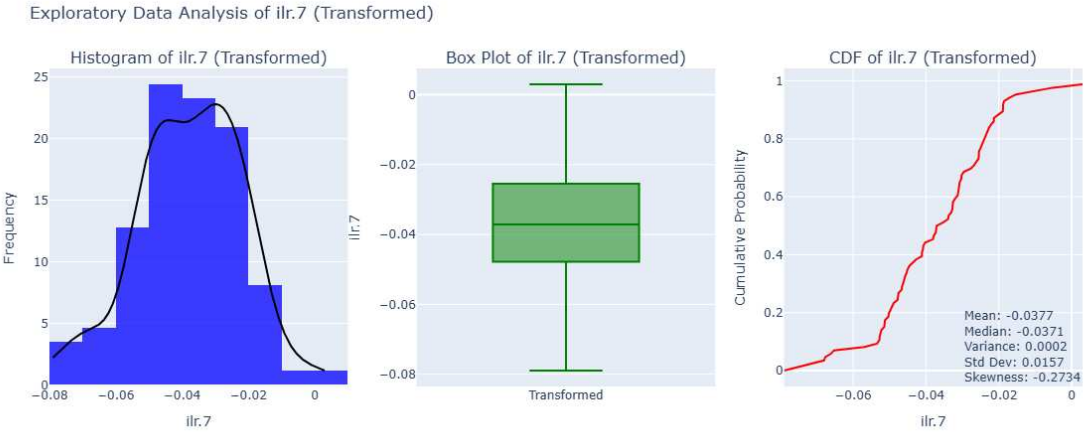


Figure A.38: EDA plot for ilr.7

Figure A.39: EDA plot for ilr.8

Figure A.40: EDA plot for ilr.9

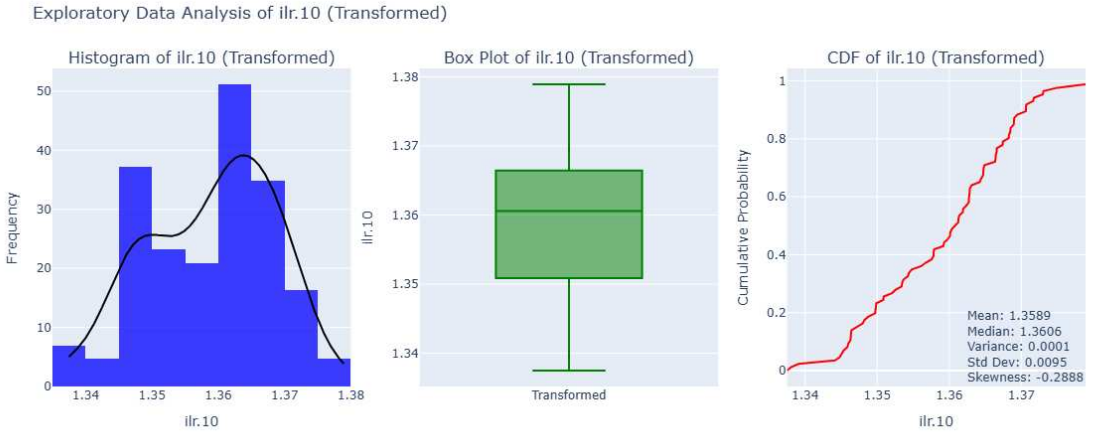


Figure A.41: EDA plot for ilr.10

Figure A.42: EDA plot for ilr.11

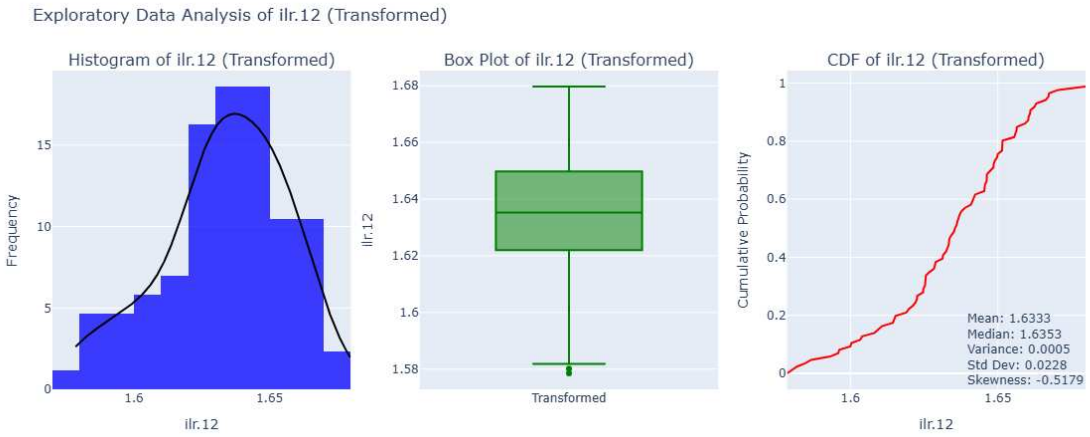


Figure A.43: EDA plot for ilr.12

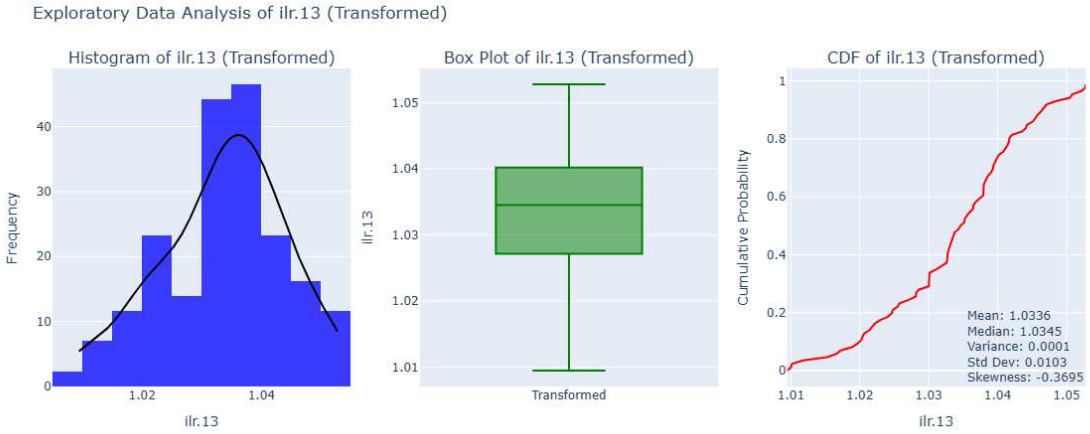


Figure A.44: EDA plot for ilr.13

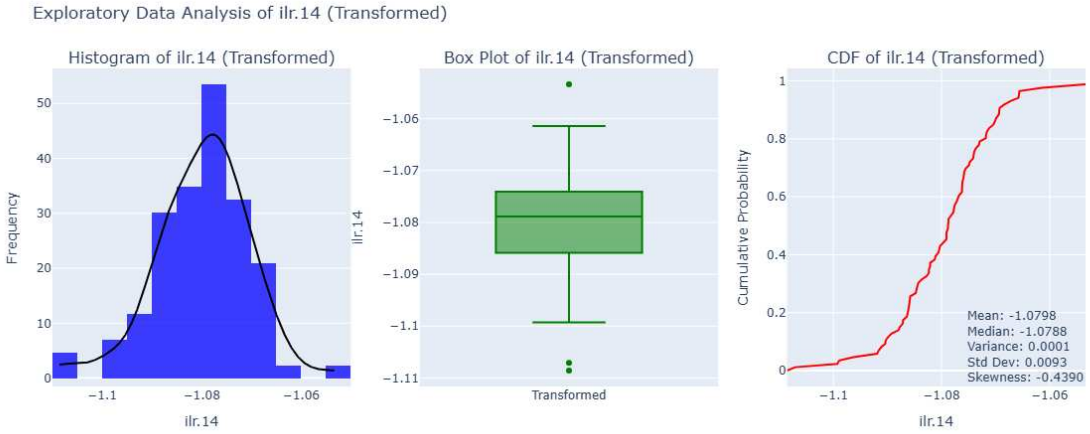


Figure A.45: EDA plot for ilr.14

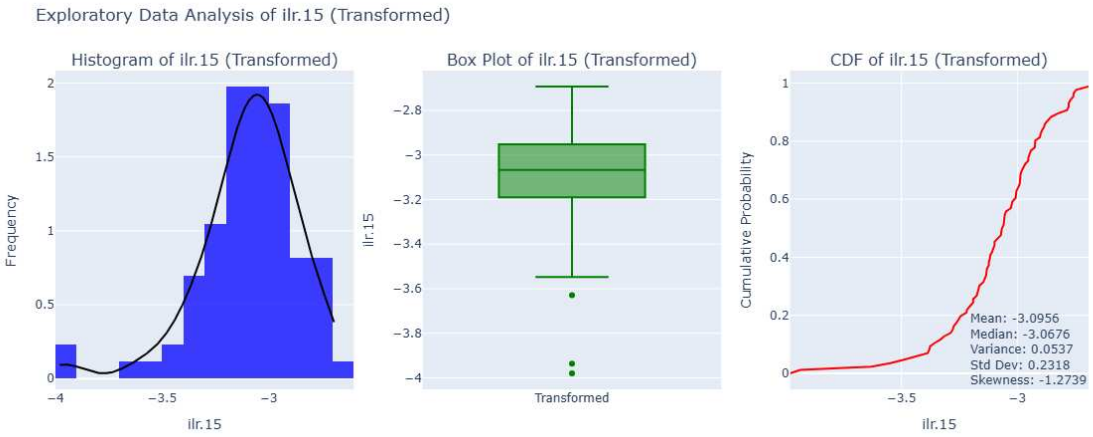


Figure A.46: EDA plot for ilr.15

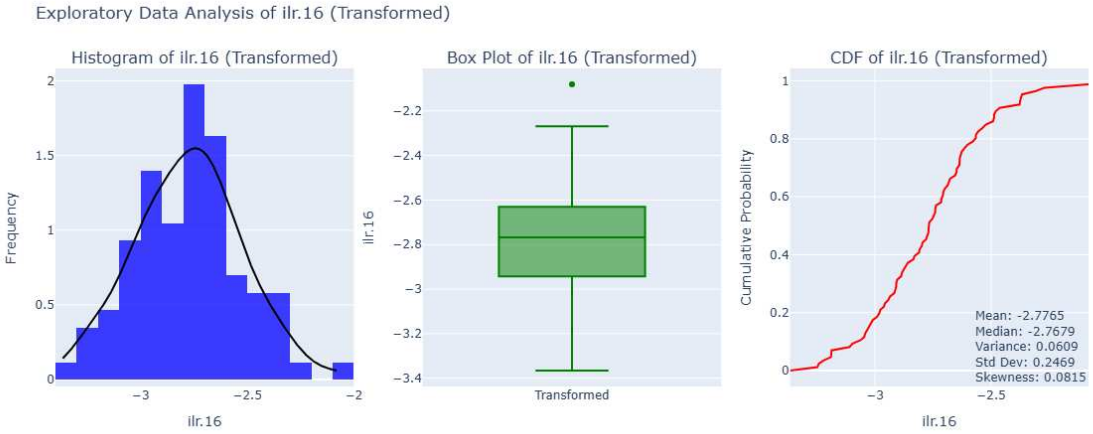


Figure A.47: EDA plot for ilr.16

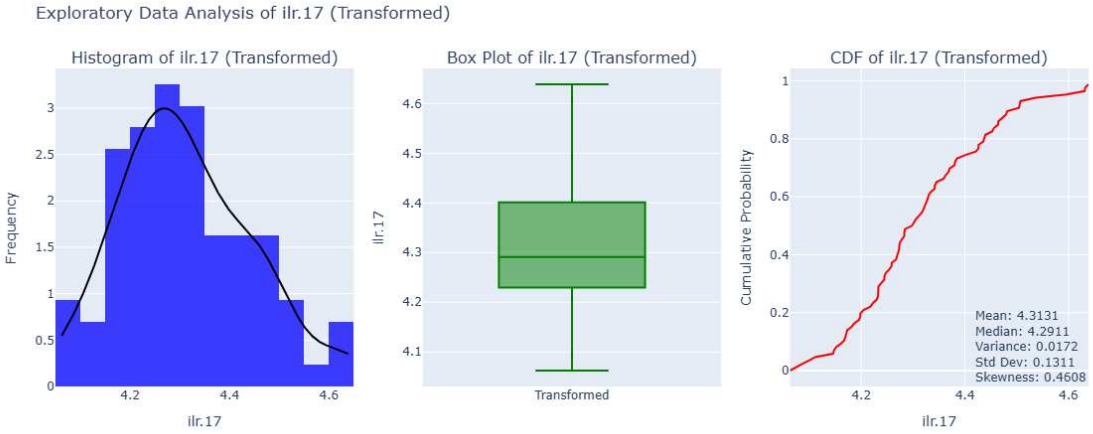


Figure A.48: EDA plot for ilr.17

Figure A.49: EDA plot for ilr.18

Figure A.50: EDA plot for ilr.19

Figure A.51: EDA plot for ilr.20

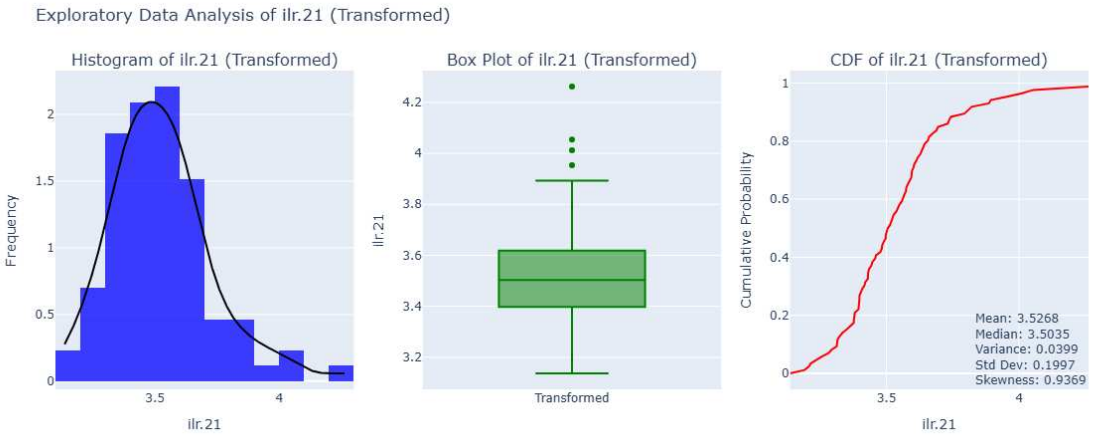


Figure A.52: EDA plot for ilr.21

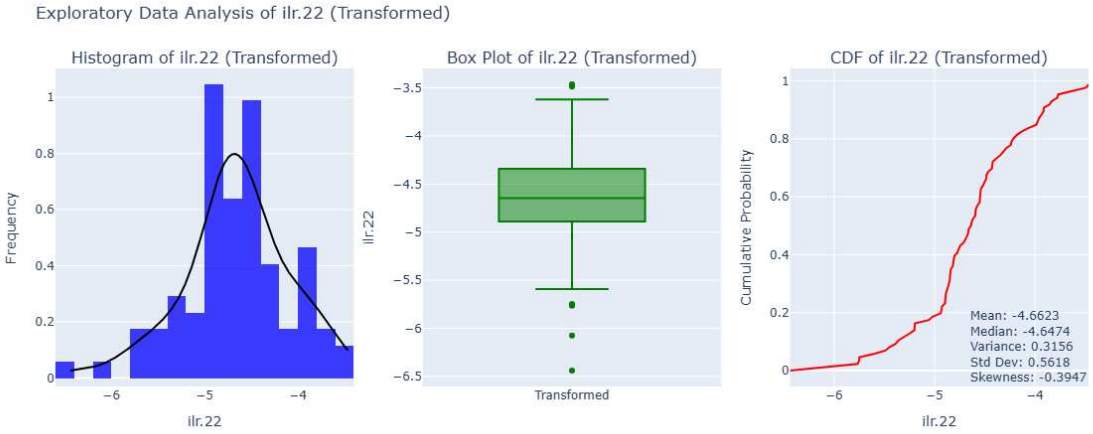


Figure A.53: EDA plot for ilr.22

Figure A.54: EDA plot for ilr.23

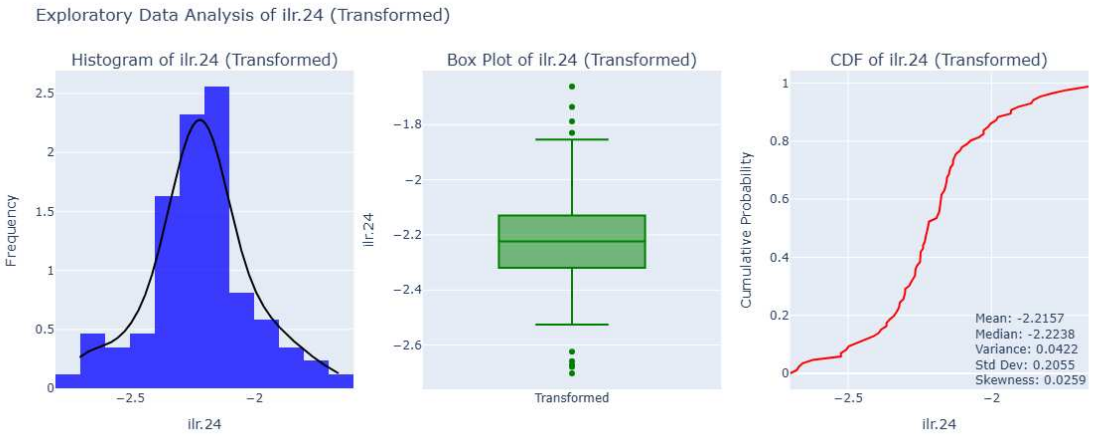


Figure A.55: EDA plot for ilr.24

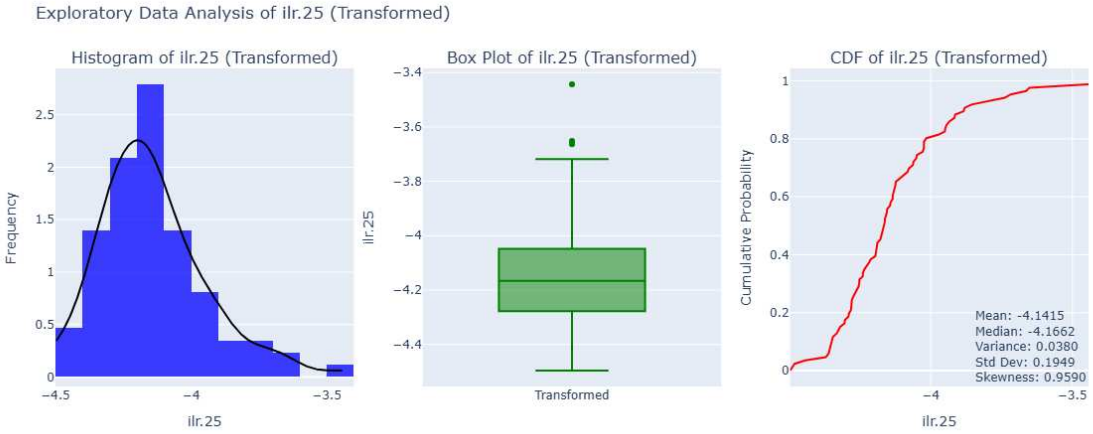


Figure A.56: EDA plot for ilr.25

Figure A.57: EDA plot for ilr.26

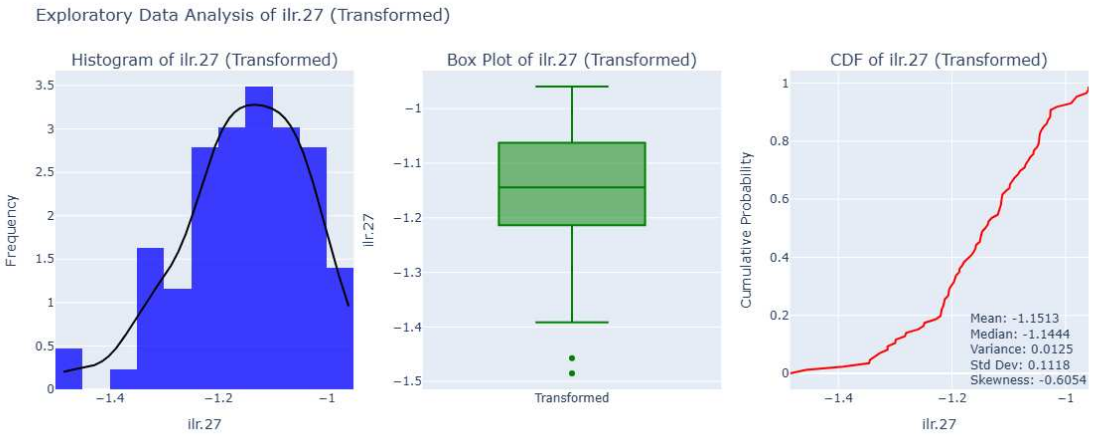


Figure A.58: EDA plot for ilr.27

Figure A.59: EDA plot for ilr.28

Figure A.60: EDA plot for ilr.29

Figure A.61: EDA plot for ilr.30

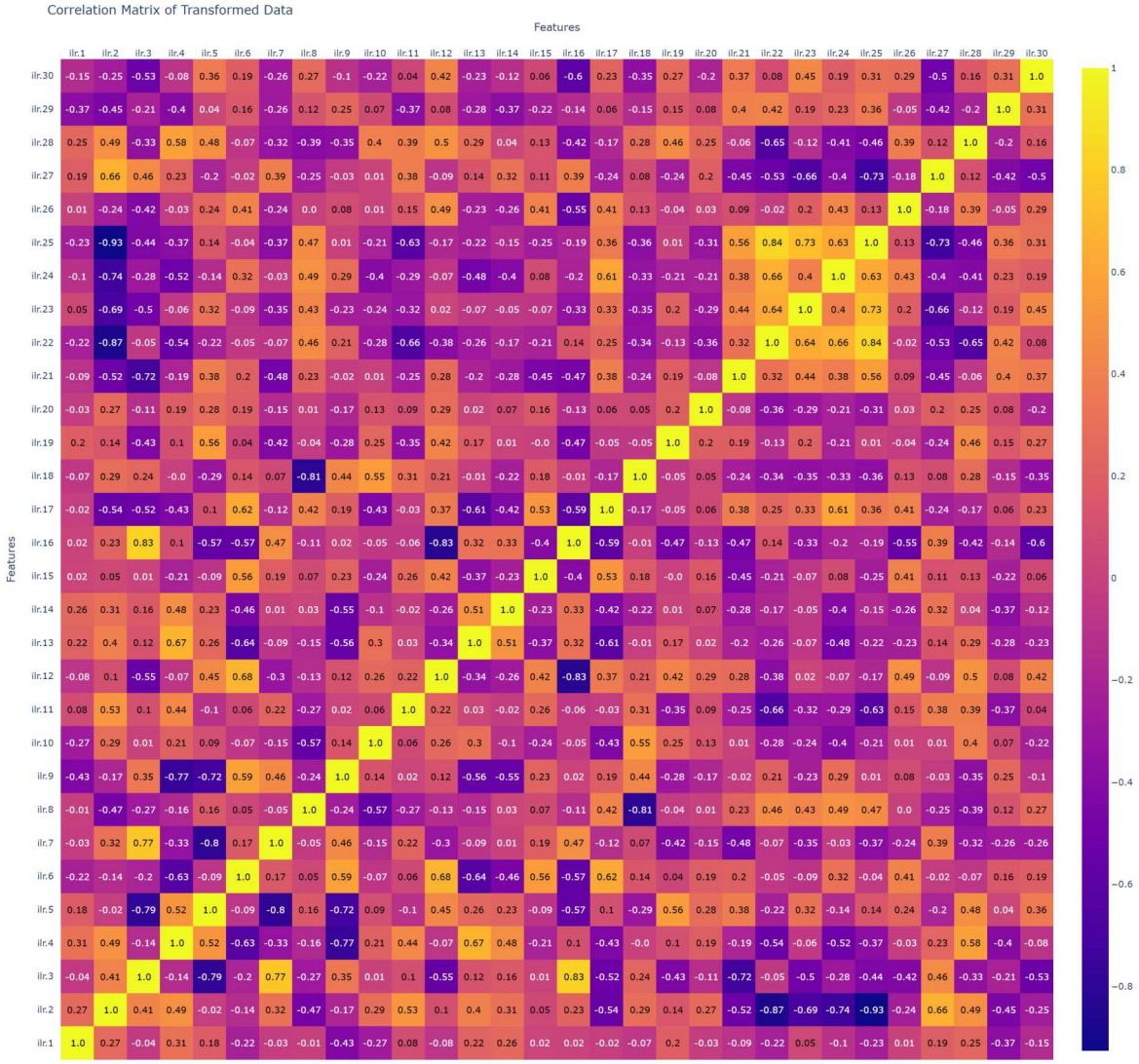


Figure A.62: Correlation matrix of ILR variables

Table A.1: Numerator and denominator elements for each ILR coordinate, as defined by the sequential binary partition matrix.

ILR coordinate	Numerator elements	Denominator elements
1	La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Y, Li, Co, Ni, Cu, Zn, Mo, Te, Na, Mg, Al, K, Ca, Ti, Mn, Fe, Others	Others
2	La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Y, Co, Mo, Te, Ti	Li, Ni, Cu, Zn, Na, Mg, Al, K, Mn, Fe
3	La, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Gd	Ce, Yb, Lu, Y, Co, Mo, Te, Ti
4	La, Gd, Tb, Dy, Ho, Er, Tm	Pr, Nd, Sm, Eu
5	La	Gd, Tb, Dy
6	La	Ho, Er, Tm
7	Ho, Er, Tm	Ho
8	Er	Tm
9	Gd, Tb	Dy
10	Gd	Tb
11	Pr, Nd, Sm, Eu	Ti
12	Pr, Nd	Sm, Eu
13	Nd	Pr
14	Sm	Eu
15	Ce, Yb, Lu, Co, Mo, Te	Y, Ti
16	Ce, Yb, Lu	Co, Mo, Te
17	Yb	Lu
18	Gd	Tb
19	Co	Mo
20	Mo	Te
21	Y	Ca
22	Li, Ni, Cu, Zn, Na, Mg, Al, K, Mn, Fe	Na, Mg, Al, K, Mn, Fe
23	Li	Al
24	Cu	Al
25	Ni, Zn	Mg, K, Mn, Fe
26	Ni	K
27	Zn	Mn
28	Mg	Mn
29	Na	Mn, Fe
30	Mn	Fe

A.3. Kriging Estimates

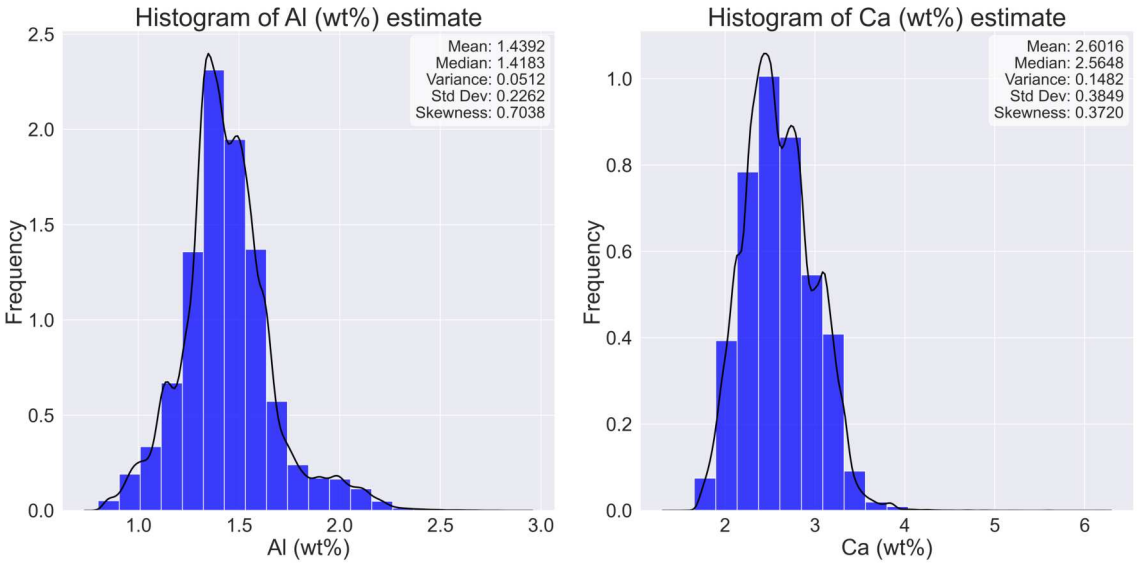


Figure A.63: EDA plots for Al (left) and Ca (right).

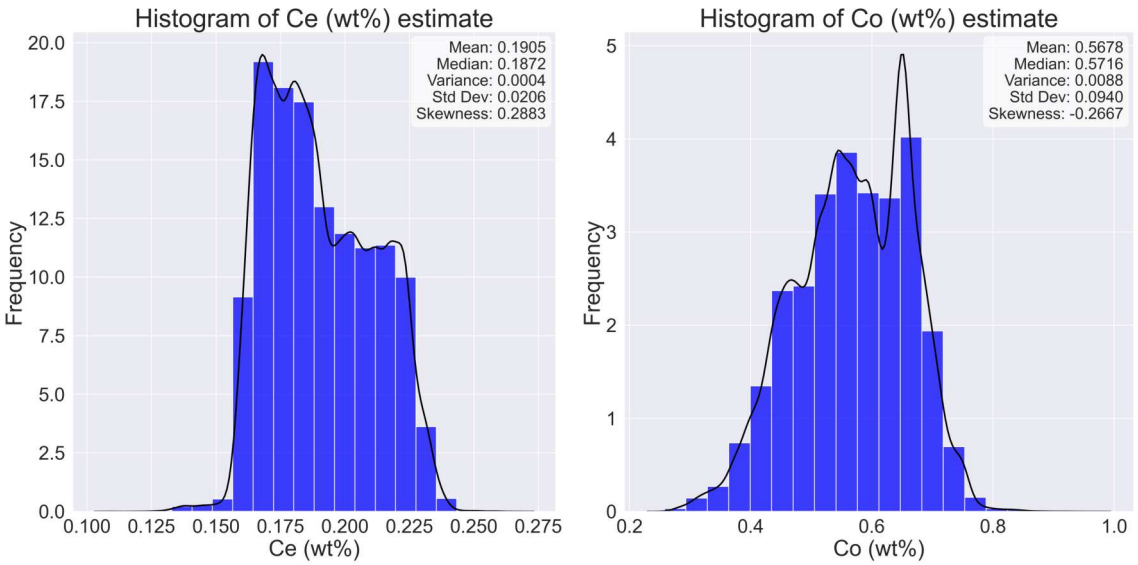


Figure A.64: EDA plots for Ce (left) and Co (right).

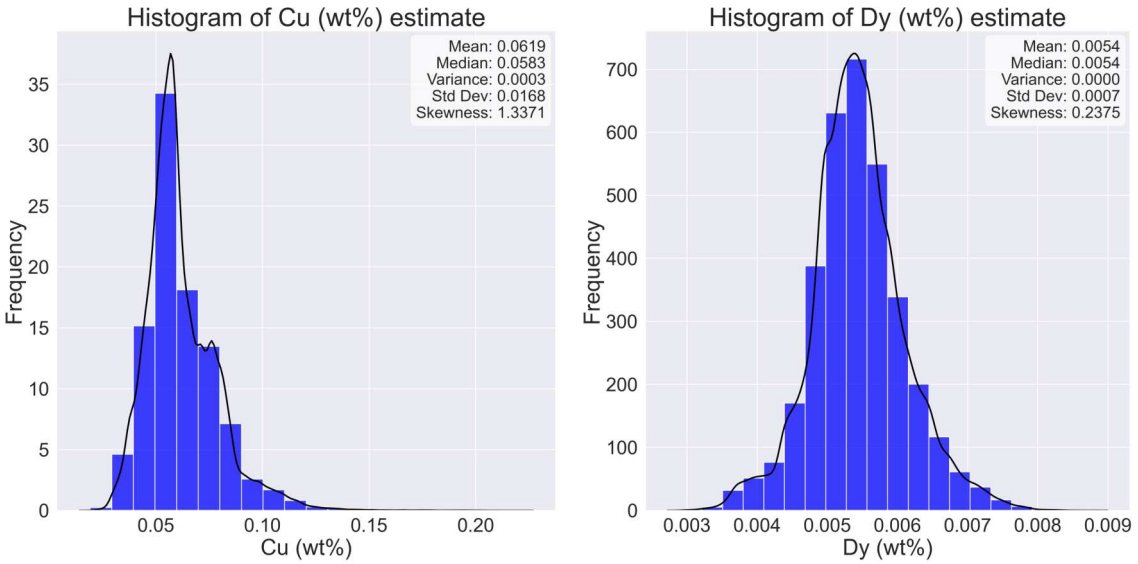


Figure A.65: EDA plots for Cu (left) and Dy (right).

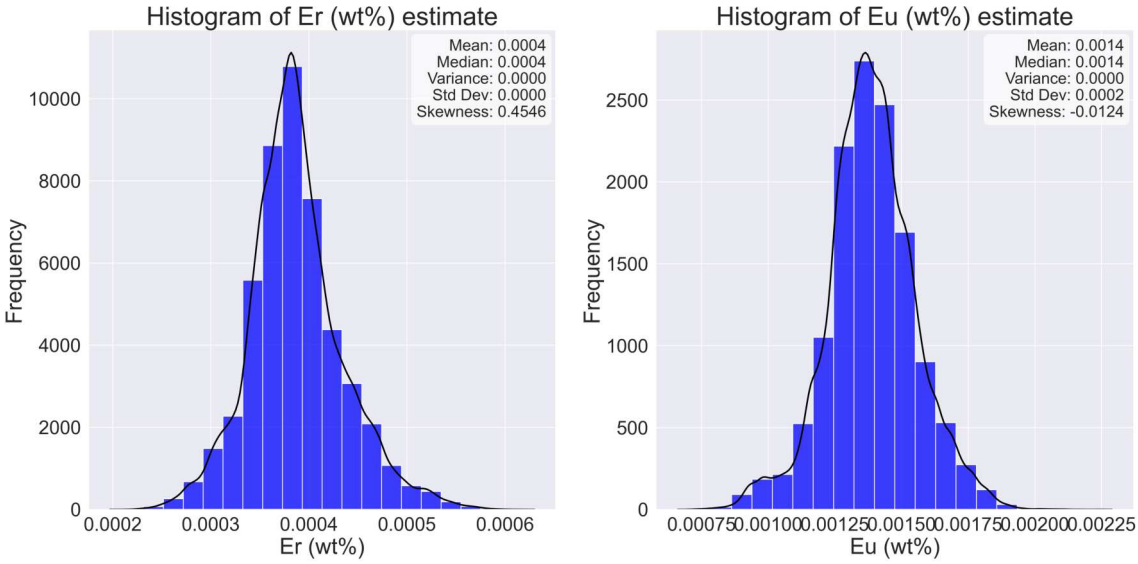


Figure A.66: EDA plots for Er (left) and Eu (right).

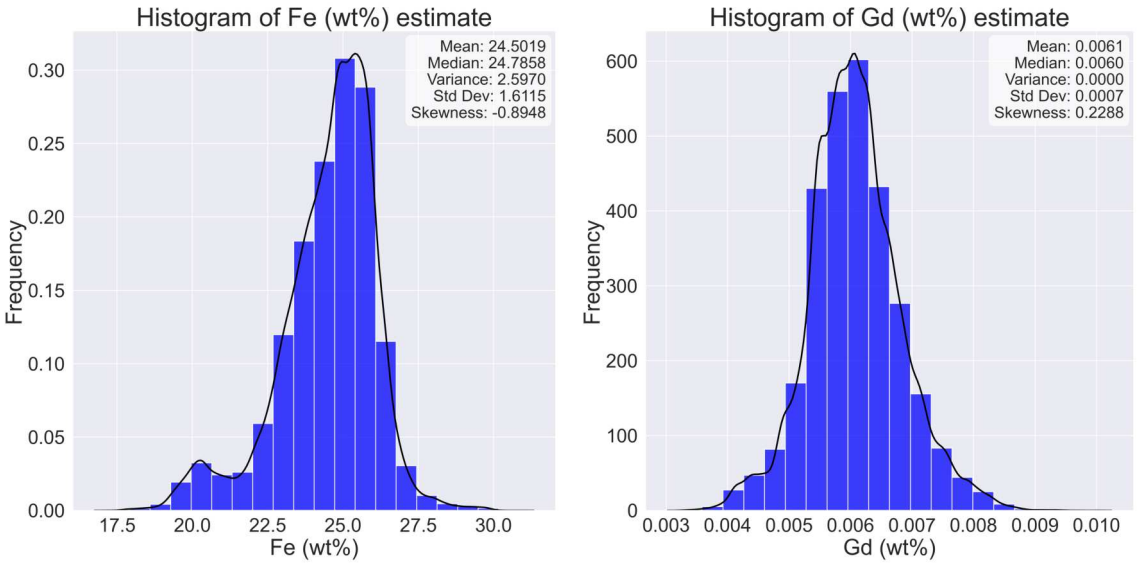


Figure A.67: EDA plots for Fe (left) and Gd (right).

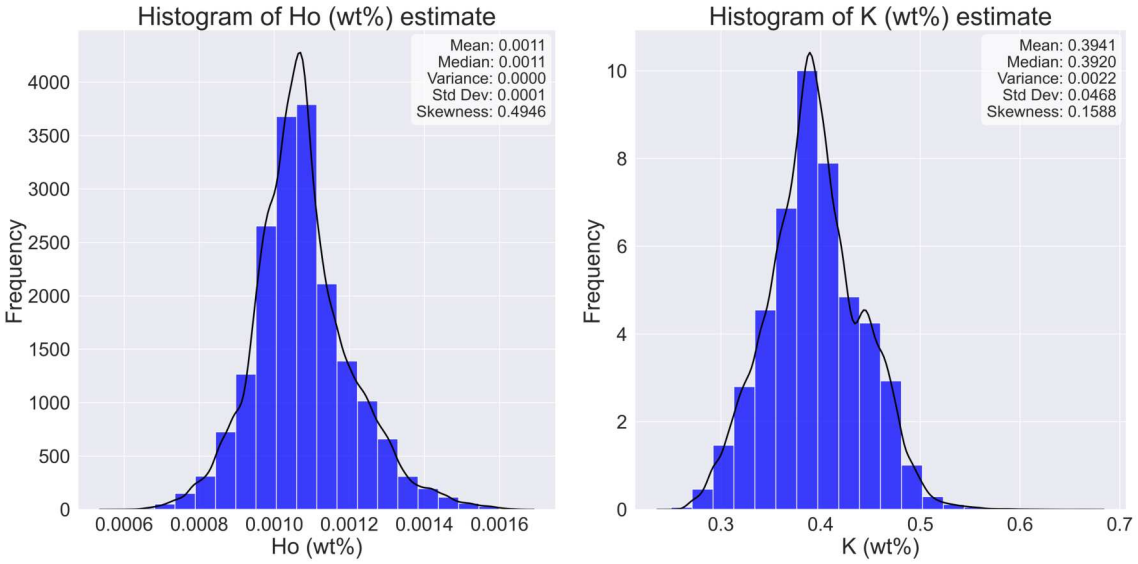


Figure A.68: EDA plots for Ho (left) and K (right).

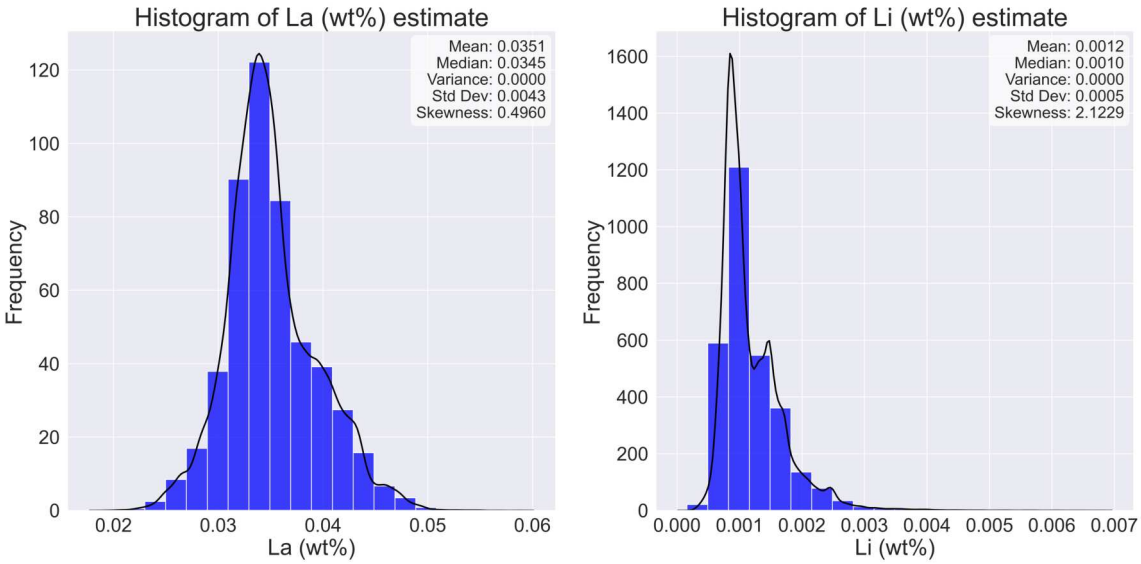


Figure A.69: EDA plots for La (left) and Li (right).

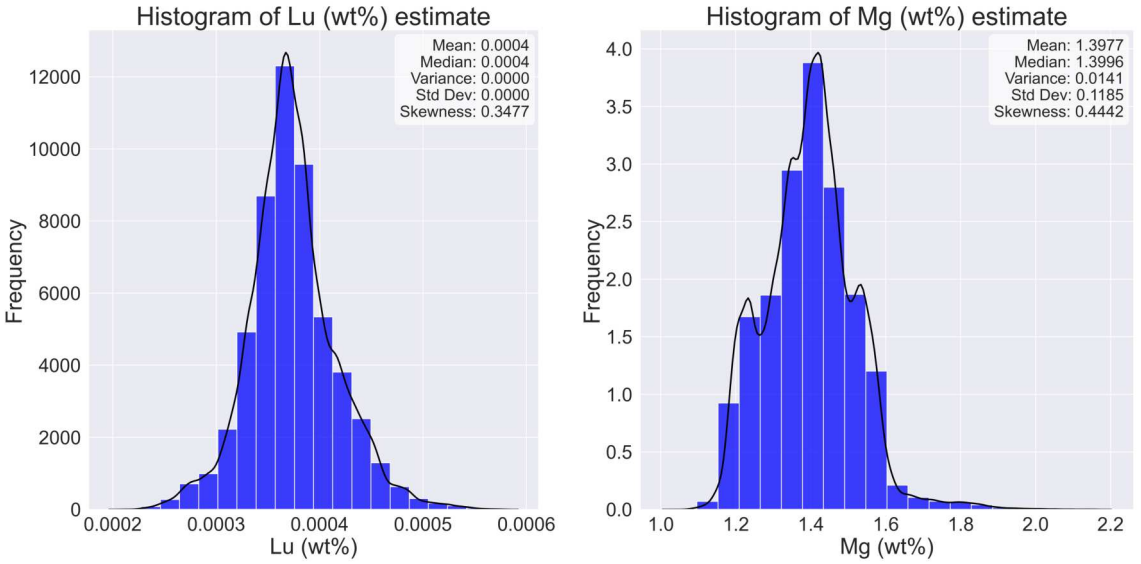


Figure A.70: EDA plots for Lu (left) and Mg (right).

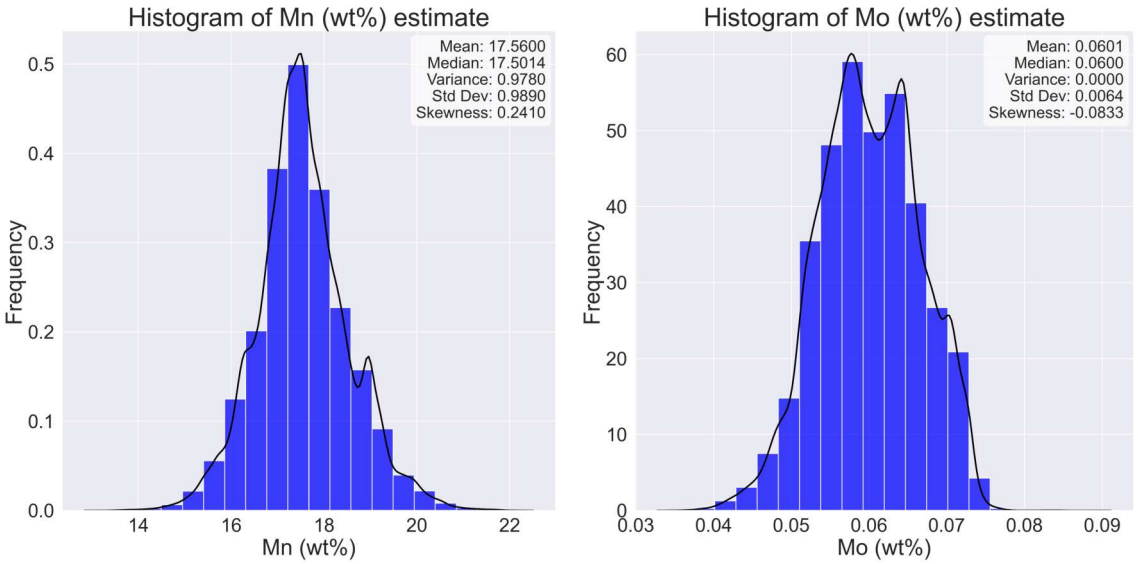


Figure A.71: EDA plots for Mn (left) and Mo (right).

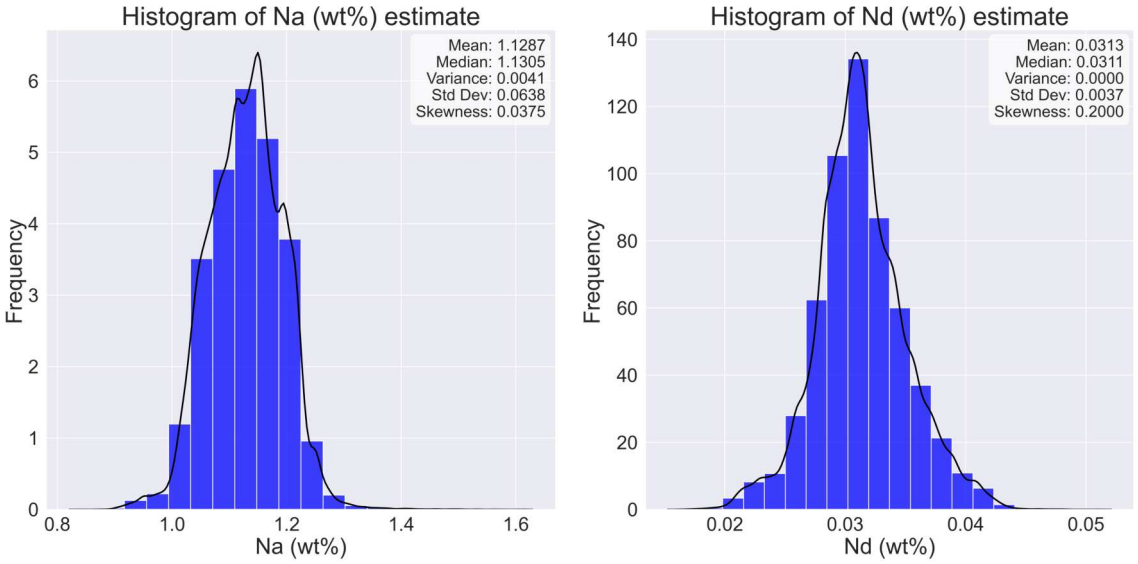


Figure A.72: EDA plots for Na (left) and Nd (right).

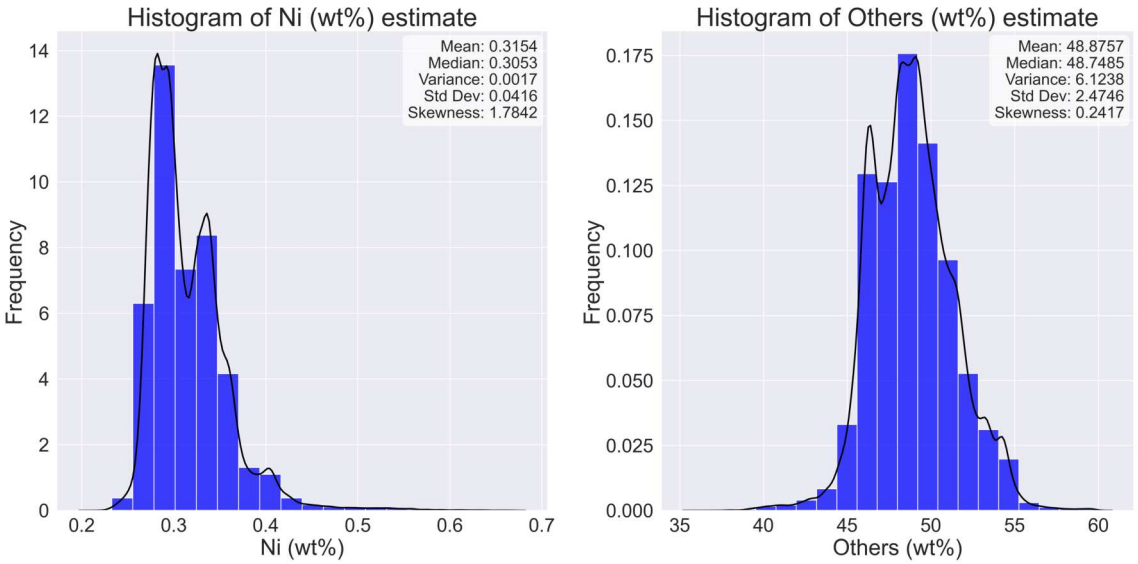


Figure A.73: EDA plots for Ni (left) and Others (right).

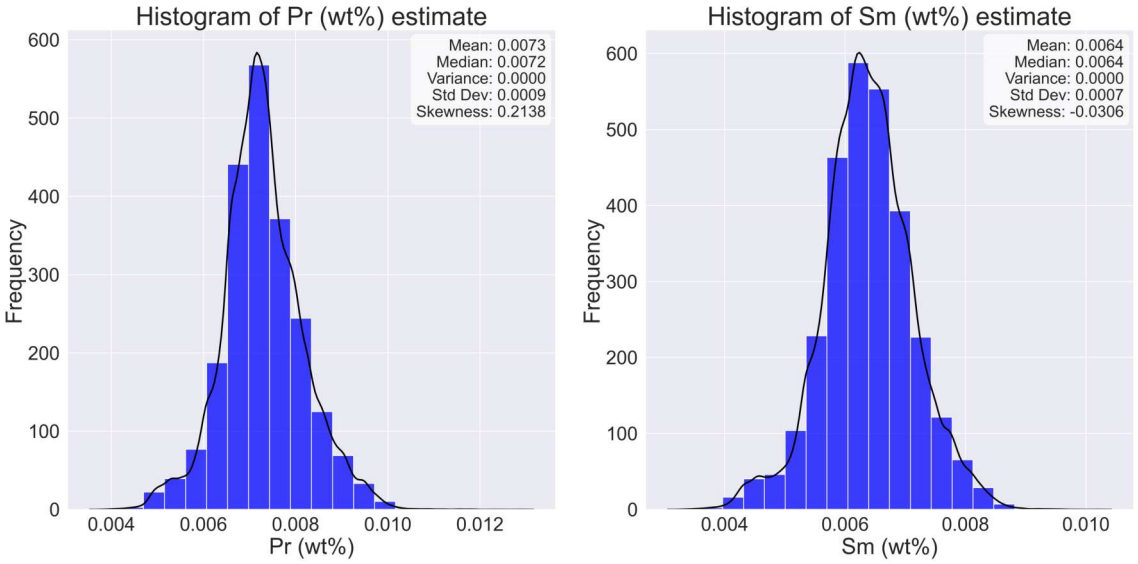


Figure A.74: EDA plots for Pr (left) and Sm (right).

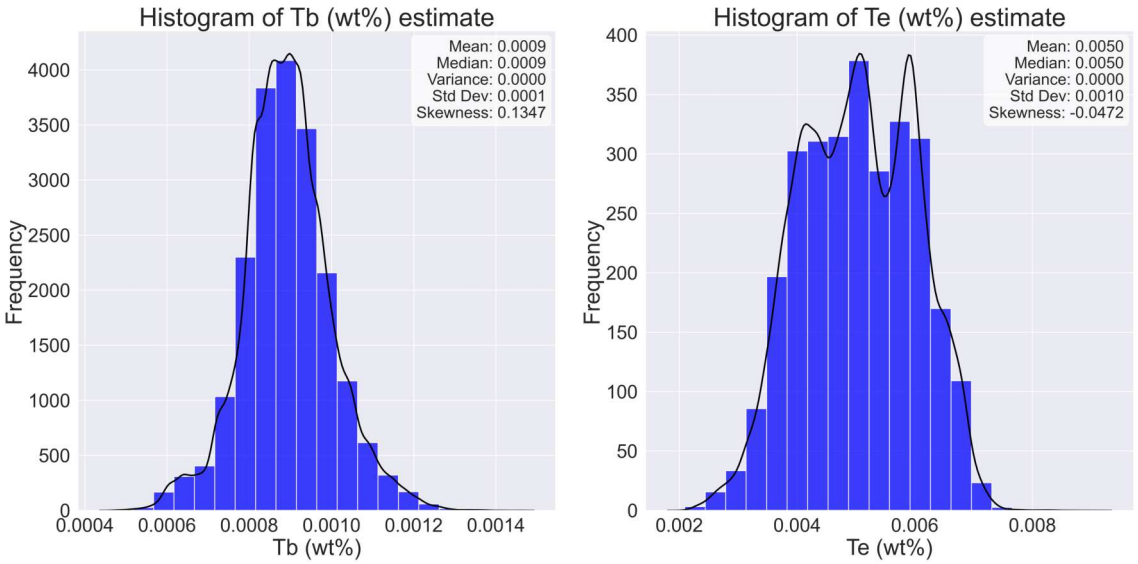


Figure A.75: EDA plots for Tb (left) and Te (right).

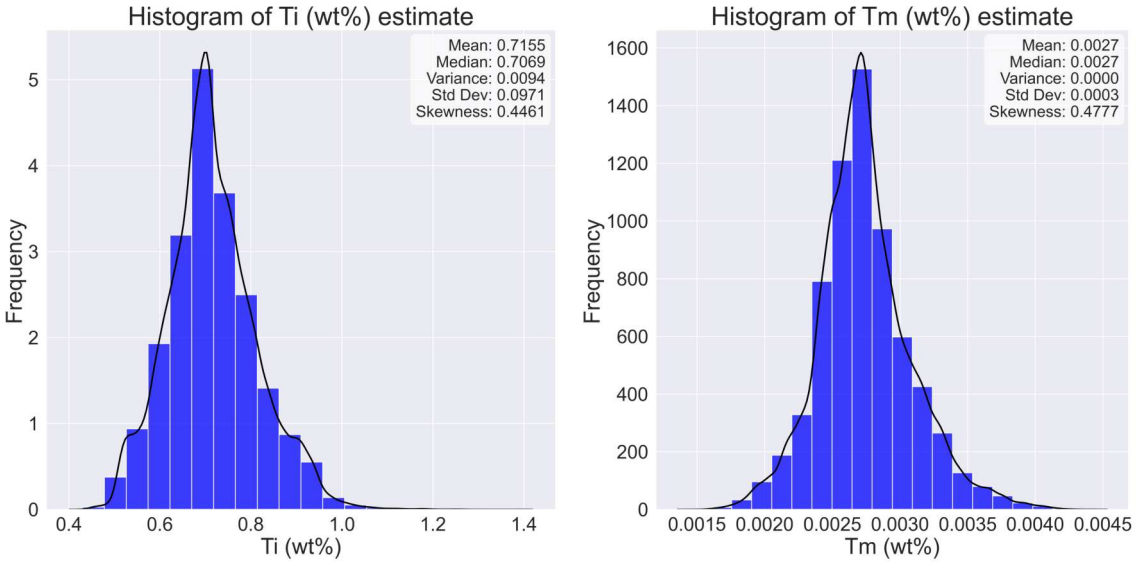


Figure A.76: EDA plots for Ti (left) and Tm (right).

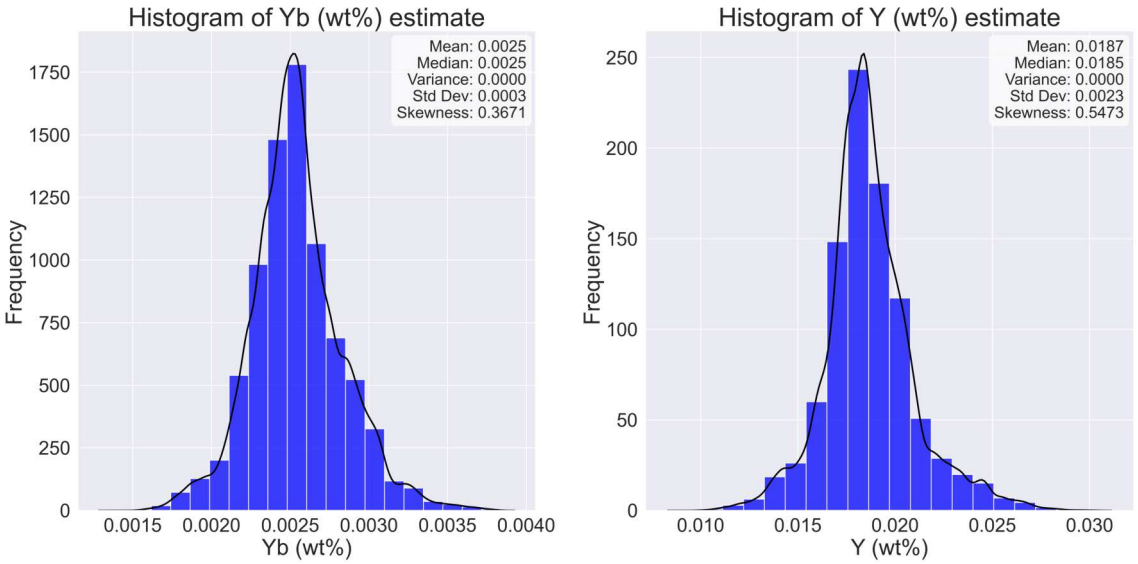


Figure A.77: EDA plots for Yb (left) and Y (right).

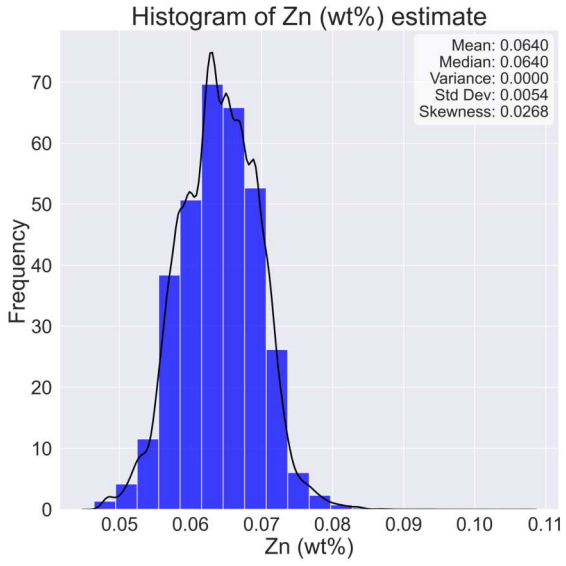


Figure A.78: EDA plot for Zn.

A.4. EDA ALR

Table A.2: CLR variance and percent of total variance explained by each element. Total CLR Variance: 2.671200

Element	CLR Variance	Percent of Total Variance
La (wt%)	0.0297	1.11
Ce (wt%)	0.0205	0.77
Pr (wt%)	0.0383	1.43
Nd (wt%)	0.0365	1.37
Sm (wt%)	0.0398	1.49
Eu (wt%)	0.0395	1.48
Gd (wt%)	0.0386	1.45
Tb (wt%)	0.0378	1.42
Dy (wt%)	0.0369	1.38
Ho (wt%)	0.0373	1.40
Er (wt%)	0.0367	1.37
Tm (wt%)	0.0338	1.27
Yb (wt%)	0.0312	1.17
Lu (wt%)	0.0316	1.18
Y (wt%)	0.0684	2.56
Li (wt%)	0.7207	26.98
Co (wt%)	0.0828	3.10
Ni (wt%)	0.1081	4.05
Cu (wt%)	0.3125	11.70
Zn (wt%)	0.0493	1.85
Mo (wt%)	0.0439	1.64
Te (wt%)	0.0914	3.42
Na (wt%)	0.0217	0.81
Mg (wt%)	0.0381	1.43
Al (wt%)	0.1402	5.25
K (wt%)	0.0697	2.61
Ca (wt%)	0.2814	10.53
Ti (wt%)	0.0473	1.77
Mn (wt%)	0.0181	0.68
Fe (wt%)	0.0195	0.73
Others (wt%)	0.0699	2.62

B

Variograms

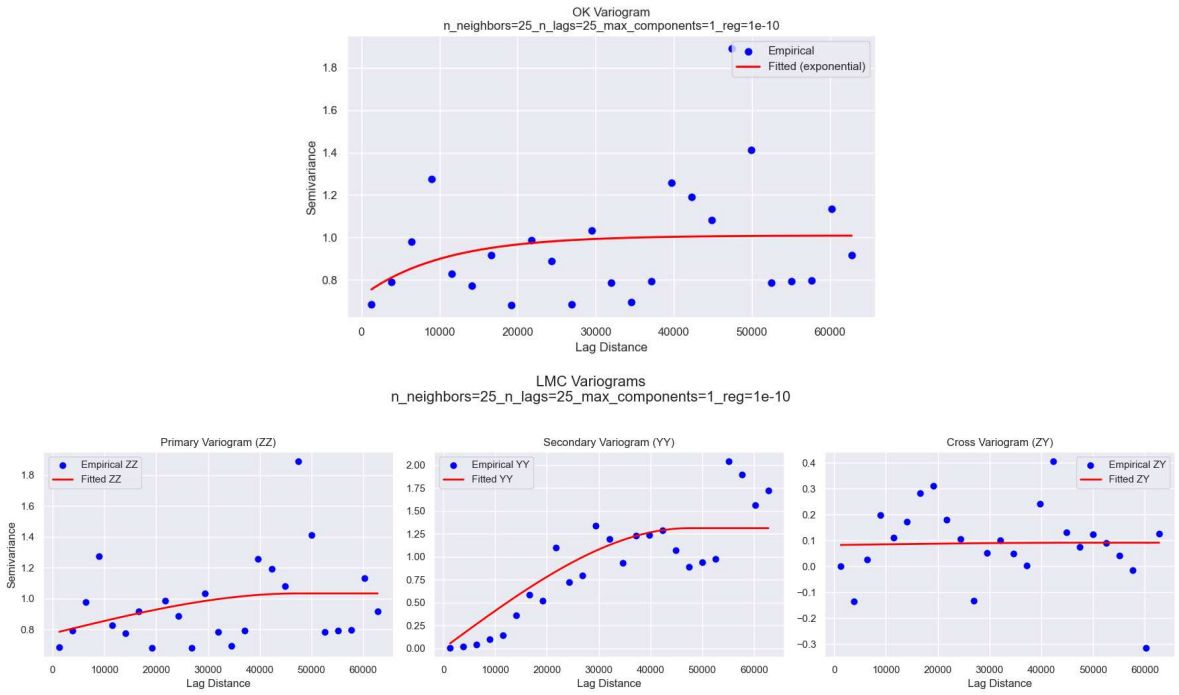


Figure B.1: OK (top) and LMC (bottom) variogram fits for ilr.1.

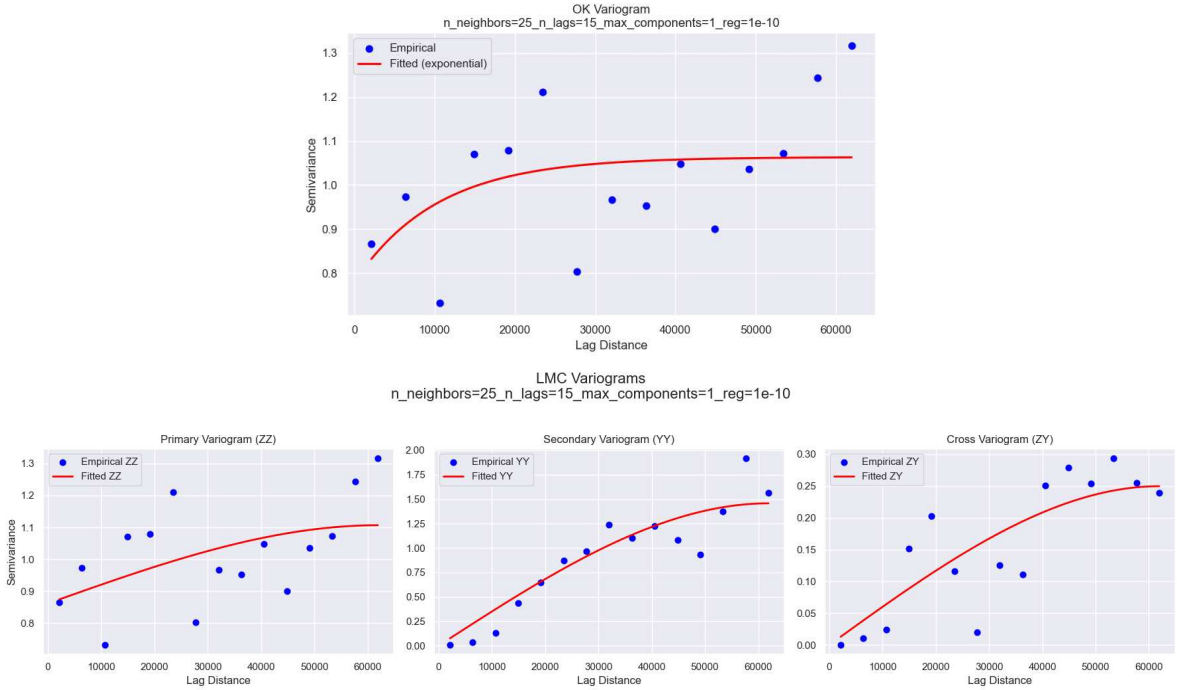


Figure B.2: OK (top) and LMC (bottom) variogram fits for ilr.2.

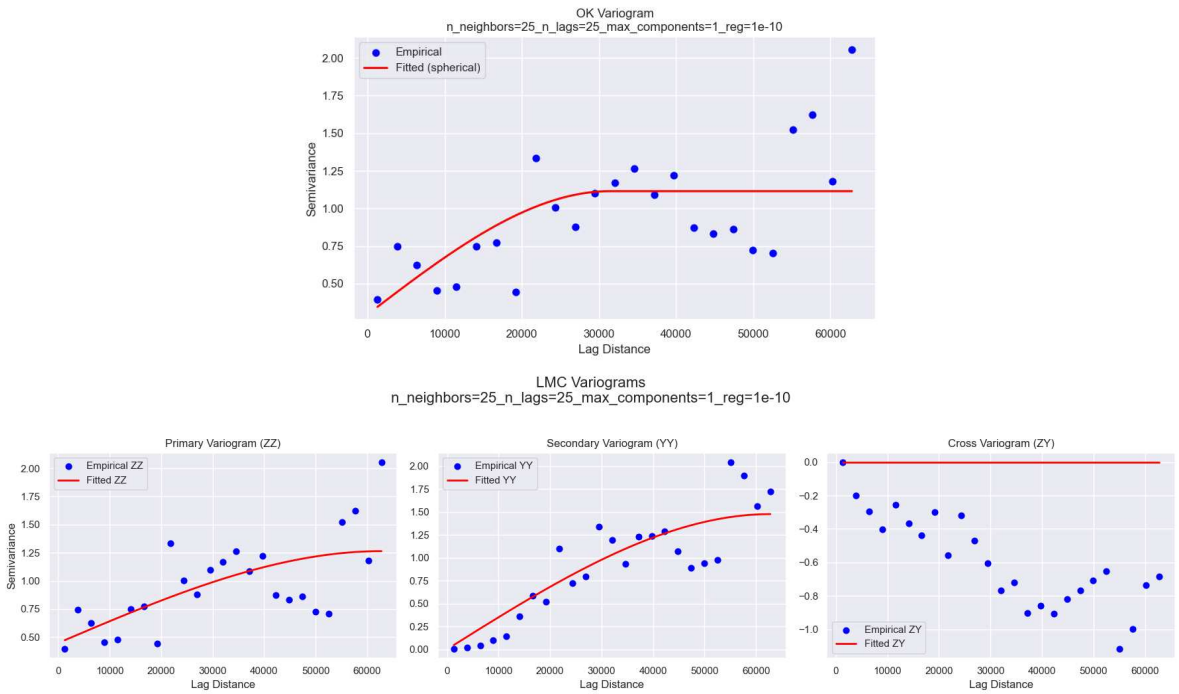


Figure B.3: OK (top) and LMC (bottom) variogram fits for ilr.3.

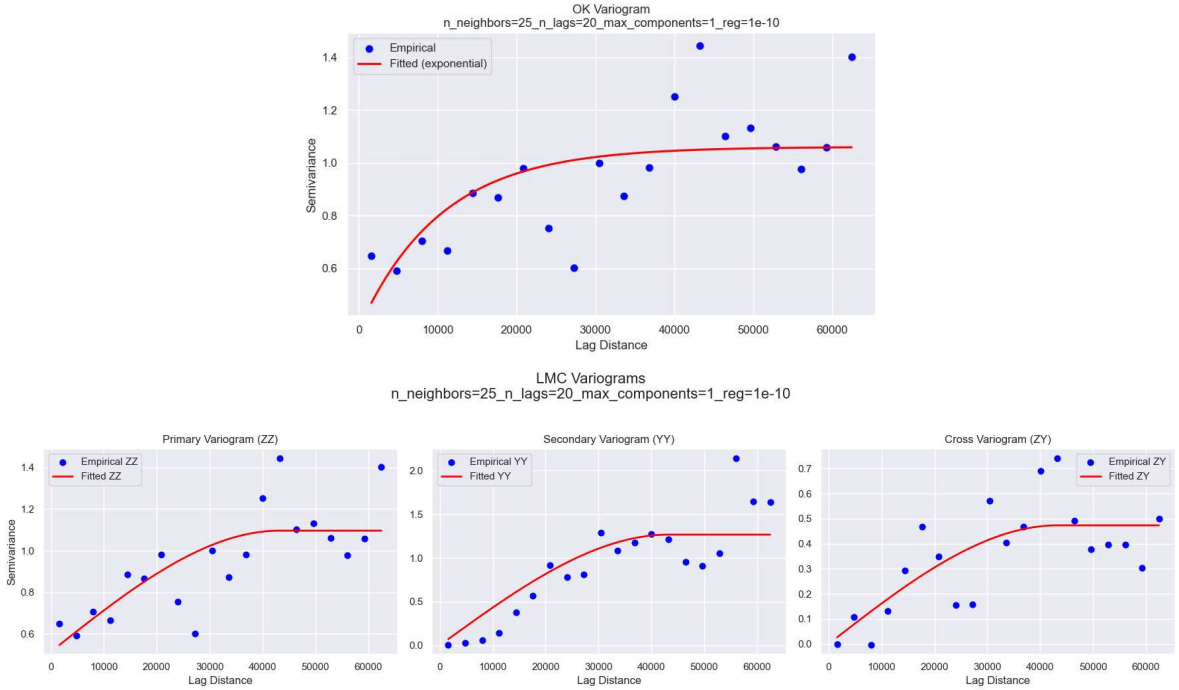


Figure B.4: OK (top) and LMC (bottom) variogram fits for ilr.4.

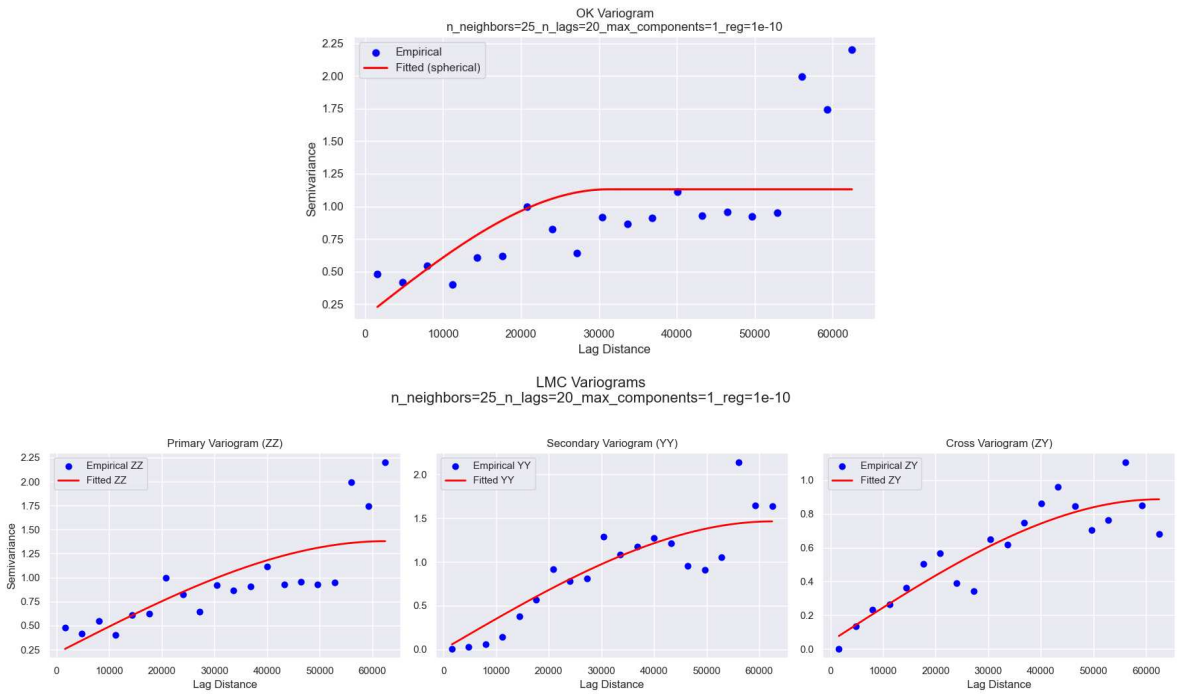


Figure B.5: OK (top) and LMC (bottom) variogram fits for ilr.5.

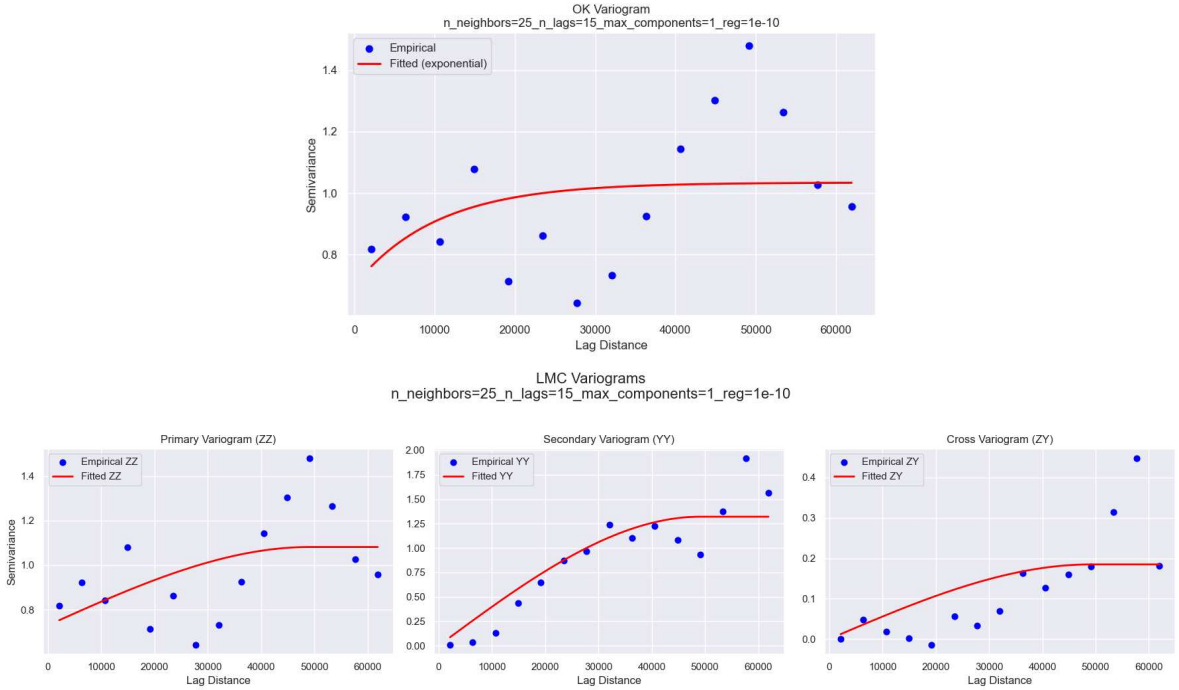


Figure B.6: OK (top) and LMC (bottom) variogram fits for ilr.6.

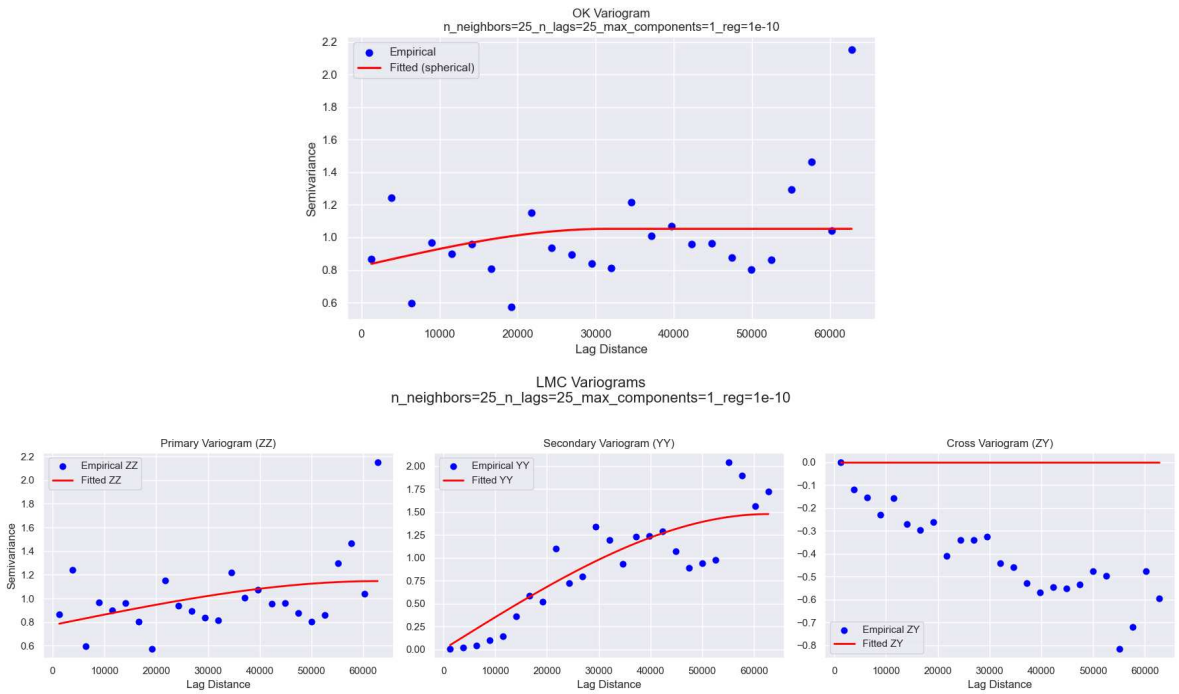


Figure B.7: OK (top) and LMC (bottom) variogram fits for ilr.7.

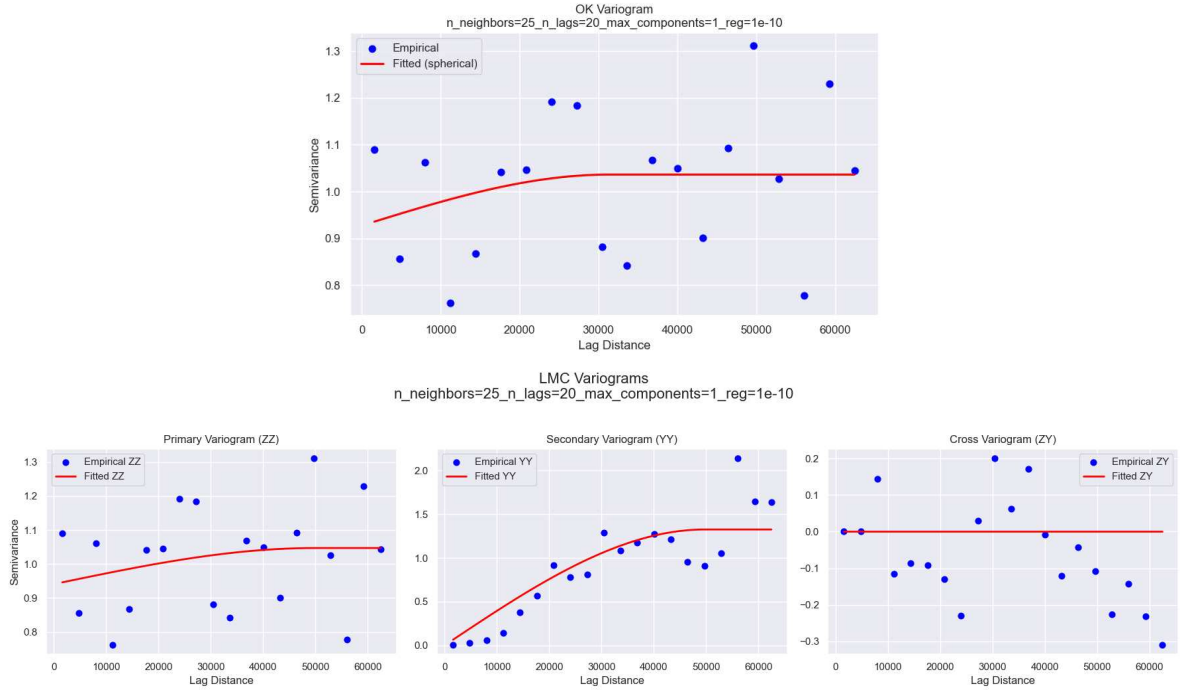


Figure B.8: OK (top) and LMC (bottom) variogram fits for ilr.8.

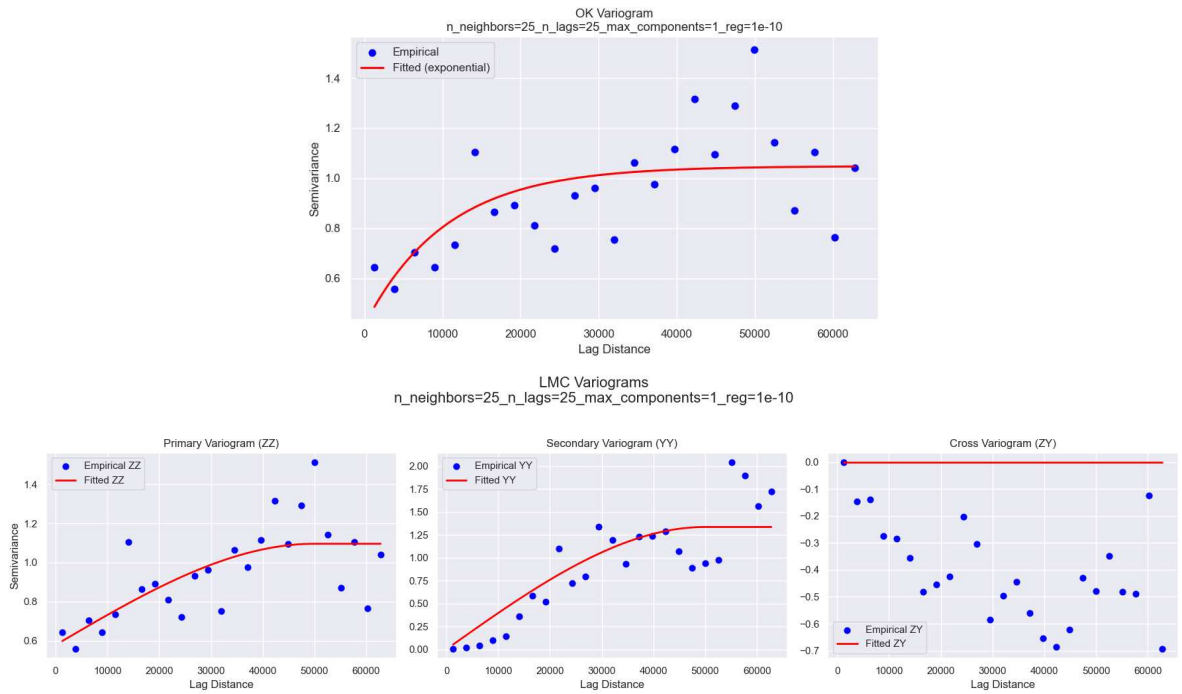


Figure B.9: OK (top) and LMC (bottom) variogram fits for ilr.9.

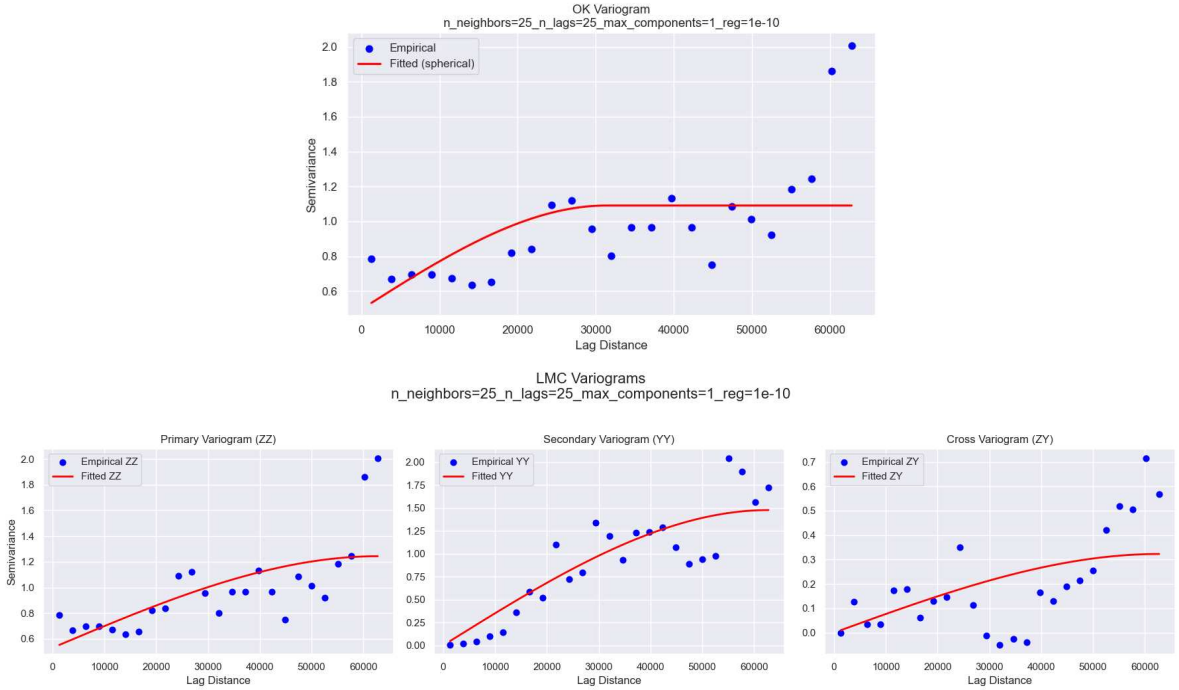


Figure B.10: OK (top) and LMC (bottom) variogram fits for ilr.10.

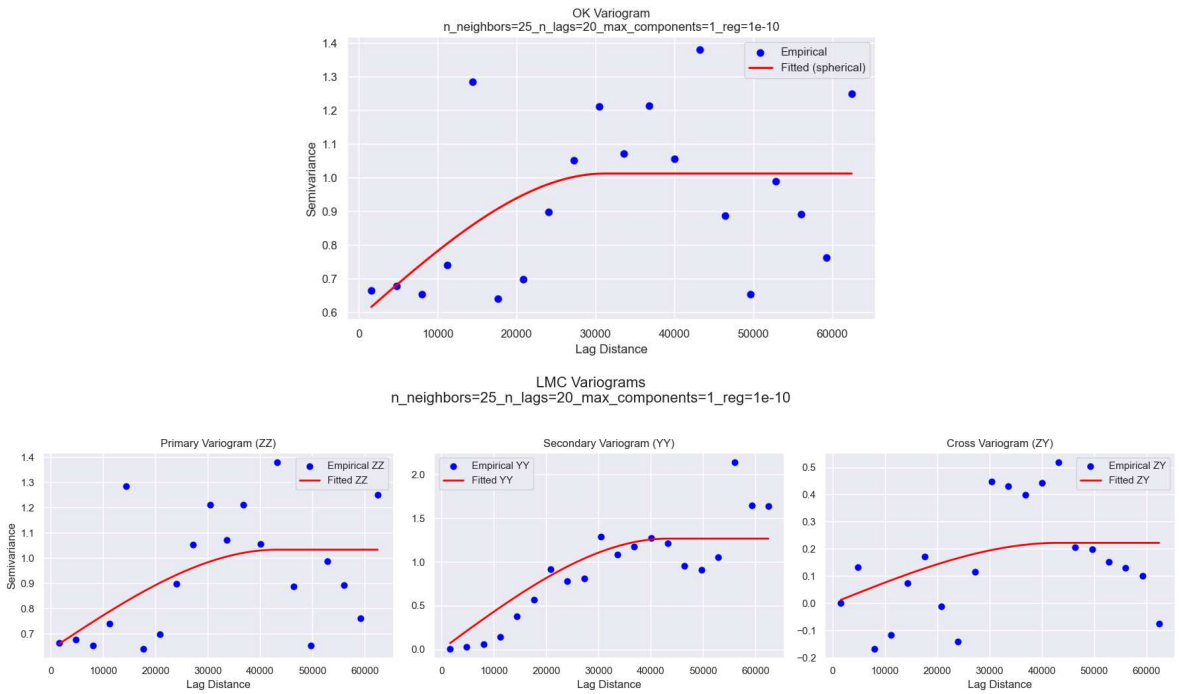


Figure B.11: OK (top) and LMC (bottom) variogram fits for ilr.11.

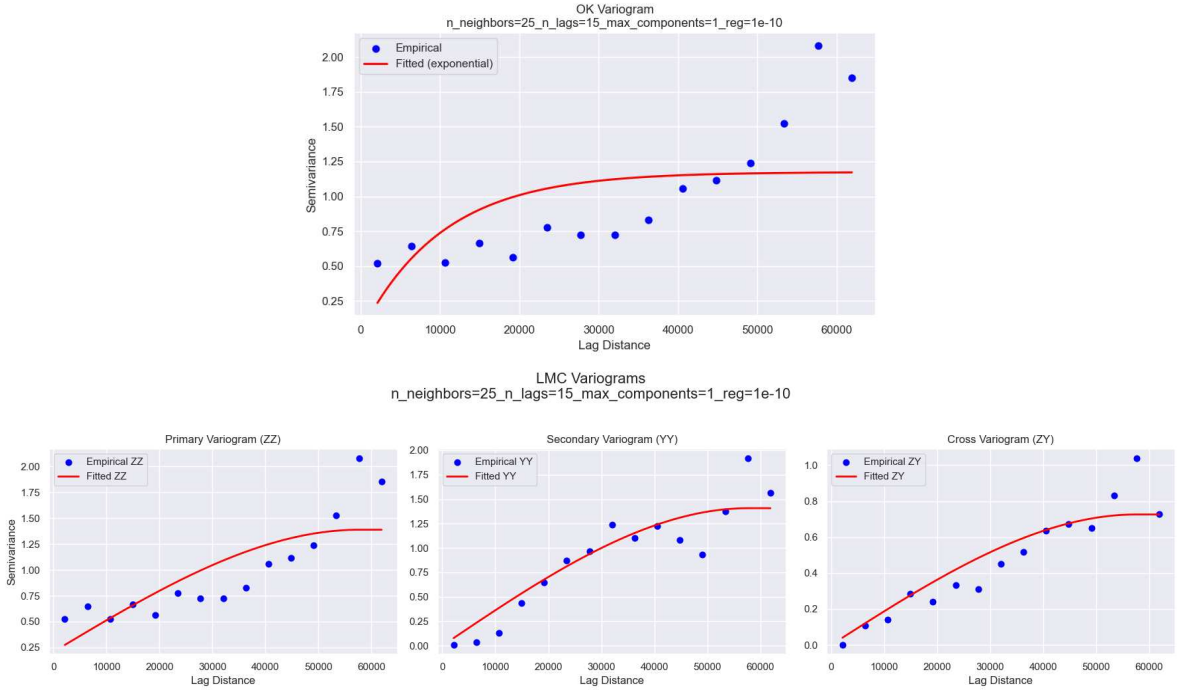


Figure B.12: OK (top) and LMC (bottom) variogram fits for ilr.12.

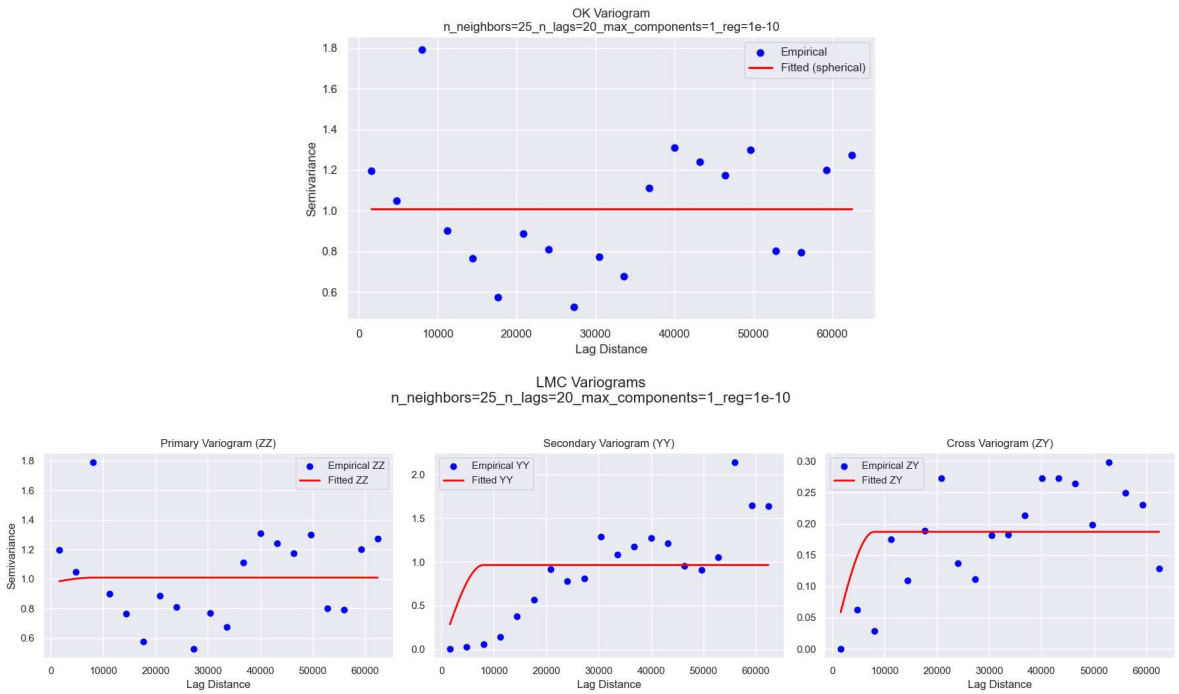


Figure B.13: OK (top) and LMC (bottom) variogram fits for ilr.13.

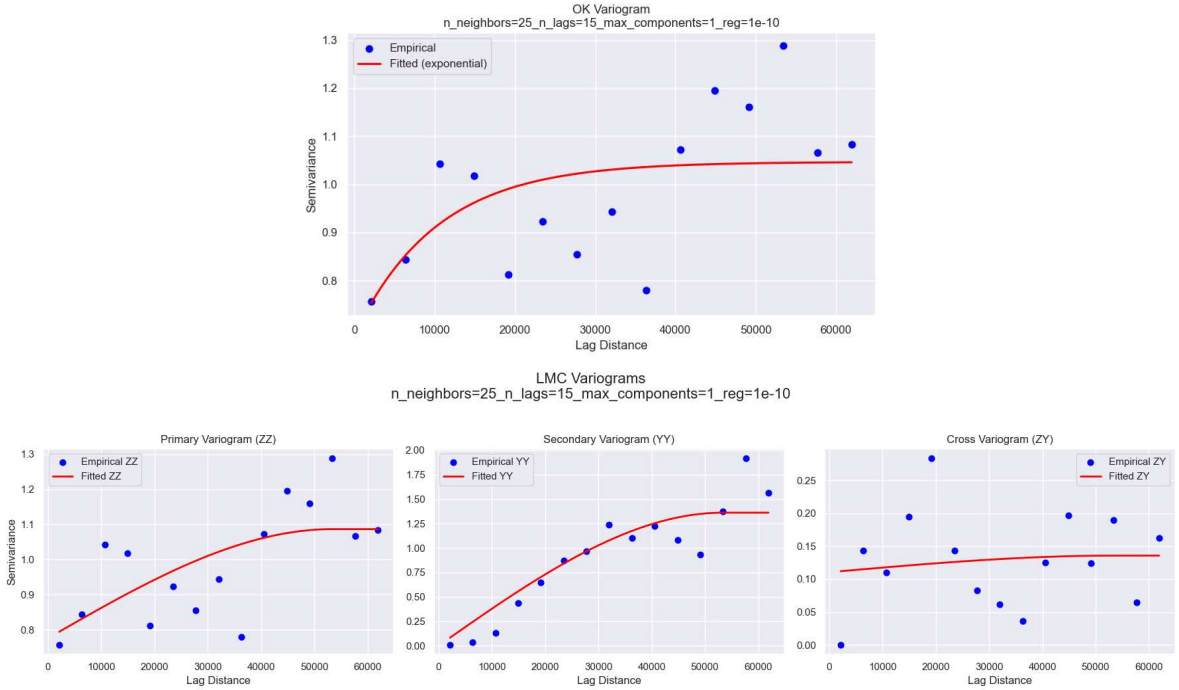


Figure B.14: OK (top) and LMC (bottom) variogram fits for ilr.14.

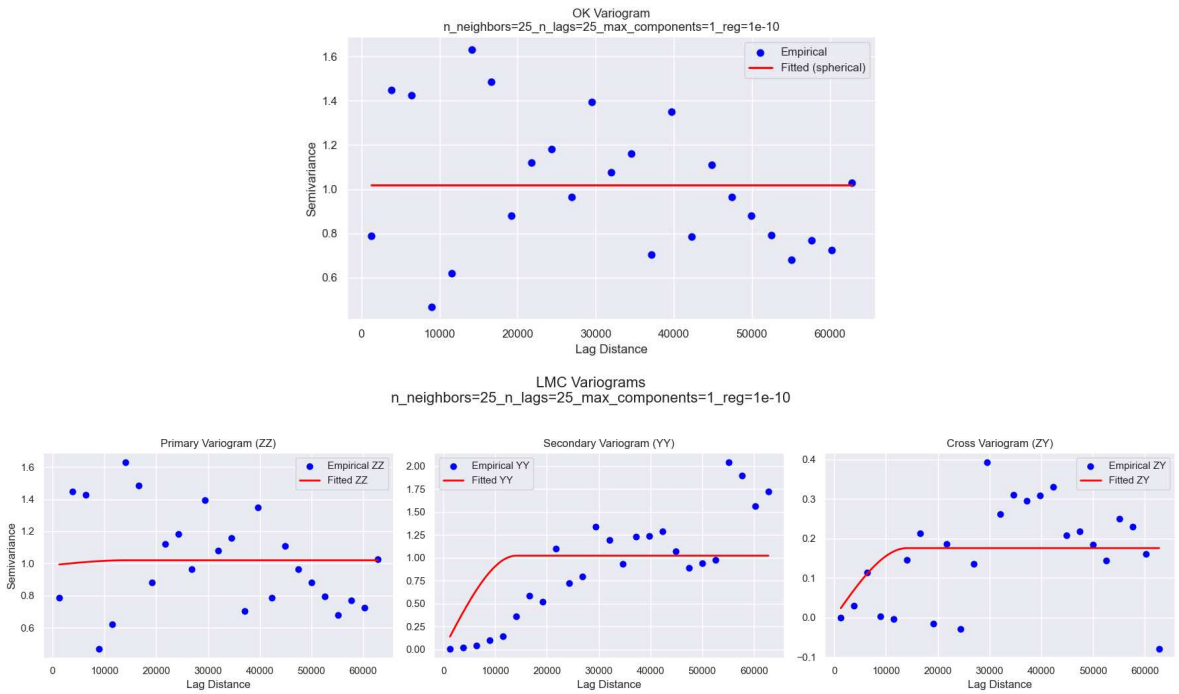


Figure B.15: OK (top) and LMC (bottom) variogram fits for ilr.15.

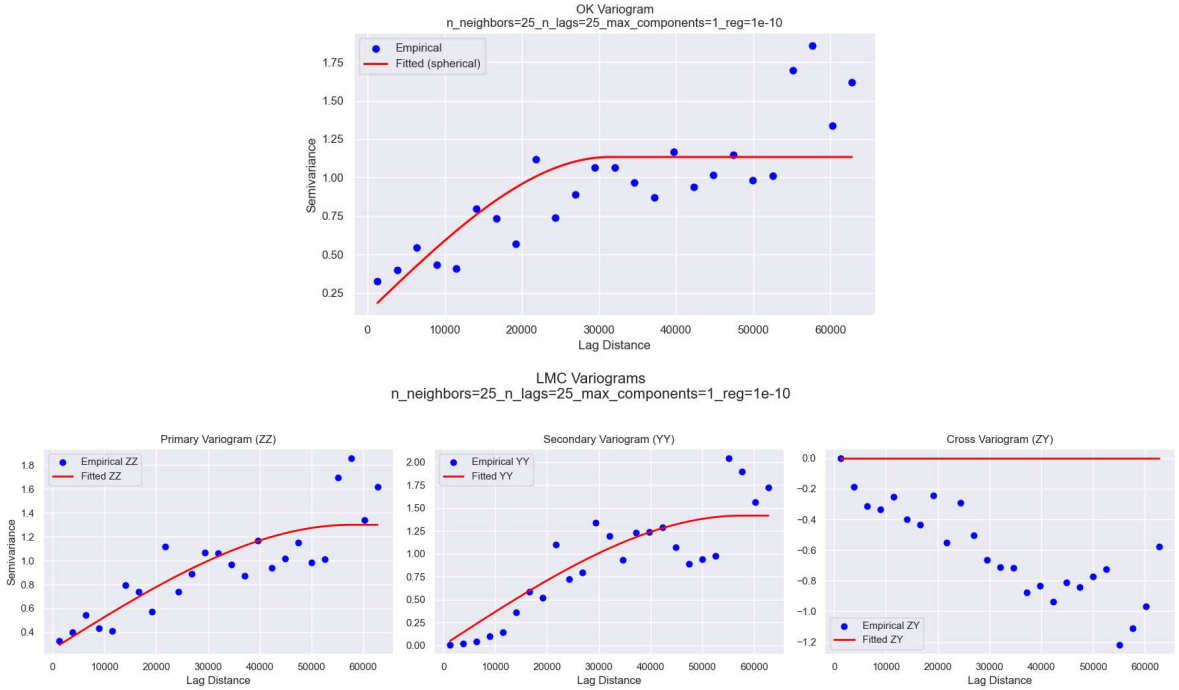


Figure B.16: OK (top) and LMC (bottom) variogram fits for ilr.16.

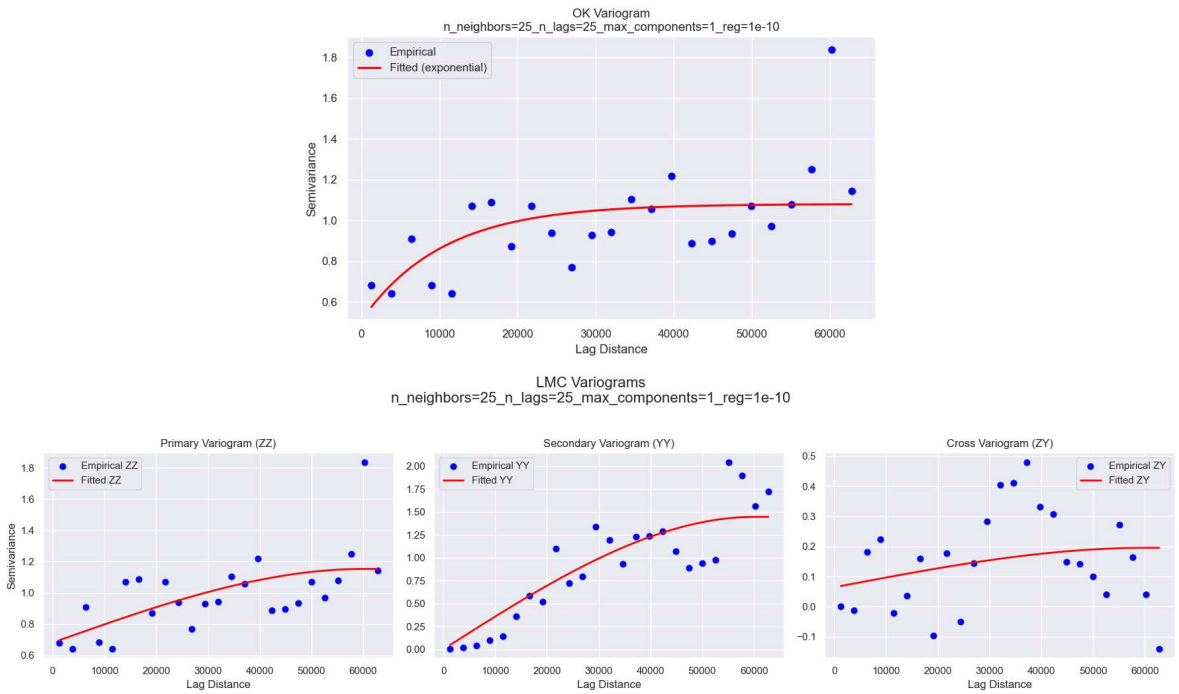


Figure B.17: OK (top) and LMC (bottom) variogram fits for ilr.17.

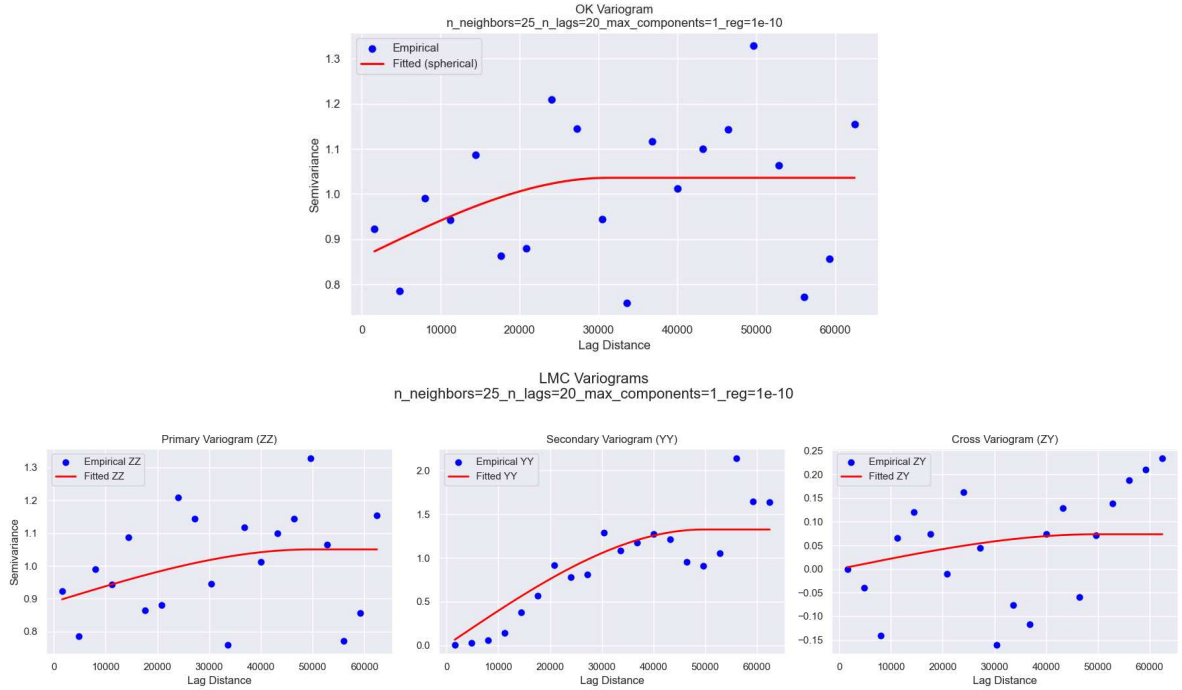


Figure B.18: OK (top) and LMC (bottom) variogram fits for ilr.18.

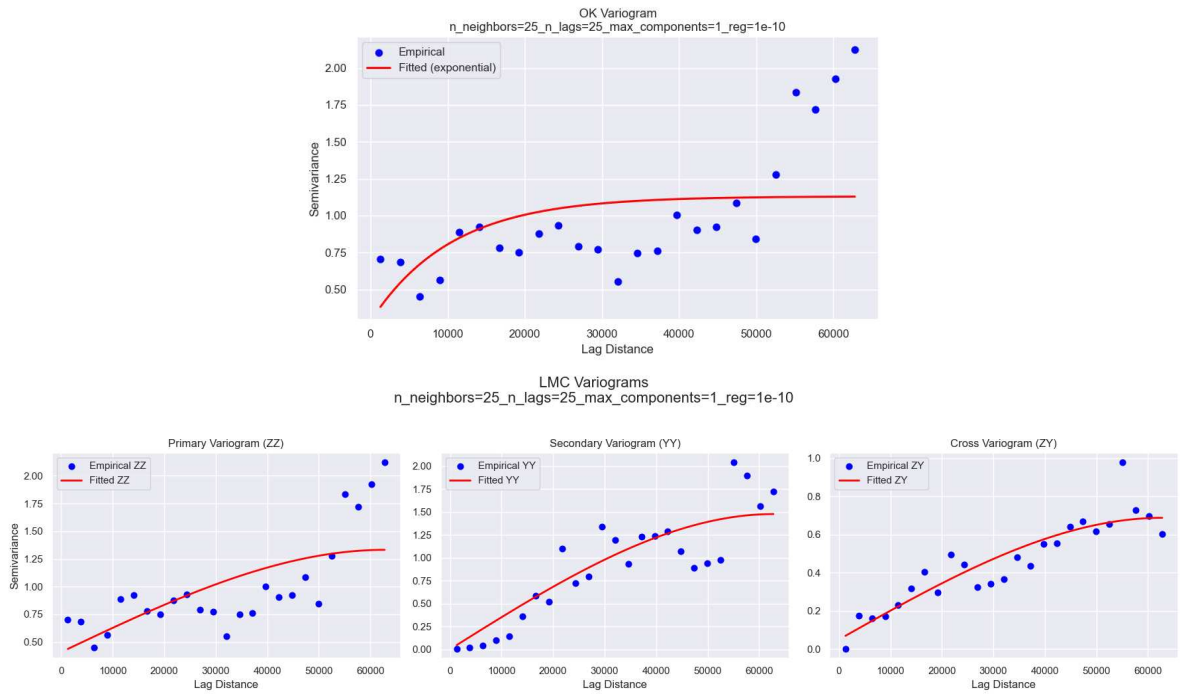


Figure B.19: OK (top) and LMC (bottom) variogram fits for ilr.19.

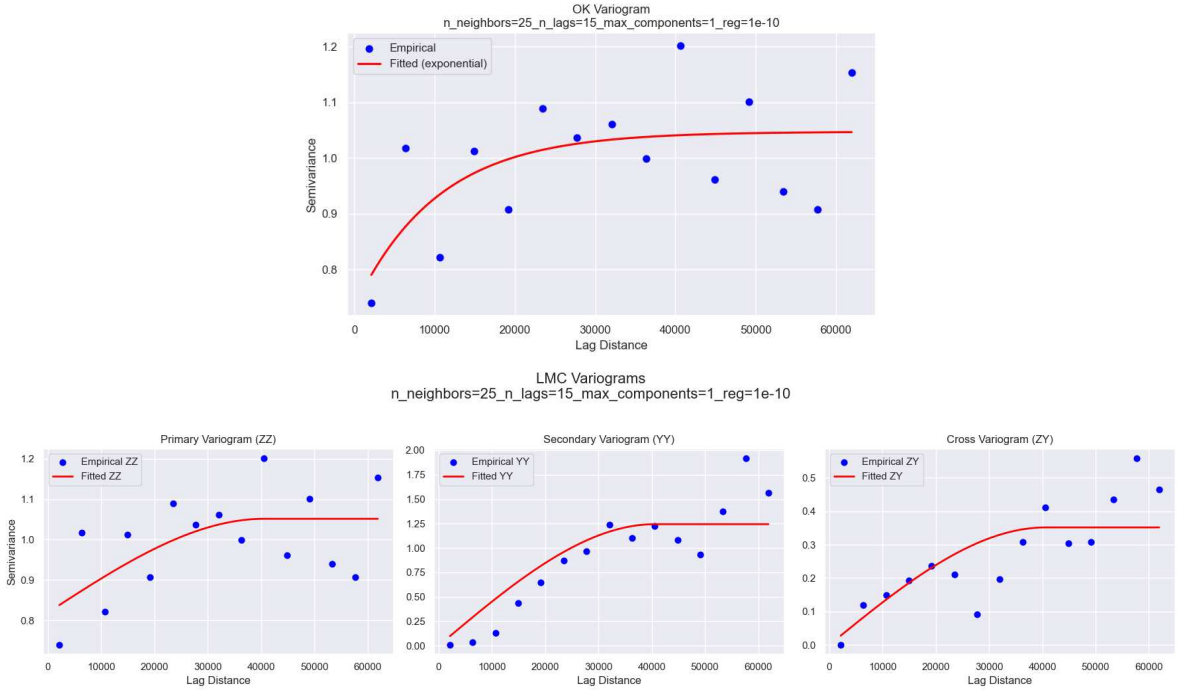


Figure B.20: OK (top) and LMC (bottom) variogram fits for ilr.20.

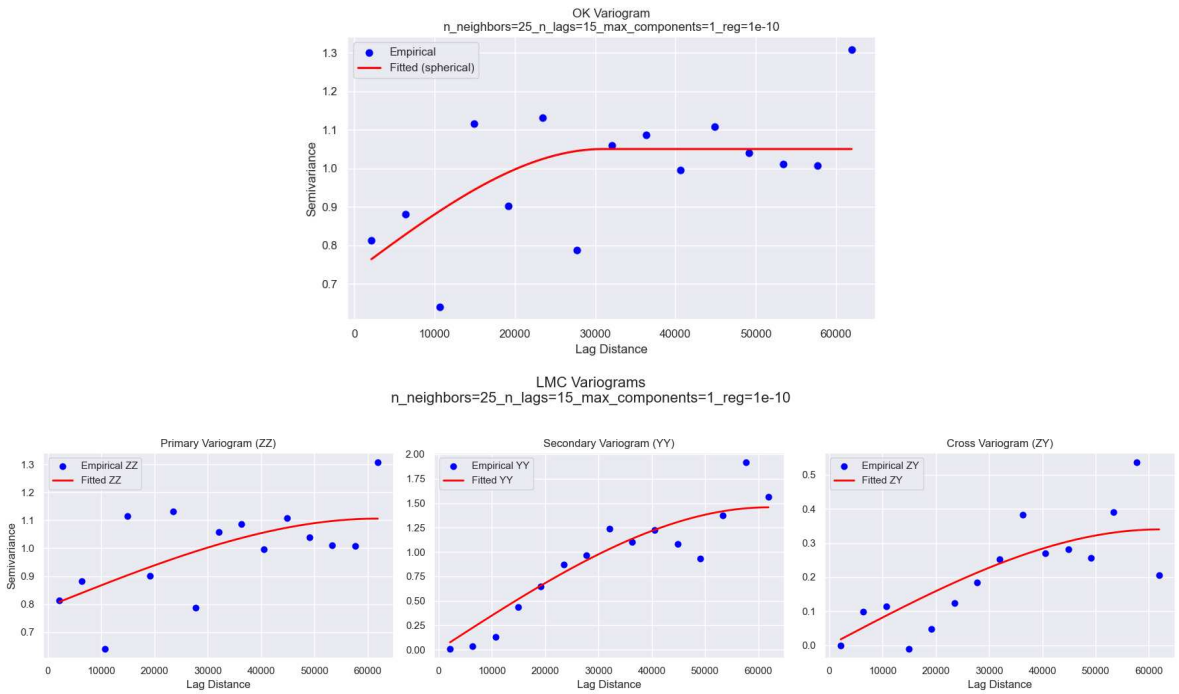


Figure B.21: OK (top) and LMC (bottom) variogram fits for ilr.21.

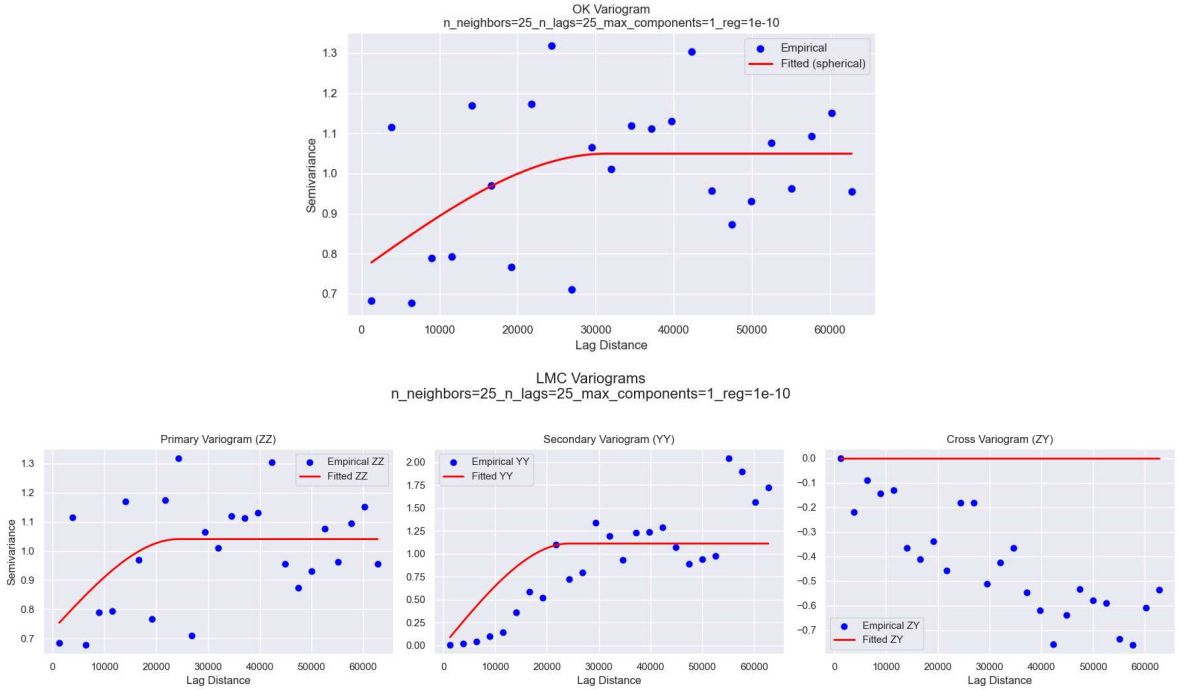


Figure B.22: OK (top) and LMC (bottom) variogram fits for ilr.22.

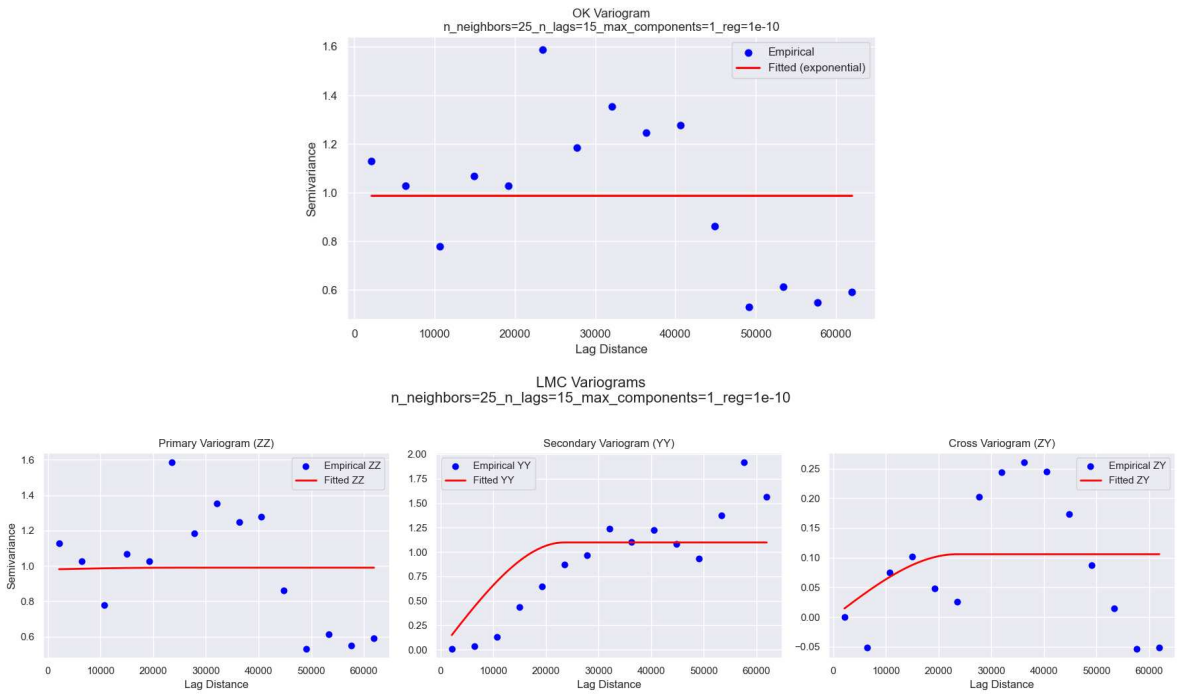


Figure B.23: OK (top) and LMC (bottom) variogram fits for ilr.23.

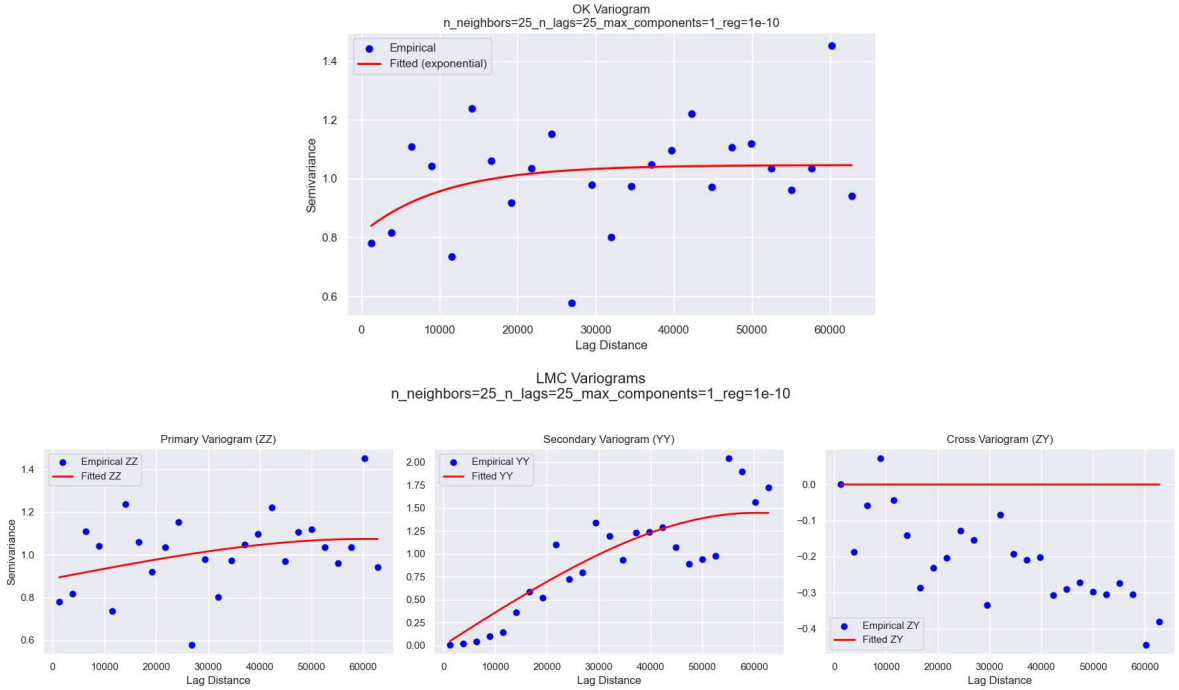


Figure B.24: OK (top) and LMC (bottom) variogram fits for ilr.24.

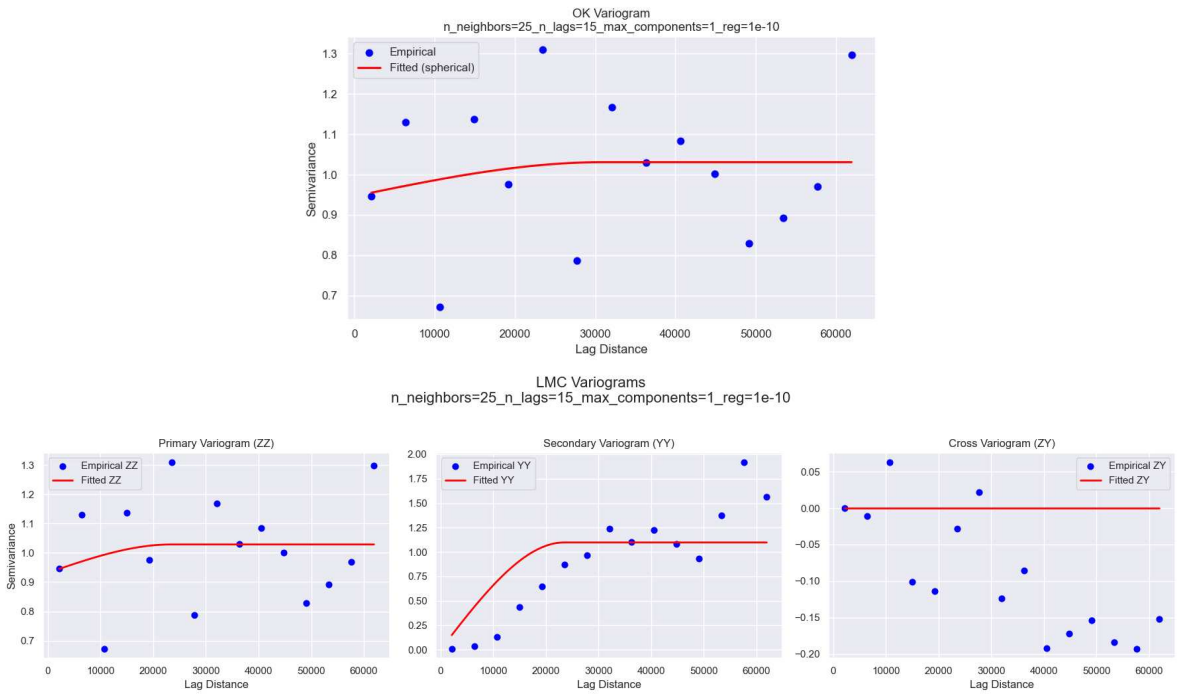


Figure B.25: OK (top) and LMC (bottom) variogram fits for ilr.25.

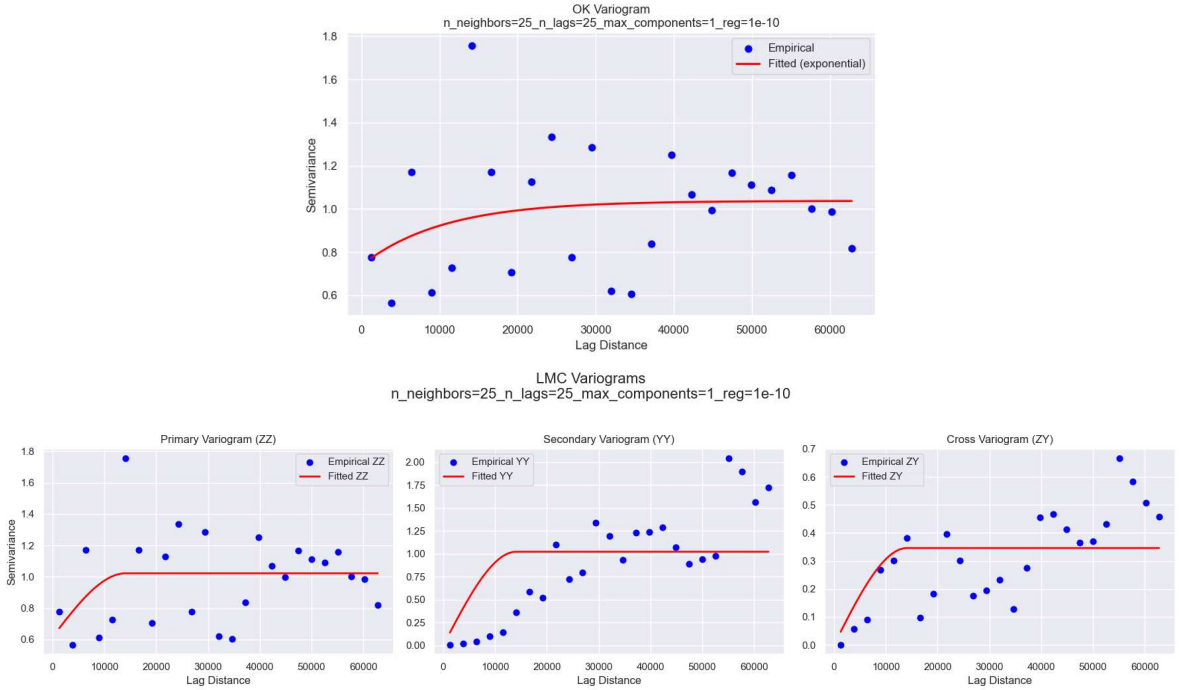


Figure B.26: OK (top) and LMC (bottom) variogram fits for ilr.26.

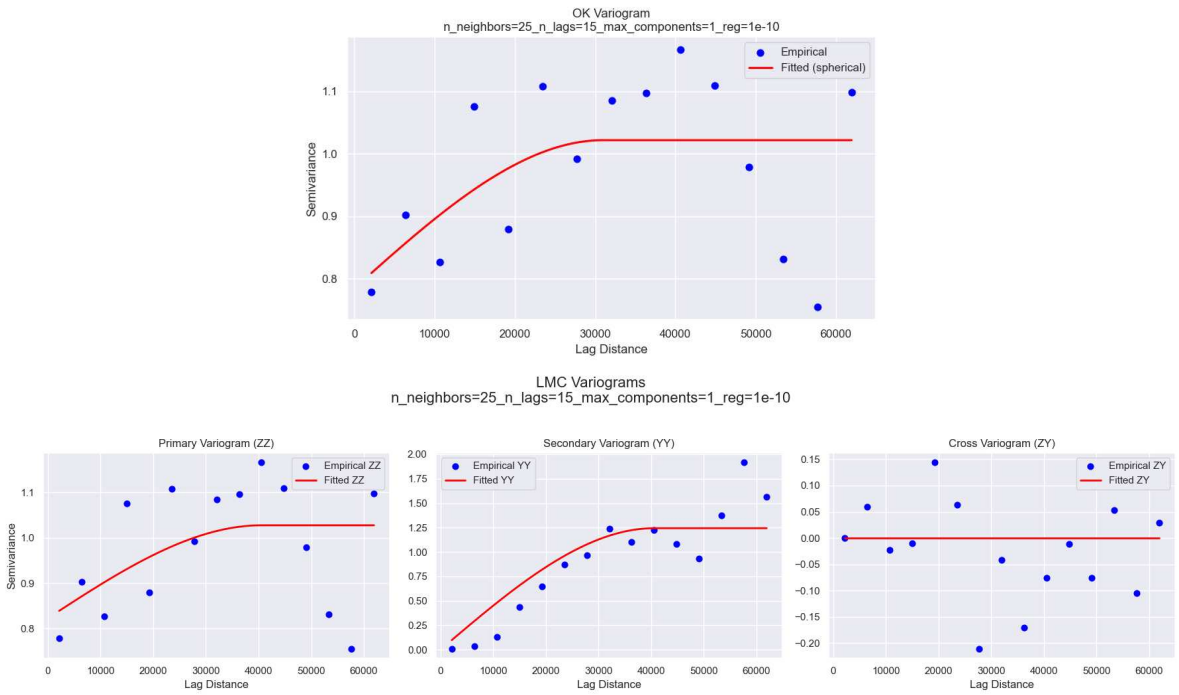


Figure B.27: OK (top) and LMC (bottom) variogram fits for ilr.27.

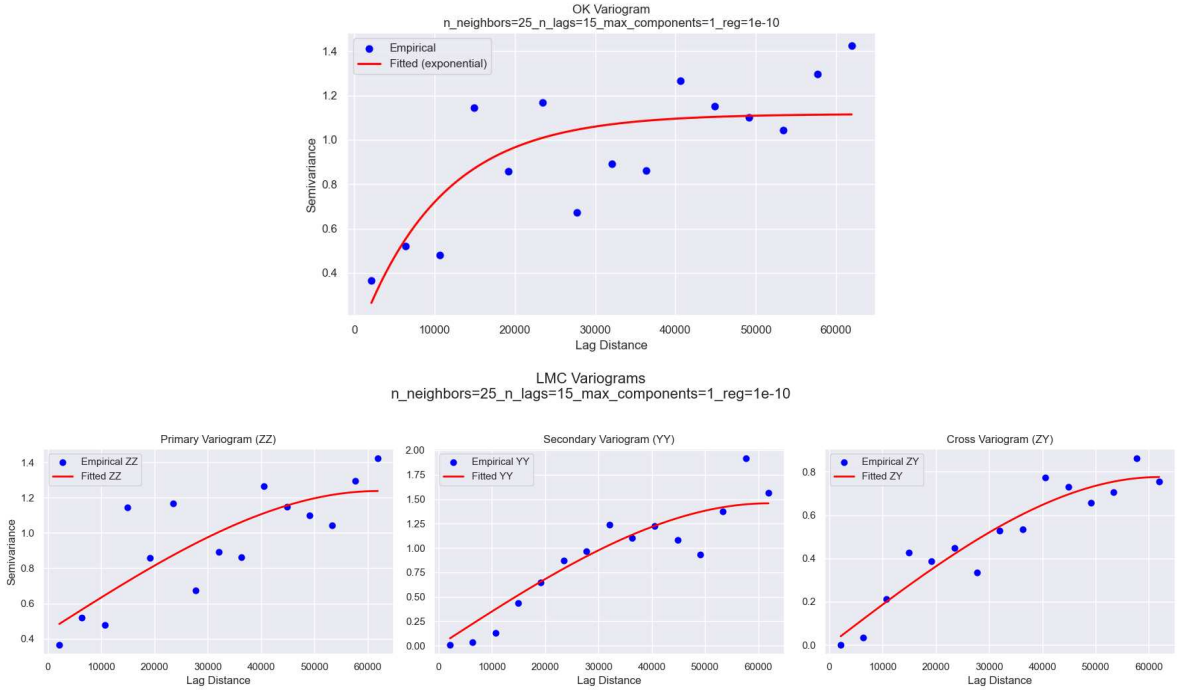


Figure B.28: OK (top) and LMC (bottom) variogram fits for ilr.28.

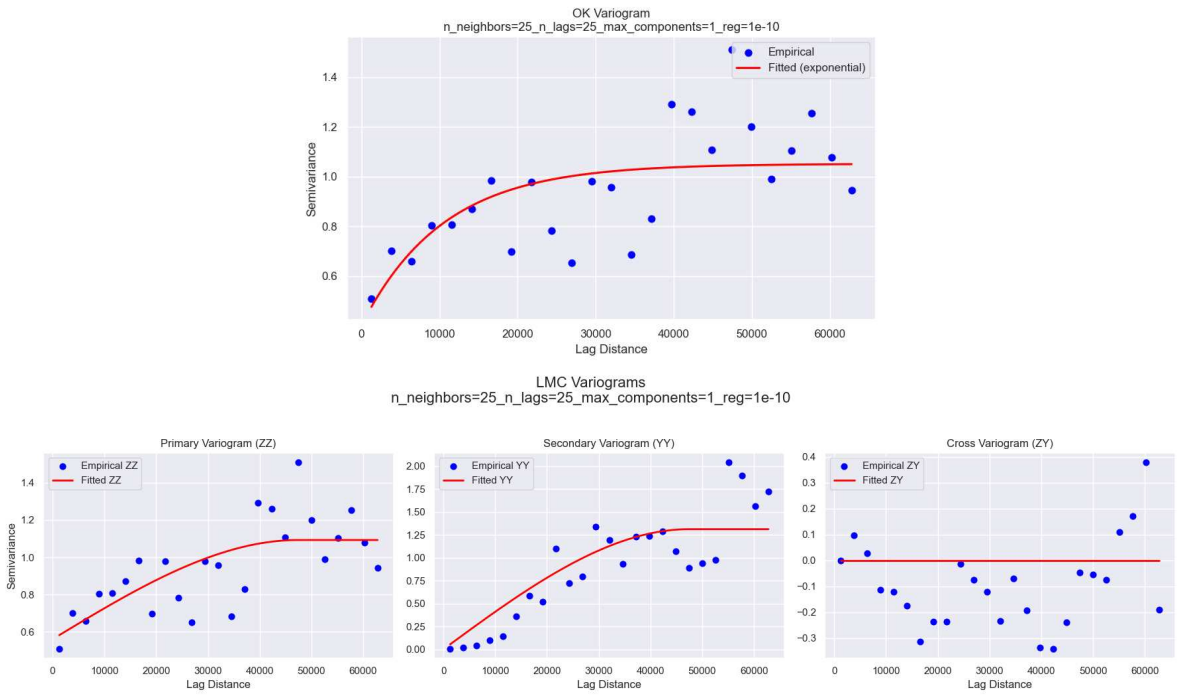


Figure B.29: OK (top) and LMC (bottom) variogram fits for ilr.29.

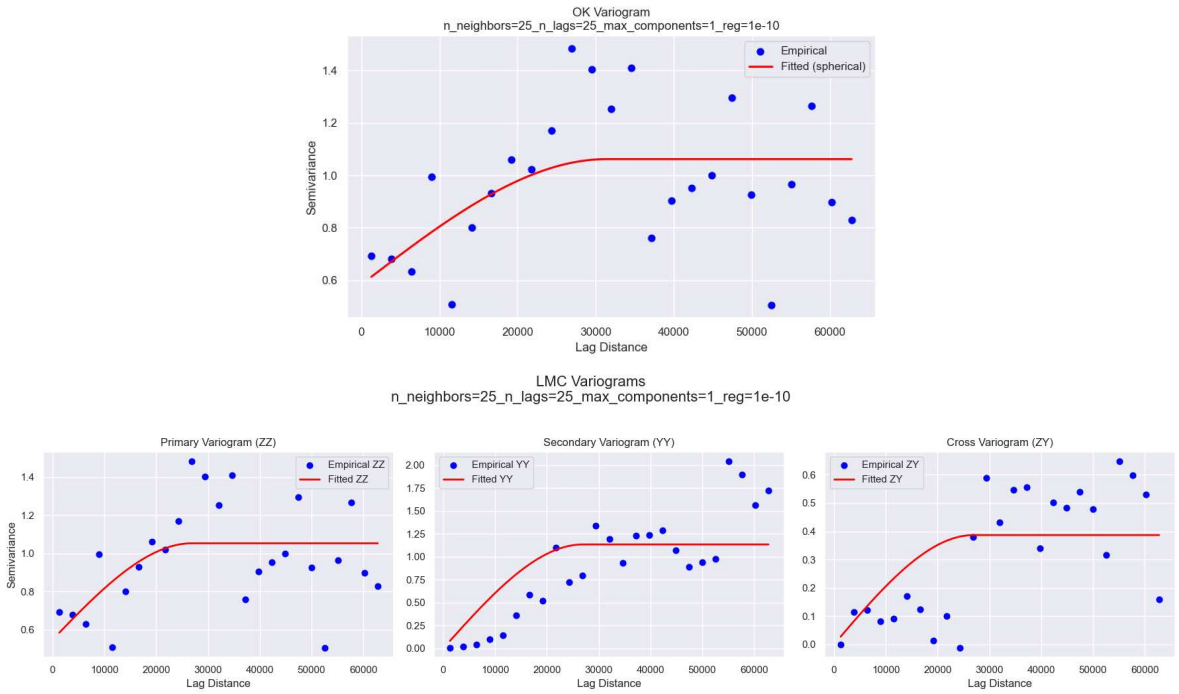


Figure B.30: OK (top) and LMC (bottom) variogram fits for ilr.30.

C

Block-wise estimate and variance plotsand other estimate plots

C.1. Estimate Plots of Light Rare Earth Elements (LREEs)

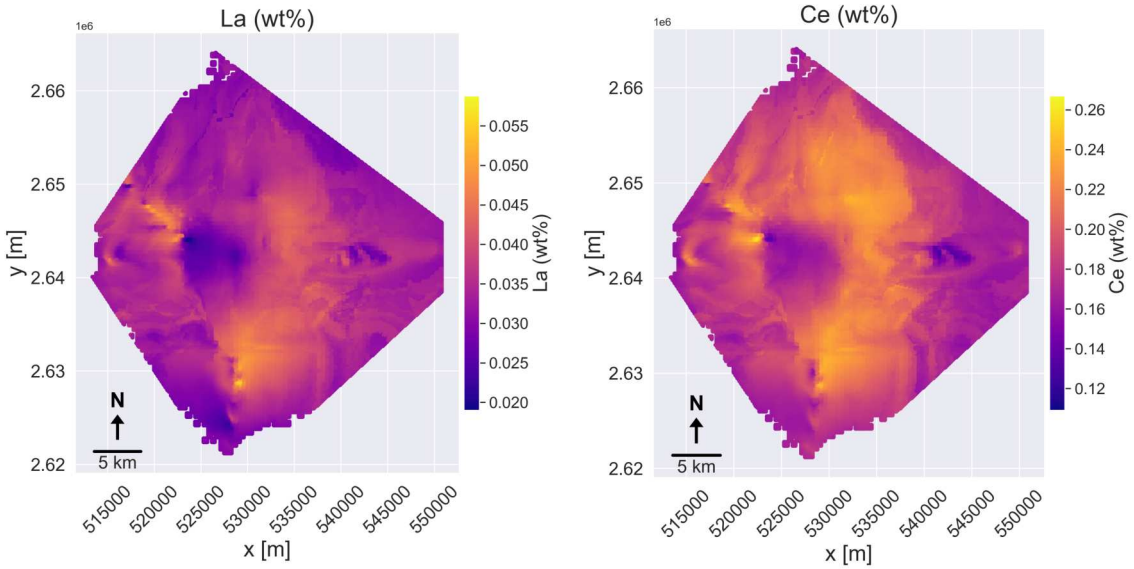


Figure C.1: Block-wise estimates of Lanthanum (left) and Cerium (right) (LREEs).

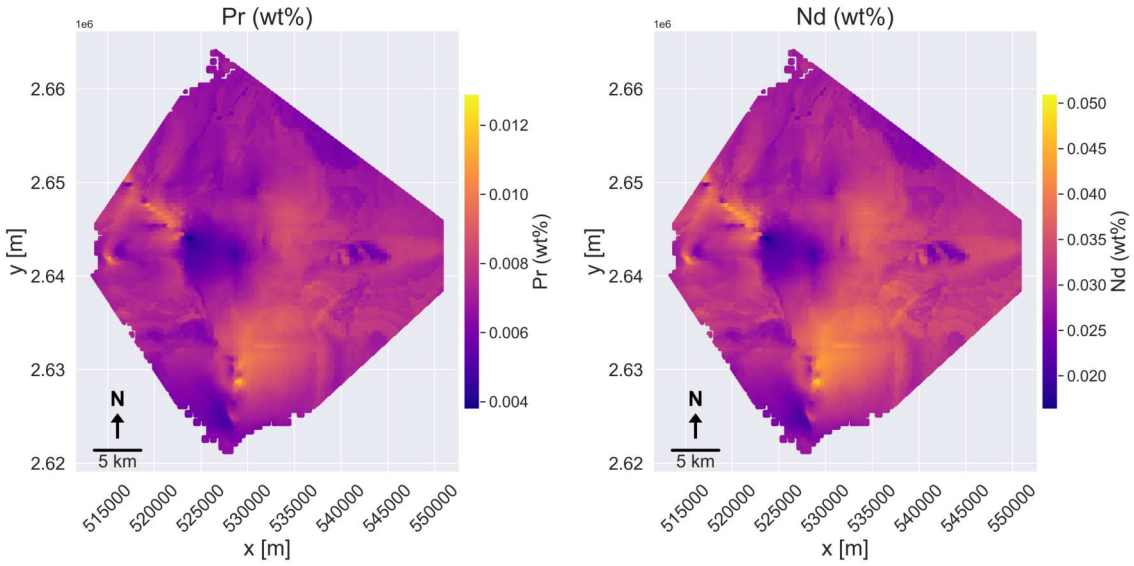


Figure C.2: Block-wise estimates of Praseodymium (left) and Neodymium (right) (LREEs).

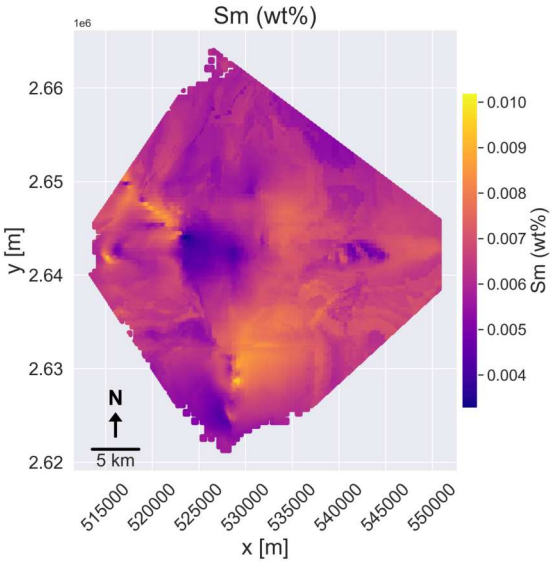


Figure C.3: Block-wise estimate of Samarium (LREE).

C.2. Variance Plots of Light Rare Earth Elements (LREEs)

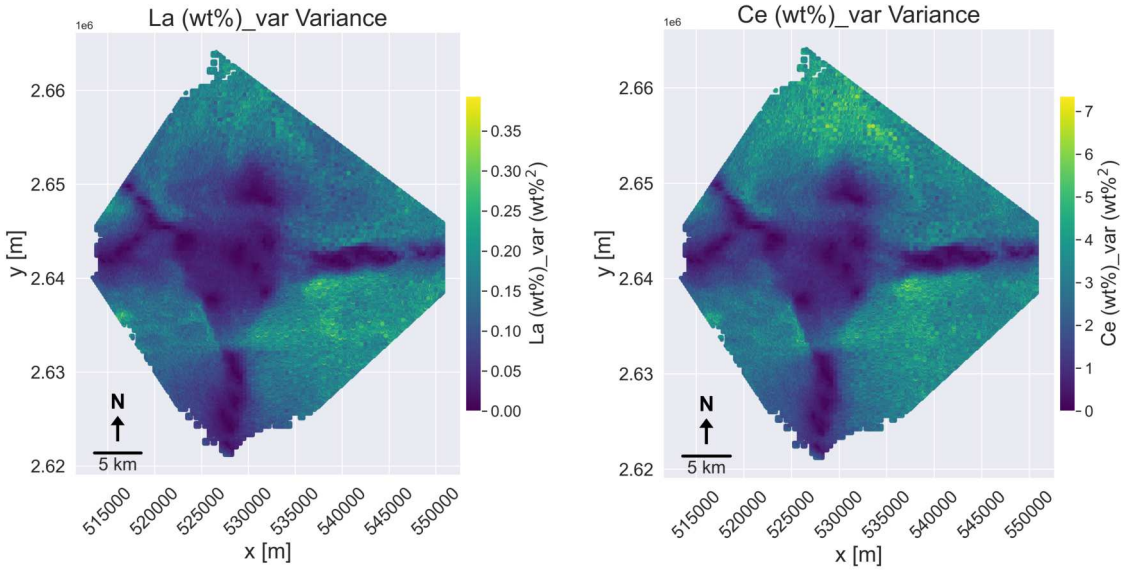


Figure C.4: Kriging variance of Lanthanum (left) and Cerium (right) (LREEs).

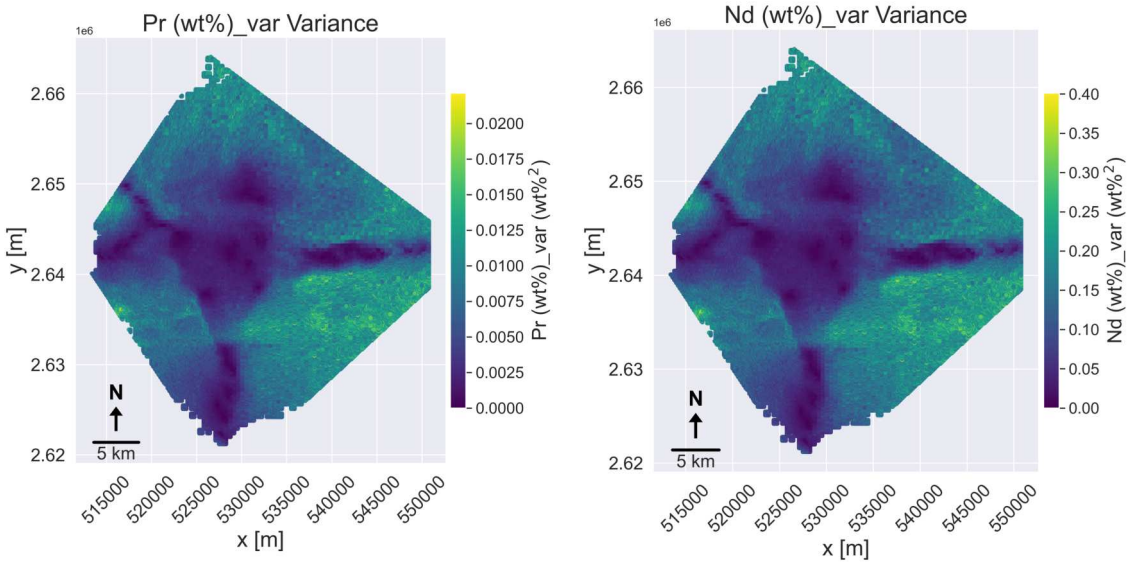


Figure C.5: Kriging variance of Praseodymium (left) and Neodymium (right) (LREEs).

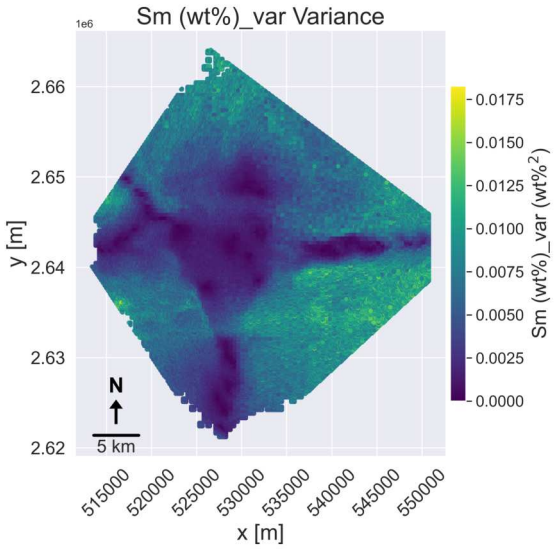


Figure C.6: Kriging variance of Samarium (LREE).

C.3. Estimate Plots of Heavy Rare Earth Elements (HREEs)

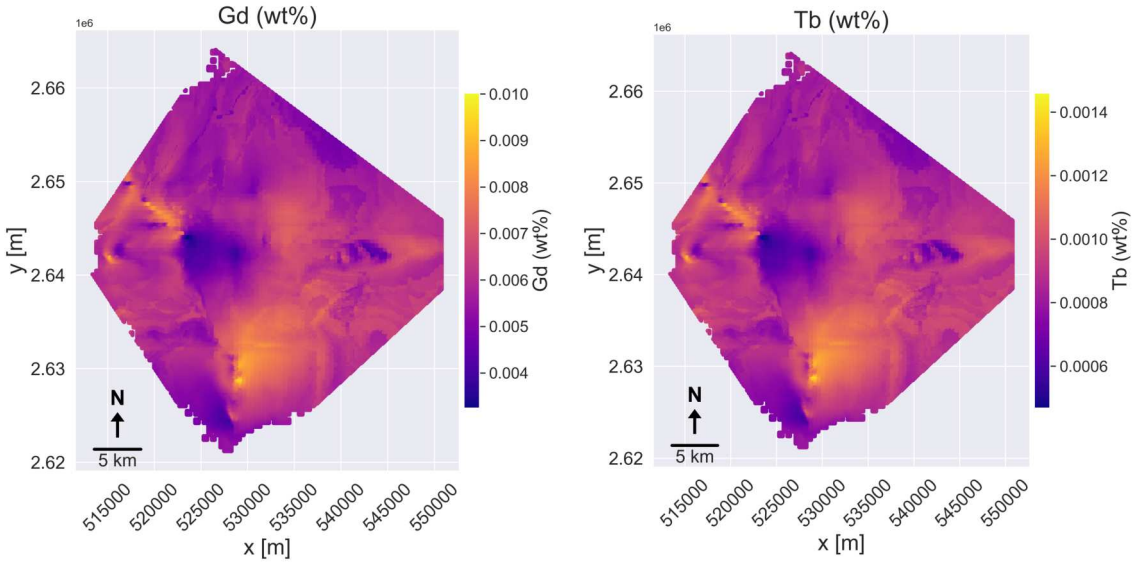


Figure C.7: Block-wise estimates of Gadolinium (left) and Terbium (right) (HREEs).

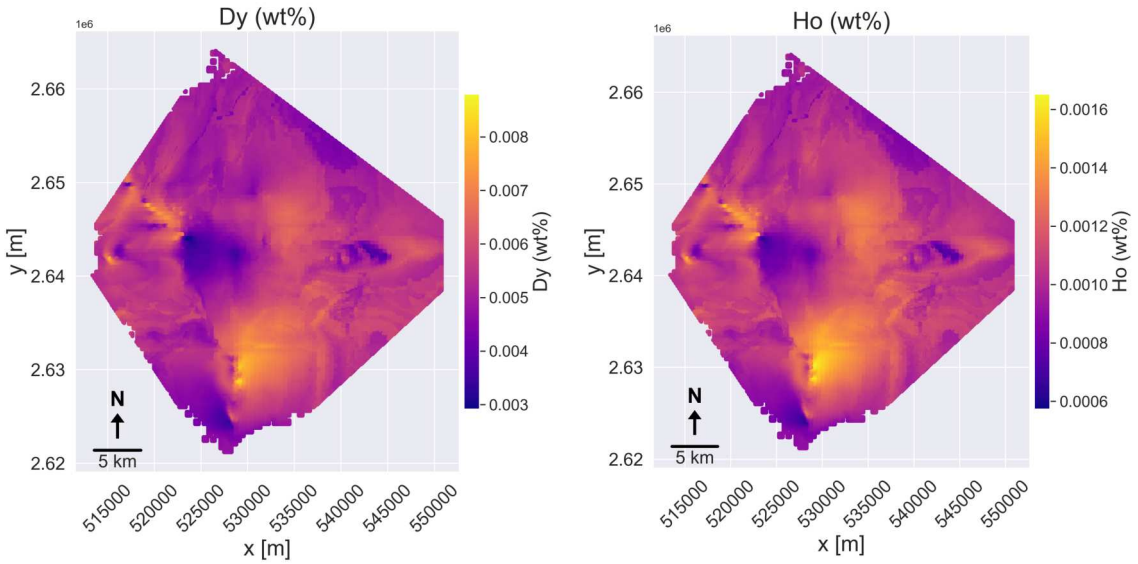


Figure C.8: Block-wise estimates of Dysprosium (left) and Holmium (right) (HREEs).

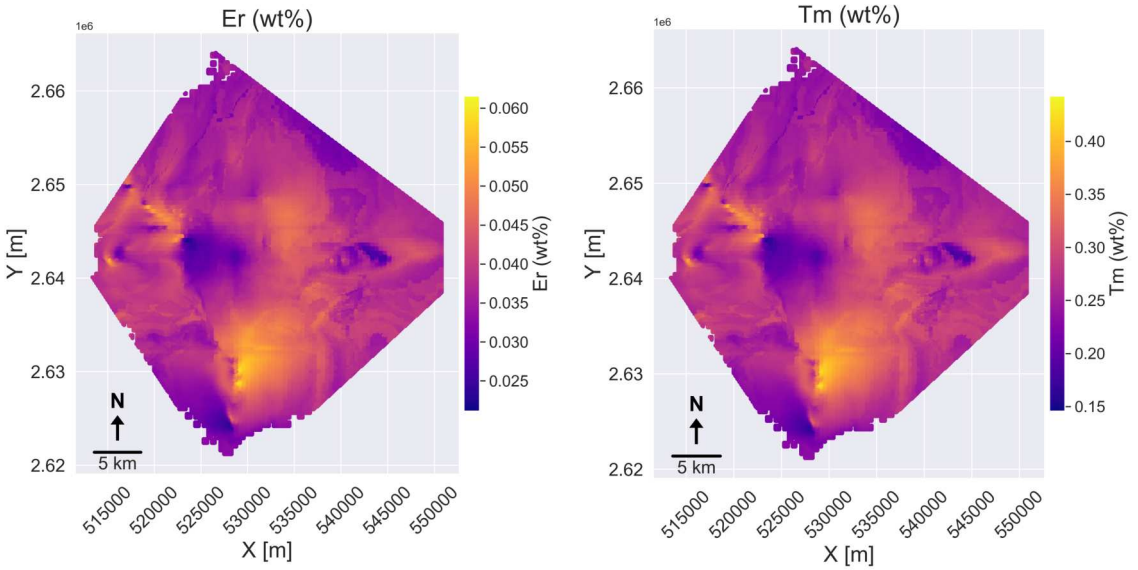


Figure C.9: Block-wise estimates of Erbium (left) and Thulium (right) (HREEs).

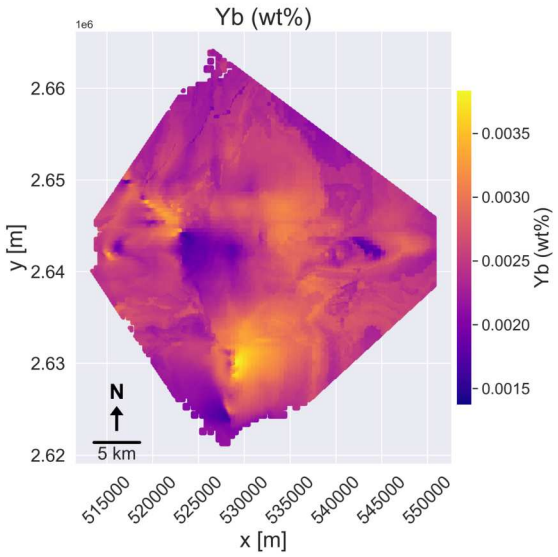


Figure C.10: Block-wise estimate of Ytterbium (HREE).

C.4. Variance Plots of Heavy Rare Earth Elements (HREEs)

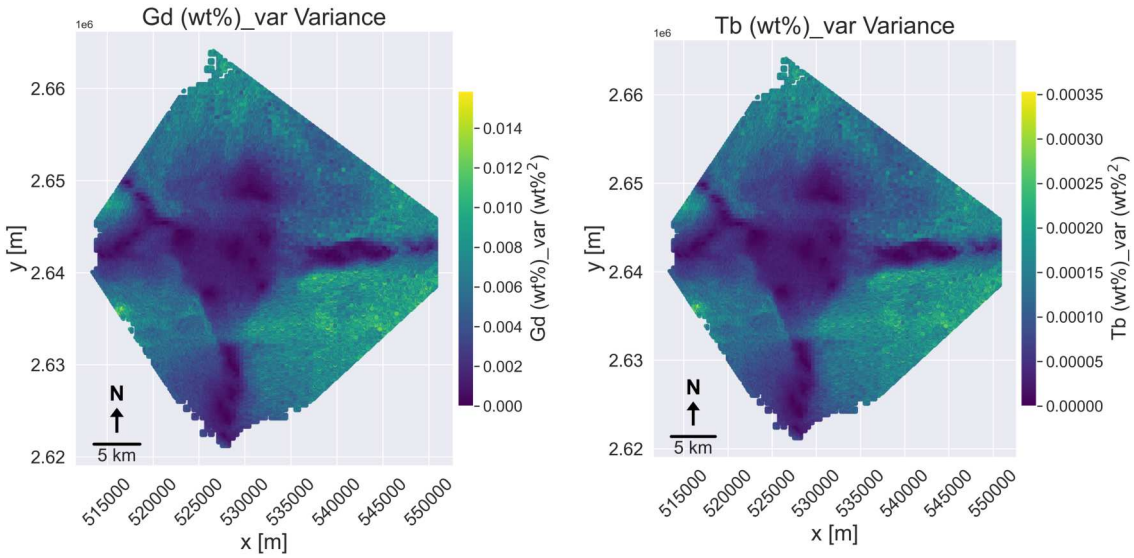


Figure C.11: Kriging variance of Gadolinium (left) and Terbium (right) (HREEs).

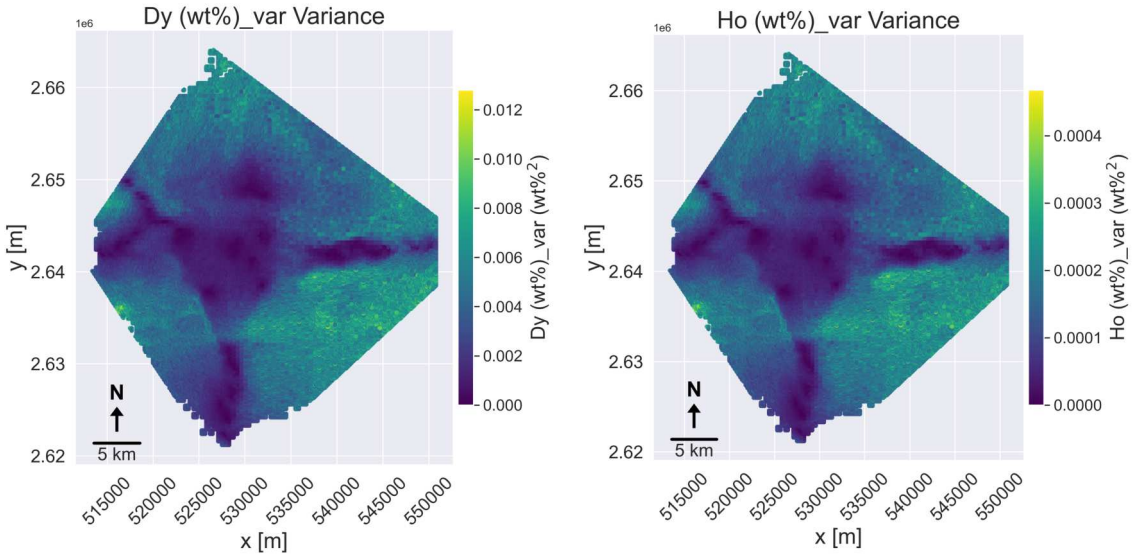


Figure C.12: Kriging variance of Dysprosium (left) and Holmium (right) (HREEs).

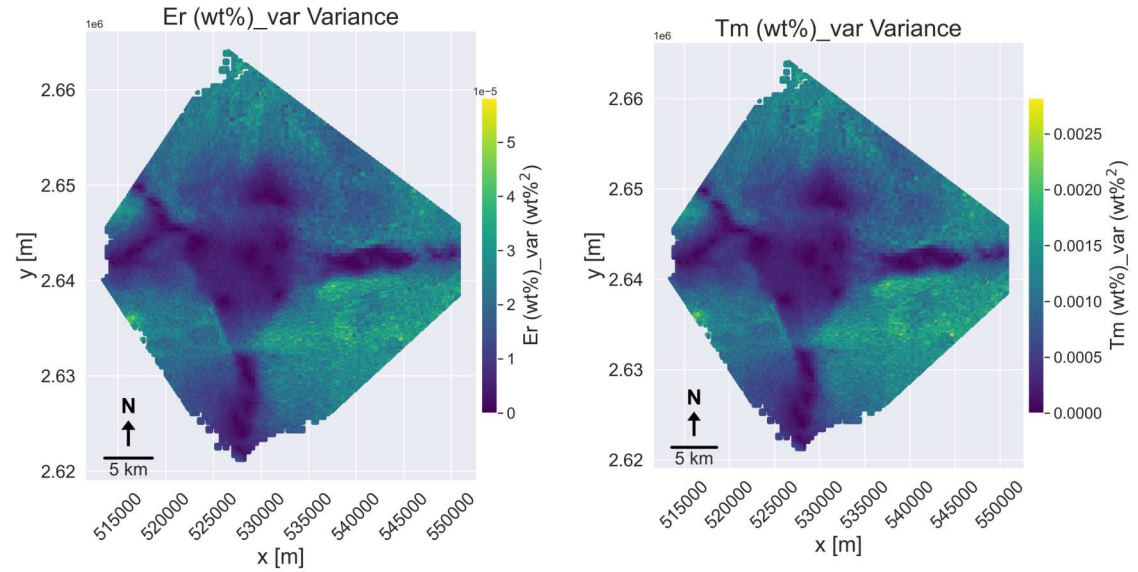


Figure C.13: Kriging variance of Erbium (left) and Thulium (right) (HREEs).

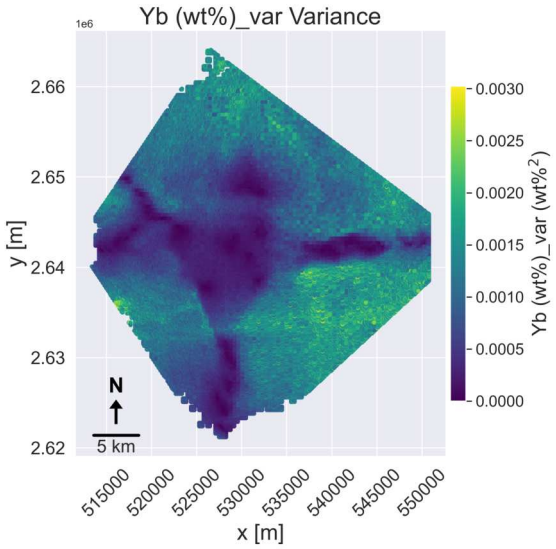


Figure C.14: Kriging variance of Ytterbium (HREE).

C.5. Estimate Plots of Non-CRM Elements

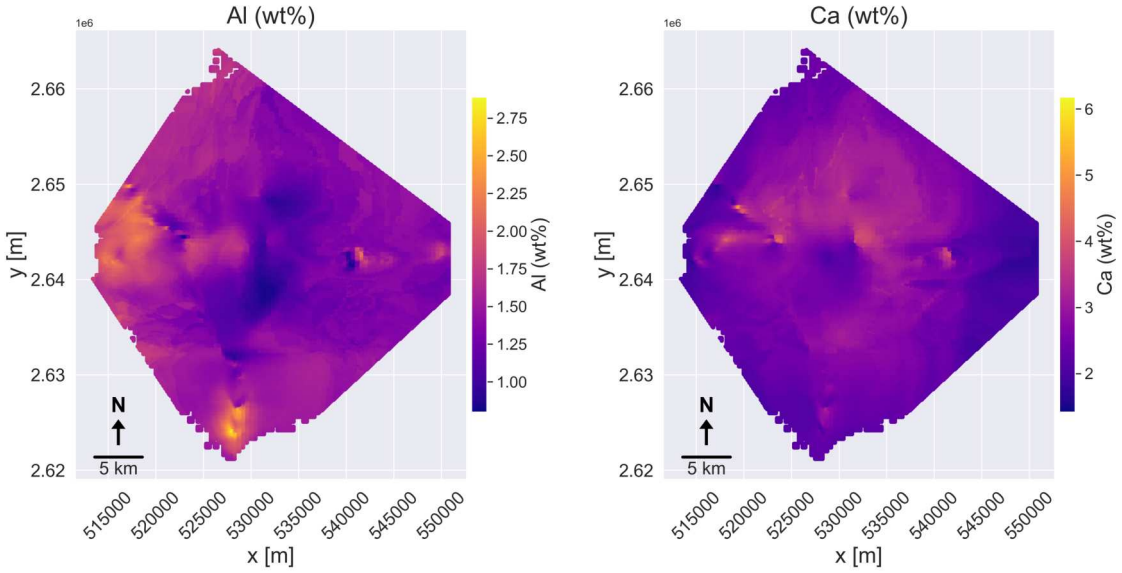


Figure C.15: Block-wise estimates of Aluminium (left) and Calcium (right).

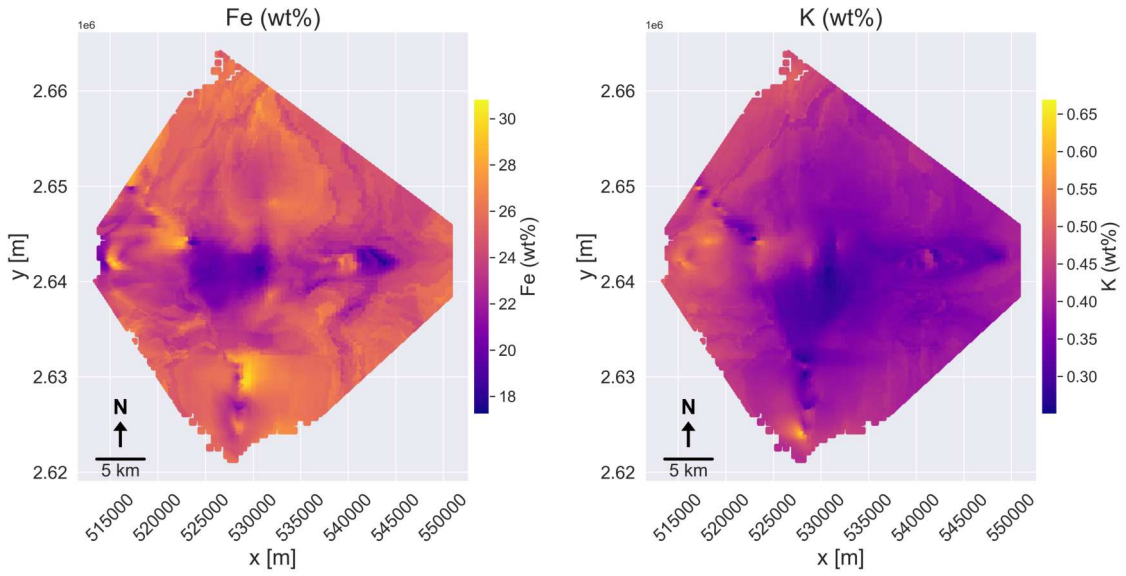


Figure C.16: Block-wise estimates of Iron (left) and Potassium (right).

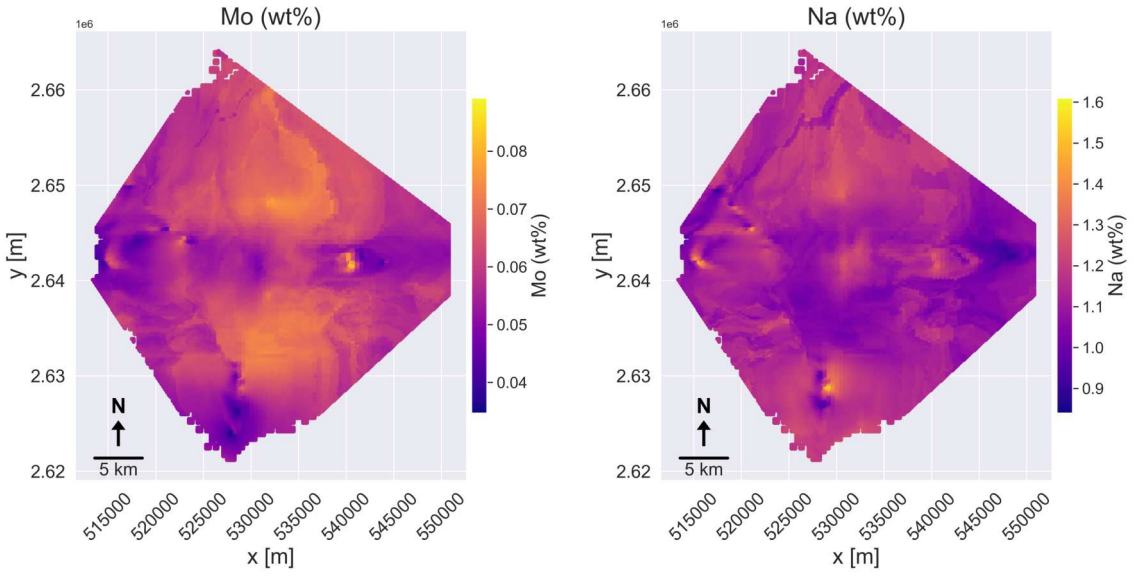


Figure C.17: Block-wise estimates of Molybdenum (left) and Sodium (right).

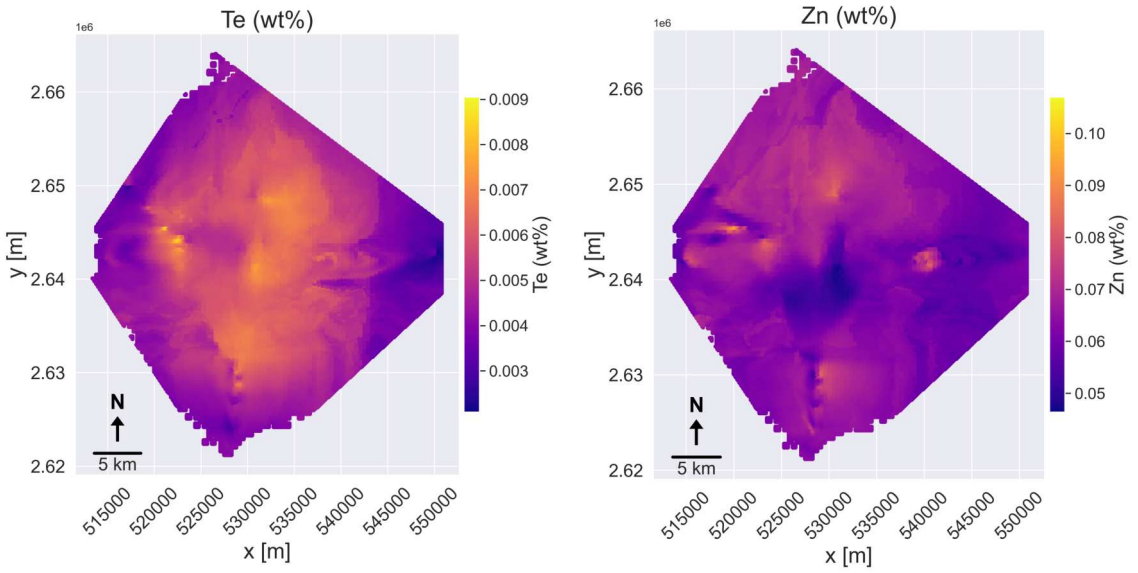


Figure C.18: Block-wise estimates of Tellurium (left) and Zinc (right).

C.6. Variance Plots of Non-CRM Elements

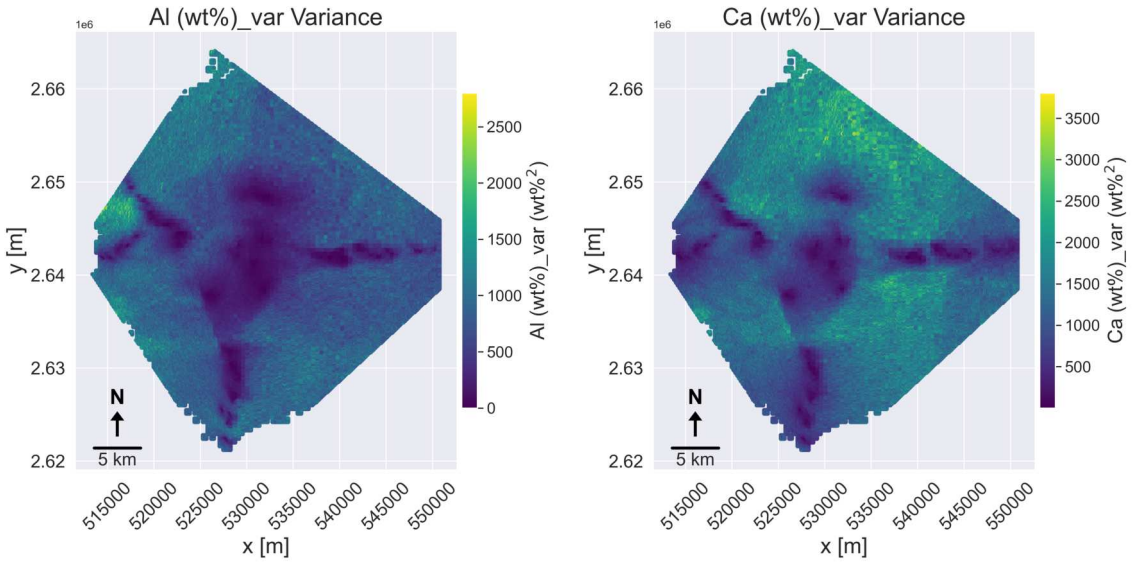


Figure C.19: Kriging variance of Aluminium (left) and Calcium (right).

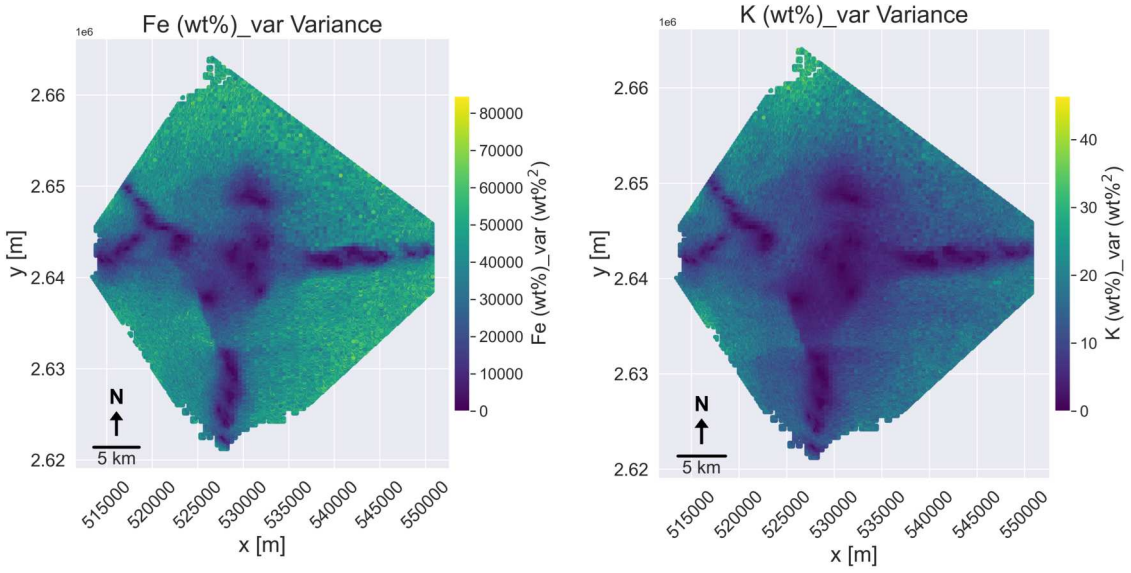


Figure C.20: Kriging variance of Iron (left) and Potassium (right).

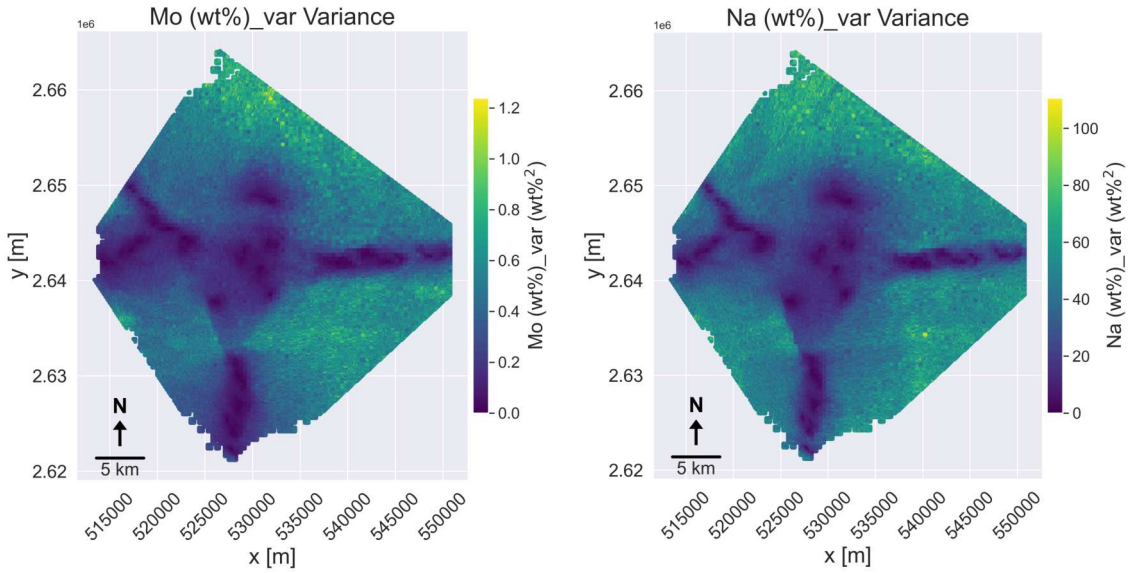


Figure C.21: Kriging variance of Molybdenum (left) and Sodium (right).

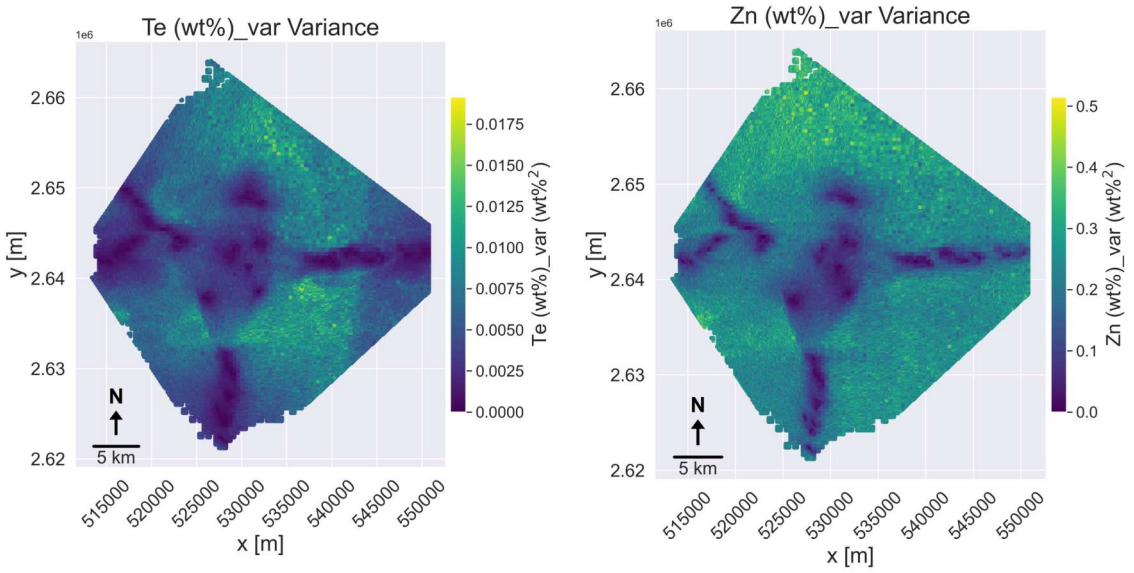


Figure C.22: Kriging variance of Tellurium (left) and Zinc (right).

C.7. Clustering

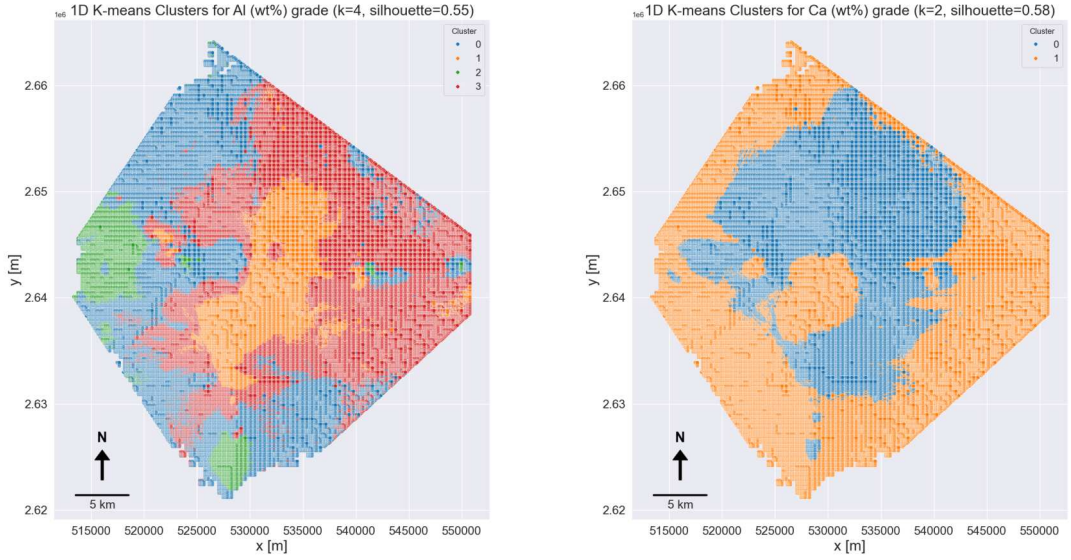


Figure C.23: 1D K-means cluster plots for Al (left) and Ca (right).

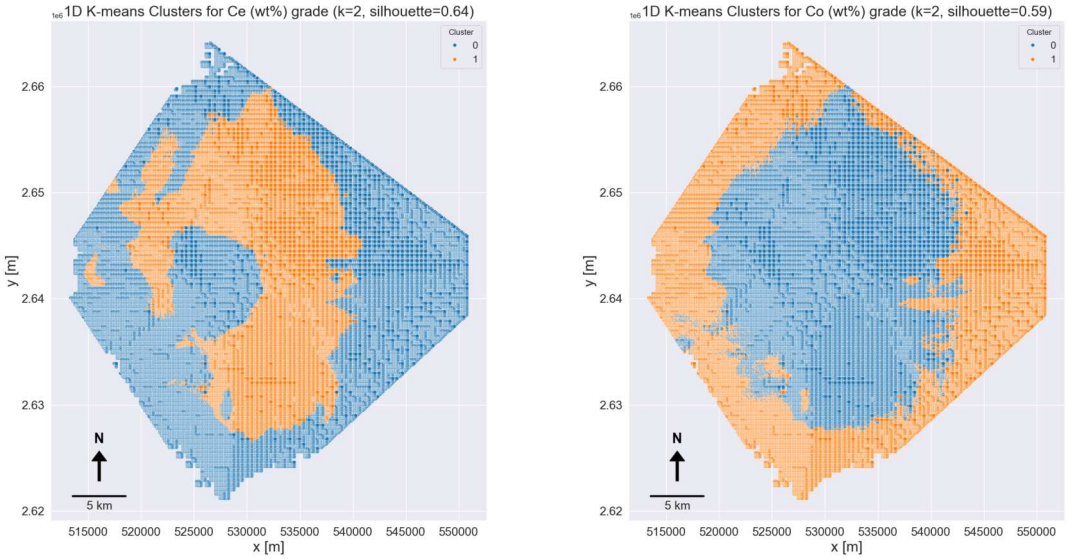


Figure C.24: 1D K-means cluster plots for Ce (left) and Co (right).

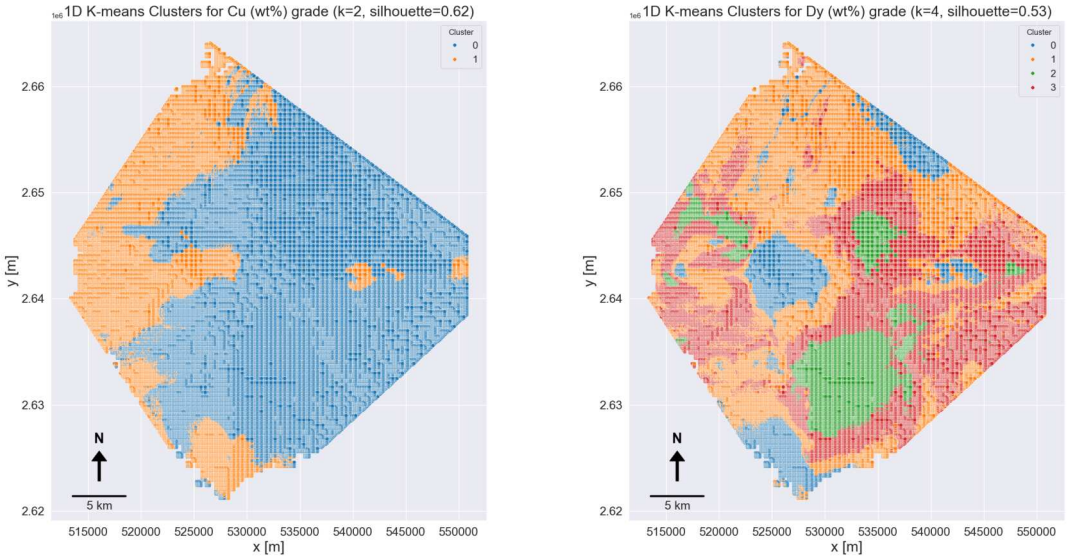


Figure C.25: 1D K-means cluster plots for Cu (left) and Dy (right).

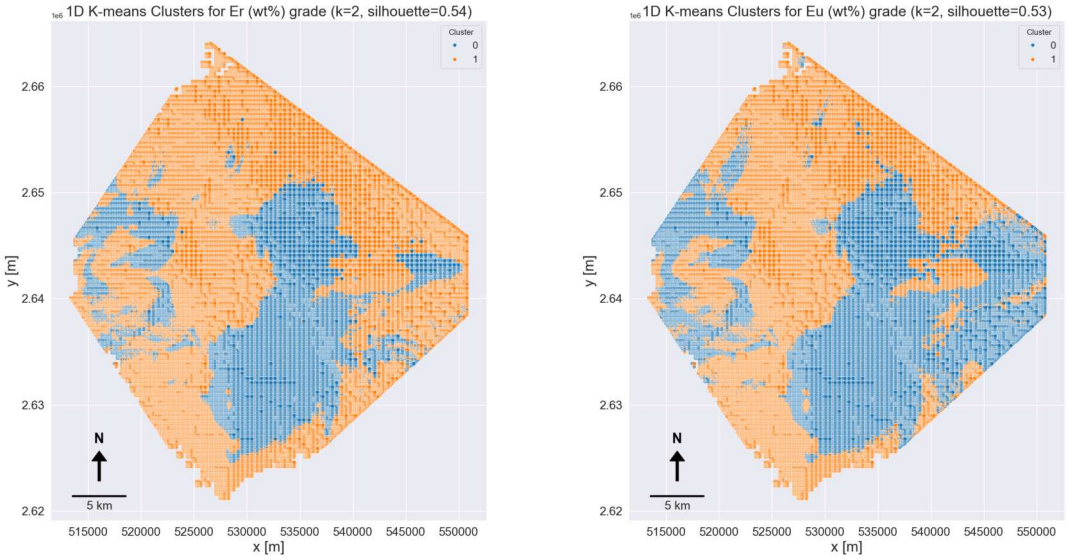


Figure C.26: 1D K-means cluster plots for Er (left) and Eu (right).

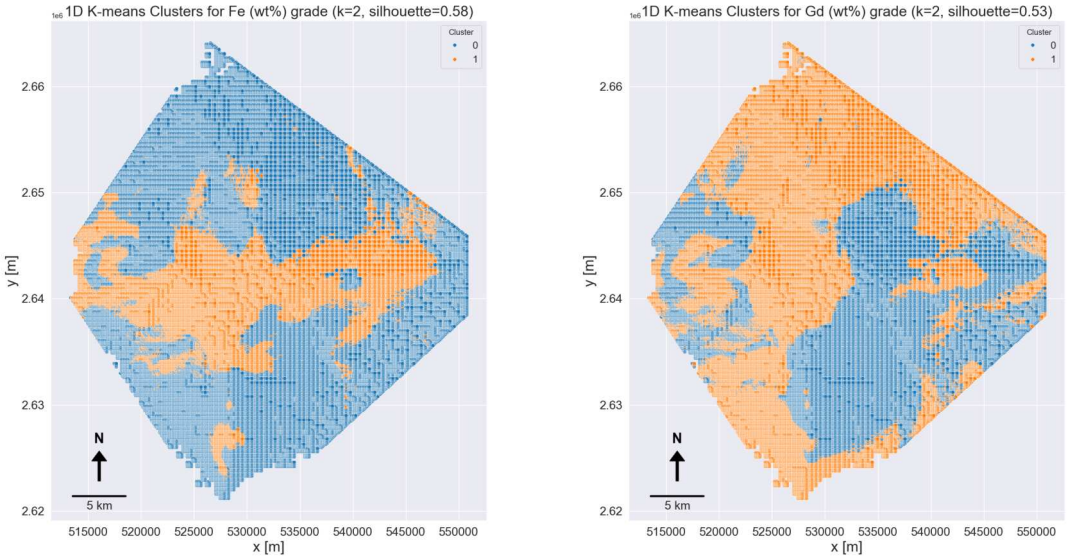


Figure C.27: 1D K-means cluster plots for Fe (left) and Gd (right).

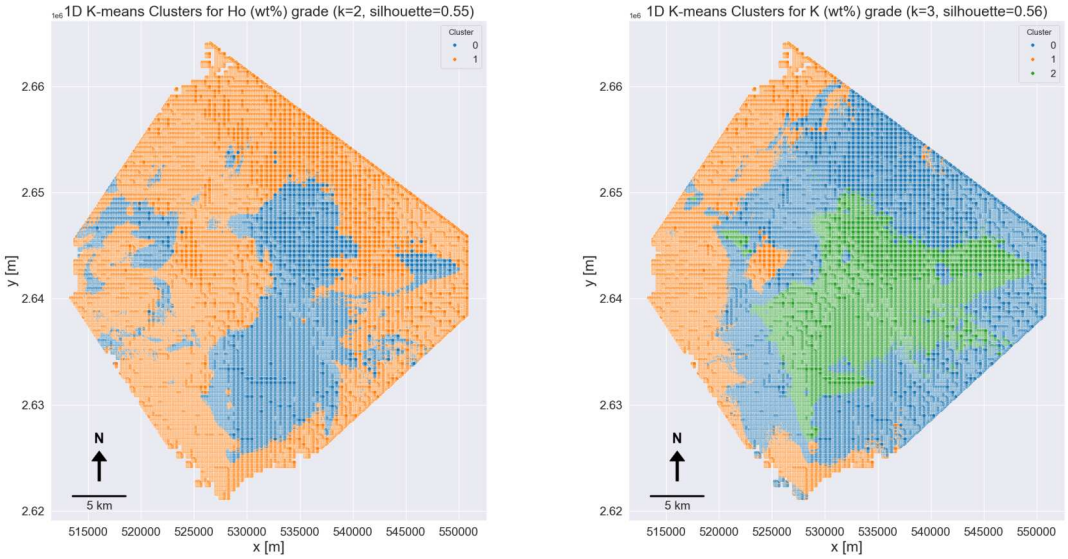


Figure C.28: 1D K-means cluster plots for Ho (left) and K (right).

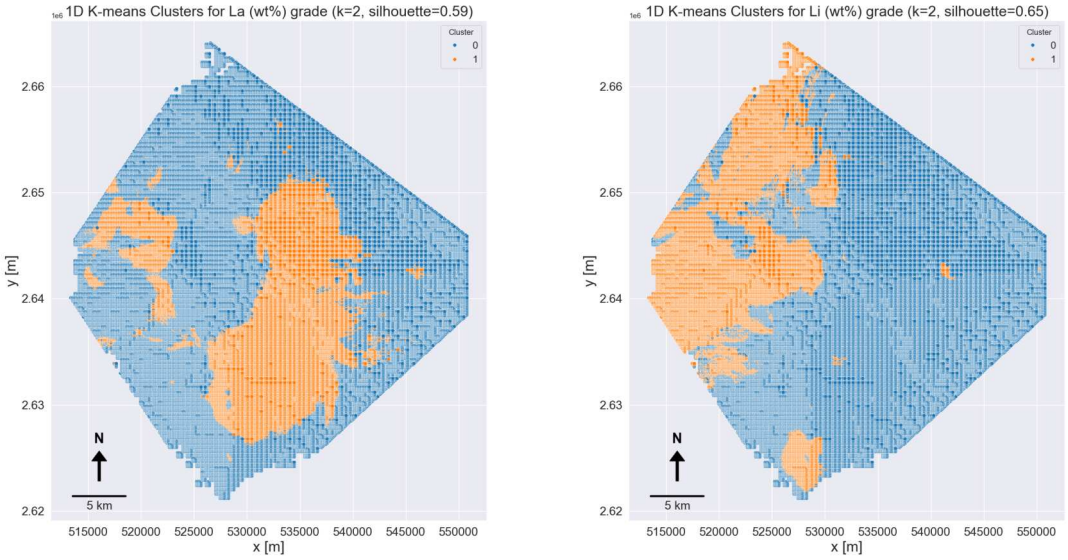


Figure C.29: 1D K-means cluster plots for La (left) and Li (right).

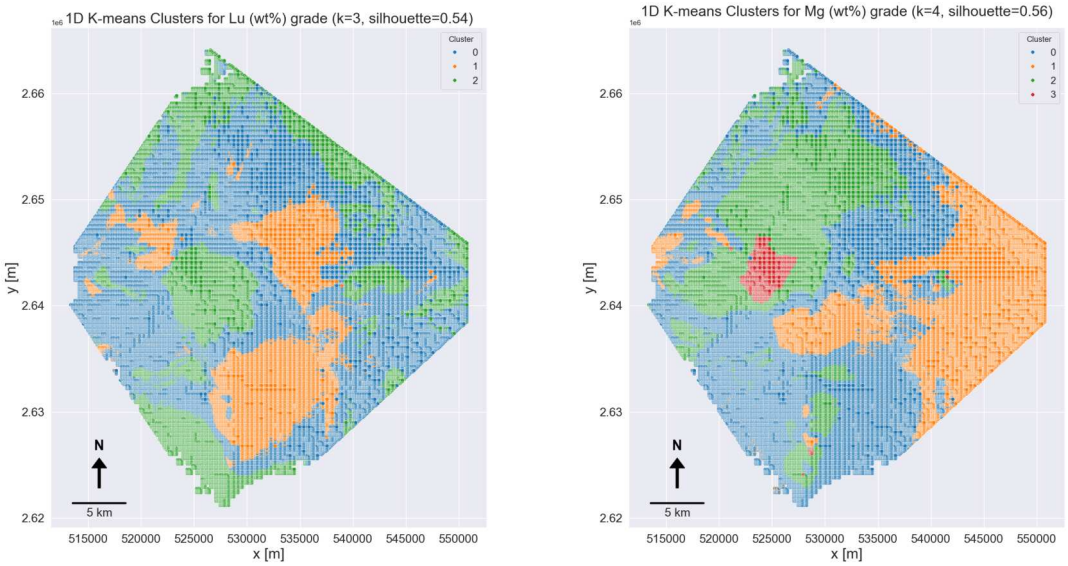


Figure C.30: 1D K-means cluster plots for Lu (left) and Mg (right).

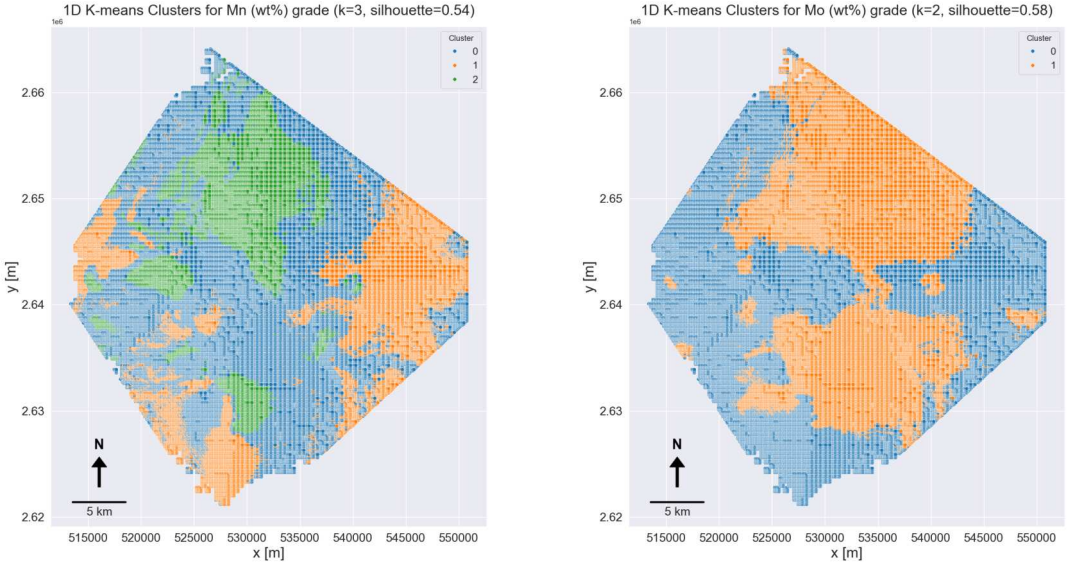


Figure C.31: 1D K-means cluster plots for Mn (left) and Mo (right).

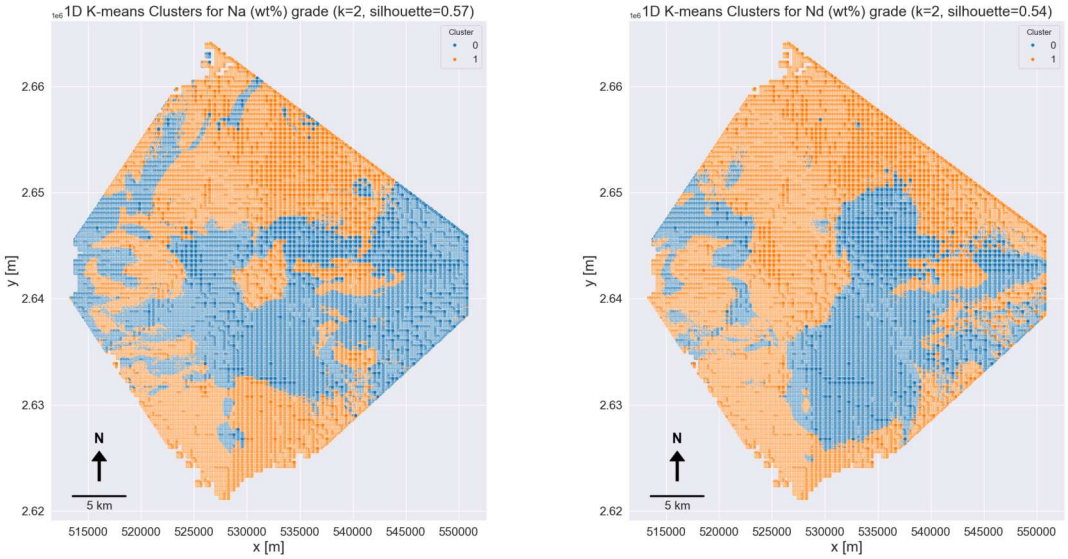


Figure C.32: 1D K-means cluster plots for Na (left) and Nd (right).

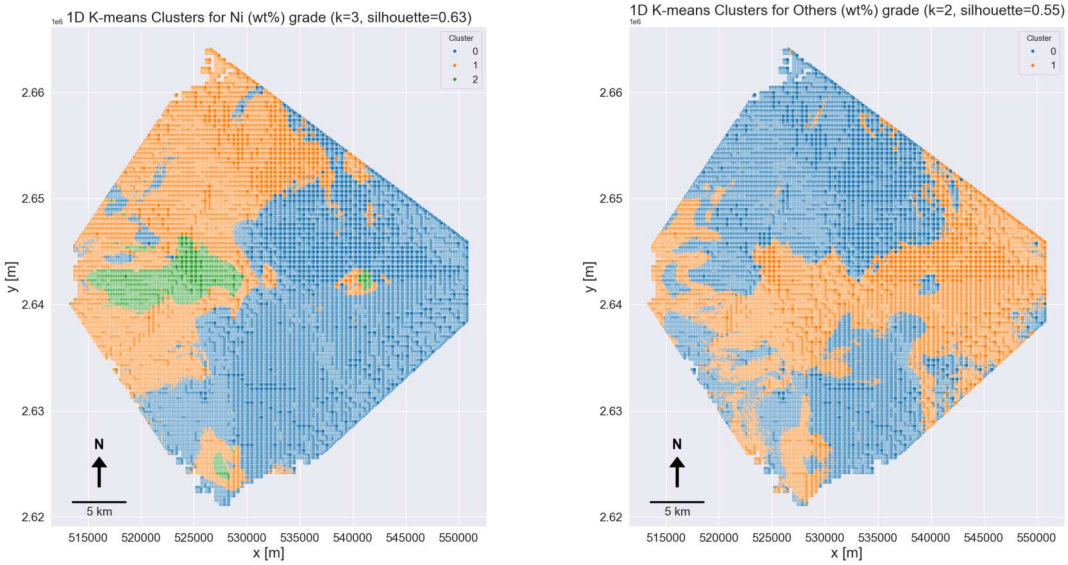


Figure C.33: 1D K-means cluster plots for Ni (left) and Others (right).

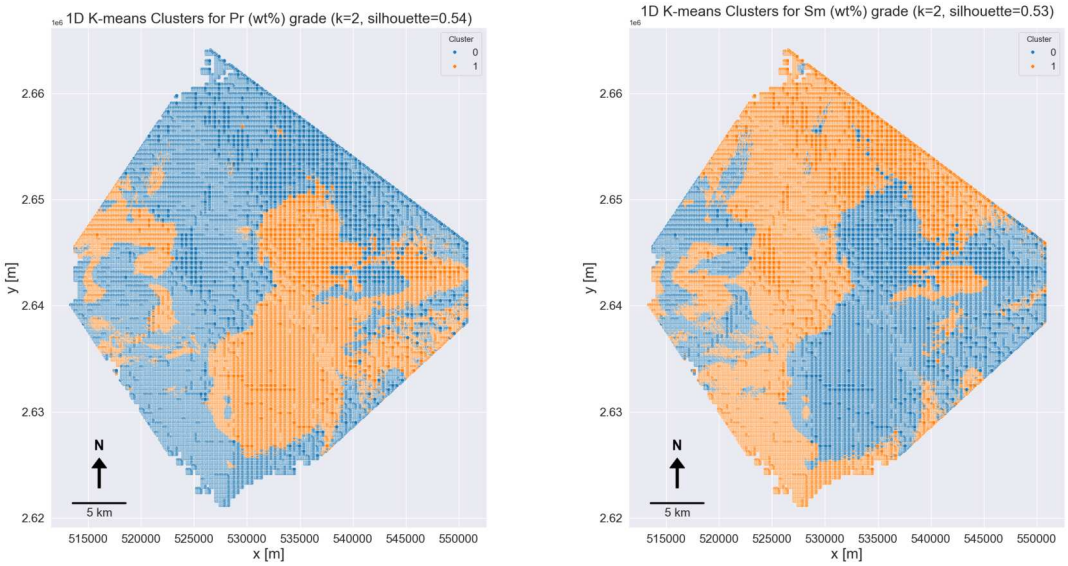


Figure C.34: 1D K-means cluster plots for Pr (left) and Sm (right).

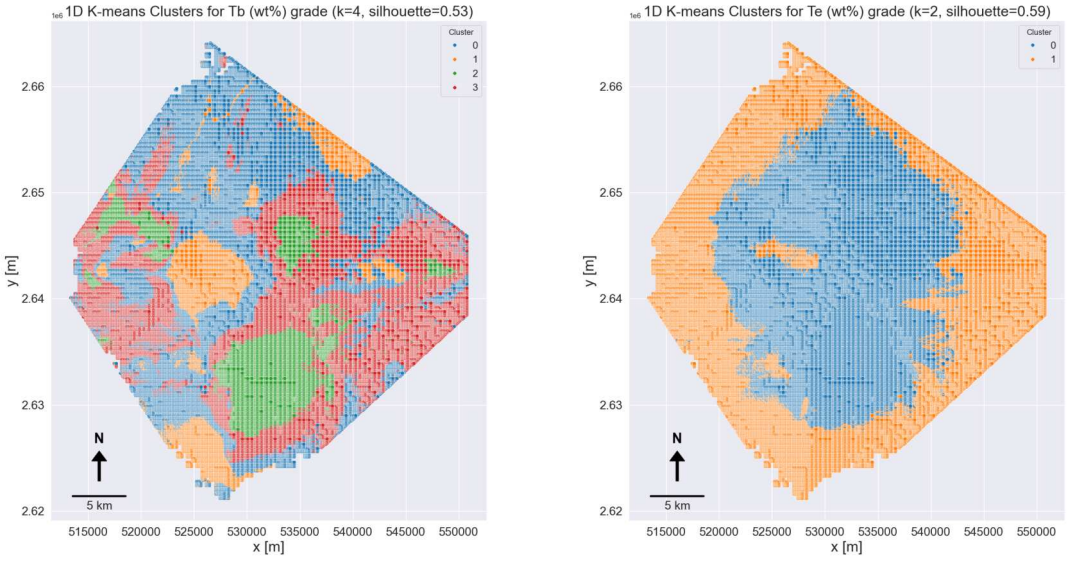


Figure C.35: 1D K-means cluster plots for Tb (left) and Te (right).

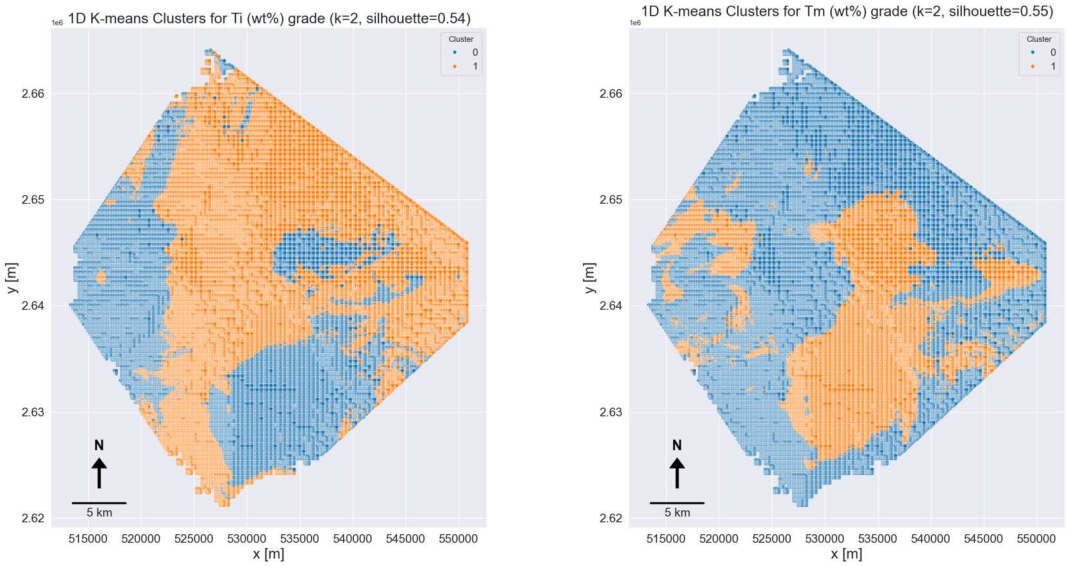


Figure C.36: 1D K-means cluster plots for Ti (left) and Tm (right).

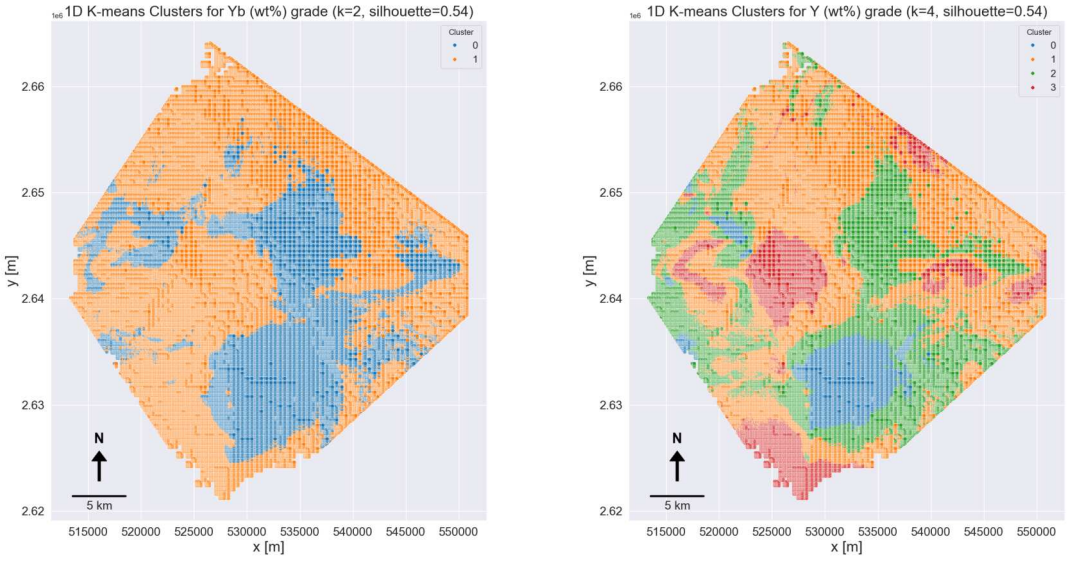


Figure C.37: 1D K-means cluster plots for Yb (left) and Y (right).

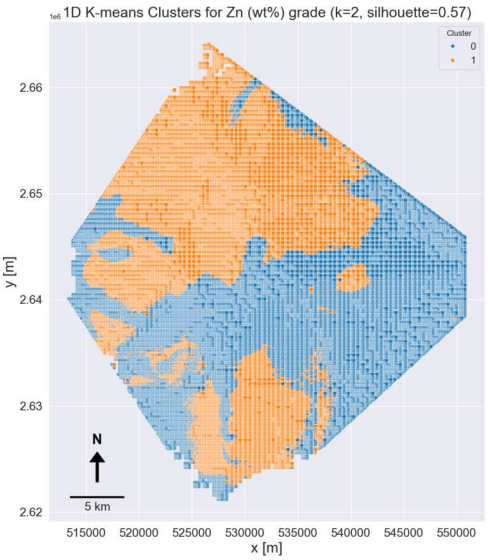


Figure C.38: 1D K-means cluster plot for Zn.

C.8. Grade tonnage curves

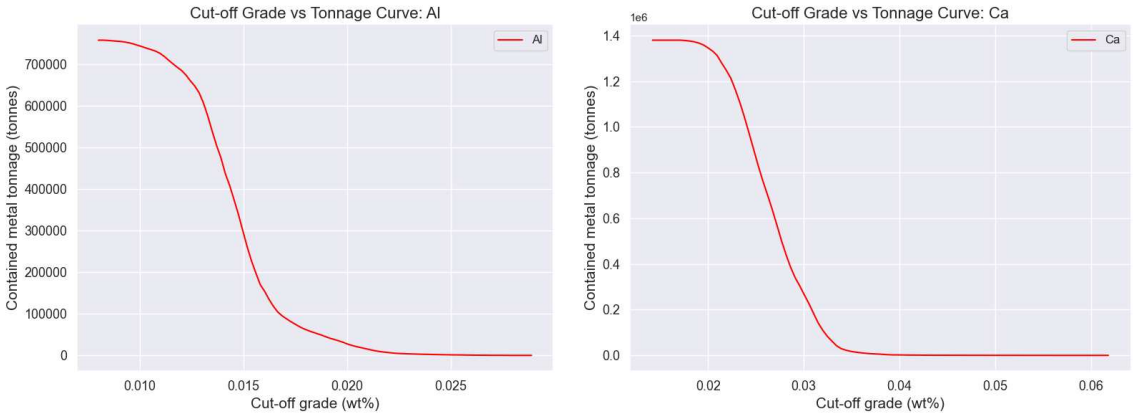


Figure C.39: Grade-tonnage curves for Al (left) and Ca (right).

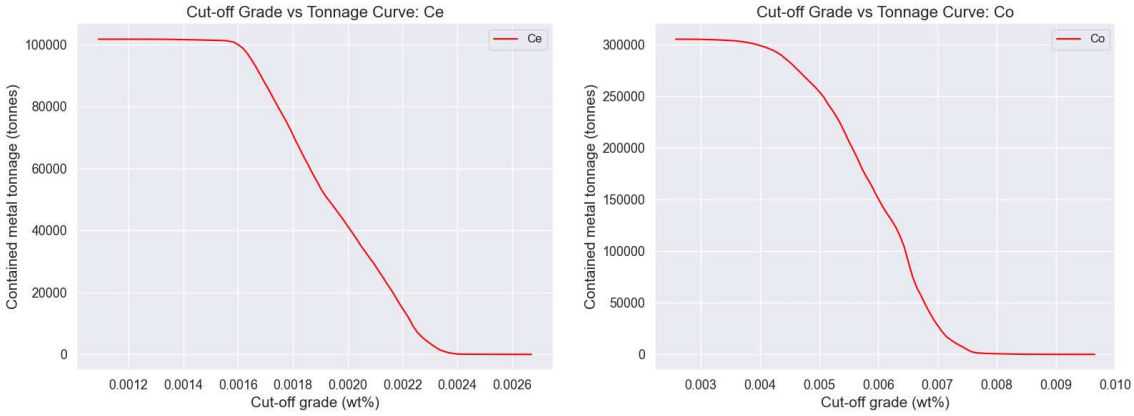


Figure C.40: Grade-tonnage curves for Ce (left) and Co (right).

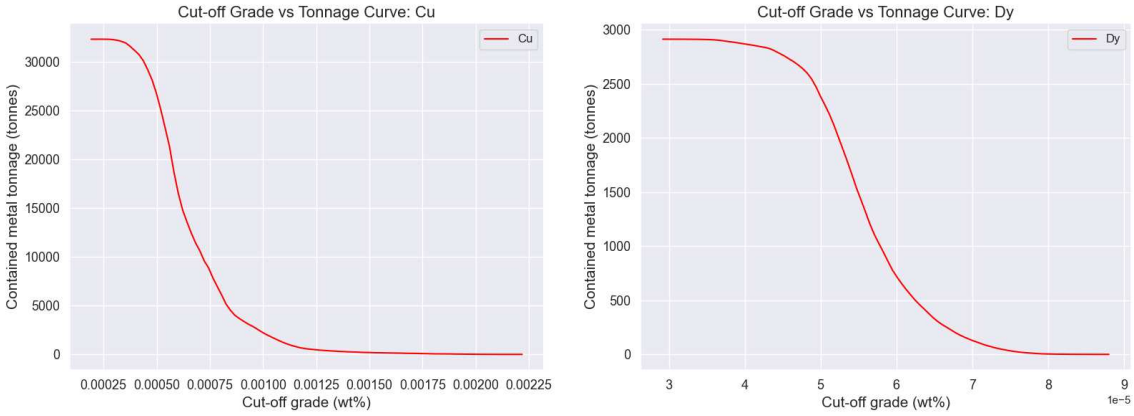


Figure C.41: Grade-tonnage curves for Cu (left) and Dy (right).

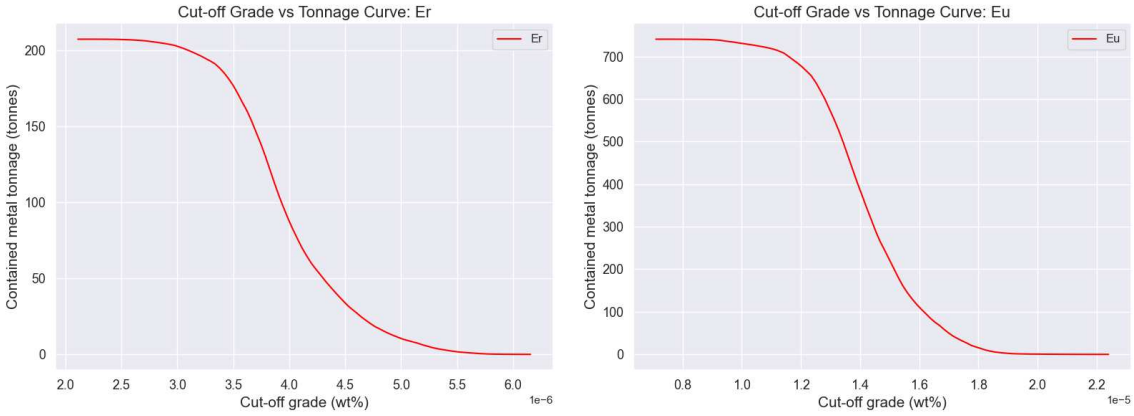


Figure C.42: Grade-tonnage curves for Er (left) and Eu (right).

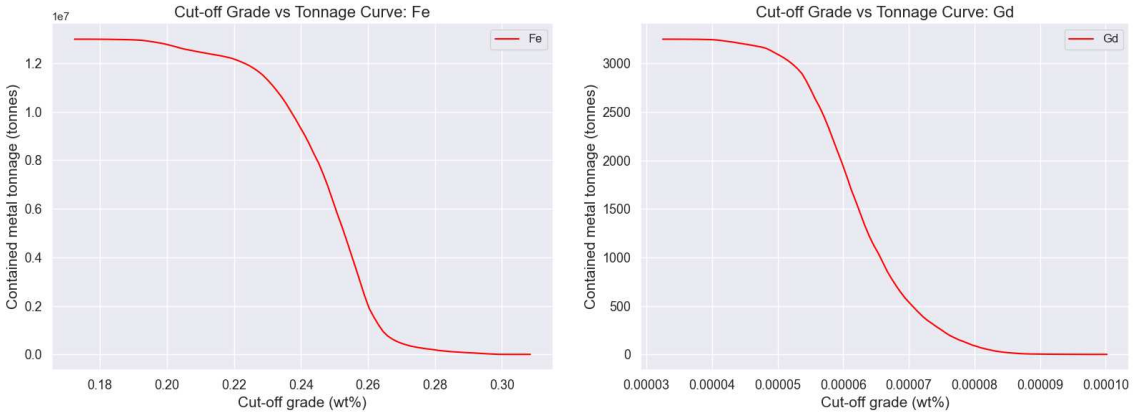


Figure C.43: Grade-tonnage curves for Fe (left) and Gd (right).

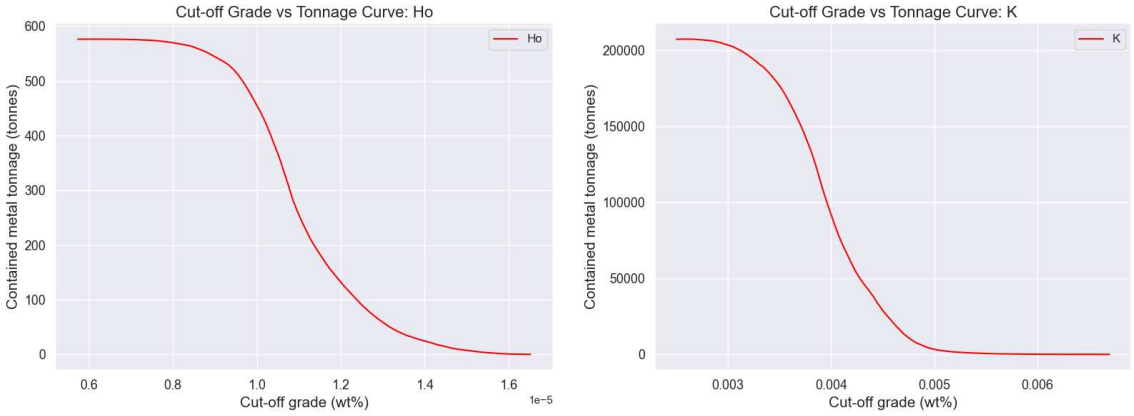


Figure C.44: Grade-tonnage curves for Ho (left) and K (right).

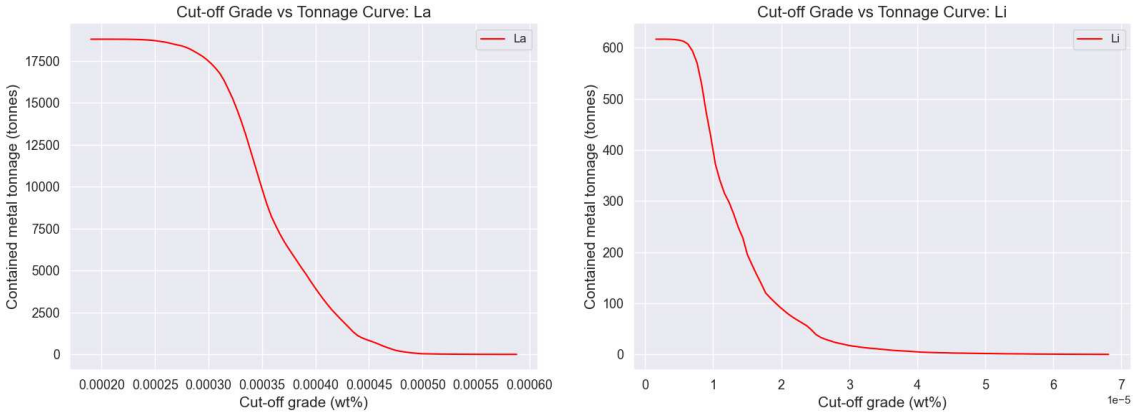


Figure C.45: Grade-tonnage curves for La (left) and Li (right).

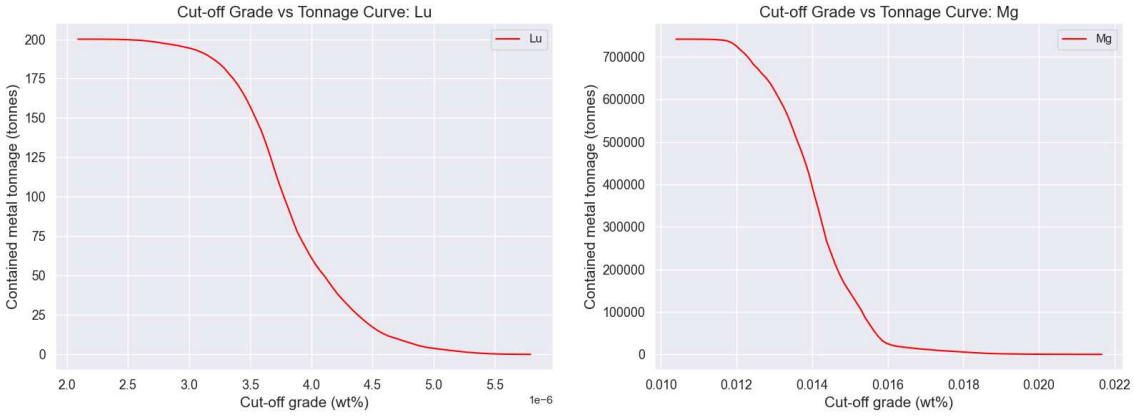


Figure C.46: Grade-tonnage curves for Lu (left) and Mg (right).

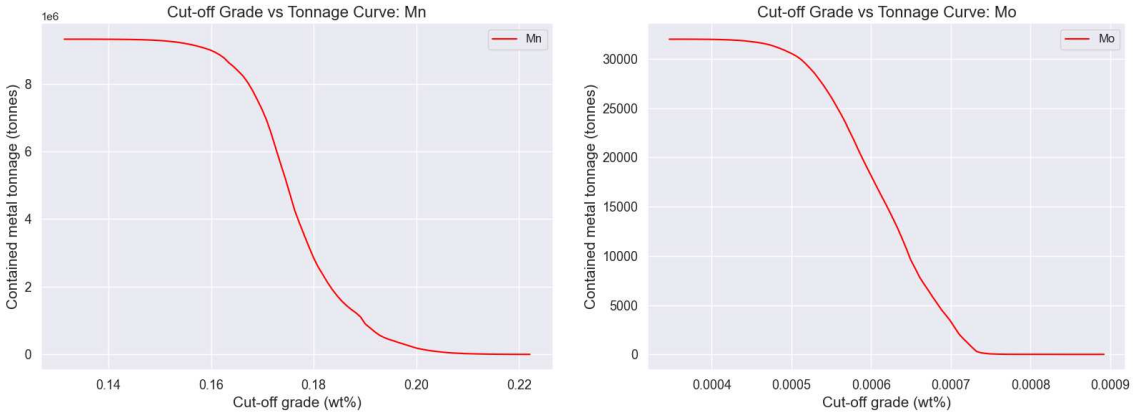


Figure C.47: Grade-tonnage curves for Mn (left) and Mo (right).

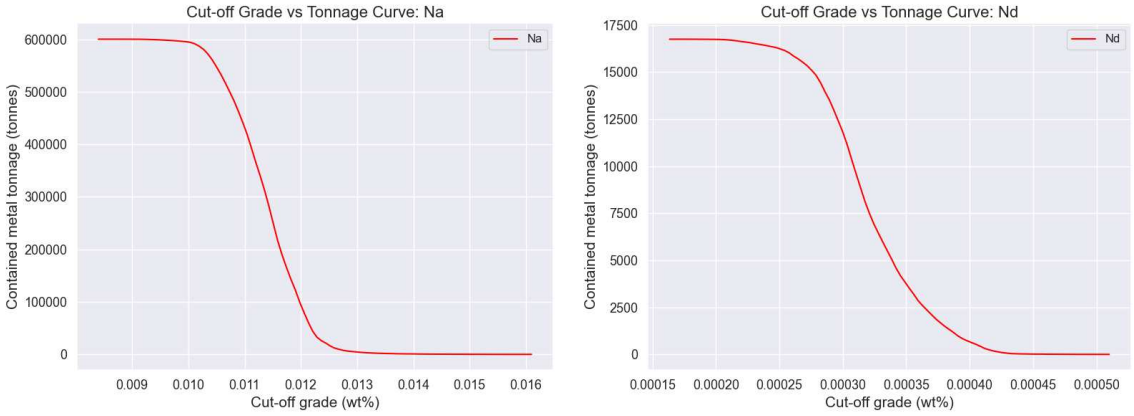


Figure C.48: Grade-tonnage curves for Na (left) and Nd (right).

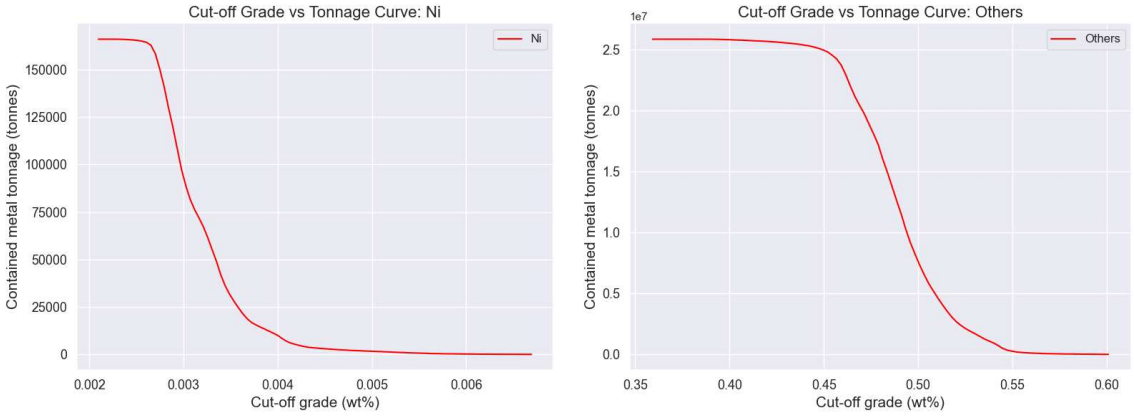


Figure C.49: Grade-tonnage curves for Ni (left) and Others (right).

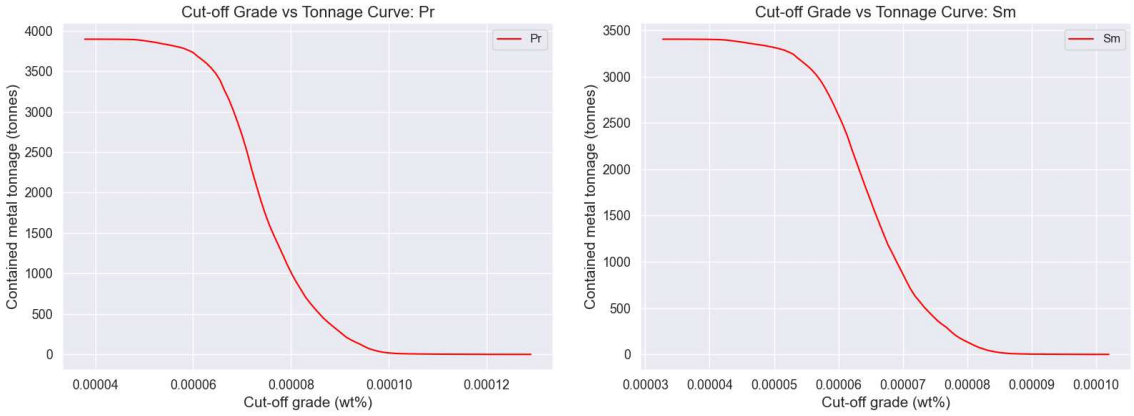


Figure C.50: Grade-tonnage curves for Pr (left) and Sm (right).

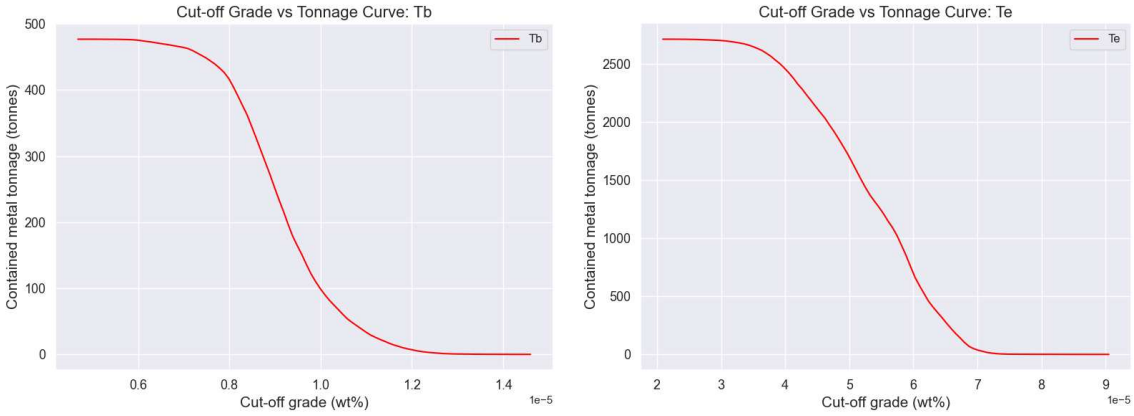


Figure C.51: Grade-tonnage curves for Tb (left) and Te (right).

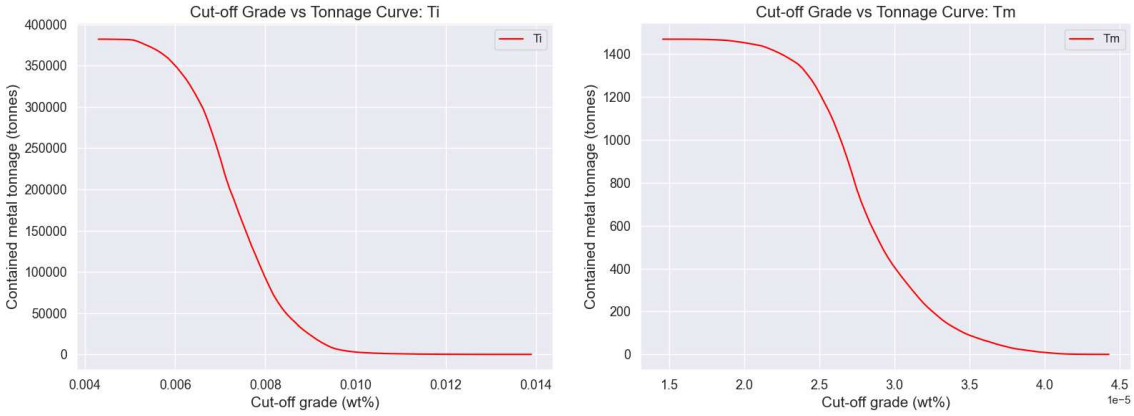


Figure C.52: Grade-tonnage curves for Ti (left) and Tm (right).

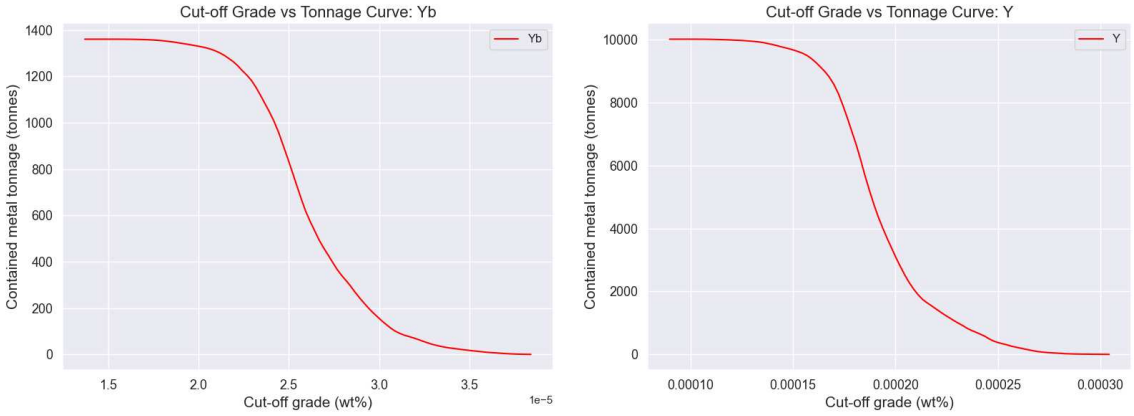


Figure C.53: Grade-tonnage curves for Yb (left) and Y (right).

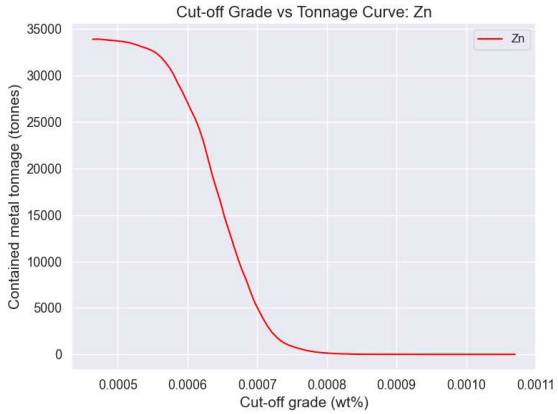


Figure C.54: Grade-tonnage curve for Zn.

D

Data

Table D.1: Overview of data.

Type of data	**File format**	**How will the data be collected (for re-used data: source and terms of use)**	**Purpose**	**Storage location**	**Who will have access to the data**
Ocean floor topography	.csv, .dxf	Provided by supervisor, collected from a PhD thesis by Sarah Atinta Howarth titled "An Investigation into the Variability of Ferromanganese Crusts in the NE Atlantic" (Howarth, 2022)	To map the ocean floor and understand its topography for exploration purposes	Project storage drive	The project team (The research student and supervisors)
Geochemical data (chemical composition and crust thickness)	.csv	Provided by supervisor, collected from the same PhD thesis	To analyze the chemical composition and crust thickness of the ocean floor for resource estimation	Project storage drive	The project team