National- to continental-scale governmental geophysical efforts for critical mineral mapping, USA

Anne E. McCafferty[1], Chelsea M. Amaral[1], and Garth Graham[1]
[1]U.S. Geological Survey, Geology, Geochemistry and Geophysics Science Center, Box 25046, Mail Stop 964, Denver Federal Center, Denver, Colorado 80225 USA

Abstract. The U.S. Geological Survey (USGS) has established robust collaborations with domestic state and international geological surveys to provide geophysical and other types of earth science data that act to underpin critical mineral research efforts across the United States, Canada, and Australia. The Earth Mapping Resource Initiative (EMRI) is a national-scale collaborative effort with state geological surveys to improve geophysical and geological data to advance our understanding of the United States' critical mineral endowment. The Critical Mineral Mapping Initiative (CMMI) is a tri-national collaboration with the federal geological surveys of Canada and Australia to conduct research that will aid in identifying new areas with potential for critical mineral deposits across all three countries. This study describes the important interplay between the EMRI and CMMI and how each act in a complementary fashion to advance critical mineral research. We present examples that illustrate how magnetic anomaly data are used to define critical mineral prospectivity for Mississippi Valley-type (MVT) Zn-Pb mineral systems and illustrate how CMMI magnetic derivative maps were considered into USGS' EMRI efforts to acquire modern high-resolution airborne geophysical data over a large area within the US Midcontinent.

1 Introduction

Within the last four years, the USGS has established two significant research collaborations with US State Geological Surveys and with federal government surveys of Canada (Geological Survey of Canada) and Australia (Geoscience Australia). In 2019, the USGS launched the Earth Mapping Resource Initiative (EMRI) in collaboration with the Association of American State Geologists (AASG). EMRI's goals are to improve our understanding of the geologic framework of the country by mapping aspects of the surface and subsurface. Interpretations of subsurface geology and architecture are being greatly improved with the acquisition of new high-resolution geophysical surveys (Day 2019). EMRI activities were accelerated in 2021 with the passage of the Bipartisan Infrastructure Law, which provides a significant increase in funding for geologic mapping and geophysical surveys to better understand the United States' critical mineral endowment. The tri-national Critical Mineral Mapping Initiative, a collaboration among the federal geological surveys of the United States (U.S. Geological Survey), Canada (Geological Survey of Canada) and Australia (Geoscience Australia) was formed in 2019. The mission of the CMMI is to conduct research to better understand critical mineral resources in known deposits, determine the geological controls on known critical mineral deposits, and identify new sources of supply through mineral prospectivity mapping and resource assessment (Kelley 2020).

Both the EMRI and CMMI are multi-faceted in their research approaches to mapping critical mineral geology and prospectivity. The CMMI is concentrating one aspect of its research on mapping mineral prospectivity for basin-hosted Zn-Pb deposits, including Mississippi Valley-type (MVT) deposits, using machine learning techniques (Lawley et al. 2022). In parallel, EMRI has identified 'focus' areas across the United States with potential for critical minerals (Dicken et al. 2022). Focus areas are defined collaboratively with the state geological surveys and are selected based on criteria such as areas undergoing active mining, areas currently or previously having been mined with by-product critical mineral production, or areas identified as prospective via exploration and research. The EMRI focus areas incorporate a wide range of system and deposit types (Hofstra and Kreiner 2020) and include known world class MVT districts. Collaborative discussions with state geological surveys define planning and collection of new data including high-resolution airborne magnetic and radiometric surveys.

Mississippi Valley-type deposits, including those in the United States, account for a significant proportion of the world's base metal production and resources. These enormous hydrothermal systems, formed from evaporative brines and hosted in sedimentary basins are also a high-potential target for a long list of critical minerals including Ba, Be, Co, F, Ga, Ge, In, Nb, Ni, Sn, Ti, and Zn as well as principal commodities such as Ag, Cu, Pb, Th, and Y (Hofstra and Kreiner 2020; Dicken et al. 2022).

This paper highlights geophysical efforts that support the goals of both projects by providing examples of how the CMMI and EMRI have worked in a complementary way to leverage legacy and modern magnetic data to identify MVT deposit potential in the southern Midcontinent of the United States.

2 Methods

Mineral prospectivity modelling requires ingestion of numerous earth science-related datasets, including surface geology, structure, and deposit location data, as well as geophysical data that can image rock properties in the subsurface to depths that can reach tens to hundreds of kilometres. These latter

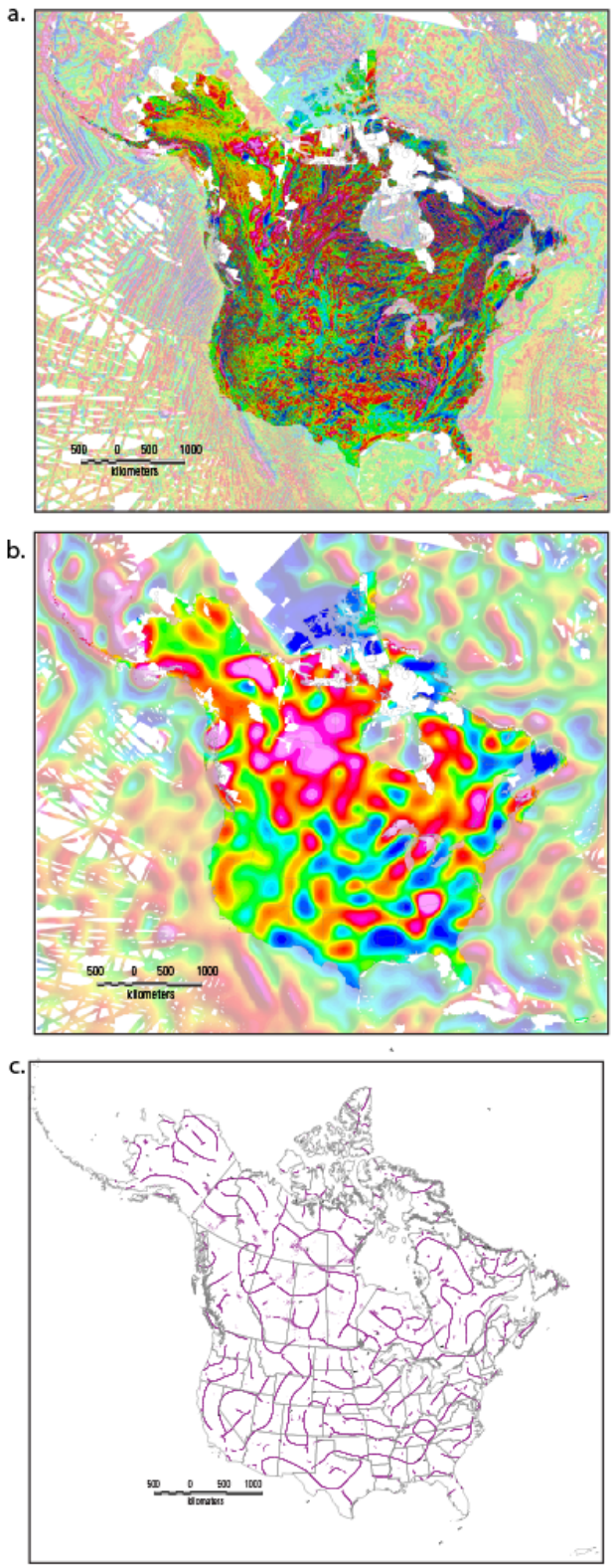

Figure 1. (a) Magnetic anomaly map of the US and Canada; (b) Long-wavelength magnetic anomaly; (c) Deep crust magnetic boundaries (McCafferty et al., 2023).

data can inform researchers on deeper crustal frameworks and former geologic processes that could have focused ore deposit formation. The significance of individual data types (e.g., magnetics) to geology and known deposits can be investigated on a layer-by-layer basis to investigate empirical relationships. The data can be then feature engineered to further enhance the empirical relationships prior to incorporating individual layers into the modelling process.

Efforts to create CMMI mineral prospectivity models (Lawley et al. 2020) involved processing of geologic deposit location, geologic, and geophysical data to evidential layers for input into a machine learning environment. National-scale magnetic anomaly data for the US, Canada, and Australia were filtered to enhance long-wavelength magnetic anomalies (McCafferty et al. 2023) for this purpose.

The rationale for emphasizing the long wavelengths in the magnetic anomaly field arose from recent studies that show geophysical data related to physical property changes within the deep lithosphere are instrumental in mapping locations of major tectonic features and craton boundaries that are spatially associated with the distribution of sedimentary hosted deposits including MVT deposits (Hoggard et al. 2020; Huston et al. 2022).

In general, deep-seated geologic sources give rise to long- wavelength anomalies. To enhance the footprint of the deep-crustal magnetic sources, the horizontal gradient magnitude (HGM) of the long-wavelength magnetic field (reduced-to-pole then transformed to pseudo gravity) was calculated. The edges outlining the magnetic source, often referred to as 'worms', track the maxima of the HGM and are interpreted to map the outer extent of the deep crustal magnetic sources (Fig. 1c).

3 Results

Processing of total magnetic field data to emphasize only long-wavelength (deep) magnetic features, as used in CMMI prospectivity modelling, permits simplification of a complex total magnetic anomaly map at the continental scale (e.g., Fig 1a) to more digestible broad scale trends (Fig.1b and 1c).

The Midcontinent of the US is host to the largest MVT Zn-Pb province in world and includes the world-class districts of the Old Lead Belt, Viburnum Trend, and Tri-State districts among others. Eighteen MVT EMRI focus areas present (Fig. 2b). Of the 18 focus areas, 12 (67%) have overlap with one of four deep magnetic boundaries. Seventeen of the 18 focus areas (94%) occur within a 40 km distance of a deep boundary.

Analyses of other non-MVT EMRI focus areas (Dicken et al. 2022) show a similar spatial relationship with the locations of the deep boundaries. This region of world class MVT deposits also hosts several other igneous mineral systems including the southeast Missouri IOA/IOCG province (i.e., Pea Ridge, MO IOA, Boss, MO IOCG), mafic magmatic systems (i.e., Glen Mountain PGE complex), and magmatic REE systems (i.e., Hicks Dome, IL and Magnet Cove, AR carbonatites). A total of 21 EMRI focus areas are mapped related to these 3 mineral systems, with 16 or 76% of the focus areas either overlapping or falling within 40 km of a magnetic boundary.

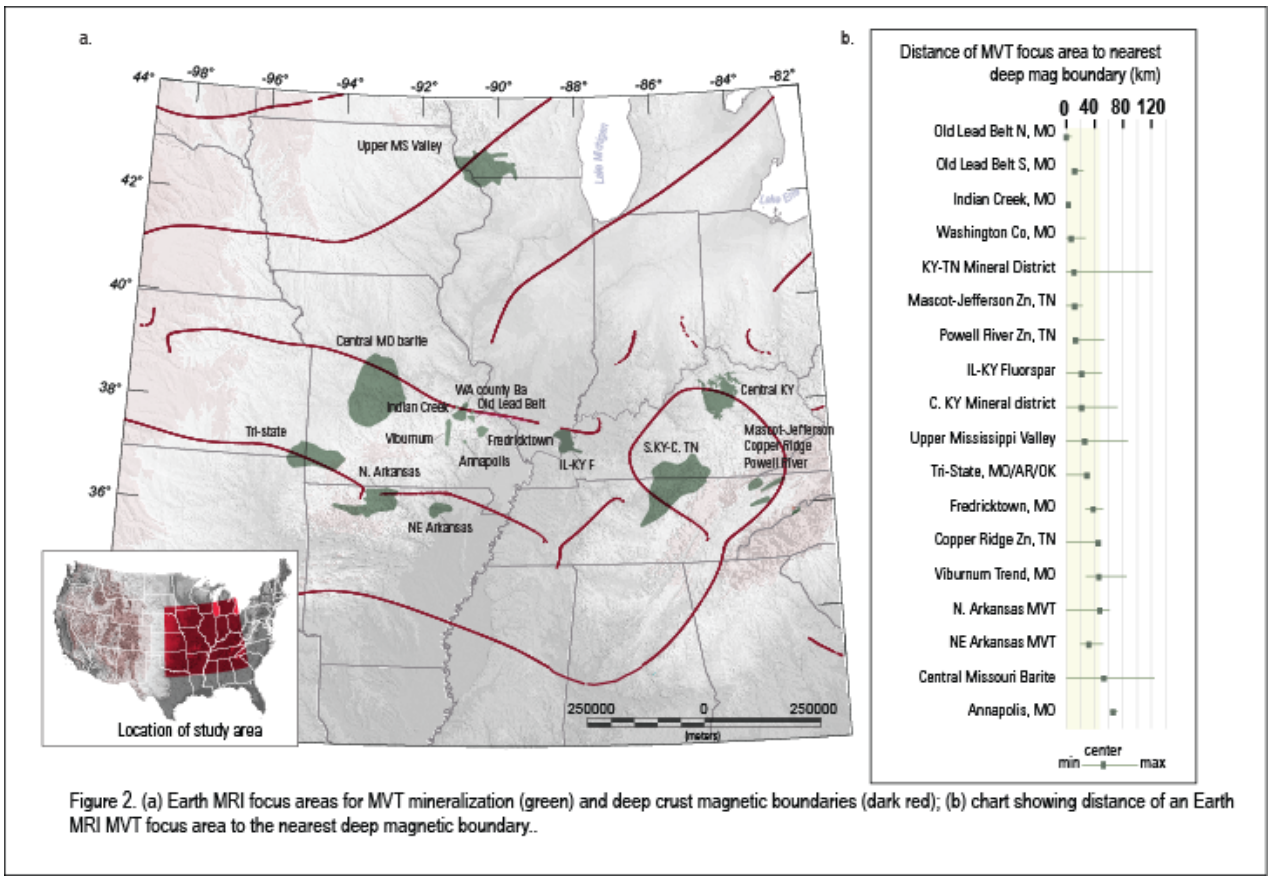

Figure 2. (a) Earth MRI focus areas for MVT mineralization (green) and deep crust magnetic boundaries (dark red); (b) chart showing distance of an Earth MRI MVT focus area to the nearest deep magnetic boundary..

4 Discussion

The deep crust magnetic boundaries are interpreted to map the edges of ancient geologic terranes that acted to preferentially control overlying sedimentary geometries and subsequent younger mineralizing events. Comparison of paleo-reconstruction models from a global terrane database of Eglington et al. (2013) show many of the magnetic boundaries parallel and lie close to the edges of ancient basement terranes.

Depth to the deep magnetic boundaries is estimated to be approximately equivalent to the depth to the Moho from studies done on Curie depth across this region. The Moho ranges in depth from 32 to 44 km in this area. We assume these depths approximate the lower boundary on magnetic susceptibility related to changes in mineralogy across the crust/mantle boundary. This assumption is supported by a study using the North America magnetic compilation (Ravat 2007) that determined the cold Archean and Proterozoic provinces within the Midcontinent of the U.S. were generally characterized by a non-magnetic mantle (Ravat and Purucker 2012).

The close coincidence of the deep crust magmatic boundaries to the locations of known MVT districts hints at a fundamental control by deep crustal boundaries on the siting of these mineralized districts within an old, broadly stable craton. We hypothesize that these deep-–seated structures influenced the sedimentary facies patterns in the overlying sedimentary basins, which may have influenced the flow paths of the brines responsible for MVT mineralization at least in the US Midcontinent. Preliminary examination of focus areas for some other deposit types (not shown) suggest that deep crustal boundaries may also control the siting of other deposit types. Further research is required to affirm causation. Nonetheless, these results suggest that, in some geological environments, deep geophysics may help to concentrate the prospective search area for ore deposits. Identifying ore-deposit locations through EMRI studies is thus consequential in advancing our understanding of our earth and improving responsible management of its resources.

5 Conclusion

National-scale magnetic anomaly data generated as part of the CMMI project, show spatial association with MVT mineral systems across the US Midcontinent. The vast majority (94%) of EMRI MVT focus areas occur within 40 kilometres of a deep magnetic boundary. The boundaries are estimated to represent geologic sources near the Moho and map the edges of ancient geologic terranes. Acquisition of modern high-resolution magnetic and radiometric surveys are being designed and flown for the EMRI program over this area of the Midcontinent. The new data will shed light on the shallow expression of magnetic anomalies that allow for refinement of the first-order

controls on mineralization provided by the deep magnetic boundary analyses.

Acknowledgements

The authors wish to thank Ani Tikku (USGS reviewer) and an SGA reviewer for constructive comments.

Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government

References

Day WC (2019) The Earth Mapping Resources Initiative (Earth MRI)—Mapping the Nation's critical mineral resources (ver. 1.2, September 2019). U.S. Geological Survey Fact Sheet 2019–2007. https://doi.org/10.3133/fs20193007

Dicken CL, Woodruff LG, Hammarstrom JM, Crocker KE (2022) GIS, supplemental data table, and references for focus areas of potential domestic resources of critical minerals and related commodities in the United States and Puerto Rico. US Geological Survey data release. https://doi.org/10.5066/P9DIZ9N8. Accessed 6 March 2023

Eglington B, Pehrsson S, Ansdell K, Lescuyer J, Quirt D, Milesi J, Brown P (2013) A domain-based digital summary of the evolution of the Palaeoproterozoic of North America and Greenland and associated unconformity-related uranium mineralization. Precambrian Research. https://doi.org/10.1016/j.precamres.2013.01.021

Hofstra AH, Kreiner DC (2020) Systems-Deposits-Commodities-Critical Minerals Table for the Earth Mapping Resources Initiative (ver. 1.1, May 2021). US Geological Survey Open-File Report 2020–1042. https://doi.org/10.3133/ofr20201042.

Hoggard MJ, Czarnota K, Richards FD, Huston DL, Jaques AL, Ghelichkhan S (2020) Global distribution of sediment–hosted metals controlled by craton edge stability. Nature Geoscience 13:504–510. https://doi.org/10.1038/s41561-020-0593-2.

Huston DL, Champion DC, Czarnota K, Duan J, Hutchens M, Paradis S, Hoggard M, Ware B, Gibson GM, Doublier M, Kelley K, McCafferty A, Hayward N, Richards F, Tessalina S, Carr G (2022) Zinc on the edge—isotopic and geophysical evidence that cratonic edges control world-class shale-hosted zinc-lead deposits. Mineralium Deposita. http://dx.doi.org/10.1007/s00126-022-01153-9.

Kelley KD (2020) International geoscience collaboration to support critical mineral discovery. US Geological Survey Fact Sheet 2020– 3035. https://doi.org/10.3133/fs20203035.

Lawley CJM, McCafferty AE, Graham GE, Huston DL, Kelley KD, Czarnota K, Paradis S, Peter JM, Haward N, Barlow M, Emsbo P, Coyan J, San Juan CA, Gadd MG (2022) Data-driven prospectivity modelling of sediment-hosted Zn-Pb mineral systems and their critical raw material. Ore Geology Reviews 141:104635. https://doi.org/10.1016/j.oregeorev.2021.104635.

McCafferty AE, San Juan CA, Lawley CJ, Graham GE, Gadd MG, Huston DL, Kelley KD, Paradis S, Peter JM, Czarnota K (2023) National-Scale Geophysical, Geologic, and Mineral Resource Data and Grids for the United States, Canada, and Australia- Data in Support of the Critical Minerals Mapping Initiative. US Geological Survey data release. https://doi.org/10.5066/P970GDD5

Ravat D (2007) Crustal magnetic fields. In: Gubbins D, Herrero-Severa E (eds) Encyclopedia of Geomagnetism and Paleomagnetism. Springer, New York, pp 140-144

Ravat D, Purucker M (2012) Unraveling the magnetic mystery of the Earth's lithosphere: The background and the role of the CHAMP Mission. In: Reigber C, Lühr H, Schwintzer P (eds) First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies. Springer, New York, pp 251-260

A GIS-based mineral prospectivity analysis of the Neoproterozoic Arabian Shield

Christophe Bonnetti[1], Arnaud Fontaine[1], Célestine Berthier[1,3], Julien Feneyrol[1], Joffrey Corbet[1], Virginie Masson[1], Rémi Bosc[1,2], Adil M. Hashim[2]
[1]Arethuse Geology EURL, 29 Allée de Saint Jean, 13710 Fuveau, France
[2]Arethuse Arabia Mining, King Abdul Aziz Street, 13315 Riyadh, Kingdom of Saudi Arabia
[3]GeoRessources, Université de Lorraine, CNRS, 54000 Nancy, France

Abstract. The Neoproterozoic Arabian Shield formed following three major tectono-magmatic events during the Cryogenian–Ediacaran Nabitah orogenic cycle, including the pre-accretion, syn-orogenic and late- to post-orogenic stages, representing fertile environment for various mineral systems related to precious, base and rare metals. At the shield scale, the mineral prospectivity analysis that was performed, based on the selective review and reclassification of multiple geological and geophysical datasets, identified a series of mineral belts correlated with suture and shear zones concentrating the majority of mineral deposits and occurrences. (i) Pre-accretion arc-related porphyry, epithermal, VMS mineral systems and magmatic deposits related to ultramafic rocks are predominantly distributed along the Nabitah, Al Amar, Bi'r Umq and Yanbu suture zones. (ii) Orogenic gold veins mainly developed in the N-trending Nabitah shear zone that is coaxial with the Nabitah and Al Amar sutures. Gold was remobilised from source rocks during this syn-collisional tectonic event associated with a peak M1 of low-grade metamorphism. Orogenic gold mineralisation also occurred sporadically along the NW-trending Najd strike-slip fault system developed during late orogenic extension and associated with a peak M2 of high-grade metamorphism, locally. (iii) Finally, magmatic-hydrothermal rare metal deposits formed in association with late- to post-orogenic alkaline, peralkaline and peraluminous granites.

1 Introduction and geological setting

The Arabian Shield (AS) is part of a larger geological Neoproterozoic assemblage, the Arabian-Nubian Shield (ANS) spreading over parts of Egypt, Eritrea, Ethiopia, Saudi Arabia, Somalia, Sudan and Yemen (Nehlig et al. 1999), which represents an area > 1,100,000 km^2 of fertile environment for various mineral systems related to precious, base and rare metals (Technip Group et al. 2015). However, the availability of geological and geophysical data is highly variable depending on the country, and historical surveys as well as exploration works are heterogeneously distributed and do not always cover areas of interest from a mineral prospectivity point of view. Although many of the geological data collected over the past decades in the AS have been digitalised into a Geographical Information System (GIS), an overall review of the major geological events that occurred through the geodynamic evolution of the shield, and their roles in providing the favourable conditions for ore genesis is currently lacking. Therefore, we reviewed and combined a broad variety of datasets collected over the past ten years into an in-house GIS database to assess the prospectivity potential of certain areas at the shield scale based on modern tectonic concepts (e.g. suture-structural-mineral belts) and the mineral systems approach, which can then be applied as exploration targeting criteria for different mineralisation styles at the belt or district scale (e.g. McCuaig et al. 2010).

The ANS evolved between ~870 and 550 Ma as one of the largest tracts of juvenile Neoproterozoic crust in the world (Johnson 2014; Figure 1). Within this domain, the AS is differentiated by a series of variably oriented sutures punctuated by ophiolite complexes (Stern et al. 2004), shear zones and fold belts (Meyer et al. 2014; Elisha et al. 2017; Figure 2). Suture zones and coaxially developed shear belts highlight the boundaries of magmatic arc remnants and micro-continental blocks (Stern and Johnson 2010; Johnson 2014) that were accreted during the Cryogenian-Ediacaran Nabitah orogeny (Nehlig et al. 2002) as a result of the Greater Gondwana assembly at 544 Ma (Stern and Johnson 2010).

The geodynamic evolution of the AS can be summarized by three main tectono-magmatic phases reflecting a 300-million-year process of continental crustal growth represented by amalgamated juvenile magmatic arcs and associated volcano-sedimentary basins, syn-orogenic intrusive bodies and molassic basins, and late to post-orogenic granitoid intrusions (Stern and Johnson 2010; Figures 1 and 2):

(i) a Late Tonian–Early Cryogenian (~880–660 Ma) pre-orogenic rifting episode that triggered Rodinia break-up, which was followed by the formation of multiple magmatic island arcs together with fore- and back-arc volcano-sedimentary basins within an intra-oceanic subduction setting (Johnson et al. 2011; Johnson 2014). This episode corresponds with a major stage of juvenile crust formation in the northern East African Orogen, between West and East Gondwana (Stern and Johnson 2010). As the convergence progressed, these arcs were progressively amalgamated to form the AS, with a peak accretion age at ca. 780 Ma (Stern et al. 2004);

(ii) a Late Cryogenian–Early Ediacaran (~690–590 Ma) syn-orogenic stage marked by the onset of the Nabitah collision or orogeny (Nehlig et al. 2002; Johnson et al. 2011, 2013), which is dominantly characterised by an early peak M1 of low grade (greenschist) metamorphism at ca. 710 Ma (Elisha et al. 2017), the development of the N-trending, dextral transpressional Nabitah shear and fold belt

(~680–640 Ma; Johnson et al. 2011), deformed syn-orogenic granitoids and post-accretion molassic basins (e.g. Murdama Group sediments; ~670–570 Ma; Johnson et al. 2013);

(iii) a Late Cryogenian–Ediacaran (~650–530 Ma) late to post-orogenic phase of extensional collapse (Blasband et al. 2000) marked by the offset of the Nabitah structures by the NW-trending, sinistral Najd strike-slip fault system (Stern and Johnson 2010; Meyer et al. 2014), with local development of a late peak M2 of high grade (> amphibolite) metamorphism at ca. 620 Ma associated with a series of gneissic domes (Johnson et al. 2013; Elisha et al. 2017), late sediments infill in molassic basins (e.g. Jibalah Group > 640 Ma; Johnson et al. 2013) and numerous intrusions of late- to post-orogenic granitoids and dykes (~650–530 Ma) with post-collisional anorogenic signatures (Eyal and Eyal 1987; Lehmann et al. 2020).

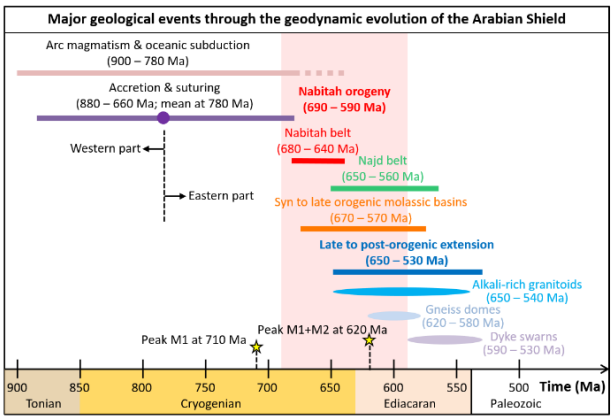

Figure 1. Chronology of major geological events through the geodynamic evolution of the Arabian Shield (modified after Eyal and Eyal 1987; Blasband et al. 2000; Nehlig et al. 2002; Stern et al. 2004; Stern and Johnson 2010; Johnson et al. 2011, 2013; Johnson 2014; Meyer et al. 2014; Elisha et al. 2017; Lehmann et al. 2020).

2 Methodology

The mineral prospectivity analysis (i.e. Carranza 2021) was performed on QGIS software by reviewing multiple datasets of geological and geophysical information including airborne magnetic and satellite imagery data, stratigraphic units, lithologies, structures, ore deposits and mineral occurrences inventory compiled from the literature and publicly available data provided by the Saudi Geological Survey. Lithologies were then classified based on their petrologic features, stratigraphic ages and related tectono-magmatic event through the Nabitah orogenic cycle. Similarly, structures were reclassified based on their nature, distribution and kinematic indicators with a particular attention in locating terrane boundaries and suture zones along which major structural belts developed. Finally, a selective review of the mineral occurrence database (Nehlig et al. 1999; Technip Group et al. 2015) was conducted to determine key criteria and favourable contexts for ore genesis to assess the prospectivity potential of the main mineral systems developed in major structural belts of the AS that are presented in the litho-tectonic and metallogenic map of Figure 2.

Mineral occurrences previously classified by commodities were reclassified according to their mineralisation styles and deposit types following the mineral system approach (McCuaig et al. 2010) and using criteria such as metal association, morphology, host rock, structural control, alteration pattern, and the local geological context. This mineral system approach was integrated into the geodynamic context as a precursor to statistical and spatial analyses that will be performed in the future.

3 Mineral prospectivity analysis

3.1 Pre-accretion, magmatic arc-related mineral systems

The pre-accretion stage of the AS was dominated by arc magmatism within an oceanic subduction domain and provided favourable geological setting for the formation of VMS, epithermal and porphyry mineral systems, as well as magmatic mineralisation related to ultramafic igneous rocks. To date, no porphyry Cu-Mo deposit has been discovered and only few occurrences hosted in granodiorite mainly located at the edge of the Ad Dawadimi terrane in eastern AS were reported. Orthomagmatic Cr-Ni-Cu deposits and occurrences are predominantly hosted in mantle-derived or oceanic crust remnants ultramafic rocks (e.g. serpentinite, gabbro, peridotite) and distributed along the Yanbu, Nabitah and Al Amar suture zones. Syngenetic VMS Cu-Zn-(±Au) deposits and occurrences are mainly hosted in arc-related volcano-sedimentary basins showing bimodal volcanic records. Hence, their spatial distribution follows the orientation trend of their host lithologies, which are highlighted by the suture zones (Nabitah, Bi'r Umq, Al Amar, Yanbu) as a result of magmatic arcs collage during the accretion phase. For instance, the giant Jabal Sayid deposit (estimated resources of 56.4 Mt at 2.2% Cu, 0.1% Zn, 0.2 g/t Au, 5.0 g/t Ag; Technip Group et al. 2015) is hosted in bimodal volcanic rocks within the accretionary complex of a forearc volcano-sedimentary basin, south of the contact with the Bi'r Umq suture zone. Finally, prominent epithermal deposits and associated occurrences are mainly hosted in the contact zone between arc-related plutonic and volcanic rocks along the Al Amar (e.g. Al Amar deposit with estimated resources of 6.8 Mt at 14.0 g/t Ag, 5.2 g/t Au, 4.5% Zn; Technip Group et al. 2015) and Bi'r Umq (e.g. Mhad Ahd Dhahab deposit with estimated resources of 3.4 Mt at 44.2 g/t Ag and 9.2 g/t Au; Technip Group et al. 2015) suture zones.

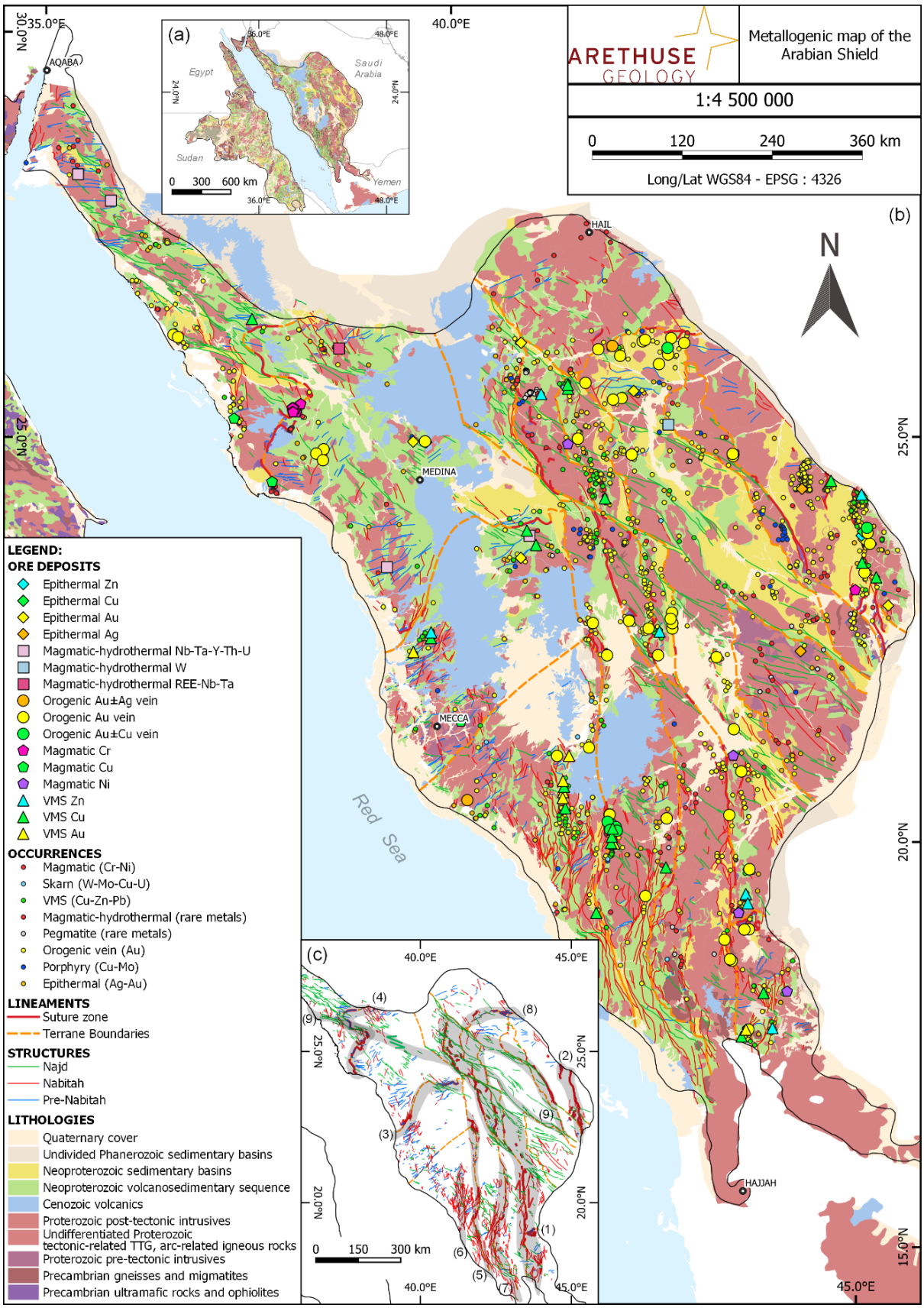

Figure 2. a Litho-tectonic map of the Arabian-Nubian Shield. **b** Litho-tectonic and metallogenic map of the Arabian Shield. **c** Structural map of the Arabian Shield showing the distribution of the main mineral belts. The grey zones correspond with the mineral belts as follow: (1) the Nabitah suture/shear zone, (2) the Al Amar suture zone, (3) The Bi'r Umq suture zone, (4) the Yanbu suture zone, (5) The Umm Farwah shear zone, (6) the Baydah shear zone, (7) the Tadj/Ibran shear zone, (8) the Hibashi suture/fault zone, (9) the Najd shear zone (modified after Nehlig et al. 1999, 2002; Blasband et al. 2000; Stern et al. 2004; Stern and Johnson 2010; Johnson et al. 2011, 2013; Johnson 2014; Meyer et al. 2014; Technip Group et al. 2015; Lehmann et al. 2020).

3.2 Syn- to late orogenic gold systems

Orogenic gold system is widespread throughout the AS and was related to two major tectono-metamorphic events: (i) the syn-collisional development of the N-trending Nabitah shear and fold belt that is coaxial with the Nabitah and Al Amar sutures across the shield and associated with parallel shear zones in its southern part. Gold deposits and occurrences (e.g. Ar Rjum deposit with estimated resources of 84.0 Mt at 1.3 g/t Au in the Zalim district; Technip Group et al. 2015) are structurally-controlled by shear zones, often overprinting pre-accretion mineral systems, and occur as auriferous quartz-carbonate-(±sulfide) veins, which are hosted in various lithologies that experienced the peak M1 of metamorphism at the origin of fluid-mediated gold remobilization from source rocks; and (ii) the development of the NW-trending Najd strike-slip fault system during the late orogenic extension associated with the peak M2 of metamorphism,. In this context, gold mineralisation is hosted in Nabitah/Nadj related shears within the contact zone of late orogenic intrusions (e.g. Ad Duwayhi deposit with estimated resources of 31.0 Mt at 2.4 g/t Au; Technip Group et al. 2015).

3.3 Late to post-orogenic magmatic-hydrothermal rare metal systems

The AS also experienced a relatively intense episode of magmatism during the late- to post-orogenic stage of the Nabitah orogenic cycle, with emplacement favoured along major crustal discontinuities represented by the Nabitah and Najd structural belts. The distribution and typology of associated rare metal magmatic-hydrothermal deposits and occurrences were mainly controlled by the chemical affinities of magma sources with Nb-Ta-REE-Th-U-Sn mineralisation related to alkaline and peralkaline granites and W-Sn-Nb-Ta-Li-Be associated with peraluminous granite and pegmatite.

4 Conclusions and perspectives

The synthesis of the main geological and metallogenic events through the geodynamic evolution of the Arabian Shield allowed the characterisation of various mineral systems that formed during the pre-accretion, syn-orogenic and late- to post-orogenic phases of the Nabitah orogenic cycle. The mineral prospectivity analysis identified a series of mineral belts that concentrate the large majority of ore deposits and occurrences with suture zones (Nabitah, Al Amar, Bi'r Umq and Yanbu) and structural belts (Nabitah and Najd shear zones) as a major pathfinder at the shield scale. This study was therefore a necessary step to define prospective areas and narrow down the prospect generation for spatial and statistical analyses at the belt or district scale to provide guidance for exploration and generate future metallogenic research studies in the AS.

Acknowledgements

The authors thank Arethuse Geology for funding this study and providing its in-house GIS database.

References

Blasband B, White S, Brooijmans P, De Boorder H, Visser W (2000) Late Proterozoic extensional collapse in the Arabian-Nubian Shield. J Geol Soc London 157:615-628

Carranza EJM (2021) Mineral Prospectivity Analysis. In: Daya Sagar B, Cheng Q, McKinley J, Agterberg F (eds) Encyclopedia of Mathematical Geosciences. Encyclopedia of Earth Sciences Series. Springer, Cham

Elisha B, Katzir Y, Kylander-Clark A (2017) Ediacaran (~620 Ma) high-grade regional metamorphism in the northern Arabian Nubian Shield: U-Th-Pb monazite ages of the Elat schist. Precambrian Res 295:172-186

Eyal Y and Eyal M (1987) Mafic Dykes in the Arabian-Nubian Shield. J Earth Sci 36:195-211

Johnson PR (2014) An Expanding Arabian-Nubian Shield geochronologic and isotopic dataset: Defining limits and confirming the tectonic setting of a Neoproterozoic accretionary orogen. Open Geol J 8:3-33

Johnson PR, Anderson A, Collins AS, Fowler AR, Fritz H, Ghebreab W, Kusky T, Stern RJ (2011) Late Cryogenian-Ediacaran history of the Arabian-Nubian Shield: A review of depositional, plutonic, structural, and tectonic events in the closing stages of the northern East African Orogen. J African Earth Sci 61:167-232

Johnson PR, Halverson GP, Kusky TM, Stern RJ, Pease V (2013) Volcanosedimentary basins in the Arabian-Nubian Shield: Markers of repeated exhumation and denudation in the Neoproterozoic accretionary orogen. Geosciences 3:389-445

Lehmann B, Zoheir B, Neymark L, Zeh A, Emam A, Radwan A, Zhang R, Moscani R (2020) Monazite and cassiterite U-Pb dating of the Abu Dabbab rare-metal granite, Egypt: Late Cryogenian metalliferous granite magmatism in the Arabian-Nubian Shield. Gondwana Res 84:71-80

McCuaig TC, Beresford S, Hronsky J (2010) Translating the mineral systems approach into an effective exploration targeting system. Ore Geol Rev 38:128-138

Meyer SE, Passchier C, Abu-Alam T, Stuwe K (2014) A strike-slip core complex from the Najd fault system, Arabian Shield. Terra Nova 26:387-394

Nehlig P, Salpeteur I, Asfirane F, Bouchot V, Eberlé JM, Genna A, Kluyver HM, Lasserre JL, Leister JM, Nicol N, Récoché G, Shanti M, Thiéblemont D (1999) The mineral potential of the Arabian Shield: A reassessment. In proceedings of the IUGS/UNESCO Meeting on the "Base and Precious Metal Deposits in the Arabian Shield", Jeddah

Nehlig P, Genna A, Asfirane F (2002) A review of the Pan-African evolution of the Arabian Shield. GeoArabia 7:103-124

Stern RJ and Johnson P (2010) Continental lithosphere of the Arabian Plate: A geologic, petrologic, and geophysical synthesis. Earth Sci Rev 101:29-67

Stern RJ, Johnson PR, Kröner A, Yibas B (2004) Neoproterozoic ophiolites of the Arabian-Nubian Shield. Develop Precambrian Geol 13:95-128

Technip Group, Arethuse Geology, brgm (2015) Comprehensive mining strategy study project – Metals inventory for the kingdom of Saudi Arabia. Technical Report 064856C001-430, unpublished

Large-scale structural controls on hot spring mineral deposits of geothermal systems (Mt. Amiata, Italy) highlighted by machine learning algorithms?

Paolo S. Garofalo[1], Lara Capitanio[1], Elisa Mariarosaria Farella[2], Simone Rigon[2], Fabio Remondino[2], Ivan Callegari[3], Daniele Rappuoli[4]
[1] Dept. of Biological, Geological and Environ. Sciences, University of Bologna, I
[2] 3D Optical Metrology Unit, Fondazione Bruno Kessler – Trento, I
[3] German University of Technology – Sultanate of Oman
[4] Parco Nazionale Museo delle Miniere del Monte Amiata, Piancastagnaio, I

Abstract. The cinnabar (±stibnite) deposits of the Mt. Amiata geothermal system and the associated hot springs and gas vents, occur along a N-S directed, narrow longitude region.
In this study, we combine a geological and geophysical dataset gathered from the early stages of geothermal exploration of the district with a multivariate statistical analysis carried out by Machine Learning (ML) algorithms to highlight possible correlations between the distribution of the geothermal expressions of Mt. Amiata and its geological/structural features. We used 5 distinct ML supervised models (Ordinary Least Squares Linear Regressor, Multilayer Perceptron Regressor, Support Vector Regressor, CatBoost, and Random Forest) to determine which set of geological or geochemical features of the dataset reproduces the distribution of the geothermal expressions of the area with sufficient accuracy.
The regressors CatBoost and Random Forest, which use decision trees for probability calculations, are the most efficient in predicting the narrow-longitude distribution of the geothermal expressions of Mt. Amiata. Also, the only combination of predictors generating probability maps that accurately reproduce the distribution of the geothermal expressions is the one considering permeability, Hg solubility, T, and distances from faults and folds. This shows that only a combination of geological/geochemical factors can explain the peculiar regional distribution.

1 Introduction

With an historical production of c. 117 kt of Hg at a grade of 0.2-8 wt% (Segreto, 1991), the cinnabar(±stibnite) deposits of Mt. Amiata (south Tuscany, Italy) form one of the largest mercury districts ever documented. Here, 14 deposits of distinct sizes and economic importance were exploited between the years 1846 and 1982. Most of these deposits occurred along with uneconomic prospects within an area that is >30 km long in the N-S direction – from the Pietrineri deposit to the N to the Catabbio deposit to the S – and c. 15 km wide in the E-W direction (Fig. 1).

The district is located within a geothermal system that was explored with a set of geophysical methods since 1953 (Cataldi, 1967). Presently, close to the cities of Piancastagnaio and Bagnore (Fig. 1) this system hosts 5 power plants having 88 MW installed capacity, which exploit two distinct wet-steam reservoirs (Barelli et al., 2010). The first is 2500-4000 m deep at c. 300-350 °C, and the second is 500-1000 m deep at 150-230 °C. In the district, CO_2-rich gas vents and hot springs were documented in detail (e.g., Frondini et al., 2009; Magi et al., 2019).

The area belongs to the Apennine thrust-and-fold belt, which consists of a stack of tectonic units detached from the Adria plate that migrated progressively eastwards from the Late Oligocene to the Early Miocene (Marroni et al., 2015). Since the Late Miocene, the migration of this deformation front was followed by a stage of extensional tectonics and post collisional magmatism.

A young volcano-plutonic system controls the geothermal system of Mt. Amiata, although only indirect evidence exists on the nature, shape, and depth of emplacement of the pluton (Gianelli et al., 1988). Several data suggest the presence of this pluton at 4-7 km depth, with an apophysis located below Monte Labbro (Fig. 1).

The Mt. Amiata volcano is a 305-231 ka old, small size volcano fed by a SSW-NNE eruptive fissure. From base to top, it is made of trachydacitic flows, olivine latitic to trachydacitic domes and flows, and olivine latites (Conticelli et al., 2015). Geochemical and isotopic compositions indicate a genesis from mixing between a high silica, high-K calc-alkaline magma and a mafic ultrapotassic magma.

The volcano-plutonic system intruded the folded and faulted stratigraphic sequence made of allochthonous flysch units deposited onto oceanic crust (Ligurian Domain), Mesozoic carbonatitic and Cenozoic terrigenous formations (Tuscan Domain), and a poorly outcropping Carboniferous sequence made of graphitic quartz-phyllites, metagreywackes, and carbonate-bearing quartz-phyllites (Gianelli et al., 1988).

The Mt. Amiata deposits consisted mostly of disseminations, massive stratabound, stockworks, and breccias cemented by cinnabar, metacinnabar, and marcasite (Arisi Rota et al., 1971). Stibnite, native Hg, pyrite, chalcopyrite, realgar, and orpiment were reported as minor phases. The most common gangue was calcite with minor celestite, fluorite, gypsum, zeolites, dawsonite, and amorphous silica (opal, chalcedony).

The deposits share many similarities with those of the near-surface hot spring deposits formed close to volcanic centers (Pirajno, 2020).

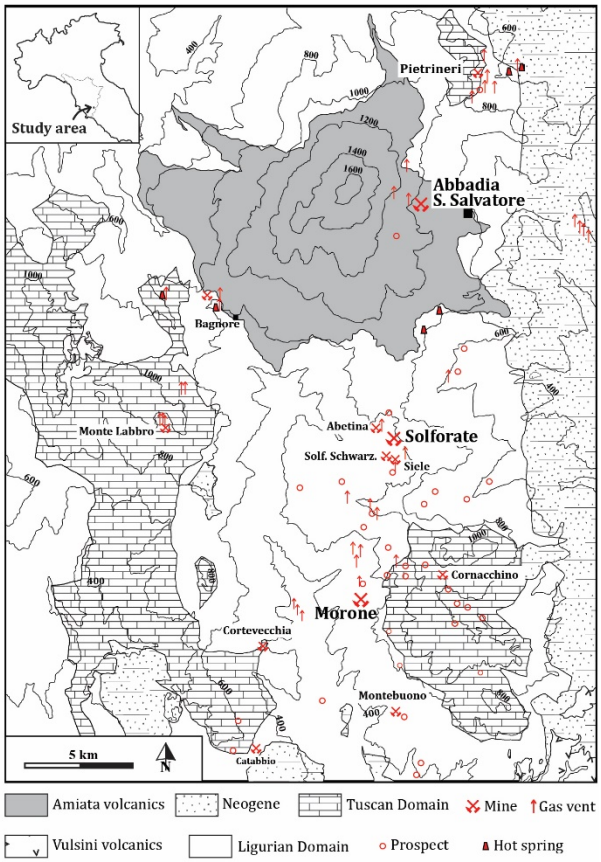

Figure 1. Geological map of the Mt. Amiata ore district and location of cinnabar(±stibnite) deposits, prospects, gas vents, and hot springs (modified from: Calamai et al., 1970). The three largest deposits are shown with larger symbols, and one of the 14 deposits of the district (Cerreto Piano) is located outside this map. Notice that all occurrences formed within a narrow longitude region.

A distinct characteristic of these deposits is their location within a wide latitude but narrow longitude region (Fig. 1). Such distribution was known since early geological documentation of the area (e.g., De Castro, 1914); however, the controlling factors of such peculiar regional distribution were never explored to a sufficient extent. For instance, recent work (Brogi et al., 2011) showed that the location of the Morone deposit (Fig. 1) was controlled by the formation of sinistral shear zones during the Pleistocene, in particular by focussing of the ore fluid within extensional jogs and pull apart structures. While very important at the deposit scale, these structures alone do not explain why the deposits formed within a narrow longitude region at the district scale, leaving unanswered a fundamental question on the genesis of these deposits.

The aim of our work is to apply Machine Learning (ML) algorithms to couple geological and geophysical data in order to explore possible correlations between the peculiar narrow-longitude distribution of the geothermal expressions of the area and its geological/structural features. Moreover, we apply a multivariate statistical analysis to predict the distribution of geothermal expressions in areas without surveys. We deliberately choose a limited dataset of surveys collected during the early stages of geothermal exploration because we want to simulate as much as possible a reconnaissance stage of mineral exploration, i.e., one in which a typical multidisciplinary mineral exploration dataset (from remote sensing, field mapping, geophysical and geochemical surveys, limited drilling) would be coupled with the statistical tools provided by ML algorithms.

2 Methodology

2.1 The dataset

The geological data used for our predictions consists of the documented lithologies of the study area (Fig. 1) and of its relevant structural data. In detail, we considered all antiforms, synforms, and faults (inverse, normal) mapped in the geothermal database (Cataldi, 1967) and in recent studies (Bonciani et al., 2005). The lithologies were grouped into 4 complexes according to their established permeability, and care was taken to determine the positions of the top of the shallow reservoir (500-1000 m deep) within the study area. These data were used to construct 2D and 3D geological models of the area.

The ore deposit data consists of the locations of all the geothermal expressions of the area, which include cinnabar (±stibnite) ore deposits, (uneconomic) prospects, gas vents, springs (cold, hot), geothermal wells, and power plants.

The physical-chemical dataset consists in the temperature (T) gradient, heat flow data, Bouguer anomaly data, and Hg solubility in the geothermal fluid, which was considered equal to the experimentally determined solubility of $Hg°_{aq}$ in water in the 0-350 °C interval (Clever et al., 1985).

This information was used to prepare a set of 24,398 points that mark the entire study area. Each point was univocally identified (via longitude, latitude, and depth) and characterized by unique predictor values of permeability, T, T gradient, Hg solubility, heat flow, Bouguer anomaly, distance from the nearest fold, distance from the nearest fault, and vertical distance from the top of the reservoir.

2.2 Machine Learning models

We calculated probability maps of the study area using five distinct ML supervised models (i.e., regressors). They are (1) Ordinary Least Squares Linear Regressor, (2) Multilayer Perceptron Regressor, (3) Support Vector Regressor, (4) CatBoost, and (5) Random Forest. These regressors differ from each other in that (1)-(2)-(3) calculate probability values by linear regression while (4)-(5) calculate probabilities by decision trees (Sun et al., 2019). Each regressor calculated the probability that a specific combination of geological or geochemical features of the dataset can reproduce the distribution of the geothermal expressions of the area. We

divided these features into two categories, namely geometric and ore predictors (Table 1), which distinguish features that identify potential structural or lithological controls (distance from folds, faults, and reservoir) from physical-chemical controls (permeability, T, Hg solubility, etc.). Kriging-based spatial interpolation was used later to calculate 2-D probability maps, which were generated at distinct depths up to 1 km.

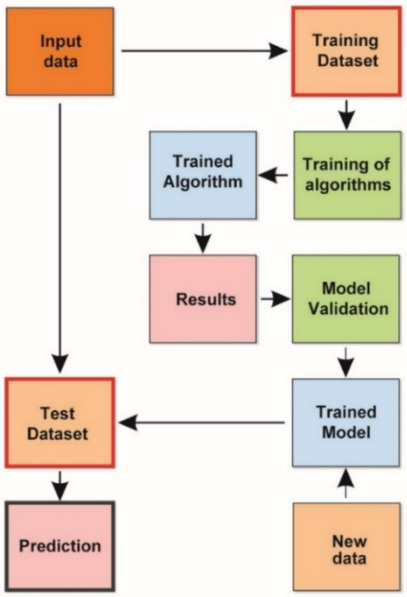

Figure 2. Block diagram showing the typical workflow of a ML algorithm.

Following typical ML methodologies, the available data were split into two subsets, namely the Training dataset and the Test dataset (Fig. 2). The first one was used to train the regressors, generate the ML model, and validate it. The Test dataset was used to carry out the predictions (Pedregosa et al., 2011). The Training dataset consisted of points where the occurrence of cinnabar deposits and/or prospects are historically known. This dataset was augmented following three established techniques (Farella et al., 2021) to improve the training of the distinct regressors when dealing with unbalanced classes of data. This generated three distinct datasets with which algorithms were trained and probability predictions were subsequently carried out.

Correlations and predictions were calculated for eight combinations of predictors (Table 1) using the Scikit-learn Python libraries (Pedregosa et al., 2011), which integrate the five regressors listed above and other state-of-the-art ML algorithms. Our combinations do not consider all possible permutations of ore and geometric predictors, but rather favor the role of permeability, Hg solubility, and faults in the formation of the geothermal expressions. Each combination of predictors was analyzed with the 5 regressors, but out of the 24 combinations of augmented datasets and regressors only seventeen provided results whose statistical significance was evaluated. We evaluated the accuracy of our prediction through three parameters: RMSE, MAE and R^2. Below, we present and discuss two representative results of this work.

3 Results

The Ordinary Least Squares Linear Regressor, the Multilayer Perceptron Regressor, and the Support Vector Regressor generate probability maps that do not reproduce the spatial distribution of the geothermal expressions of Mt. Amiata with any of the considered combination of predictors. In contrast, CatBoost and Random Forest generate maps that reproduce with good accuracy the distribution using the augmented Training dataset and a specific combination of predictors (Test 8).

Figure 3 shows two examples of interpolated probability maps generated with the described method using Random Forest. These maps highlight the different correlations that a given set of predictors generate with a given ML algorithm. Thus, the exclusive use of ore predictors (Test 1) by the Random Forest regressor generates high probability areas (Fig. 3a) that do not fit the established distribution of the geothermal expressions of Mt. Amiata (e.g., black triangles denoting deposits/prospects). Similar results are obtained by almost all other regressors using the combination of predictors presented in Table 1.

Table 1. Combination of predictors used in this study.

	Ore Predictors						Geometric Predictors		
Test no.	Permeability	Hg Solubility	Heat flux	T	T gradient	Bouguer Anomaly	Distance from fold	Distance from fault	Distance from top of C4
1	X	X	X	X	X	X			
2							X	X	X
3	X	X	X	X	X	X	X	X	X
4	X	X					X	X	X
5	X	X						X	
6	X	X	X					X	
7	X	X		X				X	
8	X	X		X			X	X	

Note: C4 denotes “Complex 4”, i.e., the permeable reservoir

The only combination of predictors that allowed CatBoost and Random Forest to generate probability maps that approximate accurately the distribution of the geothermal expressions is the one that considers permeability, Hg solubility, T, and distances from faults and folds (Fig. 3b, Test 8).

4 Interpretation

ML algorithms based on decision trees for probability calculations (i.e., CatBoost, Random Forest) prove to be the most efficient in predicting the true distribution of geothermal expressions of Mt. Amiata for this dataset. The evidence that only a combination of physical-chemical and geological

predictors (Test 8) is able to reproduce the narrow-longitude distribution of all geothermal expressions suggests that not all geological/structural factors play the same role in controlling ore precipitation, gas venting, and hot spring discharge in a geothermal field.

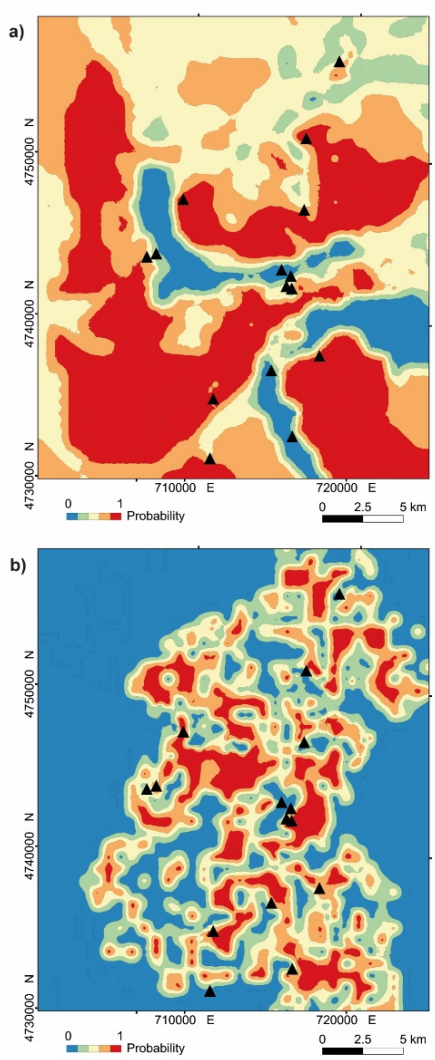

Figure 3. Correlation maps based on the Random Forest regressor and calculated for the entire study area. The black triangles are the known cinnabar (±stibnite) deposits/prospects (Fig. 1). (a) Map calculated considering only the ore predictors (Test 1, Table 1). (b) Map calculated considering a selection of ore and geometric predictors (Test 8).

Also, fault density alone is probably unable to control the regional distribution of all expressions of a geothermal system, but must combine with other factors that favor underground transport and precipitation, i.e. presence of folds, fluid T, and solubility. This is true also for the distribution at depth of geothermal expressions. Distributions of ore features similar to those reported here were documented in the Goldstrike Gold System, North Carlin Trend (Dobak et al., 2020).

References

Arisi Rota, F., Brondi, A., Dessau, G., Franzini, M., Monte Amiata SpA, Stabilimento Minerario del Siele SpA, Stea, B. and Vighi, L., 1971. I Giacimenti Minerari. In: SIMP (Editor), La Toscana Meridionale, Società Italiana di Mineralogia e Petrologia, pp. 357-571.

Barelli, A., Ceccarelli, A., Dini, I., Fiordelisi, A., Giorgi, N., Lovari, F. and Romagnoli, P., 2010. A Review of the Mt. Amiata Geothermal System (Italy), World Geothermal Congress 2010, Bali, Indonesia, pp. 1-6.

Bonciani, F., Callegari, I., Conti, P., Cornamusini, G. and Carmignani, L., 2005. Neogene post-collisional evolution of the internal Northern Apennines: insights from the upper Fiora and Albegna valleys, (Mt. Amiata geothermal area, southern Tuscany). Bollettino della Societa Geologica Italiana, Vol Spec 3: 103-118.

Brogi, A., Fabbrini, L. and Liotta, D., 2011. Sb–Hg ore deposit distribution controlled by brittle structures: The case of the Selvena mining district (Monte Amiata, Tuscany, Italy). Ore Geology Reviews, 41(1): 35-48.

Calamai, A., Cataldi, R., Squarci, P. and Taffi, L., 1970. Geology, Geophysics and Hydrogeology of the Monte Amiata Geothermal Field. Geothermics, 1(Special Issue 1): 1-9.

Cataldi, R., 1967. Remarks on the geothermal research in the region of Monte Amiata (Tuscany - Italy). Bulletin Volcanologique, 30(1): 243-269.

Clever, H.L., Johnson, S.A. and Derrick, M.E., 1985. The Solubility of Mercury and Some Sparingly Soluble Mercury Salts in Water and Aqueous Electrolyte Solutions. Journal of Physical and Chemical Reference Data, 14(3): 631-680.

Conticelli, S., Boari, E.L., Burlamacchi, L., Cifelli, F., Moscardi, F., Laurenzi, M.A., Ferrari Pedraglio, L., Francalanci, L., Benvenuti, M.G., Braschi, E. and Manetti, P., 2015. Geochemistry and Sr-Nd-Pb isotopes of Monte Amiata Volcano, Central Italy: evidence for magma mixing between high-K calc-alkaline and leucititic mantle-derived magmas. Italian Journal of Geosciences, 134(2): 266-290.

De Castro, C., 1914. Genesi dei giacimenti cinabriferi del Monte Amiata, Memorie Descrittive Carta Geologica d'Italia, pp. 1-77.

Dobak, P.J., Robert, F., Barker, S.L.L., Vaughan, J.R., Eck, D., 2020. Chapter 15: Goldstrike Gold System, North Carlin Trend, Nevada, USA, Geology of the World's Major Gold Deposits and Provinces. Society of Economic Geologists, 23: 313-334.

Farella, E.M., Özdemir, E. and Remondino, F., 2021. 4D Building Reconstruction with Machine Learning and Historical Maps. Applied Sciences, 11(4): 1445.

Frondini, F., Caliro, S., Cardellini, C., Chiodini, G. and Morgantini, N., 2009. Carbon dioxide degassing and thermal energy release in the Monte Amiata volcanic-geothermal area (Italy). Applied Geochemistry, 24(5): 860-875.

Gianelli, G., Puxeddu, M., Batini, F., Bertini, G., Dini, I., Pandeli, E. and Nicolich, R., 1988. Geological model of a young volcano-plutonic system: the geothermal region of Monte Amiata (Tuscany, Italy). Geothermics, 17(5/6): 719-734.

Magi, F., Doveri, M., Menichini, M., Minissale, A. and Vaselli, O., 2019. Groundwater response to local climate variability: hydrogeological and isotopic evidences from the Mt. Amiata volcanic aquifer (Tuscany, central Italy). Rendiconti Lincei. Scienze Fisiche e Naturali, 30(1): 125-136.

Marroni, M., Moratti, G., Costantini, A., Conticelli, S., Benvenuti, M.G., Pandolfi, L., Bonini, M., Cornamusini, G. and Laurenzi, M.A., 2015. Geology of the Monte Amiata region, Southern Tuscany, Central Italy. Italian Journal of Geosciences, 134(2): 171-199.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. and Dubourg, V., 2011. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12: 2825-2830.

Pirajno, F., 2020. Subaerial hot springs and near-surface hydrothermal mineral systems past and present, and possible extraterrestrial analogues. Geoscience Frontiers, 11(5): 1549-1569.

Segreto, L., 1991. Monte Amiata. Il mercurio italiano. Strategie internazionali e vincoli extraeconomici. Ciriec. Storie d'impresa pubblica e di pubblico interesse. Franco Angeli (Ed.), 1-355 pp.

Sun, T., Chen, F., Zhong, L., Liu, W. and Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. Ore Geology Reviews, 109: 26-49.

Imputation methods for REE and Y in zircon

Carlos Carrasco-Godoy[1], Ian H. Campbell[1]
[1]Research School of Earth Sciences, Australian National University, Australia

Abstract. The shape of chondrite normalized zircon rare earth element (REE) patterns and Cerium (Ce) and Europium (Eu) anomalies provide insights into the conditions under which a zircon was crystallized. Therefore, using these elements as variables is common during machine learning algorithms training. However, several predictive models do not tolerate missing observations. Here, we propose two new methods, based on the lattice strain theory (Chondrite-Lattice) and Onuma diagrams (Chondrite-Onuma), using chondrite normalized values instead of partition coefficient. We compiled a dataset of ~1500 zircons, with known REE + Y concentrations, and used it to test and calibrate these methods. They require analyses of as few as five REEs to impute the missing REE in incomplete or legacy datasets and to estimate La, Pr, or Ce* in magmatic zircons. This allows for magma fertility models to have more variables available for model training and testing, in addition to providing a standardized method to calculate Ce*, La and Pr in zircon.

1 Introduction

Zircon is one of the most resilient accessory minerals in nature and is present in a wide range of geological environments. The geochemical and isotopic information it contains provides insights to understand fundamental earth processes (Finch and Hanchar 2003).

The rare earth elements (REE) and Y have been widely studied in zircon. Their concentration and shape of their chondrite normalized pattern give information about their source rock (e.g. Rubatto 2002; Zhu et al. 2022), and insights into the conditions under which the zircon crystallized (e.g. redox state, Loucks et al. 2020), among others.

Several studies have attempted to establish geochemical discriminants that identify zircons that were crystallized from “fertile” magma. Here we refer to fertility as the capacity of a magma to form a porphyry copper deposit. Pizarro et al. (2020), based on traditional statistical analysis, suggested the term porphyry indicator zircon (PIZ) for a zircon with characteristics that are considered indicators of fertility (for instance, Eu/Eu* > 0.4 or Ce/Nd > 1, among others).

Recent studies have applied machine learning algorithms to discriminate fertility among zircons, giving better predictions than traditional methods (e.g., Zou et al. 2022; Zhou et al. 2022). However, a drawback of several machine learning models is that they are unable to deal with missing data (Kuhn 2020). If the amount of missing information is small, imputation techniques can be used during model training and cross-validation. Some of the common methods for imputation include the replacement of missing values for the mean or median of the variable or more complex methods that use predictive models based on the non-missing variables (Kuhn 2020). Particularly, in the case of REE, it is possible to impute using the geometric mean of adjacent chondrite-normalized REE of the element to replace. However, this method does not consider the curvature of the zircon pattern and it cannot be applied if two or more consecutive REE are missing or next to Ce and Eu.

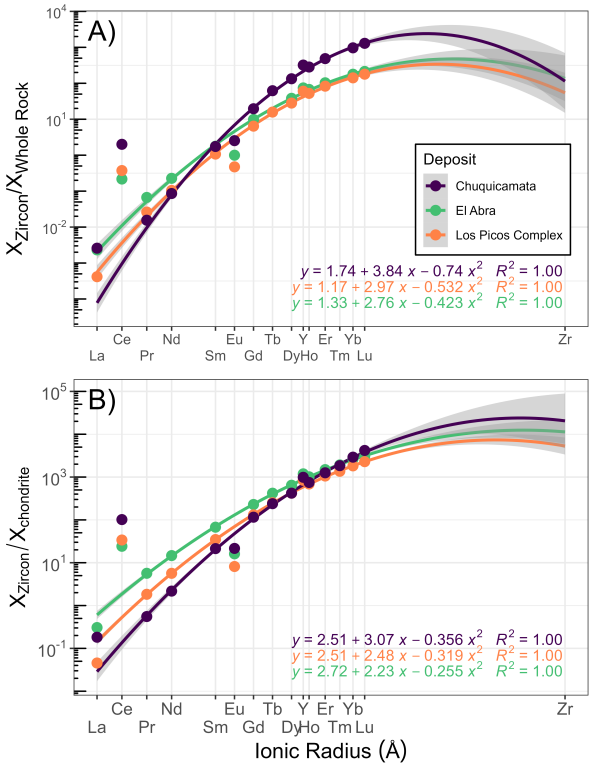

Figure 1. Onuma diagrams for three zircons from Ballard et al. (2002). Equations for each parabola are colour-coded, where y = $\log_{10}$(x). The X-axis is reversed. The shading indicates a 95% confidence interval for the quadratic fit. **A** Partition coefficient estimates vs Ionic radii. **B** Chondrite normalised values vs ionic radii. Chondrite values were taken from Palme and O’Neill (2014)

We demonstrate two methods that derive from Onuma diagrams (Chondrite-Onuma) and the lattice strain theory (Chondrite-Lattice), but using chondrite normalized values instead of partition coefficients, to impute missing REE and to calculate La and Pr where their concentrations are low; together with Ce* and Eu* anomalies in magmatic zircons.

2 Onuma diagrams and the lattice strain theory

Onuma et al. (1968) showed a relationship between an element's ionic radius and its partition coefficient

for a given mineral. All the elements of the same charge (e.g., +3 for most REE), describe a quadratic function with decreasing partition coefficient as the ionic radii are further displaced from the ionic radii of the lattice site where the substitution occurs (Figure 1A).

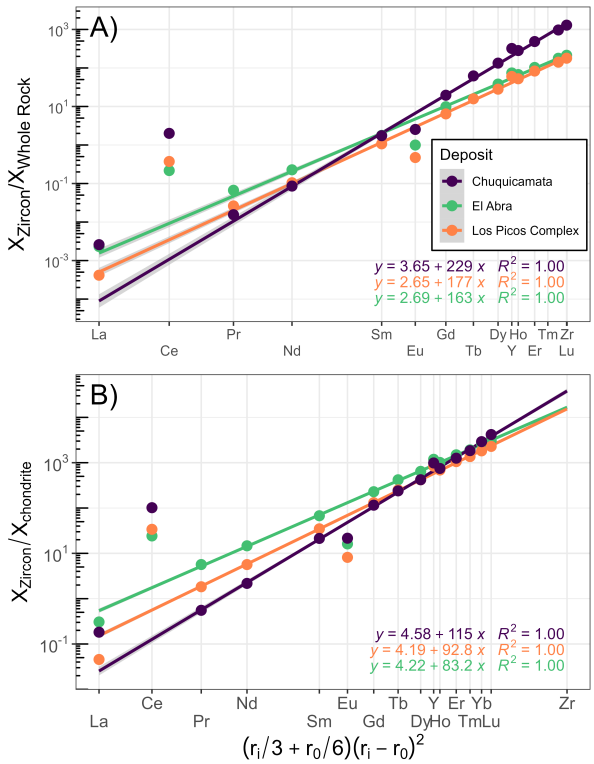

Figure 2. Lattice strain regression for the same zircons as Figure 1. Equations for each parabola are colour-coded, where y = $\log_{10}$(x). The X-axis is reversed and represents the misfit parameter from the lattice strain equation. The shading indicates a 95% confidence interval for the quadratic fit. **A** Partition coefficient estimates vs Ionic radii. **B** Chondrite normalised values vs ionic radii. Chondrite values from Palme and O'Neill (2014)

The lattice strain theory (Blundy and Wood 1994) explains the relationship between the partition coefficients and the misfit between the lattice site in the crystal and the actual ionic radius of an element occupying that space. The elements with the same charge describe a linear relationship (Figure 2A) where the partition coefficient decreases as the difference between the effective lattice site (r_0) and the cation radius increases (r_i).

These methods can be used to impute missing REE, or to calculate La or Ce* in zircons (e.g., Burnham 2020; Loader et al. 2022). However, they require precise knowledge of the melt composition, which is often unavailable or cannot be analysed (e.g., detrital zircons). However, if chondrite-normalized values are used instead of partition coefficients, they have the same quadratic and linear relationship for the Onuma diagrams and the lattice strain theory, respectively. Therefore, we have used this empirical observation to overcome the limitations of using partition coefficients.

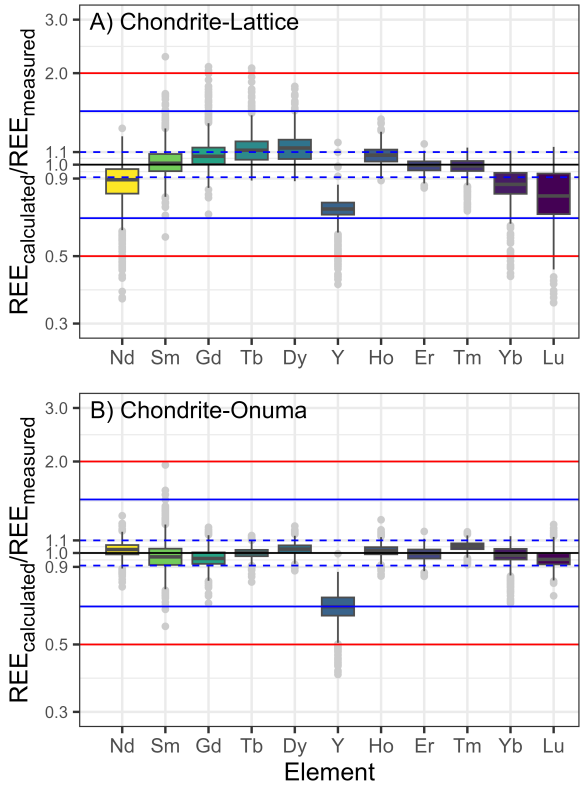

Figure 3. Boxplots for the ratio between calculated and measured concentrations for each REE, excluding La, Pr, Ce and Eu. Calculated values will plot along the y = 1 line if they are equal to the measured values. The blue segmented lines, blue continuous lines and red lines are reference lines for a discrepancy between measured and calculated values of 10%, 50% and 100%, respectively. Each box contains 50% of the data. Each box and whiskers represent 99.3% of the data. The grey dots are outliers and represent 0.7% of the data. **A** Chondrite-Lattice. **B** Chondrite-Onuma methods.

Figures 1 and 2 show the comparison of Onuma diagrams and lattice strain regressions for three zircons from Ballard et al. (2002) using partition coefficients and chondrite normalized values. We have used whole rock compositions as a proxy for melt composition and excluded Ce and Eu because can be present in more than one oxidation state. We have also excluded La and Pr due to their susceptibility to LREE contamination (Claiborne et al. 2018; Zou et al. 2019; Zhong et al. 2019; Burnham 2020). Both cases have an R^2 higher than 0.99.

3 Data Compilation

We have compiled nearly 1500 zircons from the literature that have the whole range of REE + Y analysed with no missing values (Ballard et al. 2002; Buret et al. 2016, 2017; Burnham and Berry 2017; Loader et al. 2017; Large et al. 2018, 2020, 2021; Lu et al. 2019; Zhu et al. 2020; Pizarro et al. 2020). Most of this data has been filtered for LREE-rich or titanite inclusions in their original publications. This dataset was used to evaluate the performance of the Chondrite-Onuma and Chondrite-Lattice

methods for different scenarios of missing REE. We have used a separate small dataset (Colombini et al. 2011; Taylor et al. 2015; Claiborne et al. 2018) of melt-zircon pairs to calculate La, Ce* and Pr using the lattice strain theory. The calculated values were used to compare the methods proposed in this study with those of Zhong et al. (2019). We have tested almost exclusively magmatic zircons.

4 Imputation of REE + Y

Figures 3A and 3B show the ratio between predicted and measured concentrations for the complete dataset for each REE and Y for the Chondrite-Lattice and Chondrite-Onuma methods, respectively.

Overall, the Chondrite-Lattice method (Figure 3A) gives predictions that disagree, in median, up to ~30% (Figure 3A), which is higher for the HREE section of the REE pattern. In contrast, the Chondrite-Onuma method (Figure 3B) gives predicted values that disagree by <5% with respect to the measured concentrations. Yttrium is underestimated by both methods, which is also observed in Figures 1 and 2 where the measured values fall above the regression line. In both cases, if the disagreement is systematic, the predictions can be calibrated by projecting them to the ratio = 1 line.

The cases in Figure 3 are overfitted models because all the REEs are used for prediction. We have reproduced 6 different scenarios where a different number of REEs are missing or censored (e.g., where not analysed or below detection limits). Both methods give good results if only three evenly distributed REEs, and identical results to those in Figure 3 if 4 or more REEs are used, excluding La, Pr, Ce, Eu and Y during modelling.

5 Calculated La, Ce* and Pr

Measured La and Pr concentrations in zircon are often considered unreliable (Claiborne et al. 2018; Zou et al. 2019; Zhong et al. 2019; Burnham 2020) and Ce anomalies are traditionally calculated using the geometric mean of these elements to obtain Ce*. Thus, their concentrations cannot be used as reference values. Therefore, we have used the best estimates for La and Pr, which are those derived from the lattice strain theory (using partition coefficients).

Figure 4 shows the calculated values for different methods vs the lattice strain estimates for La and Ce. The calculated Ce is equal to Ce*. Both cases tend to slightly overestimate La and Ce* when compared to the lattice strain theory.

Zhong et al. (2019) proposed a method to calculate Ce anomalies based on a logarithmic regression between the REE atomic number and their chondrite normalized concentrations. However, their method gives values that underestimate La and Ce* up to 2 and 1 orders of magnitude, respectively, compared to the lattice strain estimates. There is no difference between the methods for estimating Pr.

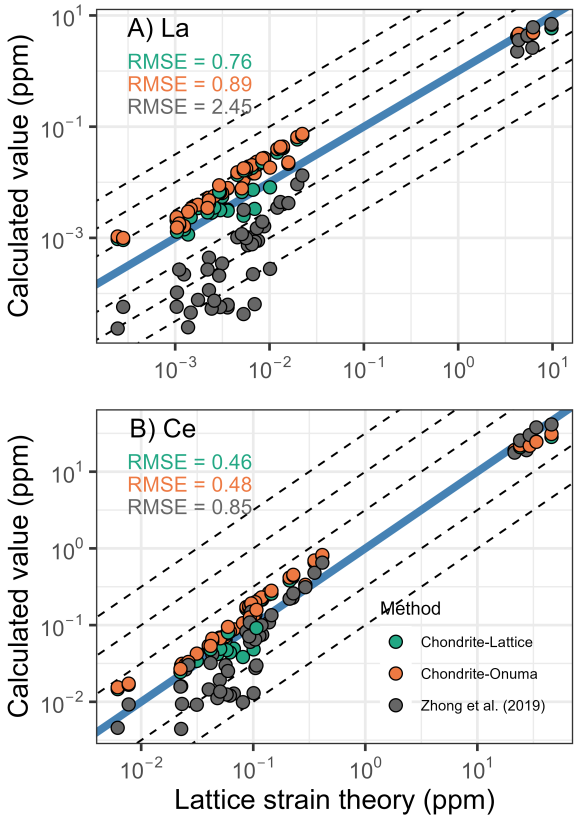

Figure 4. A La and **B** Ce concentrations obtained by the lattice strain theory vs the calculated values using the chondrite-lattice strain, chondrite-Onuma and the Zhong et al. (2019) methods. Calculated Ce is equal to Ce* in this case. The blue line is the identity line. The segmented lines are 0.5 orders of magnitude apart. Root mean square errors (RMSE) are colour coded according to the method, lower values are better. Zircon-glass pairs are from Colombini et al. (2011), Taylor et al. (2015) and Claiborne et al. (2018)

6 Implications and conclusions

The Chondrite-Lattice and Chondrite-Onuma can be used to impute missing REE in zircon, which gives complete datasets for training and testing of machine learning models. This is especially useful in the case of legacy data. Furthermore, the methods provide a standardized procedure for calculating La, Ce* and Pr so they can be used as input variables for new fertility models based on zircon geochemistry. We recommend the Chondrite-Onuma method for imputation and the Chondrite-Lattice to calculate La, Ce* and Pr.

7 Code availability

The methods used in the work have been compiled in the ImputeREE package for the R programming language. The package is accessible in the CRAN network and on the development site at https://github.com/cicarrascog/imputeREE. A

companion app is accessible from the development website.

Acknowledgements

This work was funded by the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO BECAS CHILE/2019 - 72200364.

References

Ballard JR, Palin MJ, Campbell IH (2002) Relative oxidation states of magmas inferred from Ce(IV)/Ce(III) in zircon: application to porphyry copper deposits of northern Chile. Contributions to Mineralogy and Petrology 144:347–364. https://doi.org/10.1007/s00410-002-0402-5

Blundy J, Wood B (1994) Prediction of crystal–melt partition coefficients from elastic moduli. Nature 372:452–454. https://doi.org/10.1038/372452a0

Buret Y, von Quadt A, Heinrich C, et al (2016) From a long-lived upper-crustal magma chamber to rapid porphyry copper emplacement: Reading the geochemistry of zircon crystals at Bajo de la Alumbrera (NW Argentina). Earth and Planetary Science Letters 450:120–131. https://doi.org/10/f8242b

Buret Y, Wotzlaw J-F, Roozen S, et al (2017) Zircon petrochronological evidence for a plutonic-volcanic connection in porphyry copper deposits. Geology 45:623–626. https://doi.org/10.1130/G38994.1

Burnham AD (2020) Key concepts in interpreting the concentrations of the rare earth elements in zircon. Chemical Geology 551:119765. https://doi.org/10.1016/j.chemgeo.2020.119765

Burnham AD, Berry AJ (2017) Formation of Hadean granites by melting of igneous crust. Nature Geoscience 10:457–461. https://doi.org/10.1038/ngeo2942

Claiborne LL, Miller CF, Gualda GAR, et al (2018) Zircon as Magma Monitor: Robust, Temperature-Dependent Partition Coefficients from Glass and Zircon Surface and Rim Measurements from Natural Systems. In: Moser DE, Corfu F, Darling JR, et al. (eds) Geophysical Monograph Series. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 1–33

Colombini LL, Miller CF, Gualda GAR, et al (2011) Sphene and zircon in the Highland Range volcanic sequence (Miocene, southern Nevada, USA): elemental partitioning, phase relations, and influence on evolution of silicic magma. Miner Petrol 102:29. https://doi.org/10.1007/s00710-011-0177-3

Finch RJ, Hanchar JM (2003) Structure and Chemistry of Zircon and Zircon-Group Minerals. Reviews in Mineralogy and Geochemistry 53:1–25. https://doi.org/10.2113/0530001

Kuhn M (2020) Feature engineering and selection: a practical approach for predictive models. CRC Press, Taylor & Francis Group, Boca Raton London New York

Large SJE, Buret Y, Wotzlaw JF, et al (2021) Copper-mineralised porphyries sample the evolution of a large-volume silicic magma reservoir from rapid assembly to solidification. Earth and Planetary Science Letters 563:116877. https://doi.org/10.1016/j.epsl.2021.116877

Large SJE, Quadt A von, Wotzlaw J-F, et al (2018) Magma Evolution Leading to Porphyry Au-Cu Mineralization at the Ok Tedi Deposit, Papua New Guinea: Trace Element Geochemistry and High-Precision Geochronology of Igneous Zircon. Economic Geology 113:39–61. https://doi.org/10.5382/econgeo.2018.4543

Large SJE, Wotzlaw J-F, Guillong M, et al (2020) Resolving the timescales of magmatic and hydrothermal processes associated with porphyry deposit formation using zircon U–Pb petrochronology. Geochronology 2:209–230. https://doi.org/10.5194/gchron-2-209-2020

Loader MA, Nathwani CL, Wilkinson JJ, Armstrong RN (2022) Controls on the magnitude of Ce anomalies in zircon. Geochimica et Cosmochimica Acta. https://doi.org/10.1016/j.gca.2022.03.024

Loader MA, Wilkinson JJ, Armstrong RN (2017) The effect of titanite crystallisation on Eu and Ce anomalies in zircon and its implications for the assessment of porphyry Cu deposit fertility. Earth and Planetary Science Letters 472:107–119. https://doi.org/10.1016/j.epsl.2017.05.010

Loucks RR, Fiorentini ML, Henríquez GJ (2020) New magmatic oxybarometer using trace elements in zircon. J Petrology. https://doi.org/10.1093/petrology/egaa034

Lu Y, Smithies R, Wingate M, et al (2019) Zircon fingerprinting of magmatic-hydrothermal systems in the Archean Yilgarn Craton

Onuma N, Higuchi H, Wakita H, Nagasawa H (1968) Trace element partition between two pyroxenes and the host lava. Earth and Planetary Science Letters 5:47–51. https://doi.org/10.1016/S0012-821X(68)80010-X

Pack A, Russell SS, Shelley JMG, van Zuilen M (2007) Geo- and cosmochemistry of the twin elements yttrium and holmium. Geochimica et Cosmochimica Acta 71:4592–4608. https://doi.org/10.1016/j.gca.2007.07.010

Pizarro H, Campos E, Bouzari F, et al (2020) Porphyry indicator zircons (PIZs): Application to exploration of porphyry copper deposits. Ore Geology Reviews 126:103771. https://doi.org/10.1016/j.oregeorev.2020.103771

Rubatto D (2002) Zircon trace element geochemistry: partitioning with garnet and the link between U–Pb ages and metamorphism. Chemical Geology 184:123–138. https://doi.org/10.1016/S0009-2541(01)00355-2

Taylor RJM, Harley SL, Hinton RW, et al (2015) Experimental determination of REE partition coefficients between zircon, garnet and melt: a key to understanding high-T crustal processes. Journal of Metamorphic Geology 33:231–248. https://doi.org/10.1111/jmg.12118

Zhong S, Seltmann R, Qu H, Song Y (2019) Characterization of the zircon Ce anomaly for estimation of oxidation state of magmas: a revised Ce/Ce* method. Miner Petrol 113:755–763. https://doi.org/10.1007/s00710-019-00682-y

Zhou Y, Zhang Z, Yang J, et al (2022) Machine Learning and Singularity Analysis Reveal Zircon Fertility and Magmatic Intensity: Implications for Porphyry Copper Potential. Nat Resour Res. https://doi.org/10.1007/s11053-022-10122-y

Zhu Z, Campbell IH, Allen CM, et al (2022) The temporal distribution of Earth's supermountains and their potential link to the rise of atmospheric oxygen and biological evolution. Earth and Planetary Science Letters 580:117391. https://doi.org/10.1016/j.epsl.2022.117391

Zhu Z, Campbell IH, Allen CM, Burnham AD (2020) S-type granites: Their origin and distribution through time as determined from detrital zircons. Earth and Planetary Science Letters 536:116140. https://doi.org/10.1016/j.epsl.2020.116140

Zou S, Chen X, Brzozowski MJ, et al (2022) Application of Machine Learning to Characterizing Magma Fertility in Porphyry Cu Deposits. Journal of Geophysical Research: Solid Earth 127:e2022JB024584. https://doi.org/10.1029/2022JB024584

Zou X, Qin K, Han X, et al (2019) Insight into zircon REE oxybarometers: A lattice strain model perspective. Earth and Planetary Science Letters 506:87–96. https://doi.org/10.1016/j.epsl.2018.10.031

Zircon geochemistry: insights into porphyry copper deposits fertility from machine learning applications

Carlos Carrasco-Godoy[1], Ian H. Campbell[1], Yamila Cajal[1,2]
[1]Research School of Earth Sciences, The Australian National University
[2]Centre for Ore Deposit and Earth Sciences, University of Tasmania

Abstract. Zircon is a widespread mineral in igneous rocks and its geochemistry allows the reconstruction of the age and conditions of magma formation. This makes it possible to assess how a zircon that grew from a magma that formed a porphyry deposit differs from one that did not. Several studies have proposed geochemical signatures of zircon that can be used to distinguish between ore-bearing and barren magmas, such as Eu and Ce anomalies. For this study, ca. 18,000 zircons were compiled, including zircons from more than thirty deposits, which are compared with zircons from barren intrusions and rivers. We have trained different models for predicting fertility, focusing on the insights that can be obtained from these models. An oversampled random forest gave the best results with a ROC AUC of 0.977. The results suggest that the fertility signal in zircons becomes stronger as the porphyry systems evolve. The model reveals that there are differences in the LREE content of fertile and barren zircons but not in the HREE. The results show that the changes in the Ce anomaly in zircon are controlled by changes in Pr and La rather than changes in Ce.

1 Introduction

Porphyry copper deposits are the source of ~75% of the world's copper and ~20% of its gold (Sillitoe 2010). Although global consumption of copper is expected to increase in the next 50 years, the discovery of new deposits has decreased over the last decades (Elshkaki et al. 2016; Schodde 2019). Furthermore, the use of copper has become fundamental in the transition from fossil fuels to renewable energies (Månberger and Stenqvist 2018). Therefore, it is critical to improve our current exploration methods to assure copper's future demand.

Zircon is a widespread accessory mineral present in a range of geological environments. It contains diagnostic elements and isotopes in its crystal structure, which in addition to its resistance to chemical and physical weathering (Finch and Hanchar 2003), make it suitable for a wide range of applications from geochronology (e.g., Compston and Pidgeon 1986) to identifying Earth's major periods of mountain building (Zhu et al. 2022).

Zircon can be found in igneous rocks that range in composition from intermediate to felsic and are important reservoirs for incompatible elements (e.g., U, Hf, REE, among others) in their host rock (Finch and Hanchar 2003). The content of several of these elements in the zircon lattice varies as physicochemical properties of the magma change, such as temperature (e.g. Ti, Ferry and Watson 2007), oxygen fugacity (Ce and U, Loucks et al. 2020), the magma composition or co-crystallizing minerals (Loader et al. 2017, 2022; Zhong et al. 2018)

Therefore, we consider the hypothesis that the conditions required for the formation of an economic porphyry copper deposit led to zircons with a unique trace elements geochemistry that can be used to predict fertility. Several studies have used traditional methods (e.g., univariate statistics) to define geochemical characteristics to classify fertile zircons (Dilles et al. 2015; Lu et al. 2016; Pizarro et al. 2020; Leslie et al. 2021). Pizarro et al., (2020) summarized these characteristics and suggest the term porphyry indicator zircon (PIZ) to be used for zircons with high Hf concentrations (> 8,750 ppm), high Ce/Nd (> 1), Eu/Eu* (> 0.4), (10,000xEu/Eu*)/Y (> 1), (Ce/Nd)/Y (> 0.01) ratios, intermediate Th/U ratios and low Dy/Yb (< 0.3).

Recent studies (Zou et al. 2022; Zhou et al. 2022) have used machine learning algorithms to distinguish fertility in zircon crystals. They show that univariate criteria (e.g., Pizarro et al. 2020) have accuracies, depending on the dataset, of between 80 to 90% whereas machine learning algorithms can reach up to 94% accuracy depending on the model used.

2 Data and Methods

For this study, ca. 18,000 zircons from ore-bearing and barren igneous rocks, and detrital grains from rivers, have been carefully compiled from the literature. The global dataset includes zircons from more than thirty porphyry copper deposits. The porphyry dataset contains nearly 5,000 zircons. The geological information of each grain (e.g., age, host rock composition, etc.) was carefully compiled where available. Furthermore, each zircon was labelled according to the deposit as Cu, Cu±Au, Cu±Mo or Cu±Au±Mo porphyry deposits, as well as according to their temporal distribution within the deposit as precursor or pre-, syn- or post-mineralization if indicated in the literature. The barren dataset considers 13,000 zircons from barren sources divided into three subsets: the river (n ~7,000, Zhu et al. 2020), GEOROC (n ~5,000, excluding porphyry-associated publications, Lehnert et al. 2000) and the barren subsets (n ~ 800). Considering that economic porphyry copper deposits are rare (Richards 2022) it is a reasonable assumption to consider that the detrital zircons, from Earth's major rivers, are barren. The GEOROC subset considers zircon from the GEOROC database. The barren subset considers zircon grains spatially or

temporally associated with porphyry copper deposits but that are considered barren within the district.

Cerium anomalies have been associated with an increase in the oxygen fugacity of the magma, but its calculation is difficult due to the low La and Pr content of zircon (Ballard et al. 2002; Zou et al. 2019; Zhong et al. 2019). Here, we have used a new empirical method to calculate Ce (Carrasco-Godoy and Campbell, in review), based on the lattice strain theory, to impute any missing REE and to calculate standardized La, Pr concentrations and Ce anomalies.

We have trained a random forest (Breiman 2001), decision tree and logistic regression. Each model was trained using five times repeated 10-fold cross-validation, considering raw and oversampled datasets to account for class unbalance. Each model was trained with and without centred log-ratio (CLR) transformation of the data (Aitchison 1984). The models were ranked according to their performance metrics: the area under the receiver operating characteristic curve (ROC-AUC) which indicates the possibility of a random fertile zircon ranked higher than a randomly selected barren zircon; sensitivity which is the proportion of fertile zircons correctly identified among the fertile zircons; and specificity which is the proportion of barren zircons correctly identified among the total of barren zircons. The three best models had their hyperparameters tuned using a mix of grid search and simulated annealing (Kuhn and Silge 2022).

Here, we present the results of the best model with a detailed analysis of the model predictions on individual probabilities in addition to the outcome prediction. Then, we link the results of the models to geological processes than can lead to the formation of fertile zircons.

Data processing, feature selection and model training and testing were performed in R programming language using the base (R Core Team 2022), tidyverse (Wickham et al. 2019) and tidymodels (Kuhn and Wickham 2020) metapackages. Random forest and decision trees were fitted with the Rpart and ranger packages (Wright et al. 2021; Therneau and Atkinson 2022).

3 Results and discussions

The best model was an oversampled random forest without CLR transformation. Centred log-ratio transformation has been widely applied to remove the effects of the constant sum in closed data (e.g., major elements that sum 100%, Aitchison 1984). However, we did not observe any improvement between CLR and the raw data. We attribute this to the closed nature of the zircon crystal lattice which would not allow more elements than their formula unit can hold or the non-parametric nature of Random Forest models (Nathwani et al. 2022).

The metrics during model training were the area under the receiving operating characteristic (ROC AUC) of 0.977 ± 0.05%, sensitivity of 0.87 ± 0.28% and specificity of 0.95 ± 0.11%, whereas in the testing set where a ROC AUC of 0.978, sensitivity of 0.89 and specificity of 0.94.

An additional set with zircons neither in the testing nor training set was used for external validation. Most of the data are from deposits, rivers or barren intrusions that were not present in the training or testing sets. The results are ROC AUC of 0.978, but with a lower sensitivity (0.707) and a higher specificity (0.98).

Most binary classification algorithms estimate membership probabilities and use 0.5 as a boundary. Figure 1 shows a normalized histogram of the fertile zircon membership probability according to each subset for the predictions in the testing set. Only the deposit dataset contains zircons labelled as fertile. In this histogram, each bin represents the relative proportions of each class rather than the total counts. The individual probabilities in the testing set show that most of the zircons, with more than 50% probability of being fertile, belong to the testing set and misclassified zircons fall to a minimum after the probability is higher than 90%. Therefore, depending on the degree of confidence the boundary to define fertile zircons can be adjusted. A shift of the classification boundary (red line in Figure 1) to the right would increase the confidence of a zircon being fertile at the cost of discarding more fertile zircons as barren (an increase in the specificity and a decrease in the sensitivity).

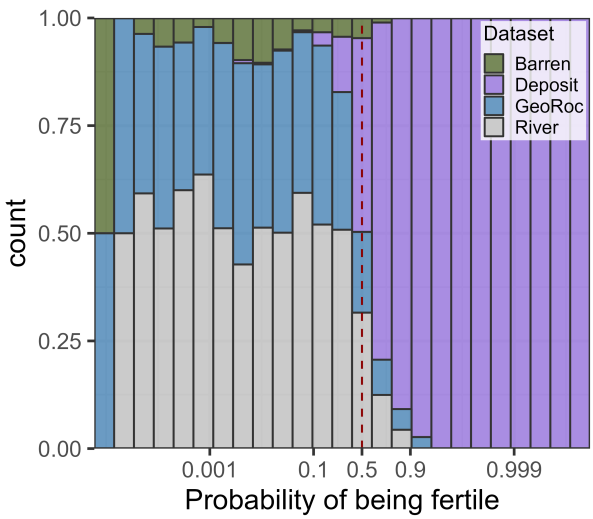

Figure 1. Normalized histogram of the predicted probabilities from oversampled random forest in the testing set. The colours indicate the different subsets from the main dataset. X-axis is logit-transformed. The red dashed line indicates a probability of 0.5. Each bin shows the proportion of zircons from each dataset that falls within that probability.

The analysis of the misclassified grains for each class membership, considering a probability boundary of 0.5, decreases from precursor (34%) to post-mineral (1.6%) which suggests an increase in the fertile signature as the porphyritic systems evolve. In contrast, barren zircons mislabelled as fertile reach 18% for the barren subset, 14% for I-type granites, 5% for detrital zircons from rivers and

1.9% for S-type granites. The misclassification percentage can be used as a benchmark of how many mislabelled zircons can be expected when applying a predictive model to unknown data.

The random forest corrected Gini impurity importance ranking (Nembrini et al. 2018) for the features included in the model training indicates that Eu anomaly and Ce anomalies are the most important variables, which is consistent with observations by other authors (Loader et al. 2017; Zou et al. 2022; Zhou et al. 2022). There is a decrease in the importance from LREE to HREE. The study of the median between ore-associated and barren zircon shows little to no difference for the HREE. In contrast, there is an increase in the median difference from pre-mineral to post-mineralization zircons as the ionic radius of the REE increases. Lanthanum and Pr have the highest variation, a factor of 6 and 4, respectively. However, these variations should be taken as indicative only, due to the difficulty of measuring LREE in zircon makes it difficult to gauge the real magnitude of these changes. In contrast, Ce concentrations are relatively constant with a maximum median variation of less than 1.6. This is consistent with the observations of Loader et al. 2022. We suggest that the variation in Ce anomaly in zircon is mainly controlled by the depletion of La and Pr, rather than the variation of Ce.

3 Conclusions and future work

We have shown that random forest models provide an improvement in predictions of zircon fertility when applied to porphyry copper deposits over traditional methods based on univariate statistics. We have compiled the largest and most complete dataset to date for training fertility models using zircon geochemistry.

Insights obtained from the model suggest that there is an increase in the fertility signal in zircons as porphyry systems evolve. Ce and Eu anomalies are the best predictors of fertility and the LREE have more weight in the prediction of fertility than HREE. Changes in Ce anomaly are likely controlled by variations in La and Pr rather than Ce concentrations.

Future work considers a model pipeline to classify detrital zircons according to the deposit type they are associated with, Cu, Cu-Au or Cu-Mo, and their temporality (pre- to post-mineralization).

Finally, each model can be tailored to increase the probability of correct classification by taking into account their geological and geographic context.

Acknowledgements

This work was funded by the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO BECAS CHILE/2019 - 72200364.

References

Aitchison J (1984) The statistical analysis of geochemical compositions. Mathematical Geology 16:531–564. https://doi.org/10.1007/BF01029316

Ballard JR, Palin MJ, Campbell IH (2002) Relative oxidation states of magmas inferred from Ce(IV)/Ce(III) in zircon: application to porphyry copper deposits of northern Chile. Contributions to Mineralogy and Petrology 144:347–364. https://doi.org/10.1007/s00410-002-0402-5

Breiman L (2001) Random Forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324

Compston W, Pidgeon RT (1986) Jack Hills, evidence of more very old detrital zircons in Western Australia. Nature 321:766–769. https://doi.org/10.1038/321766a0

Dilles JH, Kent AJR, Wooden JL, et al (2015) Zircon Compositional Evidence For Sulfur-Degassing From Ore-Forming Arc Magmas. Economic Geology 110:241–251. https://doi.org/10.2113/econgeo.110.1.241

Elshkaki A, Graedel TE, Ciacci L, Reck BK (2016) Copper demand, supply, and associated energy use to 2050. Global Environmental Change 39:305–315. https://doi.org/10.1016/j.gloenvcha.2016.06.006

Ferry JM, Watson EB (2007) New thermodynamic models and revised calibrations for the Ti-in-zircon and Zr-in-rutile thermometers. Contrib Mineral Petrol 154:429–437. https://doi.org/10/bzbzdw

Finch RJ, Hanchar JM (2003) Structure and Chemistry of Zircon and Zircon-Group Minerals. Reviews in Mineralogy and Geochemistry 53:1–25. https://doi.org/10.2113/0530001

Kuhn M, Silge J (2022) Tidy Modeling with R: A Framework for Modeling in the Tidyverse. O'REILLY MEDIA, INC, USA, S.l.

Kuhn M, Wickham H (2020) Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.

Lehnert K, Su Y, Langmuir CH, et al (2000) A global geochemical database structure for rocks. Geochemistry, Geophysics, Geosystems 1:. https://doi.org/10.1029/1999GC000026

Leslie C, Meffre S, Cooke DR, et al (2021) Complex Petrogenesis of Porphyry-Related Magmas in the Cowal District, Australia: Insights from LA ICP-MS Zircon Imaging. 24:22

Loader MA, Nathwani CL, Wilkinson JJ, Armstrong RN (2022) Controls on the magnitude of Ce anomalies in zircon. Geochimica et Cosmochimica Acta. https://doi.org/10.1016/j.gca.2022.03.024

Loader MA, Wilkinson JJ, Armstrong RN (2017) The effect of titanite crystallisation on Eu and Ce anomalies in zircon and its implications for the assessment of porphyry Cu deposit fertility. Earth and Planetary Science Letters 472:107–119. https://doi.org/10.1016/j.epsl.2017.05.010

Loucks RR, Fiorentini ML, Henríquez GJ (2020) New magmatic oxybarometer using trace elements in zircon. J Petrology. https://doi.org/10.1093/petrology/egaa034

Lu Y-J, Loucks RR, Fiorentini M, et al (2016) Zircon Compositions as a Pathfinder for Porphyry Cu ± Mo ± Au Deposits. In: Richards JP (ed) Tectonics and Metallogeny of the Tethyan Orogenic Belt. Society of Economic Geologists, p 0

Månberger A, Stenqvist B (2018) Global metal flows in the renewable energy transition: Exploring the effects of substitutes, technological mix and development. Energy Policy 119:226–241. https://doi.org/10.1016/j.enpol.2018.04.056

Nathwani CL, Wilkinson JJ, Fry G, et al (2022) Machine learning for geochemical exploration: classifying metallogenic fertility in arc magmas and insights into porphyry copper deposit formation. Miner Deposita. https://doi.org/10.1007/s00126-021-01086-9

Nembrini S, König IR, Wright MN (2018) The revival of the Gini importance? Bioinformatics 34:3711–3718. https://doi.org/10.1093/bioinformatics/bty373
Pizarro H, Campos E, Bouzari F, et al (2020) Porphyry indicator zircons (PIZs): Application to exploration of porphyry copper deposits. Ore Geology Reviews 126:103771. https://doi.org/10.1016/j.oregeorev.2020.103771
R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
Richards JP (2022) Porphyry copper deposit formation in arcs: What are the odds? Geosphere 18:130–155. https://doi.org/10.1130/GES02086.1
Schodde R (2019) Trends in exploration
Sillitoe RH (2010) Porphyry Copper Systems. Economic Geology 105:3–41. https://doi.org/10.2113/gsecongeo.105.1.3
Therneau T, Atkinson B (2022) rpart: Recursive Partitioning and Regression Trees
Wickham H, Averick M, Bryan J, et al (2019) Welcome to the Tidyverse. Journal of Open Source Software 4:1686. https://doi.org/10.21105/joss.01686
Wright MN, Wager S, Probst P (2021) ranger: A Fast Implementation of Random Forests
Zhong S, Feng C, Seltmann R, et al (2018) Can magmatic zircon be distinguished from hydrothermal zircon by trace element composition? The effect of mineral inclusions on zircon trace element composition. Lithos 314–315:646–657. https://doi.org/10.1016/j.lithos.2018.06.029
Zhong S, Seltmann R, Qu H, Song Y (2019) Characterization of the zircon Ce anomaly for estimation of oxidation state of magmas: a revised Ce/Ce* method. Miner Petrol 113:755–763. https://doi.org/10.1007/s00710-019-00682-y
Zhou Y, Zhang Z, Yang J, et al (2022) Machine Learning and Singularity Analysis Reveal Zircon Fertility and Magmatic Intensity: Implications for Porphyry Copper Potential. Nat Resour Res. https://doi.org/10.1007/s11053-022-10122-y
Zhu Z, Campbell IH, Allen CM, et al (2022) The temporal distribution of Earth's supermountains and their potential link to the rise of atmospheric oxygen and biological evolution. Earth and Planetary Science Letters 580:117391. https://doi.org/10.1016/j.epsl.2022.117391
Zhu Z, Campbell IH, Allen CM, Burnham AD (2020) S-type granites: Their origin and distribution through time as determined from detrital zircons. Earth and Planetary Science Letters 536:116140. https://doi.org/10.1016/j.epsl.2020.116140
Zou S, Chen X, Brzozowski MJ, et al (2022) Application of Machine Learning to Characterizing Magma Fertility in Porphyry Cu Deposits. Journal of Geophysical Research: Solid Earth 127:e2022JB024584. https://doi.org/10.1029/2022JB024584
Zou X, Qin K, Han X, et al (2019) Insight into zircon REE oxy-barometers: A lattice strain model perspective. Earth and Planetary Science Letters 506:87–96. https://doi.org/10.1016/j.epsl.2018.10.031

Building up a new mineral exploration indicator (MEI) system based on big data and AI technology

Junling Zhang[1], Huayong Chen[1*]
[1]Key Laboratory of Mineralogy and Metallogeny, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China

Abstract: In this paper we present a new mineral exploration indicator system based on big data. It focusses on specific types of mineral deposits and is guided by ore-forming models, driven by AI, statistical analysis, and other algorithm libraries. It intelligently extracts and integrates multi-scale and multi-type exploration indicators to form a system that can automatically evolve with exploration activities. The new system is a bridge that connects big data for mineral exploration and accurate prospect prediction, and may be an effective tool for guiding exploration.

1 Introduction of MEI

The concept of a "mineral exploration indicator" (MEI) in this paper refers to the characteristic of exploration results obtained through variable mineral exploration methods that have a mineralization-indicating effect. It is the integration of prospecting signs and exploration methods, which can fully reflect the relationship between prospecting signs, exploration activities, and exploration methods.

MEI has three key features: (1) spatial features, i.e., all exploration indicators are associated with specific spatial locations; (2) exploration method, i.e., all exploration indicators are obtained through specific exploration activities, within a certain range of exploration scales, and through the use of certain exploration methods; and (3) mineralization-indicating effect, i.e., the essential features of all MEI is that they have direct or indirect mineralization-indicating effects.

2 Classification of MEI

MEI can be subdivided in detail according to different classification criteria such as exploration methods, mineralization-indicating effect, spatial dimensions, data types, exploration scales, and types of mineral deposits.

(1) According to the exploration methods, they can be divided into geological indicators, mineral indicators, geophysical indicators, geochemical indicators, remote sensing indicators. (2) According to the mineralization-indicating types, they can be classified into abundance indicators, proximity indicators, and anomaly indicators (Yousefi et al., 2019). (3) According to the spatial dimensions, they can be classified into one-dimensional indicators, two-dimensional indicators, three-dimensional indicators, and four-dimensional spatiotemporal indicators. (4) According to the information expression types, they can be divided into qualitative indicators, morphological indicators, and quantitative indicators. (5) According to different exploration scales, they can be classified into global indicators, mineralization domain indicators, metallogenic province indicators, metallogenic belt indicators, ore field indicators, and deposit indicators. (6) According to the types of mineral deposit, they can be classified into MEI for porphyry copper deposits, skarn-type deposits, ion-adsorption type rare earth deposit, orogenic gold deposit.

3 Extraction of MEI

The key point of the method for extracting MEI is to analyse and judge whether the exploration data has a mineralization-indicating effect. The extraction methods for different types of MEI have similarities but also significant differences.

The extraction method for abundance indicators mainly includes statistical analysis methods such as ore grade calculating and information quantity method. The extraction method for proximity indicators mainly includes adjacency analysis methods such as buffer zone analysis and nearest neighbour analysis, as well as distance field analysis method. The extraction method for anomaly indicators mainly includes traditional statistical methods, probability plot method, multivariate statistical methods, geological morphology analysis method, multifractal method, wavelet analysis method, machine learning method (Mokhtari and Sadeghi, 2021; Cracknell and Reading, 2014; Chen *et al.*, 2023). However, the extraction methods for anomaly indicators vary for different exploration methods, such as probability density distribution method and two-dimensional empirical mode method for extracting geophysical anomaly indicators, interference removal and principal component thresholding method for extracting remote sensing anomaly indicators.

4 Construction of MEI system

The mineralization-indicating effect of a single MEI is limited, and prospect prediction is a complicated process that requires the integration of multiple exploration indicators to form a system that can jointly improve the accuracy and precision of ore prediction. The MEI system proposed in this paper is a combination of multi-scale exploration indicators according to the type of ore deposit (Fig. 1).

Therefore, the MEI system consists of two major elements: (1) a set of multi-scale exploration indicators with different exploration scales, exploration methods, and deposit types; (2) integration methods for combining the exploration indicators in the system, which can be divided into knowledge-driven methods, data-driven methods (Yousefi et al., 2021), and AI-driven methods.

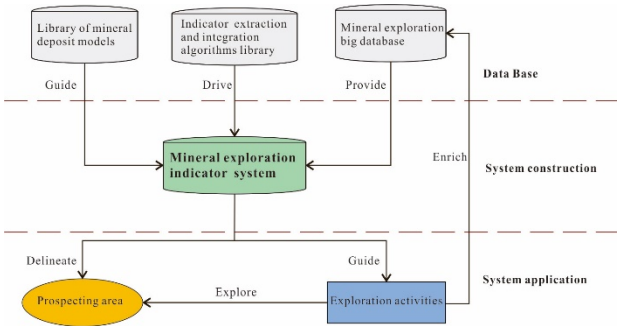

Fig. 1 The construction and application process of mineral exploration indicator system

The knowledge-driven method mainly relies on expert experience to drive the integration of indicators and is suitable for blind ore prediction in areas with low levels of mineral exploration and limited exploration data. It includes methods such as Boolean logic model, index overlay method, fuzzy logic model, and evidence belief method. The data-driven method mainly uses traditional data statistics algorithms to drive indicator integration. It is suitable for prediction in regions with high levels of exploration and rich exploration data. It includes methods such as statistical analysis, comprehensive information value method, evidence weight method, and Bayesian network classifier.

The AI-driven method mainly uses big data and artificial intelligence methods to drive indicator integration and is currently a hot and cutting-edge research topic, including methods such as logistic regression analysis, random forest method, support vector machine method, neural network method, and deep learning method (Barak et al., Wang et al., 2020; Chen et al., 2022; Roshanravan et al., 2023).

5 Future of MEI system

With the continuous development of the theoretical disciplines of ore deposits, mineral exploration, computer science, information science, and the continuous improvement of technologies such as big data, artificial intelligence, and exploration methods, the development of the MEI system research will also enter a new stage. The future research trend of the MEI system will generally develop from "simple dispersion" to "intelligent integration", which will be specifically manifested as: (1) the development of simple indicator towards complex indicator; (2) the development of two-dimensional indicator towards three/four-dimensional indicator; (3) the development of single-element indicator towards multi-element indicator; (4) the development of fuzzy indicator towards precise indicator; (5) the development of manual extraction indicator towards intelligent extraction indicator; and (6) the development of data-driven indicator integration towards intelligent-driven indicator integration.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (42230810).

References

Barak, S., Imamalipour, A., Abedi, M., Bahroudi, A. and Khalifani, F.M. (2021): Comprehensive modeling of mineral potential mapping by integration of multiset geosciences data; Geochemistry, Mineral exploration: a journey from fieldwork, to laboratory work, computational modelling and mineral processing, V.81, P.125824.

Chen, G.X., Huang, N., Wu, G.P., Luo, L., Wang, D.T. and Cheng, Q.M. (2022): Mineral prospectivity mapping based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province; Ore Geology Reviews, V143, P.104765.

Chen, Z.Y., Xiong Y.H., Yin .BJ., Sun B.J. and Zuo R.G. (2023): Recognizing geochemical patterns related to mineralization using a self-organizing map; applied geochemistry, p.105621.

Cracknell, M.J. and Reading, A.M.(2014): Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information; Computers & Geosciences, V.63,P. 22–33.

Mokhtari, Z. and Sadeghi, B. (2021): Geochemical anomaly definition using multifractal modeling, validated by geological field observations: Siah Jangal area, SE Iran; Geochemistry, Mineral exploration: a journey from fieldwork, to laboratory work, computational modelling and mineral processing, V.81, P.125774.

Yousefi, M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M.A., (2019): Exploration information systems – A proposal for the future use of GIS in mineral exploration targeting; Ore Geology Reviews 111, 103005.

Roshanravan, B., Kreuzer, O. P., Buckingham, A., Keykhay-Hosseinpoor, M. & Keys, E. (2023): Mineral potential modelling of orogenic gold systems in the granites-tanami Orogen, Northern Territory, Australia: A multi-technique approach; Ore Geology Reviews, V.152, P.105224.

Xiong, Y.H. and Zuo, R.G. (2020): Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine; Computers & Geosciences, V.140, P.104484.

Wang, J., Zuo, R.G. and Xiong, Y.H. (2020): Mapping Mineral Prospectivity via Semi-supervised Random Forest.; Natural Resources Research, V.29(1), P.189–202.

Yousefi, M., Carranza, E.J.M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M.A. and Mihalasky, M.J. (2021): Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: State-of-the-art and outlook.; Journal of Geochemical Exploration, V.229, P.106839.

Plate convergence, orogenic uplift, and tectonic preconditioning for giant porphyry copper formation in the central Andes

Alexander D. Farrar[1], Matthew J. Cracknell[1], David R Cooke[1], Thomas Schaap[1]
[1] Centre for Ore Deposit and Earth Sciences (CODES), University of Tasmania

Abstract. The central Andes is the world's highest altitude cordilleran mountain belt and contains nearly half the world's reserves of copper – a critical mineral for global clean energy technologies. While previous studies have debated the relative importance of deep-seated structural corridors, plate motion, and cordilleran development in the formation of porphyry copper deposits and the Andes, a formal, data-driven approach that assesses these relationships has not been undertaken. To address this gap, we conduct time-series and statistical analysis of published orogenic proxies in the central Andes and compare them to different plate convergence models. We identify both linear and causal tectonic processes that are occur prior to and simultaneously with the formation of these deposits. By identifying the most suitable plate motion models, we provide a novel perspective into the complex interplay between plate motion, orogenic uplift, and porphyry Cu mineralisation in the central Andes. Our data-driven results advance our understanding of how plate convergence and orogenic processes in the central Andes interact and provide insight into tectonic preconditioning processes that are required for the formation of giant porphyry copper ore deposits.

1 Introduction

The central Andes represents the type example of a cordilleran orogenic system. Because of this, there are a plethora of studies that have investigated the orogenic evolution of the central Andes, based on structural geology and basin evolution, petrology, geochronology, thermochronology, paleoelevation, seismology, and geodynamic modelling, and there have been many models produced for the plate tectonic motions of the Farallon-Nazca plate (NAZ) and South American plate (SAM). Given the myriad different models that exist for the tectonic evolution of the central Andes, how does one decide which model(s) are most the useful for further tectonic analysis related to porphyry mineralisation processes in the central Andes?

In this study, we employ data-driven analysis to compare competing plate motion models to orogenic processes in the central Andes since the late Cretaceous using the Pearson correlation coefficient (*r*), and we apply Granger causality, a robust method for identifying causal linkages between time-series variables that exhibit temporal lag. Granger causality has been widely employed in economic, medical, and climatic studies, however, its application in geology has remained largely unexplored. Using Granger causality, we investigate temporal linkages between plate convergence parameters and orogenic variables, and we assess their statistical correspondence.

Our data-driven method yields critical insight into the relationships between plate tectonic and cordilleran orogenic processes, including the necessary tectonic conditions for metallogenic episodes, which form giant porphyry copper deposits (PCDs; >3 Mt contained Cu) at the intersections of continental-scale structural corridors during high strain episodes (Farrar et al. In Press).

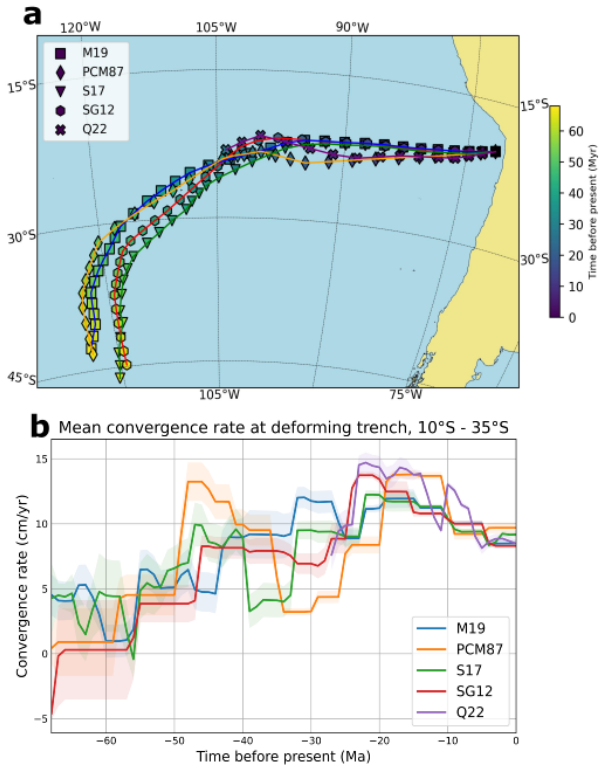

Figure 1. a) Retro-projected trajectory for a point on the Nazca plate relative to South America. We selected a seed point that is currently entering the trench at 20°S and retro-projected it to 68 Ma, using the five plate models analysed in this study (see Methodology). b) Mean convergence rate per Myr along the trench for each plate motion model, shading represents one standard deviation of the mean.

2 Methodology

Five plate tectonic motion models representing a diversity of available models for SAM-NAZ convergence (Pardo-Casas and Molnar 1987; Somoza and Ghidella 2012; Schepers et al. 2017; Müller et al. 2019; Quiero et al. 2022; hereafter referred to as PCM87; SG12; S17; M19; Q22; Figure 1) were output from pyGPlates in 1 Myr age bins. Linear correlation analysis of these models

was conducted at 1 Myr temporal intervals. The mean and standard deviation of convergence obliquity and convergence rate for each plate tectonic motion model, every 1 Myr was calculated for each trench sample point.

Time-series data representing orogenic proxies for the central Andes since the Late Cretaceous were compiled (Figure 2). These consist of paleoelevation of geomorphological domains (Boschman 2021) tectonic stress (Sr/Y) and crustal thickening (La/Yb) proxies from unaltered volcanic rocks (Loucks 2021), flat slab subduction events (Ramos and Folguera 2009) and exhumation rate history (Stalder et al. 2020) and the cumulative contained Cu associated with giant porphyry copper deposits (this study) per metallogenic epoch (Sillitoe and Perello 2005), Figure 3).

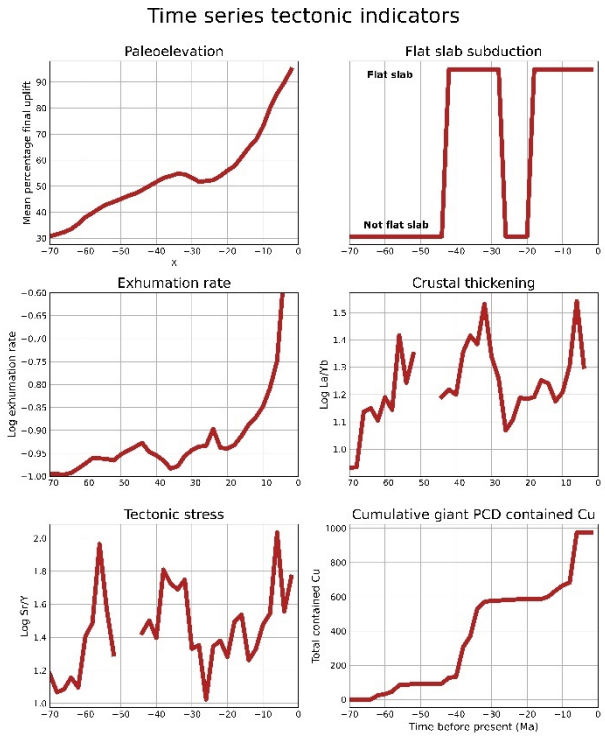

Figure 2. Time series tectonic indicators

To enable a direct temporal comparison of orogenic and plate tectonic datasets, we binned model proxies into two-million-year time intervals for linear correlation analysis (Figure 4) and G-causality analysis (Figure 5). Due to positive skewness in Sr/Y, La/Yb, and exhumation data, we log10 transformed these variables, resulting in approximately Gaussian distributions. We conducted autocorrelation and Augmented Dickey Fuller (ADF) tests on the time series datasets using the python statsmodels ADF package and we performed the G-causality tests using the statsmodels grangercausalitytests package in Python (Seabold and Perktold 2010).

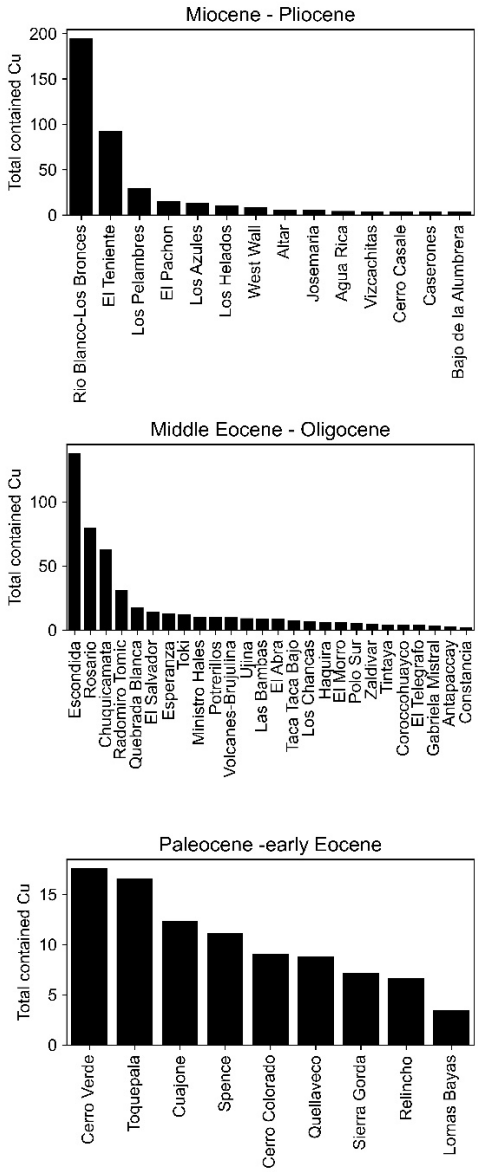

Figure 3. Pareto charts of the total contained Cu of giant porphyry copper deposits, per Cenozoic metallogenic epoch (Sillitoe and Perello 2005) in the central Andes.

3 Results

3.1 Linear correlation analysis

Our investigation of orogenic proxy pairs demonstrates significant positive correlation (r > 0.7) between tectonic stress and crustal thickening as well as between exhumation rate and paleoelevation (Figure 4). These results indicate that as tectonic stress increases, so does crustal thickening, and changes in paleoelevation are broadly synchronous with changes in exhumation rate. These results are consistent with visual analysis of Figure 3, which illustrates the cyclic nature of crustal thickening processes over the temporal range.

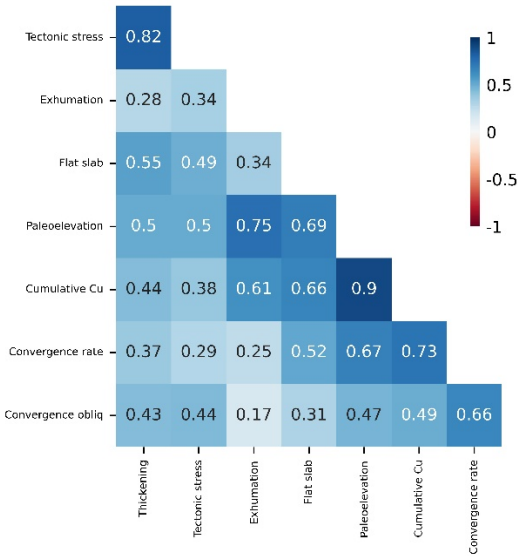

Figure 4. Pearson correlation of the means of orogenic proxies, SG12 convergence rate, PCM87 convergence obliquity, cumulative giant PCD tonnage, since 70 Ma.

Incorporating plate convergence parameters and cumulative Cu tonnage allows for the simultaneous assessment of non-orogenic proxies. Figure 4 shows that paleoelevation and cumulative Cu tonnage exhibit a significant linear relationship (r = 0.9) and convergence rate and cumulative Cu tonnage are also significantly linearly correlated (r = 0.73; Figure 5).

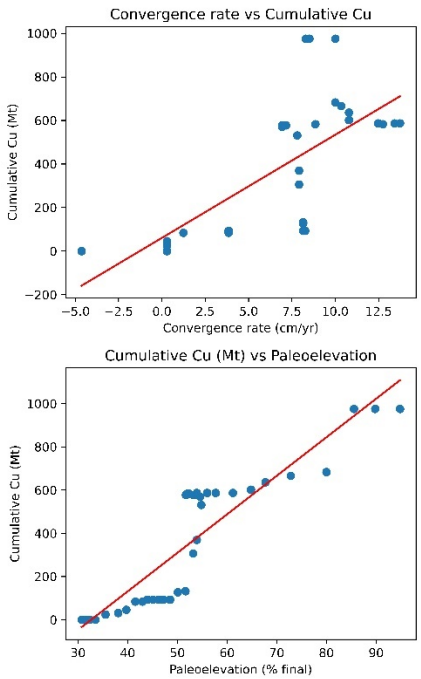

Figure 5. Scatter plots of cumulative Cu tonnage and Paleoelevation against SG12 convergence rate. Red line is the linear correlation coefficient for each pair.

3.2 Granger causality

Our investigation using Granger causality analysis (Figure 6) revealed several significant lagged causal relationships that were not detected in the linear correlation analysis for the orogenic proxies (Figure 4). These relationships include a causal effect of crustal thickening on paleoelevation, exhumation rate, and flat slab state, as well as a feedback loop where paleoelevation and exhumation rate both affect crustal thickening. Our results also indicate that changes in tectonic stress drive variations in paleoelevation, exhumation rate, and crustal thickening, while flat slab subduction leads to increased paleoelevation and exhumation rate.

Incorporating the five plate motion models to the G-causality analysis with the orogenic uplift proxies enables us to investigate causal relationships between plate motion and orogenesis. The results showed whilst many models exhibit causal linkages with orogenic proxies, the SG12 convergence rate model exhibits 7 causal linkages with orogenic proxies, indicating its suitability for further metallogenic analysis.

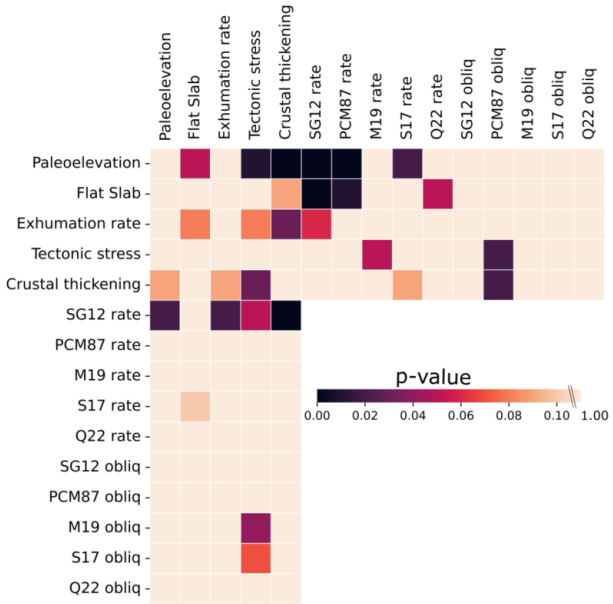

Figure 6. G-causality matrix of plate motion and orogenic proxy pairs. H_0 = x-axis variable does not G-cause the y-axis variable, H_0 is rejected if p-value of the F-test <0.10. p-values of variable pairs, non-background-colored intersections represent variable pairs that exhibit statistically significant G-causality (p-value < 0.10).

The SG12, PCM87, and S17 convergence rate models were found to strongly G-cause paleoelevation responses, while the SG12, PCM87, and Q22 models G-caused flat slab events. Additionally, the SG12 model was found to G-cause exhumation, while the M19 model was found to G-cause tectonic stress. Interestingly, the orogenic proxies themselves were found to exhibit strong causality on SG12 convergence rate, with paleoelevation, exhumation rate, crustal thickening and tectonic stress proxies all exhibiting feedback

with SG12 (Figure 6). The PCM87 convergence obliquity was found to strongly G-cause changes in crustal thickening and tectonic stress proxies and was thus defined as the most useful convergence obliquity model (Figure 6).

4 Conclusions

Our study provides a novel approach for the evaluation of plate motion models based on their relationship with convergent margin tectonics. Our data driven analysis shows that there is a significant linear relationship between increasing Cu tonnage in metallogenic epochs and increasing rate of convergence and increasing Andean elevation. There are also lagged correlations that are not able to be determined by linear analysis alone. By determining the direction of temporally lagged variables using Granger causality analysis, causal relationships between plate motion and orogenic processes can be detected that are not identified using linear correlation techniques.

Future research will build on this tectonic process model, as well as incorporating the best correlated plate motion models with orogenic proxies, to examine the causal spatial and temporal relationships between giant PCDs and plate tectonic processes in the central Andes.

Acknowledgements

This research forms part of a PhD by AF at the University of Tasmania (CODES). AF acknowledges the financial support provided by First Quantum Minerals to carry out this research. We thank G. Begg and J. Hronsky for discussions that improved this study.

References

Balázs, A., Faccenna, C., Ueda, K., Funiciello, F., Boutoux, A., Blanc, E. J. P., and Gerya, T., 2021, Oblique subduction and mantle flow control on upper plate deformation: 3D geodynamic modeling: Earth and Planetary Science Letters, v. 569.

Boschman, L. M., 2021, Andean mountain building since the Late Cretaceous: A paleoelevation reconstruction: Earth-Science Reviews, v. 220.

DeCelles, P. G., Ducea, M. N., Kapp, P., and Zandt, G., 2009, Cyclicity in Cordilleran orogenic systems: Nature Geoscience, v. 2, p. 251-257.

Farrar, A. D, Cooke, D. R., Hronsky, J. M. A., Wood, D. G., Benavides, S. B., Cracknell, M. J., Banyard, J. F., Gigola, S., Ireland, T., Jones, S. M., Piquer, J., (In Press) A model for the lithospheric architecture of the central Andes and the localization of giant porphyry copper deposit clusters. Economic Geology.

Iaffaldano, G., Bunge, H.-P., and Dixon, T. H., 2006, Feedback between mountain belt growth and plate convergence: Geology, v. 34.

Loucks, R. R., 2021, Deep entrapment of buoyant magmas by orogenic tectonic stress: Its role in producing continental crust, adakites, and porphyry copper deposits: Earth-Science Reviews, v. 220.

Meade, B. J., and Conrad, C. P., 2008, Andean growth and the deceleration of South American subduction: Time evolution of a coupled orogen-subduction system: Earth and Planetary Science Letters, v. 275, p. 93-101.

Müller, R. D., Zahirovic, S., Williams, S. E., Cannon, J., Seton, M., Bower, D. J., Tetley, M. G., Heine, C., Le Breton, E., Liu, S., Russell, S. H. J., Yang, T., Leonard, J., and Gurnis, M., 2019, A global plate model including lithospheric deformation along major rifts and orogens since the Triassic: Tectonics, v. 38, p. 1884-1907.

Pardo-Casas, F., and Molnar, P., 1987, Relative motion of the Nazca (Farallon) and South American plates since Late Cretaceous time: Tectonics, v. 6, p. 233-248.

Quiero, F., Tassara, A., Iaffaldano, G., and Rabbia, O., 2022, Growth of Neogene Andes linked to changes in plate convergence using high-resolution kinematic models: Nat Commun, v. 13, p. 1339.

Ramos, V. A., and Folguera, A., 2009, Andean flat-slab subduction through time: Geological Society, London, Special Publications, v. 327, p. 31-54.

Schepers, G., van Hinsbergen, D. J. J., Spakman, W., Kosters, M. E., Boschman, L. M., and McQuarrie, N., 2017, South-American plate advance and forced Andean trench retreat as drivers for transient flat subduction episodes: Nat Commun, v. 8, p. 15249.

Seabold, S., and Perktold, J., 2010, Statsmodels: Econometric and statistical modeling with python: Proceedings of the 9th Python in Science Conference, 2010, p. 10.25080.

Sillitoe R, Perello J (2005) Andean copper province: Tectonomagmatic settings, deposit types, metallogeny, exploration, and discovery. Economic Geology 100th Anniversary Volume 100:845-890.

Somoza, R., and Ghidella, M. E., 2012, Late Cretaceous to recent plate motions in western South America revisited: Earth and Planetary Science Letters, v. 331, p. 152-163.

Stalder, N. F., Herman, F., Fellin, M. G., Coutand, I., Aguilar, G., Reiners, P. W., and Fox, M., 2020, The relationships between tectonics, climate and exhumation in the Central Andes (18–36°S): Evidence from low-temperature thermochronology: Earth-Science Reviews, v. 210.

Suárez, R. J., Guillaume, B., Martinod, J., Ghiglione, M. C., Sue, C., and Kermarrec, J.-J., 2022, Role of convergence obliquity and inheritance on sliver tectonics: Insights from 3-D subduction experiments: Tectonophysics, v. 842.

The importance of hierarchical data structures for the interpretation of mineral trace-element data

Max Frenzel[1]
[1]*Helmholtz-Zentrum Dresden-Rossendorf, Institute Freiberg for Resource Technology, Germany*

Abstract. Recent years have seen a sharp increase in the generation and use of mineral trace-element data in geological research. However, while much new data is being generated and published, relatively little work has been done to develop appropriate methods for statistical analysis and interpretation. Several characteristic features of mineral trace-element data require careful consideration during evaluation and interpretation to avoid biased results. In particular, the generally hierarchical structure of the data must be considered. Unfortunately, this is not done in most current studies. This contribution provides a brief overview of what hierarchical data structures are, and what consequences they have for statistical analysis and data interpretation.

1 Introduction

Modern laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) systems enable the rapid, spatially resolved collection of mineral trace-element data at relatively high sensitivity and low cost. This has led to their widespread use in geological research (Sylvester and Jackson 2016).

Unfortunately, the accompanying increase in the generation and use of mineral trace-element data has not been accompanied by a commensurate increase in understanding of how best to use and interpret it. Specifically, the time lag between the capabilities for data generation and interpretation seems to be chiefly due to a lack of appreciation by many workers for the key mathematical features of the data and their consequences for statistical analysis.

While the requirements arising from the compositional nature of trace-element have already been discussed in detail elsewhere (van den Boogaart and Tolosana-Delgado 2013, Frenzel et al. 2016), this contribution focusses on hierarchical data structures. After a short description of what is meant by this term, different approaches for dealing with these data structures are described. An example is then given to illustrate the biases that may be introduced into the analysis of a dataset by ignoring them. Finally, recommendations are made for future work.

2 What are hierarchical data structures?

In hierarchical data structures, each datapoint is characterized by multiple attributes, each referring to a different level of organization. In mineral trace-element datasets, such data structures typically arise from both the nature of the data as well as the sampling and analysis processes (Dimitrijeva et al. 2018, Godefroy-Rodriguez et al. 2020).

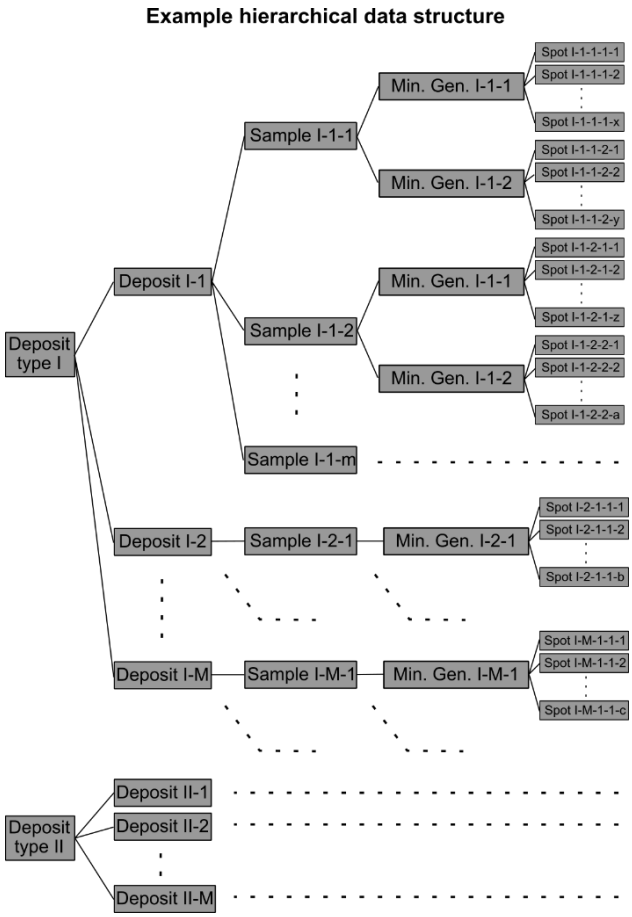

Figure 1. Example hierarchical data structure for mineral trace-element data collected on different deposit types. A more detailed description is provided in the main text.

For example, in a high-level LA-ICP-MS study each data point may be described by the following attributes, in order of increasing level of organization: “analysis spot” < “mineral generation” (if several are present) < “sample” < “deposit” < “deposit type”. This is also illustrated in Fig. 1 and has several consequences for the expected mathematical properties of the data.

Most importantly, one would expect two datapoints from such a dataset to be more similar, the more attributes they have in common. That is, two datapoints from the same sample and same mineral generation would be expected to be more similar than two points from different samples but from within the same deposit, which would in turn be expected to be more similar than datapoints collected on samples from different deposits, and so on. The chief reason for this is that the further two datapoints are separated in the data hierarchy, the further they are also separated in space and time,

and the greater should be the differences in the geological conditions they reflect.

One can mathematically express these ideas by a formula describing the observed variability in trace-element concentrations in terms of the different effects caused by differences at each of the hierarchical levels. For the data structure described above and shown in Fig. 1, one may write:

$$\log(c^A_{ijklmn}) = Ref^A + Type^A_i + Deposit^A_{ij} + Sample^A_{ijk} + Generation^A_{ijl} + \varepsilon^A_{ijklm} \quad (1)$$

Where c^A_{ijklmn} is the concentration of trace-element A measured in analysis spot m, in mineral generation l on sample k, from deposit j, belonging to deposit type i; Ref^A is a reference value for the concentration of A, e.g., a global mean; $Generation^A_{ijl}$ is the mean effect of mineral generation l in deposit j in district i on $log(c^A)$, expressed as an additive value, and corrected for variations between samples, deposits, and deposit types; $Sample^A_{ijk}$ is the mean effect of sample k in deposit j of type i, and so on. Finally, ε^A_{ijklm} is the residual value of the specific analysis spot. Note that this description assumes that the identified mineral generations are only consistently identifiable within the same deposit ij, and that the effects of the different factors, or hierarchical levels are statistically independent. Alternative models are possible, where consistent identification of mineral generations is feasible over smaller or larger scales, and where interactions between the different factors occur (cf. Winter 2013, Dmitrijeva et al. 2018).

If one assumes that the dataset is balanced, i.e., the same number of observations are available for each unique combination of the different attributes, then the total variance of the observed data following model (1) would be:

$$var[\log(c^A_{ijklmn})] = var(Type^A_i) + var(Deposit^A_{ij}) + var(Sample^A_{ijk}) + var(Generation^A_{ijl}) + var(\varepsilon^A_{ijklm}) \quad (2)$$

This is just another way of expressing the ideas described above, namely, that knowing the value of one measurement with a set of attributes $ijklm$ already narrows the expected range of values for the next measurement with the same, or some of the same attributes. This may seem obvious to most geologists. However, it has dramatic consequences for the statistical analysis of the data. Namely, it means that individual datapoints are not expected to be statistically independent (cf. Dmitrijeva et al., 2018). Nor is it obvious that individual sets of observations $ijklmn$ can be assumed to be identically distributed, i.e., to follow the same probability distribution. For instance, the mean effects of the samples, $Sample^A_{ijk}$, within one deposit should have different mean and variance than those from the next deposit, resulting in a different distribution of the corresponding $log(c^A)$.

However, both statistical independence and identical distribution (iid) are key assumptions in virtually all statistical methods. To further complicate matters, mineral trace-element data is often unbalanced, i.e., different numbers of observations are available for each specific combination of attributes. The combination of these features means that standard statistical methods cannot be sensibly applied to the raw spot data. Biases are introduced into data analysis if this is done, as illustrated below.

Finally, we note that hierarchical data structures and unbalanced datasets also occur frequently in other areas of geochemistry. Some attention had been paid to this in the past, e.g., in the hierarchical estimation of Clarke values describing crustal abundances (Ketris and Yudovich, 2009). However, this has unfortunately not entered universal practice.

3 Dealing with hierarchical data

To sensibly apply standard statistical methods to hierarchically structured data, one must find a way to modify this data such that the iid assumption generally required for data analysis is satisfied. This can be achieved by aggregation of the data to the hierarchical level relevant for the analysis.

Consider the case where one is interested in the differences in trace-element signatures between deposit types for a dataset with the same structure as our example in Fig. 1. The relevant hierarchical level for analysis would be that of individual deposits. The mean trace-element concentrations for the deposits can relatively safely be assumed to be independent from each other and follow a simple probability distribution, i.e., to be iid. The main task then is to infer the probability distribution of deposit means for each deposit type from the available data and use this to answer any question(s).

Different methods exist to achieve the necessary aggregation of the data to the desired hierarchical level. In the present example, the simplest way of doing so would be to compute hierarchical means for the individual deposits and use these means for further analysis (cf. Ketris and Yudovich, 2009). Hierarchical computation in this case would mean, that the mean for each mineral generation on each sample is first calculated from individual analysis spots, then the mean for each sample is estimated from the means per mineral generation, and finally the deposit mean is calculated from the sample means (cf. Fig. 1). This removes the biases due to the different numbers of observations available from each sample, deposit, generation etc.

However, this approach is cumbersome. The unbalanced nature of the data also means that some hierarchical means are more uncertain than others, i.e., means will generally be more certain for deposits where more samples were taken. This can be dealt with in the further statistical analysis by giving weights to each of the means to reflect its uncertainty. However, the relevant uncertainties themselves are not always easy to quantify, particularly where only single observations are

available, e.g., where only one sample has been taken for a given deposit. Such cases are in fact relatively common (cf. Frenzel et al. 2016).

A more sophisticated way of performing the hierarchical estimation of mean values and the corresponding uncertainties is to fit a model of the form of eq. (1) to the data. The relevant class of models for this purpose are linear mixed effects (LME) models (Winter, 2013; Dmitrijeva et al., 2018). These models can be used to simultaneously make unbiased estimates of the mean effects (including uncertainties) of the different attributes at each hierarchical level. Such estimates can then be used in further data analysis.

In fact, LME models are much more versatile than this, and can be used to analyse many different problems. Where the capabilities of LME models are suitable to address a specific question, it is therefore best to apply them directly to a given dataset. This may remove the need for cumbersome hierarchical aggregation of the data prior to analysis.

4 Effects of ignoring hierarchical data structures

Disregard for hierarchical data structures is typically expressed in mineral trace-element studies by the treatment of individual datapoints from the lowest hierarchical level of a dataset as iid observations. Thus, in an LA-ICP-MS dataset all individual spot measurements belonging to one deposit type may be taken to represent this type, regardless of how many deposits the data covers, or how many samples were analysed per deposit.

What does this do to the analysis and interpretation of the data? It will have two major effects. First, it will suggest that there are far greater numbers of independent observations, and thus greater statistical power, than are actually present in the dataset. Second, it will introduce bias into the analysis whenever the data is unbalanced, i.e., nearly always. Finally, it will introduce artefacts to the shapes of the observed data distributions. These effects are illustrated graphically in Figure 2 using an example from the literature.

Figure 2a shows a PCA biplot from Bélissont et al. (2014) indicating different “fields” of sphalerite composition for different types of Pb-Zn deposits, based on a “large” set of LA-ICP-MS point analyses. On the other hand, Fig. 2b shows the same data reduced to its relevant hierarchical level, i.e., the hierarchical means for individual deposits, including their associated uncertainties. Several features are apparent from this comparison.

First, the data distribution in Fig. 2a shows many blob-like features, or clusters. Second, each of these clusters appears to be defined by many datapoints. Since the statistical power of a dataset increases with $1/\sqrt{n}$ (think of the standard error of the mean), where n is the number of available iid observations, Fig. 2a would suggest that such complex distributional shapes reflect the real distribution of the data for the different deposit types. For instance, if one was to analyse a new sample from another MVT deposit not represented in the original dataset, surely it would plot inside one of the MVT fields delineated in Fig.2a.

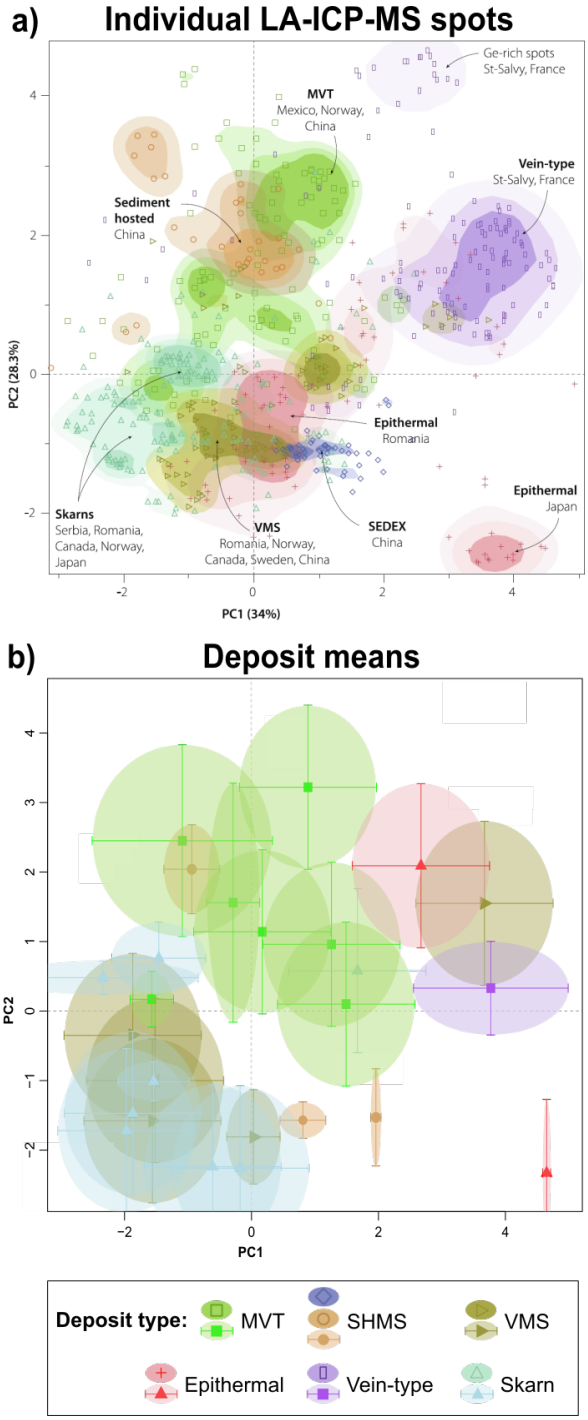

Figure 2. Graphical illustration of the typical effects of treating individual LA-ICP-MS analysis spots as independent observations: a) reproduction of Fig. 13c of Bélissont et al. (2014), a PCA-biplot of several hundred individual analysis spots of sphalerite from different types of Pb-Zn deposits; b) reconstruction of a) showing deposit means and associated 95% confidence intervals. Note that the somewhat imperfect reconstruction in panel b) is chiefly due insufficient documentation in Bélissont et al. (2014) regarding the assumptions made to run PCA. They do not describe whether they used scaled or unscaled input variables, nor how they treated missing values (i.e., below detection limit or missing at random).

In fact, nothing could be further from the truth. As Fig. 2b shows, most of the clusters observed in Fig. 2a in fact appear to reflect individual deposits, although the reconstruction is not perfect (cf. explanation in figure legend). This reduction also highlights how very few iid observations are really available. The total number of deposits in the dataset is only 26. That is, the statistical power of the dataset is in fact so low, that no reliable distinction between deposit types is possible. This clearly illustrates the major effect of ignoring hierarchical data structures: it creates overconfidence in potentially biased results.

4.1 Overfitting of data by ML methods

A specific type of overconfidence in erroneous results occurs when disregard for hierarchical data structures (and sometimes other features) is coupled to the use of machine learning (ML) methods, e.g., for classification problems. As illustrated in Fig. 2a, disregard for hierarchical data structures will generally create datasets with apparently complex, clustered data distributions. ML methods are excellent at picking out the irregular, high-dimensional boundaries between such clusters for classification or regression purposes (Bishop 2006). Thus, any ML models fitted to such data will generally have a much greater degree of complexity than is justified by the true nature of the data.

To make matters worse, the false assumption that individual observations are iid also short-circuits the key quality-control measure typically used to assure the reliability of the ML results: application of the fitted model to a test dataset randomly subsampled from the original data, and therefore assumed to be independent of it. However, because the test data will again contain only individual spot analyses, which must necessarily come from the same clusters already included in the training data (samples/deposits), the assumption of independence will be violated. Thus, the apparent classification accuracy is usually greatly overestimated, providing exaggerated confidence in the potentially flawed results.

While this may seem like a rather specialized issue, the recent surge in the popularity of both ML methods and mineral trace-element data has led to the publication of many articles suffering from this problem (e.g., Sun and Zhou 2022, Li et al. 2023). Given the potentially complex interactions between high-dimensional data structures and the classification algorithms typically used (random forests, neural nets etc.), it is difficult to say which of the results reported in such studies are reliable.

5 Conclusions and future work

Hierarchical data structures are currently ignored by most workers using mineral trace-element data. This introduces biases into data analysis and may result in conclusions that are not justifiable given the data. To avoid such issues in the future, appropriate methods must be used for data analysis. Hierarchical aggregation and the use of LME models for data analysis both provide adequate approaches. In fact, the use of these methods offers significant potential for interesting discoveries regarding the nature and causes of the often-substantial variance observed in the trace-element signatures of many of the common and less-common minerals occurring in mineral deposits. Specifically, the quantitative understanding of the variance structure of the data via LME models may be useful in this regard (e.g., Dmitrijeva et al. 2018, Frenzel et al. 2022).

Acknowledgements

I would like to thank all those colleagues, particularly Raimon Tolosana-Delgado, Marija Dmitrijeva, Nigel Cook, and others from Freiberg, Adelaide, and elsewhere, who have over the past 10 years made significant contributions to my understanding of this research area by many fruitful, though sometimes heated, discussions. Without their intellectual input, I would not have been able to write this contribution.

References

Belissont R, Boiron M-C, Luais B, Cathelineau M (2014) LA-ICP-MS analyses of minor and trace elements and bulk Ge isotopes in zoned Ge-rich sphalerites from the Noailhac – Saint-Salvy deposit (France): Insights into the incorporation mechanisms and ore deposition processes. Geochim Cosmochim Acta 126:518-540.

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York, 738 p.

Dmitrijeva M, Metcalfe AV, Ciobanu CL, Cook NJ, Frenzel M et al. (2018) Discrimination and variance structure of trace element signatures in Fe-oxides: a case study of BIF-mineralisation from the Middleback Ranges, South Australia. Math Geosci 50:381-415.

Frenzel M, Hirsch T, Gutzmer J (2016) Gallium, germanium, indium, and other minor and trace elements in sphalerite – A meta-analysis. Ore Geol Rev 76:52-78.

Frenzel M, Voudouris P, Cook NJ, Ciobanu CL, Gilbert S, Wade BP (2022) Evolution of a hydrothermal ore-forming system recorded by sulfide mineral chemistry: A case study from the Plaka Pb-Zn-Ag deposit, Lavrion, Greece. Miner Depos 57:417-438

Godefroy-Rodriguez M, Hagemann S, Frenzel M, Evans NJ (2020) Laser ablation ICP-MS trace element systematics of hydrothermal pyrite in gold deposits of the Kalgoorlie district, Western Australia. Miner Depos 55:823-844

Ketris MP, Yudovich YaE (2009) Estimations of Clarkes for carbonaceous biolithes: World averages for trace element contents in shales and coals. Int J Coal Geol 78:135-148.

Li X-M, Zhang Y-X, Li Z-K, Zhao X-F, Zuo R-G et al. (2023) Discrimination of Pb-Zn deposit types using sphalerite geochemistry: New insights from machine learning algorithm. Geosci Front 14:101580.

Sun G-T, Zhou J-X (2022) Application of machine learning algorithms to classification of Pb-Zn deposit types using LA-ICP-MS data for sphalerite. Minerals 12:1293.

Sylvester PJ, Jackson SE (2016) A brief history of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). Elements 12:307-310.

Van den Boogaart KG, Tolosana-Delgado R (2013) Analyzing compositional data with R. Springer, Berlin, 258 p.

Winter B (2013) Linear models and linear mixed effects models in R with linguistic applications. arXiv preprint https://arxiv.org/abs/1308.5499.

The role of data science in modern mineral exploration and mining: adding machine learning tools to the geoscientist's toolbox

Dina Klimentyeva[1], Britt Bluemel[1], McLean Trott[1]
[1] ALS GoldSpot Discoveries Ltd., Vancouver, Canada

Abstract. Geodatascience is an emerging field that combines traditional geoscience expertise with the (data) science of artificial intelligence and machine learning. The pace and volume of data acquisition is rapidly increasing in mineral exploration campaigns, at mining operations, and in near-mine environments, leading to the accumulation of large datasets that can be challenging to process and interpret using conventional methods. Machine learning and data science techniques add speed and consistency to interpretation of large datasets, aid in the amalgamation of new and historic datasets, and facilitate the integration of disparate data types with varying resolutions. All these factors help shorten the gap between discovery and development, and companies across the entire mining value chain, in a variety of commodities, are realizing the value of incorporating machine learning workflows and machine-assisted modelling to assist in the discovery and development of ore bodies.

1 Machine learning in mineral exploration and mining

1.1 Significance of ML tools for exploration

Machine learning (ML) and data science are increasingly gaining acceptance as exploration tools, with applications ranging from core image analysis (Acosta et al. 2019), prospectivity mapping (Carranza 2010, Sun et al. 2019) to chemostratigraphy (Bluemel 2021), and large language models helping to query the corpus of geoscientific literature (Deng et al. 2023).

1.2 Practical applications of machine learning for exploration

The most important components of any successful mineral exploration campaign are a robust geological map and a realistic geological model that represent the synthesis of field observations with interpretations from fundamental datasets such as geochemistry, mineralogy, and geophysics.

1.3 Interpretability of ML results

To integrate traditional geological interpretations with results obtained from ML models, it is necessary to understand the entire ML process from start to finish. This includes selecting fit-for-purpose data types, choosing appropriate transformations and data pre-processing, and utilizing appropriate algorithms. The results must then be critically evaluated and integrated with geological insights and field observations, to ensure the final result most closely resembles geological reality.

1.4 Linking ML results with geological reality

We can pinpoint several examples where machine learning algorithms can be easily interpreted and linked to geological insights, thereby providing a good starting point for increased acceptance and adoption of machine learning processes in exploration, for example:

- Dimensionality reduction applied to geochemical data extracts insights about different styles of mineralization. For instance, principal component analysis (PCA), which is a dimensionality reduction technique, allows the integration of statistics with geology by illustrating which geochemical elements exhibit similar behaviour, thereby adding clarity to the interpretation of datasets from new jurisdictions.

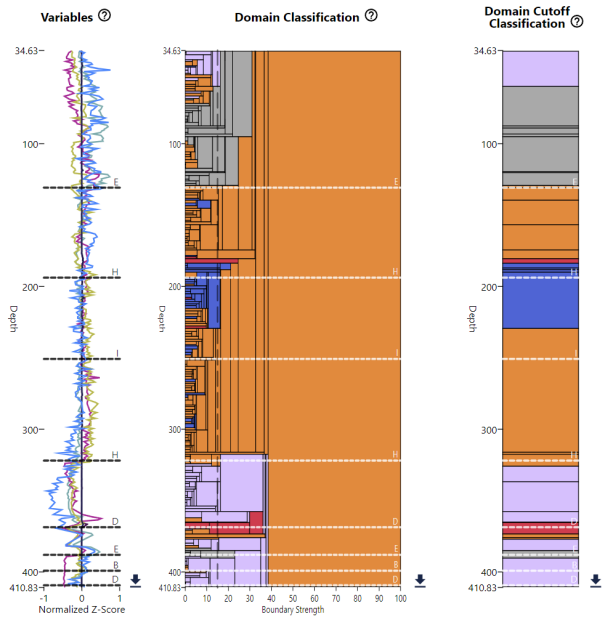

Figure 1. Drillhole domaining based on geochemistry, using Al (blue), Hf (dark blue), Ti (navy blue), Y (violet), Zr (dark red) and Nb (light green) as input signals. Zr and Hf curves are overlapping. Colours of the domains refer to inferred geological units. The dataset is processed by the ALS Goldspot Tessellation app (https://tessellation.app.goldspot.ca/), with dataset of Halley (2020) as an example

- Drillhole domaining based on geochemical information (Fig. 1). By selecting relevant

input signals, it is possible to routinely classify drillhole samples for the purpose of defining lithology, alteration, or mineralization styles.

- Reconciliation of clustering results and logged lithology or alteration labels (Fig. 2) helps derive objective criteria which can be utilized by logging geologists to differentiate and classify rock types and alteration assemblages,
- Assessment of relative importance of geochemical signal for the prediction of stratigraphy (Fig. 3) by calculating and plotting the SHapley Additive exPlanations (SHAP) values. The SHAP values represent the importance of each feature and are calculated by comparing the model's predictions with and without the involvement of each input variable (Lundberg et al. 2020). The comparison of SHAP values for different input variables can assist in selecting the most fit-for-purpose assay techniques

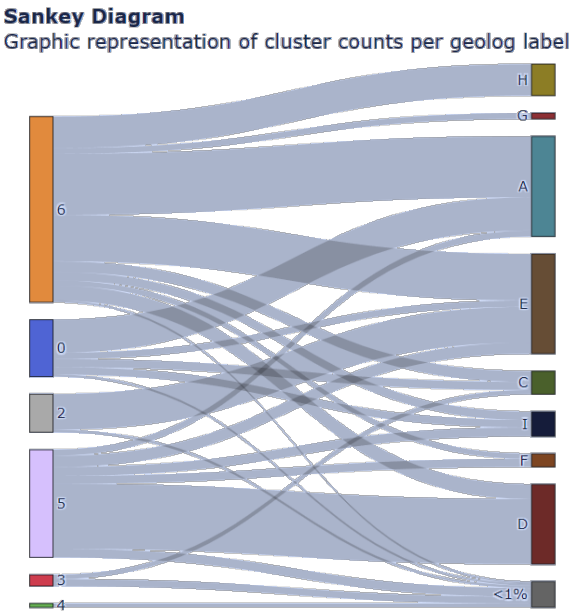

Figure 2. Relating geochemical information (left, also illustrated in the downhole plot of Figure 1) to logged alteration labels (right), using Al, Hf, Ti, Y, Zr and Nb as input signals. The dataset is processed by the ALS Goldspot Tessellation app (https://tessellation.app.goldspot.ca/), with dataset of Halley (2020) as an example

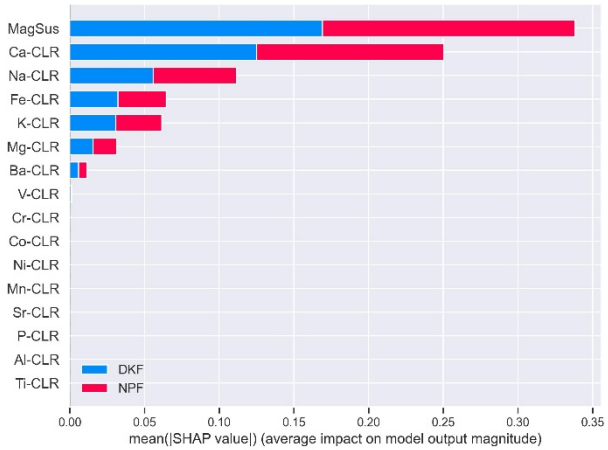

Figure 3. Defining a quantitative metric for the importance of geochemistry for predicting the stratigraphy (SHAP curve). Centred log-ratio transformed geochemical data and magnetic susceptibility (MagSus) were used as the model's inputs

1.5 Data limitations and considerations

Machine learning techniques are becoming more prevalent as computational costs continue to lower, and traditional barriers to entry like the necessity of in-depth knowledge of an object-oriented programming language (such as Fortran, C++, or Python) are overtaken by the rapid increase of the availability of GUI-based applications, such as Orange Data Mining. It is now possible to create robust ML models that can produce accurate results by memorizing the input data, and this creates the illusion of a good fit to the test dataset, but lacks the capacity to be successfully deployed and generalize to the new data! However, the diligent data scientist can recognize that these high accuracies are the result of data leakage and model overfitting.

By understanding the limitations of machine learning algorithms, and combining robust exploratory data analysis, transparent scaling and transformation procedures, we can ensure the successful deployment of many state-of-the-art ML techniques. Meticulous attention to cataloguing of metadata, consistent logging of categorical variables, and robust treatment of missing data can ensure that the data is fit-for-purpose, and properly cleaned and homogenized before use in the machine learning model. The resulting model is robust and flexible, and produces realistic results that are more easily interpretated.

2 Case study

This presentation will showcase various tools, techniques, and case studies where Artificial Intelligence, Data Science, and expert geoscientific approaches are combined to add understanding to the geological system with the goal of discovery. By enhancing the ability of an exploration team to interpret rock textures based on core photos, as well as integrating geochemistry, textural information,

and geophysics to improve understanding of already known orebodies, we can leverage our knowledge from well-defined systems to increase our understanding as we interpret data from new mineral systems. This case study combines structured data (extracted from drillcore photography) with geochemical and petrophysical data to create ML predictions of the presence of mineralization, which can be modelled in 3D. This case study provides a workflow for exploration targeting when dealing with challenges like complex deposit models, and subtle differences in geochemical or textural signal. This case study will also highlight the importance of a strong geological framework to underpin the successful deployment of machine learning algorithms.

Acknowledgements

We gratefully acknowledge the client company for permission to share the results of the project.

References

Acosta ICC, Khodadadzadeh M, Tusa L, Ghamisi P (2019) A Machine Learning Framework for Drill-Core Mineral Mapping Using Hyperspectral and High-Resolution Mineralogical Data Fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12:4829–4842. https://doi.org/10.1109/JSTARS.2019.2924292

Bluemel B (2021) Chemostratigraphic discrimination assisted by machine learning to enhance 3d modelling. In: Goldschmidt2021 abstracts. European Association of Geochemistry, Virtual

Carranza EJM (2011) From Predictive Mapping of Mineral Prospectivity to Quantitative Estimation of Number of Undiscovered Prospects. Resource Geology 61:30–51. https://doi.org/10.1111/j.1751-3928.2010.00146.x

Deng C, Zhang T, He Z, et al (2023) Learning A Foundation Language Model for Geoscience Knowledge Understanding and Utilization. arXiv preprint arXiv:2306.05064.

Halley S (2020) Mapping Magmatic and Hydrothermal Processes from Routine Exploration Geochemical Analyses. Economic Geology 115:489–503. https://doi.org/10.5382/econgeo.4722

Lundberg SM, Erion G, Chen H, et al (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:56–67. https://doi.org/10.1038/s42256-019-0138-9

Sun T, Chen F, Zhong L, et al (2019) GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. Ore Geology Reviews 109:26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003

ALS Goldspot Tessellation app (2022) https://tessellation.app.goldspot.ca/. Accessed 10 August 2023

Predicting mineral abundances from geochemistry in a heavy mineral sand deposit

Kat Lilly[1], Michael Gazley[2,3], Muhammet Kartal[4], Tom Ritchie[5]
[1]RSC, 245 Stuart St, Dunedin, New Zealand
[2]RSC, 225 Thorndon Quay, Wellington, New Zealand
[3]School of Geography, Environment and Earth Science, Victoria University of Wellington, New Zealand
[4]RSC, 13 Rheola St, West Perth, Australia
[5]Hardie Pacific, 57 Leith St, Dunedin, New Zealand

Abstract. Scanning electron microscopy (SEM)-based automated mineralogical studies were undertaken on samples from a heavy mineral sands project on West Coast, South Island, New Zealand, to characterise the mineral assemblage and to quantify the abundance of garnet and ilmenite. These data were used as a training dataset to build linear regression models that predict garnet and ilmenite abundances from major element geochemistry from X-ray fluorescence spectroscopy (XRF) data. The model-performance metrics indicate that the models robustly predict the abundances of these minerals, which allows us to rapidly and inexpensively derive garnet and ilmenite abundances which can be used as an input for subsequent mineral resource estimates (MREs).

1 Introduction

Automated, quantitative analysis of a mineral assemblage in an SEM with energy-dispersive X-ray spectroscopy (EDS) can provide a precise and accurate measurement of the abundance of mineral phases within a sample. However, it is also time-consuming, relatively expensive, and requires very careful sample preparation to ensure that the sample presented to the SEM is representative. XRF analysis is relatively cheap, rapid, and requires less complicated and time-consuming sample preparation.

A statistical model can be built to accurately predict the mineral abundances from the XRF geochemistry. This makes sense theoretically – the geochemistry is directly related to the mineralogy – and does indeed work well in practice; furthermore, for mineral deposits with simple mineralogy (Ritchie et al., 2019, Tay et al., 2021), this can be effective with a limited number of training data.

2 Methodology

2.1 Sample collection and preparation

A set of 30 samples that were representative (based on geochemistry) of the heavy mineral sand deposit were chosen for automated mineralogy, 28 of which were selected from hand auger samples to cover the full variability in garnet and ilmenite abundance across the deposit, and two of which were processing-plant concentrate to provide high-abundance samples. Samples were sieved to be between 53 µm and 2 mm; and a split was taken for automated mineralogy and another taken for pulverisation and XRF analysis.

Care was taken to ensure that the ~10 g of sample taken for each SEM analysis was representative of the original sample. Samples were mounted in 25 mm epoxy rounds, and then cut in half and remounted to present the cut faces in 30 mm epoxy rounds so as to minimise bias in the sample caused by differential settling by grain size and density.

2.2 XRF analysis

The pulverised portion of the sample was analysed by SGS Westport, New Zealand, by flux fusion XRF on a Bruker S8 TIGER instrument resulting in a dataset of 11 major elements reported in wt.% oxide.

2.3 Automated Mineralogy

The 30-mm epoxy rounds were analysed in a Hitachi 3900SU SEM using 2 Bruker EDS detectors and Bruker's Advanced Mineral Analysis and Characterization System (AMICS) software.

More than 99% (by cross-sectional area) of mineral grains were successfully classified. As a test on the quality of the AMICS results, a comparison was made between the major element chemistry as measured by XRF, and the inferred chemistry calculated from the measured abundances of the AMICS-classified minerals. This test work showed a high level of agreement between the two methods, verifying that the SEM sample preparation and analysis methods are robust.

3 Linear Regression Modelling

A multiple linear regression model was built in Python to predict both garnet and ilmenite concentrations from the XRF chemistry, using the SEM-derived mineral abundances as training data. The following elements were used as model inputs: Si, Al, Fe, Ca, Mn and Ti.

The performance of the models were evaluated by holding out a random 30% of the samples, and using bootstrap resampling on the remaining training data. The performance of the models on this unseen test data is presented in Figures 1 and 2.

The garnet model reports a root mean square error (RMSE) of 1.9 wt.%, and the ilmenite model a

RMSE of 2.7 wt.%. We consider this model performance to be fit for purpose to provide inputs for mineral resource estimates (MREs), and it is consistent with the performance of similar models that we have built for similar heavy mineral sand projects on West Coast.

An initial model such as that presented here, coupled with an examination of the geochemistry of the entire dataset, provides an approach to optimise sampling. That is to say that additional samples can be located to summarise both the geochemical variability of the dataset, and to in-fill any gaps in the mineral abundances. For example, in the dataset presented here, particular attention should be given to samples that have a predicted garnet abundance of 12–20 wt.% and a predicted ilmenite abundance of 8–20 wt.% as these samples are missing in the dataset. Care should also be taken to ensure that there is adequate sample support around the cut-off grade of any subsequent MRE; this is likely in the 1–3 wt.% garnet and ilmenite range which are not adequately sampled here.

4 Summary

In a heavy mineral sand project, where the mineralogy is quite simple, it is possible to build statistically robust models to predict garnet and ilmenite abundance with limited training data. These models can be validated by selection of additional samples to analyse by automated mineralogy based on an initial model – such as that presented here. Additional sampling and analysis is currently underway, based on the strategy we have used at other West Coast heavy mineral sand projects.

Planned future work involves building models to predict the output of the processing plant directly from the whole-sample major element geochemistry of the raw starting material. The relationship between geochemistry of the raw material and the mineral abundances in the concentrate is less direct, but it is still possible to model the latter from the former.

We also plan to do this with portable XRF results that can be acquired on site and within a matter of hours, without the need to send batches of samples away for laboratory XRF analysis.

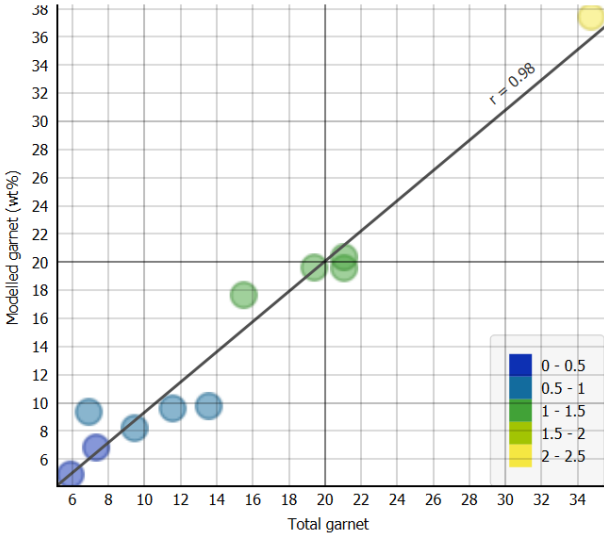

Figure 1. Modelled garnet vs measured garnet for a 30% holdout test dataset. These samples were excluded from the training dataset for the purposes of evaluating model performance, and the model trained on the remaining dataset using bootstrap resampling. Colour scale shows MnO wt.%.

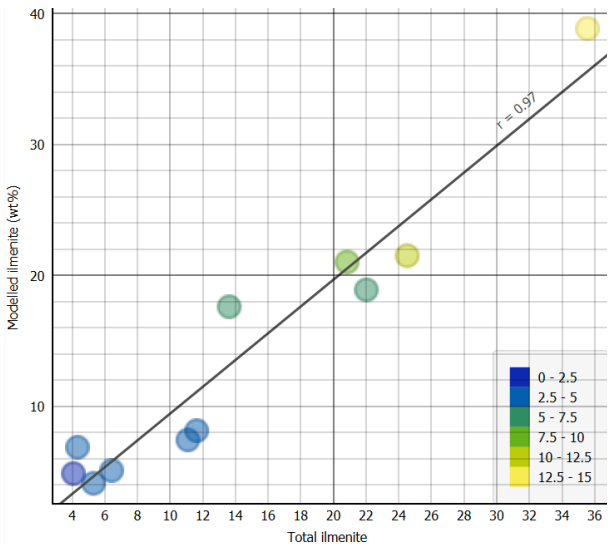

Figure 2. Modelled ilmenite vs measured ilmenite for a 30% holdout test dataset. These samples were excluded from the training dataset for the purposes of evaluating model performance, and the model trained on the remaining dataset using bootstrap resampling. Colour scale shows TiO_2 wt.%.

References

Ritchie, T. W., Scott, J. M., & Craw, D. (2019). Garnet compositions track longshore migration of beach placers in western New Zealand. Economic Geology, 114:513-540.

Tay, S. L., Scott, J. M., Palmer, M. C., Reid, M. R., & Stirling, C. H. (2021). Occurrence, geochemistry and provenance of REE-bearing minerals in marine placers on the West Coast of the South Island, New Zealand. New Zealand Journal of Geology and Geophysics, 64: 89-106.

Machine learning in mineral prospectivity mapping and target generation for critical raw materials

Vesa Nykänen
Information Solutions, Geological Survey of Finland, P.O. Box 77, FI-96101 Rovaniemi, Finland

Abstract The aim of mineral prospectivity mapping (MPM) is to delineate areas that are favourable for certain mineral deposit types. This can be based on prior knowledge using a so-called empirical or data-driven approach or by translating expert knowledge into a mathematical formula by using a conceptual or knowledge-driven approach. Both approaches can benefit from machine learning methods using advanced computer algorithms that can learn from data. This learning can either be supervised or unsupervised. Geographical information systems (GIS) provide a perfect platform for conducting MPM, as in these systems, we can automate and build complex systems to construct models that can be used to predict where the best exploration terrains are hidden. This paper aims to describe how machine learning methods can be utilized in MPM in various steps. This is demonstrated via examples of several past and ongoing research and innovation projects.

1 Mineral prospectivity mapping

A mineral prospectivity mapping (MPM) process may be split into several steps (Fig. 1) (Nykänen et al. 2023). It starts from mineral systems modelling (Step 1), in which the important ingredients of a mineral system that formed the ore body are defined. These critical factors are then translated into mappable parameters that can later be used in GIS for MPM. Then, based on the characterization of the mineral system model, the data are either acquired from existing databases or from the field (Step 2). In the data pre-processing step (Step 3), data are transformed to represent proxies for critical parameters of the mineral systems. This is quite often the most time-consuming part of an MPM-related project if the data acquisition part is not considered. Then follows the actual mineral prospectivity analysis (Step 4), in which two main approaches (or a combination of them) can be used. The final phase in MPM is model validation (Step 5), when statistical methods are applied to test how well the model has performed.

1.1 Data-driven (empirical) approach

The first approach in MPM is data driven (empirical), where prior knowledge of mineral deposits or occurrences is used to train the models. These techniques include many traditional MPM methods, as well as advanced machine learning and deep learning techniques requiring large amounts of training data to be successful. Weights of evidence (WoE) is a traditional statistical technique that is often used in data analysis and modelling for MPM (Bonham-Carter 1994). It is not considered as a form of machine learning, however, as it does not involve the use of algorithms that can learn patterns from data. Logistic regression, another popular classical data-driven MPM method, belongs to the machine learning category, and can be used for classification tasks. It is a statistical method that applies a logistic function to model the likelihood of a binary or categorical result based on one or more input features. Logistic regression is a supervised learning procedure, which means that this method requires labelled, i.e., previously known training data to learn the relationships connecting the input features and the outcome. It is a linear model, which means that it assumes a linear relationship between the input features and the log-odds of the outcome. An artificial neural network (ANN) can be seen as a form of machine learning that is constructed based on the structure and function of the human brain (Tsoukalas and Uhrig 1997; Looney 1997; Nykänen 2008; Cracknell and Reading 2014). ANNs are designed to recognize patterns in multidimensional data, learn from these patterns, and make estimates or conclusions that are derived from this learning. ANNs are comprised of joined nodes, or neurons, processing and transmitting data through a series of layers. The input layer receives the data, which is then passed through one or more hidden layers before reaching the output layer, where the final prediction or decision is made. ANNs can be used for both supervised and unsupervised learning tasks, and they can handle complex non-linear relations between the variables. Artificial neural networks are a form of machine learning, specifically deep learning algorithms, that are used to identify arrays in data and make predictions or decisions based on this learning.

Deep learning is a subfield of machine learning involving the use of ANNs with multiple layers to model and analyse complex relationships in data (LeCun et al. 2015). Deep learning algorithms are constructed to learn from large and complex datasets by automatically extracting features and patterns from the input data. Convolutional neural networks (CNNs) are common examples of deep learning model architectures. The advantage of using deep learning is its ability to learn hierarchical descriptions of the data so that each successive layer within the network learns increasingly from the features. This may be computationally intensive and may also require large amounts of training data, which can limit the applicability of CNNs in data-poor areas. Furthermore, as with all ANNs, deep learning methods also tend to be “black box” in nature, so it can be difficult to interpret their decisions and to understand the reasoning behind their predictions.

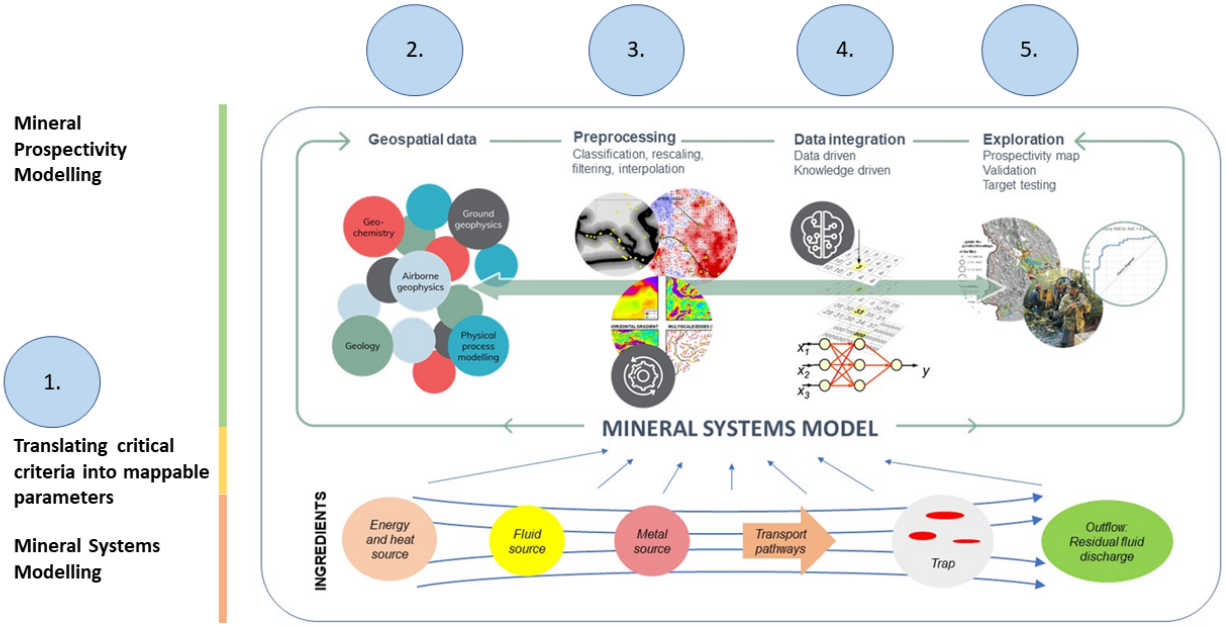

Figure 1. The mineral prospectivity mapping workflow (modified from Nykänen et al. 2023).

Applicable fields of use of ANNs include those where there is a need to analyse complex relationships in data and apply the technique to tasks, e.g., pattern recognition in images for structural geology.

Self-organizing map (SOM) is a machine learning method that uses unsupervised learning to create a low-dimensional representation of high-dimensional data (Kohonen 1990). It can be used for data visualization and exploration, feature extraction, clustering, and pattern recognition assignments. SOM is a powerful and widely used machine learning method that can be applied to different data analysis purposes, and it has also recently been applied to mineral prospectivity mapping (Chudasama et al. 2022b). In MPM, SOMs can be used to identify spatial relationships and patterns in geological, geochemical, and geophysical data that may be indicative of mineralization. The SOM technique can use as input a large exploration dataset, including geological, geochemical, and geophysical data, and find clusters of similar data within the multidimensional feature space. When these datapoints also have spatial information (i.e., coordinates), the resulting map can then be used to identify areas that are most likely to contain mineral deposits based on the patterns and relationships within the data.

1.2 Knowledge-driven (conceptual) approach

The second approach in MPM is knowledge driven (conceptual), where expert knowledge is translated into a mathematical formula or model, and it does not require any known deposit within the study area to be used as training sites (McKay and Harris 2016). Conceptual or knowledge-driven mineral prospectivity methods rely on understanding of geological processes and mineral deposit models to identify areas that are most likely to contain mineral occurrences or deposits. These methods assume that certain geological, geochemical, and geophysical features, and especially an explicit combination of them, are commonly associated with specific types of mineral deposits, and that by mapping these features from various exploration datasets, areas of high mineral potential can be identified.

Fuzzy logic is one example of a conceptual or knowledge-driven approach. It is a method that deals with problem solving and decision making under uncertainty, having no crisp boundaries between sets (true and false). It is based on fuzzy set theory (Zadeh 1965). It is not a subset of machine learning, but it can be used in machine learning as a form of reasoning, allowing a computer code to make decisions based on data. Fuzzy logic overlay is used in geospatial analysis and decision making. It involves the integration of multiple layers of data, each of which represents a different variable or factor that is relevant to a specific decision or analysis. In MPM, these factors are related to the critical mineral systems parameters.

The Boolean logic method or index overlay method uses a set of geological rules or constraints to create a model of the geological environment that is favourable for mineralization. The rules are based on expert knowledge and geological concepts, and the model is then used to identify areas of high mineral potential.

The expert system method uses a set of rules and decision trees based on expert knowledge to identify areas of high mineral potential. The rules are based on geological concepts and the decision trees are used to guide the user through the prospectivity mapping process.

2 Tools developed for public use

The Geological Survey of Finland (GTK) has been maintaining and updating a toolbox called ArcSDM, which was originally established by the U.S. Geological Survey and the Geological Survey of Canada (Sawatzky et al. 2009) and includes some of the key methodologies described by Bonham-Carter (1994). This toolbox can be freely downloaded from GitHub (Geological Survey of Finland 2023a). The tools were updated in the project Mineral Prospectivity Modeller (MPM), funded by the Finnish Funding Agency for Technology (Tekes) (Geological Survey of Finland 2023b). The same project also developed an online web service called MPM Online Tool (Fig. 2) (Geological Survey of Finland 2023c), which can be used to build simple Fuzzy logic overlay models using public geodata from Northern Finland on a web browser-based platform.

Later, in 2018–2021, GTK developed a SOM toolbox, GisSOM (Geological Survey of Finland 2023d), in an EU-funded project entitled NEXT. This toolbox can be used to cluster and visualize data, as described earlier. These SOM tools are currently being further developed in an on-going EIT RawMaterials-funded project entitled DroneSOM (DroneSOM 2023).

The most recent development concerning MPM tools at GTK is the EU-funded project Exploration Information System (EIS), in which the project team is developing new geomodels of selected mineral systems and novel, fast, and cost-efficient spatial data analysis tools for mineral exploration on top of an open GIS platform (EIS 2023). This work is being conducted together with 17 partners from leading research institutes, academia, service providers, and the mining industry. The tools created will also eventually be freely downloadable from GitHub. The project duration is from May 2022 to April 2025.

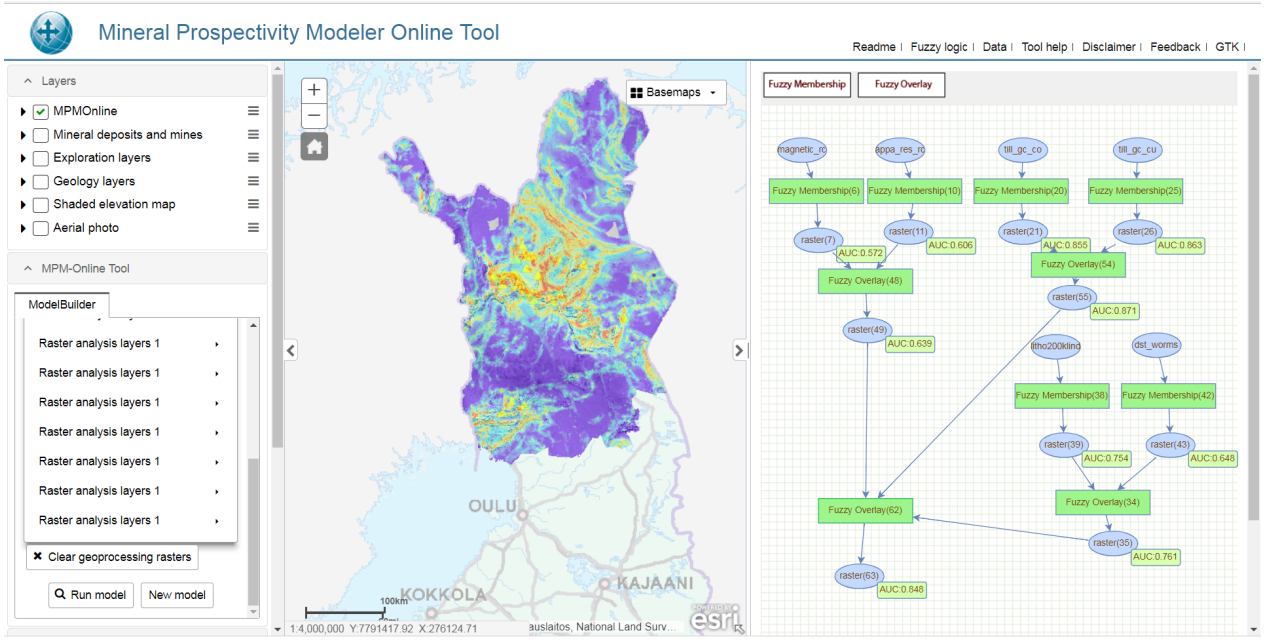

Figure 2. The MPM online tool (Geological Survey of Finland 2023c).

3 Case studies

While developing tools for MPM, the team at GTK has also tested the tools on numerous mineral deposit types using real exploration data from GTK's public databases (Geological Survey of Finland 2023e). The main target test area has been Northern Finland, which is potential for many deposit types, including orogenic gold (Nykänen 2008, Nykänen et al 2008a, Niiranen et al. 2015; Niiranen et al. 2019), orogenic gold with an atypical metal association (Chudasama et al. 2022a, 2022b), IOCG (Nykänen et al. 2008b), magmatic Ni–Cu (Nykänen et al. 2015), and various cobalt-bearing deposit types (Nykänen et al. 2023). In addition to peer-reviewed scientific papers and conference papers, some of these published models are available from the Mineral Deposits and Exploration web map service MDaE of GTK (Geological Survey of Finland 2023f).

Acknowledgements

The MPM project was funded by the Finnish Funding Agency for Technology (Tekes). The NEXT project was funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 776804–H2020-SC5-2017. The EIS project is funded by the European Union's Horizon Europe research and innovation program under grant agreement no. 1010557357, and DroneSOM is funded by EIT RawMaterials.

References

Bonham-Carter GF (1994) Geographic information systems for geoscientists–modelling with GISComputer Methods in the Geosciences 13. Pergamon, Oxford, 398 p

Chudasama B, Torppa J, Nykänen V, Kinnunen J, Lerssi J, Salmirinne H (2022a) Target-scale prospectivity modeling for gold mineralization within the Rajapalot Au-Co project area in northern Fennoscandian Shield, Finland. Part 1: Application of knowledge-driven- and machine learning-based-hybrid- expert systems for exploration targeting and addressing model-based uncertainties. Ore Geology Reviews, 147, https://doi.org/10.1016/j.oregeorev.2022.104937.

Chudasama B, Torppa J, Nykänen V, Kinnunen J (2022b) Target-scale prospectivity modeling for gold mineralization within the Rajapalot Au-Co project area in northern Fennoscandian Shield, Finland. Part 2: Application of self-organizing maps and artificial neural networks for exploration targeting, Ore Geology Reviews, 147, https://doi.org/10.1016/j.oregeorev.2022.104936.

Cracknell MJ, Reading AM (2014) Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences 63, 22-33.

DroneSOM (2023) DroneSOM accessed April 3rd, 2023, from https://dronesom.com/

EIS 2023 Exploration Information System. Accessed April 3rd, 2023, from https://eis-he.eu/

Geological Survey of Finland (2023a) ArcSDM. Spatial Data Modeler for ArcGIS. Accessed March 29th, 2023, from https://github.com/gtkfi/ArcSDM

Geological Survey of Finland (2023b) Mineral Prospectivity Modeller, MPM project. http://projects.gtk.fi/mpm/index.html

Geological Survey of Finland (2023c) MPM Online Tool http://projects.gtk.fi/mpm/online_tool/

Geological Survey of Finland (2023d) GisSOM. Accessed April 3rd, 2023, from https://github.com/gtkfi/GisSOM.

Geological Survey of Finland (2023e) Hakku. Spatial Data Products. Accessed April 4th, 2023, from https://hakku.gtk.fi/en

Geological Survey of Finland (2023f) Mineral Deposits and Exploration – MdaE. Accessed April 3rd, 2023, from https://gtkdata.gtk.fi/mdae/index.html

Kohonen T (1990) The Self-Organizing Map. Proceedings of the IEEE, 78 (9), pp. 1464-1480. doi: 10.1109/5.58325

LeCun Y, Bengio Y, Hinton, G (2015) Deep learning. Nature 521, 4326-444.

Looney CG (1997) Pattern recognition using neural networks: theory and algorithms for engineers and scientists: Oxford University Press, New York, 458 p

McKay G, Harris JR (2016) Comparison of the Data-Driven Random Forests Model and a Knowledge-Driven Method for Mineral Prospectivity Mapping: A Case Study for Gold Deposits Around the Huritz Group and Nueltin Suite, Nunavut, Canada. Natural Resources Research, 25, DOI: 10.1007/s11053-015-9274-z

Niiranen T, Lahti I, Nykänen V (2015) The orogenic gold potential of the Central Lapland greenstone belt, northern Fennoscandian shield. In Maier, W.; Lahtinen, R. and O'Brien (eds). Mineral Deposits of Finland. Elsevier. pp. 733-752.

Niiranen T, Nykänen V, Lahti I (2019) Scalability of the mineral prospectivity modelling – An orogenic gold case study from northern Finland. Ore Geology Reviews, Volume 109, 2019, Pages 11-25, ISSN 0169-1368, https://doi.org/10.1016/j.oregeorev.2019.04.002.

Nykänen V (2008) Radial Basis Functional Link Nets Used as a Prospectivity Mapping Tool for Orogenic Gold Deposits Within the Central Lapland Greenstone Belt, Northern Fennoscandian Shield. Natural Resources Research, 17, DOI: 10.1007/s11053-008-9062-0

Nykänen V, Groves DI, Ojala VJ, Gardoll SJ (2008a) Combined conceptual/empirical prospectivity mapping for orogenic gold in the northern Fennoscandian Shield, Finland. Australian Journal of Earth Sciences 55 (1), 39-59.

Nykänen V, Groves DI, Ojala VJ, Eilu P, Gardoll SJ (2008b) Reconnaissance-scale conceptual fuzzy-logic prospectivity modelling for iron oxide copper-gold deposits in the northern Fennoscandian Shield, Finland. Australian Journal of Earth Sciences 55 (1), 25-38.

Nykänen V, Törmänen T, Niiranen T (2023) Cobalt prospectivity using a conceptual fuzzy logic overlay method enhanced with the mineral systems approach. Natural Resources Research (Manuscript submitted for publication).

Sawatzky DL, Raines GL, Bonham-Carter GF, Looney CG (2009) Spatial DataModeller (SDM): ArcMAP 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural networks http://arcscripts.esri.com/details.asp?dbid=15341.

Tsoukalas LH, Uhrig RE (Eds.) (1997) Fuzzy and neural approaches in engineering: John Wiley & Sons, Inc., New York, 587 p

Zadeh LA (1965) Fuzzy sets 8. Institute of Electric and Electronic Engineering, Information and Control, pp. 338–353.

Navigating the complexities of decision-making for Critical Mineral Exploration Campaigns: Insights form AI-based Geological Prospectivity and Risk Models

Mohammad Parsa
[1]Geological Survey of Canada, Ottawa, Ontario, K1A 0E8, Canada

Abstract. This study presents an integrated framework for interpreting geological prospectivity models, which are central to decision-making and land-use planning for critical mineral exploration campaigns. Besides geological prospectivity, there are other factors that are essential to policymaking. Different uncertainties linked to prospectivity models are of factors affecting geological prospectivity and, therefore, their interpretation. In addition, mineral deposits usually form in clusters and follow certain spatial patterns, making spatial distribution another important factor for the interpretation of geological prospectivity models. Herein, an integrated approach to interpreting geological prospectivity models is presented. This approach combines geological prospectivity, uncertainty, and spatial distribution to help make informed decisions while narrowing down the search space for mineral exploration. An example of using this approach is further demonstrated for national-scale delineation of exploration targets for REEs in Canada.

1 Introduction

Critical minerals are essential for renewable energy technologies and play an inevitable role in transitioning to a carbon-free economy. The demand for renewable energy sources continues to surge, leading to an ever-increasing demand for critical minerals. Ensuring a secure and sustainable supply of critical minerals is, thus, crucial for the transition to a carbon-free economy.

Geological prospectivity models can help policymakers make informed decisions about critical mineral exploration campaigns. These models can help identify areas where critical minerals are likely to be discovered, thereby helping understand the potential supply of critical minerals and make decisions about where to invest in exploration and mining activities.

Geological prospectivity models are mostly derived by the application of various supervised regression techniques. These are probability models in which high probability values pertain to favourable zones for discovering a given type of mineral deposits. These models, therefore, are continuous models that are devoid of interpretation. One must, thus, assign a threshold value to these models for demarcating exploration targets.

Methods used for interpreting geological prospectivity models range from subjective assigning of a threshold value to objective interpretation of these models. The former method entails intrinsic problems. There is a chance of overestimating or underestimating exploration targets while setting a subjective value for interpreting geological prospectivity models. Turning to the latter methods, abrupt changes in the probability values (Porwal et al. 2003), spatial distribution of probability values (Parsa et al. 2017), and risk-return analysis (Parsa and Pour 2021) have been applied to objective interpretation of geological prospectivity models. Although these objective solutions address the problem of subjective bias, there is a need for a holistic approach to interpreting geological prospectivity models that considers spatial and statistical distribution of prospectivity values together with uncertainties linked to prospectivity models.

Herein, an integrated methodology is proposed that considers the above aspects while interpreting geological prospectivity models. This methodology has been applied to national-scale geological prospectivity models of a suite of critical minerals, helping select high priority exploration targets.

2 Methodology

There are data- and model-related uncertainties that affect the results of geological prospectivity models. An open-source framework for measuring different uncertainty types for geological prospectivity mapping is presented in this study.

This framework starts with selecting random sub-samples from the labelled samples. Each set of random sub-samples is fed into different machine and deep learning algorithm, leading to a set of geological prospectivity models. These models are derived with different labelled samples and different regression models, helping measure the data- and model-related uncertainties. This is followed by the application of risk-return analysis and spatial measurements for objectively interpret the geological prospectivity models. This framework was exploited for generating exploration targets of several critical minerals, including REEs.

The proposed framework is an open-source, versatile approach allowing for addition of algorithms or datasets.

Acknowledgements

I am grateful for the invitation of the organizers of the machine learning session, Chetan Nathwani, Francisca M. Maepa, and Daniel Gregory.

References

Parsa M, Maghsoudi A, Yousefi M (2017) An improved data-driven fuzzy mineral prospectivity mapping procedure; cosine amplitude-based similarity approach to delineate exploration targets. International journal of applied earth observation and geoinformation. 1;58:157-67.

Parsa M, Pour AB. A (2021) simulation-based framework for modulating the effects of subjectivity in greenfield mineral prospectivity mapping with geochemical and geological data. Journal of Geochemical Exploration. 1;229:106838.

Porwal A, Carranza EJ, Hale M (2003) Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping. Nat Resour RES. 12:1-25.

Assessing tourmaline as an indicator mineral using multivariate statistics

Eduardo Valentin dos Santos[1], Georges Beaudoin[1], Bertrand Rottier[1]
[1]Département de géologie et génie géologique, Université Laval, Quebec, Canada

Abstract. Tourmaline chemistry from different geological environments, including granite, Li-rich and -poor pegmatite, porphyry Cu-Mo, granite-related Sn, volcanogenic massive sulphide (VMS), unconformity U, orogenic gold, epithermal Au-Ag, and metapelite, were analysed by electron probe micro-analyser (EPMA) and laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). The data was processed and analysed using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). Most tourmaline from the majority of the investigated geological environments straddle along dravite-schorl, with the exception of unconformity U (Mg-foitite), Li-bearing pegmatite, and some granite-related Sn (elbaite-liddicoatite). LA-ICP-MS trace element PCA analysis results in good separation of Li-rich pegmatite, and unconformity U deposits. Granite-related Sn deposits tend to plot between Li-pegmatite and other magmatic rocks and magmatic-hydrothermal deposits on the first, second, and third components. Both Li-pegmatite and granite-related Sn are inversely correlated to orogenic gold deposits. PLS-DA analysis results in good separation of Li-rich and Li-poor pegmatite, and unconformity U. There is considerable overlap between other classes using PCA and PLS-DA. Further data collection and classification using machine learning (Random Forest) methods are the next steps of this project, as they will likely allow better discrimination of tourmaline from the investigated geological environments.

1 Introduction

Tourmaline is a common mineral in several geological environments and mineral deposits (Slack 1996, Trumbull et al. 2020). It has one of the largest stability ranges of crustal minerals and is characterized by low volume diffusion rates, so it usually preserves its original chemical composition and zonings reflecting the physicochemical conditions of its crystallization environments (van Hinsberg et al. 2011, Slack and Trumbull 2011).

Tourmaline compositional data can be utilized as a pathfinder for different types of deposits. Sciuba et al. (2021) demonstrated that the tourmaline composition from orogenic gold deposits is controlled by the fluid composition, metamorphic facies, and composition of the country rocks, and is overall rich in Sr, V, and Ni and poor in Li, Be, Ga, Sn, Nb, Ta, U, and Th compared to tourmaline from other deposit types and geological environments.

Nonetheless, the published tourmaline trace element datasets are often inconsistent and incomplete in the number of analyzed elements, making it difficult to compare different deposit types using statistical and machine learning methods. In this context, this project aims to build a homogeneous database and develop criteria for

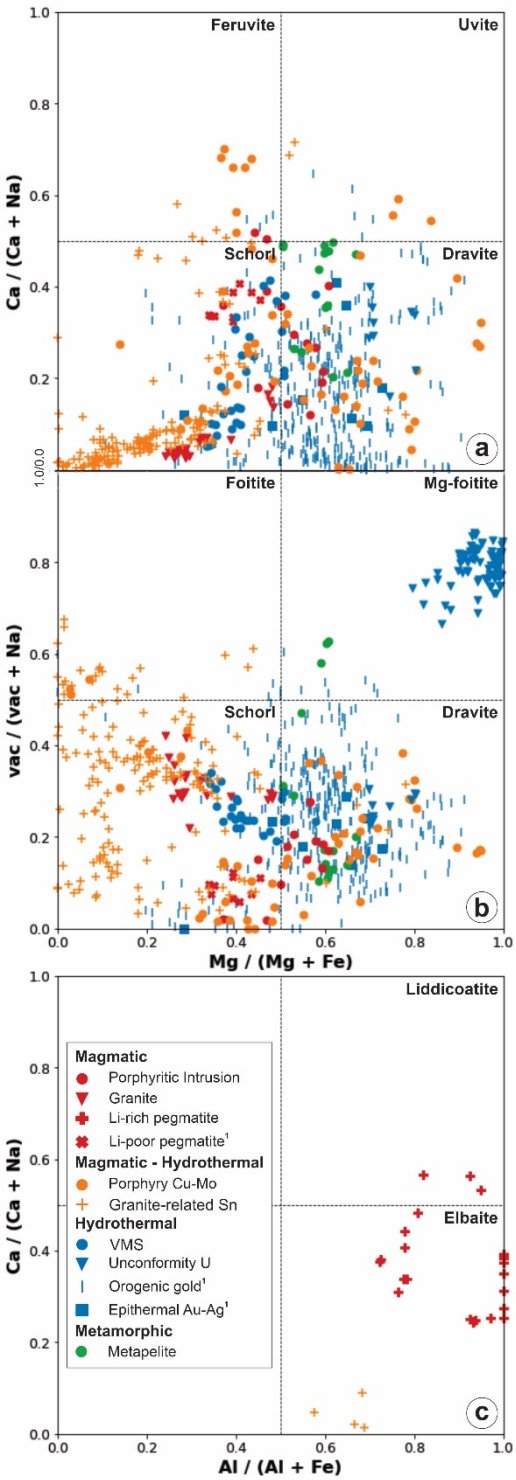

Figure 1. *Composition of tourmaline from different geological environments. Tourmaline classification diagrams from Henry et al. (2011).* ***a.*** *Mg/(Mg+Fe) vs. Ca/(Ca+Na),* ***b.*** *Mg/(Mg+Fe) vs. X-vac/(X-vac+Na), and* ***c.*** *Ca/(Ca+Na) vs. Al/(Al+Fe). 1 – Sciuba et al. (2021).*

using tourmaline chemistry as a geochemical prospecting tool.

2 Methods

2.1 Electron probe micro-analysis (EPMA)

Major and minor elements were measured with a CAMECA SX-100 electron probe micro-analyzer equipped with five WDS spectrometers at Université Laval. Silicon, Ti, Al, V, Sc, Fe, Mg, Mn, Zn, Cu, Ni, Co, Ca, Sr, Na, K, F, and Cl were analysed using a 5 µm beam, 15 kV acceleration voltage, and 20 nA current, counting 10s on background and 20s on peak.

Tourmaline structural formula were calculated according to Henry et al. (2011), on the basis of 31 anions, 29 oxygen atoms, and 3 apfu B.

2.2 Laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS)

Major, minor, and trace elements were measured using an Agilent 8900 ICP-QQQ-MS coupled with a RESOlution-SE 193nm excimer at Université Laval. Acquisition parameters were 38 µm lines, at a 5 µm/s line speed, 10 Hz laser frequency, and fluence of 4.67~8.25 J.cm^{-2}.

The Si concentration measured by EPMA was used as internal standard. The reference materials NIST-610 (^{7}Li, ^{9}Be, ^{23}Na, ^{27}Al, ^{44}Ca, ^{47}Ti, ^{53}Cr), NIST-612 (^{39}K, ^{45}Sc, ^{59}Co, ^{60}Ni, ^{71}Ga, ^{85}Rb, ^{86}Sr, ^{89}Y, ^{107}Ag, ^{111}Cd, ^{115}In, ^{118}Sn, ^{133}Cs, ^{137}Ba, ^{139}La, ^{140}Ce, ^{141}Pr, ^{146}Nd, ^{152}Sm, ^{153}Eu, ^{155}Gd, ^{159}Tb, ^{163}Dy, ^{165}Ho, ^{166}Er, ^{169}Tm, ^{172}Yb, ^{175}Lu, ^{178}Hf, ^{181}Ta, ^{182}W, ^{197}Au, ^{232}Th, ^{238}U), and GSE-1g (^{11}B, ^{24}Mg, ^{51}V, ^{55}Mn, ^{56}Fe, ^{57}Fe, ^{65}Cu, ^{66}Zn, ^{92}Zr, ^{93}Nb, ^{95}Mo, ^{207}Pb) were used as primary standards depending on the element. When not used for quantification, NIST-610, NIST-612, GSE-1g, KL2-G, and ML3B-G were used as secondary standards to control data quality.

Data reduction was carried out using the Iolite package for Igor Pro software (Paton et al. 2011).

2.3 Multivariate statistical analysis

Prior to PCA and PLS-DA, the dataset was processed for variables (elements) with values below the detection limit (censored values). Elements with more than 40% censored values were excluded. For the remaining, censored values were imputed using the log-ratio Expectation-Maximization (lrEM) algorithm (R package zCompositions; Palarea-Albaladejo and Martín-Fernández 2015). This algorithm ensures that censored values are replaced by imputed values between zero and the detection limit. After imputation the dataset was transformed using centered log-ratios to overcome the closure effect in compositional data (Aitchison 1986).

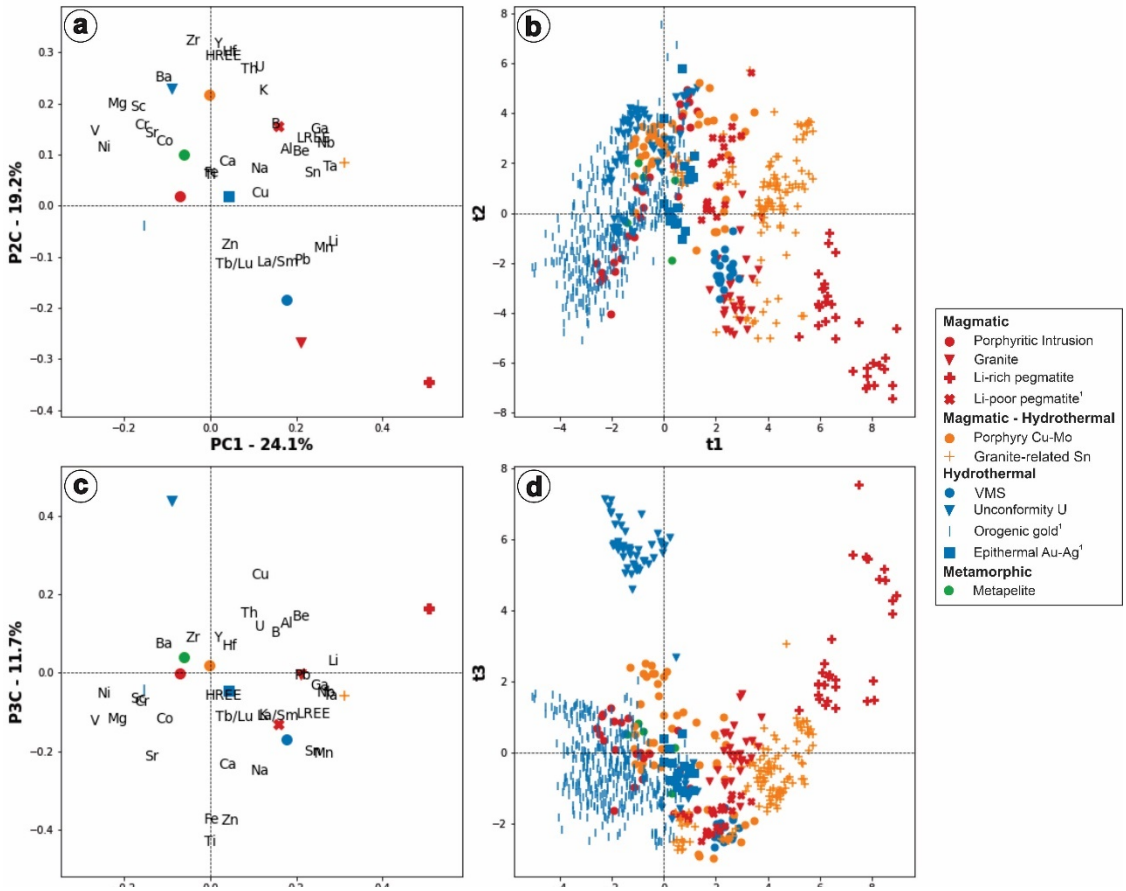

Figure 2. *PCA diagrams of tourmaline trace elements measured by LA-ICP-MS. Loadings (PC1-PC2) and scores (t1-t2) on the first and second components are shown on* ***a*** *and* ***b****. Loadings (PC1-PC3) and scores (t1-t3) on the first and third components are shown on* ***c*** *and* ***d****. 1 – Sciuba et al. (2021).*

The PCA is an unsupervised method used to reduce a larger set of variables to a smaller number of uncorrelated variables called principal components (PC). Each PC explains part of the variance of the data, with the first (PC1) capturing the greatest variance, followed by the second (PC2), and so forth. The PCA loadings are the correlation coefficients between original variables and PCs, and provide information on the impact of a variable on a given PC. The PCA Scores are composite values for each sample on each PC, calculated using the original variable values and factor weights (Makvandi et al. 2019 and references therein).

The PLS-DA, on the other hand, is a supervised classification method that combines partial least squares regression using a continuous data matrix X (elements), and discriminant analysis, using a categorical outcome matrix Y (different classes). The PLS components (scores; t1, t2, etc.) and loadings (wq*1, wq*2, etc.) are among the main PLS-DA outcomes. Another significant output is the variable importance on the projection (VIP) plot (Fig. 3e), where elements with VIP values larger than 1.0 are the most influential variables in the model, variables between 0.8 and 1.0 are moderately influential, and values smaller than 0.8 do not contribute significantly in the sample classification (Makvandi et al. 2021 and references therein).

For both PCA and PLS-DA plots, positively correlated variables are shown grouped, whereas negatively correlated variables plot diametrically opposed. The location of variables is dependent on their contribution to discrimination. Variables near the origin contribute weakly to classification, whereas the outer variables are highly variable between classes (Caraballo et al. 2022).

3 Results and discussion

Thirty tourmaline-bearing samples from granite, porphyritic intrusion, Li-rich pegmatite, porphyry Cu-Mo, granite-related Sn, VMS, unconformity U, and metapelite were investigated by EPMA and LA-ICP-MS. The Sciuba et al. (2021) dataset was added to this study, because the same set of elements were analysed.

Tourmaline major element composition shows large compositional ranges, reflected in different tourmaline species (Fig. 1). Most of the investigated geological environments present tourmaline that ranges from schorl (Fe-rich, sodic) to dravite (Mg-rich, sodic). Porphyritic intrusion, granite-related Sn, porphyry Cu-Mo, and orogenic gold have foitite (Fe-rich, X-site vacant), feruvite (Fe-rich, calcic), and uvite (Mg-rich, calcic), but these represent minor members of a dominantly schorl-dravite population of the same deposit types. Unconformity U deposits and Li-rich pegmatite are Mg-foitite (Fe-rich, X-site vacant) and elbaite (Li-rich, sodic) or liddicoatite (Li-rich, calcic), respectively.

The trace element compositions, variance, and correlations along different classes were analyzed by PCA and PLS-DA. From the PCA score and

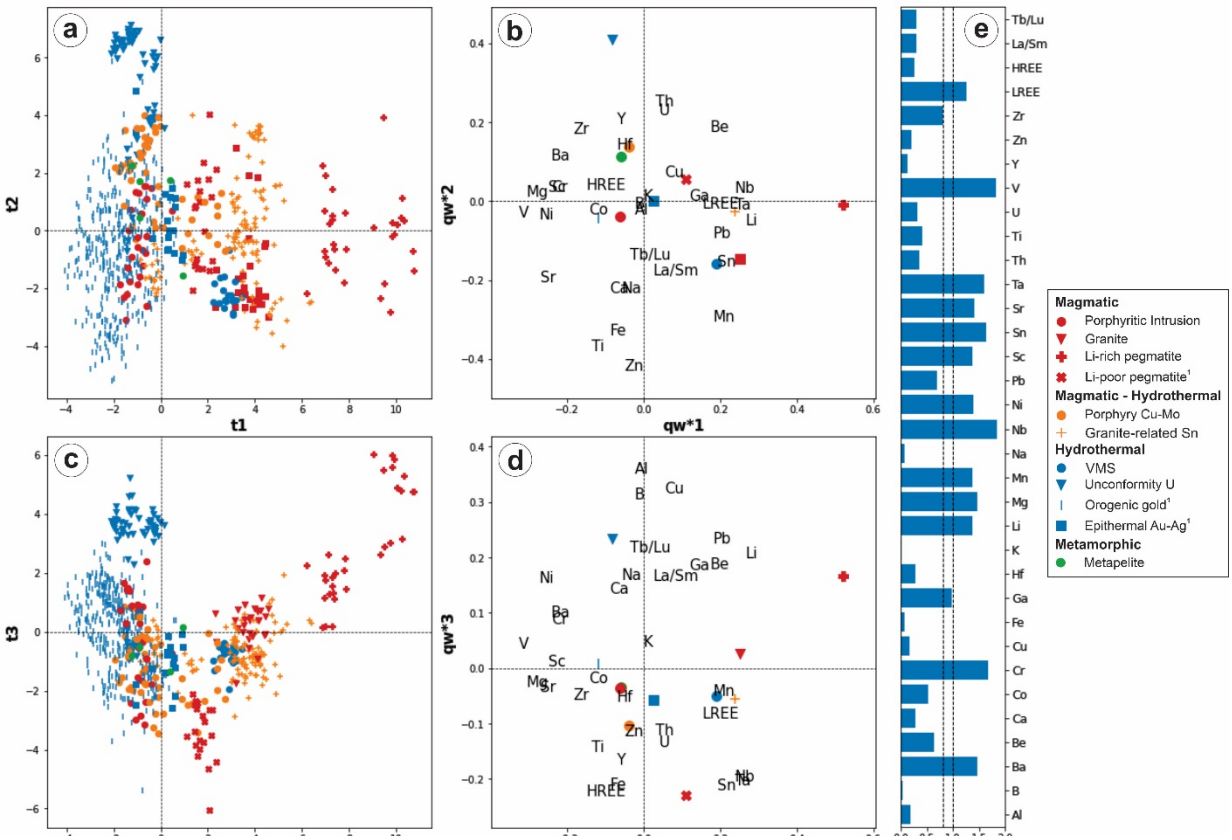

Figure 3. *PLS-DA diagrams of tourmaline trace elements measured by LA-ICP-MS. Loadings (qw*1-qw*2) and scores (t1-t2) on the first and second components are shown on* ***a*** *and* ***b****. Loadings (qw*1-qw*3) and scores (t1-t3) on the first and third components are shown on* ***c*** *and* ***d****. The VIP plot is shown on* ***e****. 1 – Sciuba et al. (2021).*

loadings plots (Fig. 2), unconformity U and Li-rich pegmatite are well separated from the other classes on the t1 vs. t3 (Fig. 2b) and t1 vs. t2 and t3 (Fig. 2b and 2d) score plots.

On the loadings plot of the first two principal components (Fig. 2a), Li-rich pegmatite is influenced by high contents of the elements on the lower right quadrant (i.e., Li, Mn, Pb, La/Sm, Tb/Lu, Zn), and by extremely low contents of the elements on the opposed quadrant (i.e., Ni, V, Cr, Co, Sc, Sr, Mg, Ba, Zr). On the first and third loadings plot (Fig. 2c), Li-rich pegmatite is influenced by a similar set of elements, with the addition of Be, Al, B, U, Th, Cu, Y, and Hf. Nonetheless, unconformity U deposits are characterized by a low concentration of most elements, especially Fe, Zn, and Ti.

All other groups are largely overlapping, but granite-related Sn tends to plot between Li-rich pegmatite and granite, Li-poor pegmatite, and VMS on both diagrams. Granite-related Sn deposits are positively correlated with LREE, Na, Al, Ga, Sn, Nb, and Ta.

Epithermal Au-Ag, metapelite, porphyritic intrusion, and porphyry Cu-Mo tend to plot near the origin. The first slightly tends toward elements enriched in felsic rocks, whereas the rest tends to the elements enriched in mafic rocks. Orogenic gold deposits are inversely correlated to Li-rich pegmatite and characterized by high Sr, V, Ni, Co, Cr, Mg, and Sc concentrations.

The PLS-DA score and loading plots (Fig. 3a, 3b, 3c, and 3d) highlight the separation of Li-rich pegmatite and unconformity U on the first, second, and third components. Lithium-poor pegmatite is well-defined on the first and third components. The Li-rich pegmatite class is evidenced by high concentrations of Be, Li, Mn, Nb, Pb, Sn, Ta, and LREE, and low concentrations of Ba, Cr, Mg, Ni, Sc, Sr, and V. Unconformity U is characterized by high Th, U, Y, and Zr and low Ca, Fe, Mn, Na, Sn, Sr, Ti, and Zn. Li-poor pegmatite (Fig. 3f) is well separated by the third component and is evidenced by high Fe, Nb, Sn, Ta, and U and low Al, B, Cu, Ni, and V.

The VIP plot (Fig. 3e) highlights the importance of Ba, Cr, Li, Mg, Mn, Nb, Ni, Sc, Sn, Sr, Ta, V, and LREE for the PLS-DA model.

4 Conclusions

Both PCA and PLS-DA are good at classifying Li-rich pegmatite and unconformity U. However, these classes are different from the other geological environments by their major element composition since they are elbaite-liddicoatite and Mg-foitite.

Both models suggest that tourmaline from orogenic gold deposits has a high concentration of elements enriched in mafic rocks. In contrast, Li-rich pegmatite and granite-related Sn deposits tourmaline have a high concentration of elements enriched in evolved felsic rocks. This suggests that tourmaline chemistry can record the nature of the magmatic source of hydrothermal fluids for magmatic-hydrothermal deposits, or the composition of buffering rocks for metamorphic fluids, as shown by Sciuba et al. (2021).

Further data collection, PLS-DA algorithm tuning with the best-performing elements, and classification using machine learning (Random Forest) methods are the next steps of this project, as they will likely allow better discrimination of tourmaline from the investigated geological environments.

Acknowledgements

This research project was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), Agnico Eagle Mines Ltd. (AEM), and the Ministère des Ressources Naturelles et des Forêts du Québec (MRNF).

References

Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, London

Caraballo E, Dare S, Beaudoin G (2022) Variation of trace elements in chalcopyrite from worldwide Ni-Cu sulfide and Reef-type PGE deposits: implications for mineral exploration. Mineral Deposita 57:1293-1321

Henry DJ, Novak M, Hawthorne FC, Ertl A, Dutrow BL, Uher P, Pezzotta F (2011) Nomenclature of the tourmaline-supergroup minerals. Am Mineral 96:895-913

Makvandi S, Beaudoin G, McClenaghan MB, Quirt D, Ledru P (2019) PCA of Fe-oxides MLA data as an advanced tool in provenance discrimination and indicator mineral exploration: case study from bedrock and till from the Kiggavik U deposits area (Nunavut, Canada). J Geochem Explor 197:199-211

Makvandi S, Huang X, Beaudoin G, Quirt D, Ledru P, Fayek M (2021) Trace element signatures in hematite and goethite associated with the Kiggavik-Andrew Lake structural trend U deposits (Nunavut, Canada). Mineral Deposita 56:509-535

Palarea-Albaladejo J, Martín-Fernández JA (2015) zCompositions – R package for multivariate imputation of left-censored data under a compositional approach. Chemometr Intell Lab Syst 143:85-96

Paton C, Hellstrom J, Paul B, Woodhead J, Hergt J (2011) Iolite: Freeware for the visualisation and processing of mass spectrometric data. J Anal At Spectrom 26:2508-2518

Sciuba M, Beaudoin G, Makvandi S (2021) Chemical composition of tourmaline in orogenic gold deposits. Mineral Deposita 56:537-560

Slack JF (1996) Tourmaline associations with hydrothermal ore deposits. In: Grew ES, Anovitz LM (eds) Boron: mineralogy, petrology and geochemistry. Rev mineral, vol 33, pp 559-643

Slack JF, Trumbull RB (2011) Tourmaline as a recorder of ore-forming processes. Elements 7 321-326

Trumbull RB, Codeço MS, Jiang S-Y, Palmer MR, Slack JF (2020) Boron isotope variations in tourmaline from hydrothermal ore deposits: a review of controlling factors and insights for mineralizing systems. Ore Geol Rev 125:103682

van Hinsberg VJ, Henry DJ, Dutrow BL (2011) Tourmaline as a petrologic forensic mineral: a unique recorder of its geologic past. Elements 7: 327-332

