Artificial Intelligence in Geosciences 2 (2021) 60–75

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

Machine learning-based prediction of trace element concentrations using data from the Karoo large igneous province and its application in prospectivity mapping

Steven E. Zhang [a], Glen T. Nwaila [b,*], Julie E. Bourdeau [c], Lewis D. Ashwal [b]

[a] SmartMin Limited, 39 Kiewiet Street, Helikon Park, 1759, South Africa
[b] School of Geosciences, University of the Witwatersrand, 1 Jan Smuts Ave, Johannesburg, 2000, South Africa
[c] Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, K1A 0E8, Canada

ARTICLE INFO

Keywords:
Machine learning
Predictive modelling
Compositional data
Prospectivity mapping
Anomaly detection
Karoo igneous province

ABSTRACT

In this study, we present a machine learning-based method to predict trace element concentrations from major and minor element concentration data using a legacy lithogeochemical database of magmatic rocks from the Karoo large igneous province (Gondwana Supercontinent). Wedemonstrate that a variety of trace elements, including most of the lanthanides, chalcophile, lithophile, and siderophile elements, can be predicted with excellent accuracy. This finding reveals that there are reliable, high-dimensional elemental associations that can be used to predict trace elements in a range of plutonic and volcanic rocks. Since the major and minor elements are used as predictors, prediction performance can be used as a direct proxy for geochemical anomalies. As such, our proposed method is suitable for prospective exploration by identifying anomalous trace element concentrations. Compared to multivariate compositional data analysis methods, the new method does not rely on assumptions of stoichiometric combinations of elements in the data to discover geochemical anomalies. Because we do not use multivariate compositional data analysis techniques (e.g. principal component analysis and combined use of major, minor and trace elements data), we also show that log-ratio transforms do not increase the performance of the proposed approach and are unnecessary for algorithms that are not spatially aware in the feature space. Therefore, we demonstrate that high-dimensional elemental associations can be modelled in an automated manner through a data-driven approach and without assumptions of stoichiometry within the data. The approach proposed in this study can be used as a replacement method to the multivariate compositional data analysis technique that is used for prospectivity mapping, or be used as a pre-processor to reduce the detection of false geochemical anomalies, particularly where the data is of variable quality.

1. Introduction

Geochemical data is a staple of geosciences and has been used in various forms to study geodynamics, crustal processes, surface processes and mineral systems. Outside of systematic surveying and exploration programs, extensive collections of geochemical data do exist, but accessing and using them requires significant user effort. We refer to such data as 'legacy geochemical data'. One example of this type of data is the collection of geochemical analyses in the hands of various researchers worldwide, which are variable in quality, production methodology and intent (Adcock et al., 2013; Ashwal 2021; Chen et al., 2020). Data-driven techniques present new opportunities to repurpose legacy geochemical data, extract additional insights and create predictive modelling methodologies (Karpatne et al., 2018; Chen et al., 2020). Four main classes of methods have been used to analyze geochemical data, some of which incorporate data-driven approaches; these are: (a) geostatistics, (b) fractal/multi-fractal models, (c) compositional data analysis, and (d) machine learning.

In this paper, we illustrate the usefulness of machine learning for legacy geochemical data analysis by creating a specific methodology that uses major (>1 wt %) and minor (1–0.1 wt %) element geochemistry to predict trace (<0.1 wt %) elements. The dataset contains both volcanic and plutonic units of the Karoo large igneous province (Gondwana Supercontinent) and is an aggregation of highly variable data in terms of analytical methods, accuracy and precision, as well as data levelling. This amount of data variability would present serious challenges for existing

* Corresponding author.
E-mail addresses: ezhan053@uottawa.ca (S.E. Zhang), Glen.Nwaila@wits.ac.za (G.T. Nwaila).

https://doi.org/10.1016/j.aiig.2021.11.002
Received 26 September 2021; Received in revised form 29 November 2021; Accepted 30 November 2021
Available online 2 December 2021

compositional data analysis methods, which uniformly assume high quality and mapping-ready geochemical data (e.g. Grunsky and de Caritat, 2019). Volcanic rocks are particularly interesting for prospectivity mapping method development, as they do not necessarily exhibit mineral-based chemical stoichiometry and therefore are not ideal for current multivariate compositional data-analysis workflows that specifically interpret and leverage stoichiometric compositions (e.g. Grunsky and de Caritat, 2019). In this paper, we show that our method can be used to detect geochemical anomalies and therefore be used for prospectivity mapping. Compared to a well-established methodology (e.g. Grunsky, 2013; Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018; Grunsky and de Caritat, 2019), there are multiple benefits to our approach, which includes: (a) a more streamlined workflow that does not require log-ratio transformations and associated data preprocessing tasks; (b) no reliance or assumptions on the existence of stoichiometry in samples; (c) a high explanatory power of regional geochemical variability, which should decrease false positivity rate of geochemical anomaly detection; and (d) a tolerance of low quality, variably sourced data (secondary data unsuitable for traditional prospectivity mapping). Data variability is important for machine learning methodology development, in order to maximise the generalisability of the new method to other data and applications (Therrien and Doyle, 2018; Hyontai, 2018). In addition, since the approach proposed in this study is independent of any specific machine learning algorithm, it would be easy to improve the performance of the proposed approach by using newer machine learning algorithms, should they become available.

2. Spatial and predictive techniques for mineral resources evaluation and prospectivity mapping

Many geochemical data analysis techniques have been trialled and applied in academic studies and geochemical exploration projects. In general, properties of geochemical data should be taken into account, such as: (a) statistical distributions that deviate from normality (Reimann and Filzmoser, 2000); (b) heteroscedasticity (i.e. unequal variances or heterogeneity of variances indicating different modes of distribution (Thompson and Howarth 1973, 1976a, b, 1978; Thompson 1973, 1982; Fletcher 1981; Stanley and Sinclair 1986; Stanley 2003); (c) spatial non-stationarity (e.g. Ellefsen and Van Gosen, 2020); and (d) compositional properties of the data (e.g. the closure problem, Aitchison, 1982).

The first approach for processing large datasets is matheronian geostatistics, which began in the 1950s (Krige 1951, 1952, 1955; Matheron, 1962), followed by the seminal works of Journel (1974, 1980), Isaaks and Srivastava (1989), Gómez-Hernández (1991), Deutsch and Journel (1992, 1997), Goovaerts (1994) and Pyrcz and Deutsch (2014). Geostatistics is devoted to the analysis and interpretation of possible spatial distributions of geophysical properties and their uncertainties using spatial interpolation, often to create maps or volumes. It assumes that some properties are more similar between samples at shorter distances. Geostatistics is used in many fields, e.g. hydrogeology, hydrology, meteorology, oceanography, geochemistry, geography, soil sciences, forestry and landscape ecology (Isaaks and Srivastava, 1989). However, in some applications, the use of geostatistics is challenging, for example: (a) where the nugget effect is significant; (b) where the simultaneous use of high-dimensional numerical and categorical types of data to estimate the target, or; (c) where non-traditional data, such as descriptive geological data can be useful during resource estimation (Nwaila et al., 2020; Zhang et al., 2021).

The second approach uses fractal models of geochemical systems. Cheng et al. (1994) first proposed fractal and multifractal models in geochemistry for the detection of geochemical anomalies. This is achieved through considering the frequency and spatial variances of geochemical data and has become known as the concentration–area model or the Cheng-Agterberg model. Further development in this approach includes the introduction of the spectrum–area multi-fractal model (Cheng et al., 2000) and local singularity analysis (Cheng, 2007). Like any other mathematical and statistics-based techniques, uses of fractal modelling can be limited by mathematically complexity. More recently, fractal models have been added to geographic information system software libraries to provide easy end-user access to aid analysis of large geochemical exploration and environmental pollution data (Carranza, 2009; Chen et al., 2017; Yu et al., 2019).

The third technique involves discipline-specific and often geospatially-tagged compositional data, which are components of samples measured as proportions of the whole, and therefore, each component only carries relative information (Aitchison, 1982). Most lithogeochemical and mineral chemistry data for major, minor and trace elements are subject to the 'constant sum' or 'closure' problem and are restricted in their numerical range. This results in unreliable results from correlation measures and Euclidean geometry-based analysis. Aitchison (1982) introduced log-ratio based data transformations under the principle that for compositional data, concentrations of individual components are irrelevant because information lies in their relative proportions. Several transformations are used to date, including the: centred log-ratio (CLR), additive log-ratio (ALR) and isometric log-ratio (ILR) (Pawlowsky-Glahn et al., 2015). These transformations map the strictly positive parts of a composition to the real number set. The CLR and ILR transformations preserve the correct representation of compositional data in Euclidean space. Although the ILR transformation also creates full-ranked covariance matrices that are useful for some traditional multivariate analysis techniques, the transformed data may be difficult to interpret, depending on the choice of the orthonormal basis (Egozcue et al., 2003; Flzmoser and Hron, 2009). Compositional geochemical data that includes major and minor oxides (SiO_2, TiO_2, Al_2O_3, FeO_{total}, CaO, MgO, Na_2O, K_2O, and P_2O_5) are typically transformed before their use in multivariate analyses (less often for trace element data), often using the CLR transformation (e.g. Grunsky, 2013; Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018; Grunsky and de Caritat, 2019). For some tasks, the ALR transformation may also work (Templ, 2021); however its asymmetric nature in parts of the composition necessitates manual exploration of the choice of denominator in the transformation. It is worth noting that the effects of closure on the prediction of trace elements using predictive modelling are yet to be established (Filzmoser et al., 2009; McKinley et al., 2016) and the interaction between this particular type of data transformation and data-driven predictive modelling techniques are largely unexplored.

The fourth technique for processing large geochemical datasets involves primarily machine learning (e.g. Lawley et al., 2021). Machine learning is a branch of artificial intelligence (AI) and computer science, which focuses on the use of data and algorithms to imitate learning (Burkov, 2020; Chen et al., 2020). Unlike static models, machine learning models are capable of self-improvement through exposure to additional data (Karpatne et al., 2018). Arthur Samuel is credited for coining the term "machine learning" with his research around the game of checkers (Samuel, 1959). It is also worth mentioning that geosciences has dealt with big data long before other industries, for example, through large geophysical surveys (e.g. seismic data acquisition) in exploration for mineral resources. In the last decade, there has been a proliferation of machine learning in many disciplines. Compared to geostatistics, fractal/multi-fractal analysis and compositional data analysis, there is far less research and publications on the use of machine learning for geochemical analysis, although the technique is theoretically very powerful for all scales of geochemical mapping (Chen et al., 2020). The power of machine learning primarily lies in its ability to recognize and leverage high-dimensional patterns to perform inferences. However, additional and sometimes key benefits to the use of machine learning lies in its ability to tolerate some level of data variability, for example, those introduced through the use of multiply sourced data (Therrien and Doyle, 2018; Hyontai, 2018), and whose effects are countered through algorithm selection and explicit controls on model overfitting. Machine learning workflows require application-specific developments and the typical task in machine learning is to create a methodology that

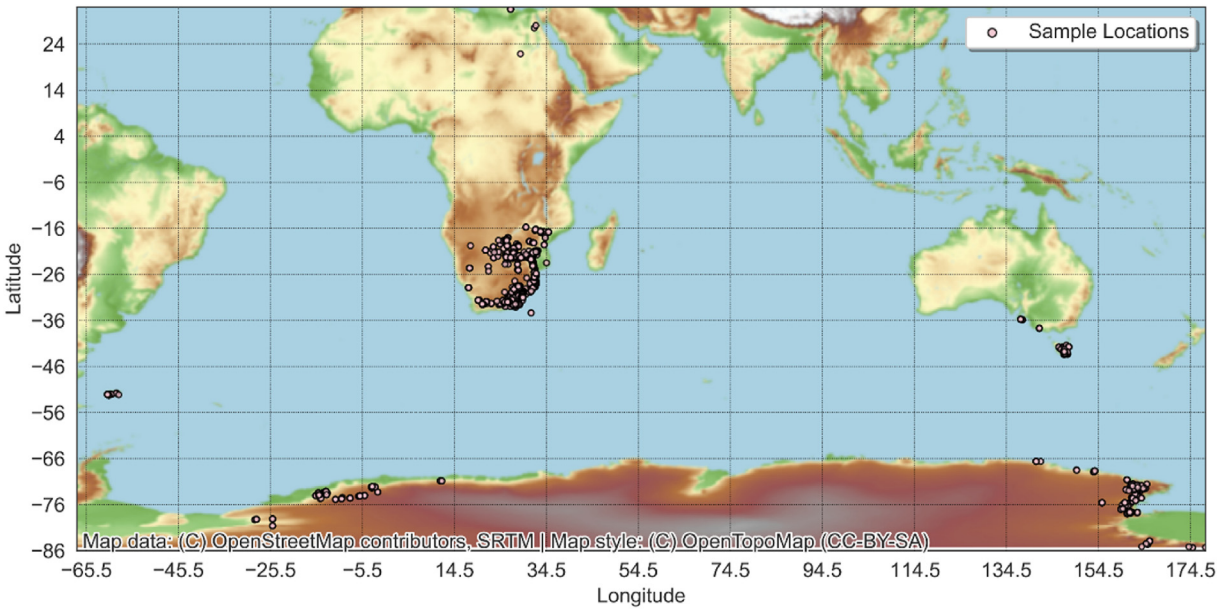

Fig. 1. Sampling sites of geochemical data of the Karoo Supergroup overlain on a shaded relief map. The total number of samples is 7917.

maximizes some notion of return – e.g. the performance.

3. Geological setting

The Karoo Supergroup rocks (300–180 Ma) are found in several basins located in southern South America, the Falkland Islands, southern Africa and Antarctica, which are altogether used to define the southwestern part of Gondwana during the late Palaeozoic (Fig. 1; Smith et al., 1993). The main Karoo Basin of South Africa is the largest and contains a well-preserved and almost continuous record of continental sedimentation, spanning 100 Ma (Smith et al., 1993; Catuneanu et al., 2005). Given the extent and state of preservation of the main Karoo Basin, it is used as a representative for the Karoo Supergroup stratigraphic succession.

The stratigraphic succession of the Karoo is divided into five groups: (a) Dwyka (ca. 300–290 Ma), (b) Ecca (290–255 Ma), (c) Beaufort (255–237 Ma), (d) Stormberg (230–183 Ma), and (e) Drakensberg (183–180 Ma) (SACS, 1980). The creation of the main Karoo Basin began with the rejuvenation of the Cape trough, causing the uplift and erosion of the Cape Supergroup to the south and the creation of a foreland basin to the north (Smith et al., 1993). The Dwyka Group (800 m thick) is composed of diamictites and associated fluvioglacial sediments, recording a time when Gondwana drifted over the southern pole, resulting in glaciation (Visser, 1991). The Ecca Group (<3000 m thick) is characterized by mudstones, siltstones, sandstones and minor conglomerates, reflecting the creation of a shallow sea after glaciation, followed by the gradual pro-gradation of deltas into that sea (Smith et al., 1993; Johnson et al., 1996). The Beaufort Group (<7000 m thick) is a fluvially derived succession of alternating mudstones, siltstones and sandstones caused by the coalescence of the pro-gradation deltas into broad alluvial plains (Johnson et al., 1997; Rubidge et al., 2000). Afterwards, upward-fining siltstones and sandstones of the Stormberg Group (1200 m thick) reflect the progressive aridification of the basin, leading to an aeolian sand-dune landscape (Johnson, 1994). Finally, sedimentation was replaced with wide-spread volcanism with the extrusion of the Drakensberg food basalts and abundant intrusive sills and dykes (up to 6600 m thick), which altogether define the Karoo large igneous province (SACS, 1980; Smith et al., 1993; Catuneanu et al., 2005; Svensen et al., 2012). The origin of the magma is related to the ascent of a deep mantle plume associated with the break-up of Gondwana (Storey, 1995; Storey and Kyle, 1997; Buiter and Torsvik, 2014). Textures vary from aphyric to porphyritic to coarsely crystalline and gabbroic in appearance, with coarser grain sizes found at the centre of some of the thicker sills. Some of the sheet intrusions are layered, caused by magmatic differentiation (Smith et al., 1993). Within southern Africa, the Drakensberg flood basalts are divided into four provinces, namely (a) tholeiitic lavas (main Karoo and Aranos basins), (b) olivine-poor basalts and rhyolitic to dacitic lavas (Lebombo Basin), (c) olivine-rich basalts (Lebombo and Zimbabwean basins), and (d) silica-rich basalts (Huab Basin) (Duncan et al., 1984). The Drakensberg flood basalts initially covered most of southern Africa by the Late Jurassic, but today are only preserved in association with the Karoo basins (Du Toit, 1954).

4. Data and methods

4.1. Karoo legacy geochemical data

Igneous rocks of the Karoo in the dataset feature 6650 volcanic and 1266 plutonic rock samples (Fig. 1 and Supplementary Data S1). The database was compiled over many years from other primary publications and databases and was first used in Ashwal et al. (2021). The samples in the database were analysed for major, minor and trace elements. The volcanic and plutonic rocks were sampled from drill cores and outcrops for a range of purposes that did not include large-scale mapping. Literature survey from the primary data sources cited by Ashwal et al. (2021) shows that samples were variably prepared and analysed. Although data in this condition would be unsuitable for traditional geochemical mapping using multivariate analysis, it is ideal for machine learning, especially in a method development setting, as algorithms that are trained on variably sourced data, where sufficient data exist, produce models that are more generalizable to new settings (Gong et al., 2019; Therrien and Doyle, 2018; Hyontai, 2018). A detailed description of the sampling and analytical methods is impossible for all sources. Instead, we summarise the general approach used for many of the primary data sources in Ashwal et al. (2021). Samples were washed with distilled water, dried and examined visually. Samples that were free of veins, signs of weathering and/or mineralisation were crushed into coarse fragments (3–10 mm). Fragments were then handpicked to remove surfaces that were contaminated by sampling tools (i.e. the original drill core surface). The remaining fragments were then washed with distilled deionised water (several times to remove attached dust), dried and powdered in an agate mill. Between each milling run, coarse quartz sand was processed to

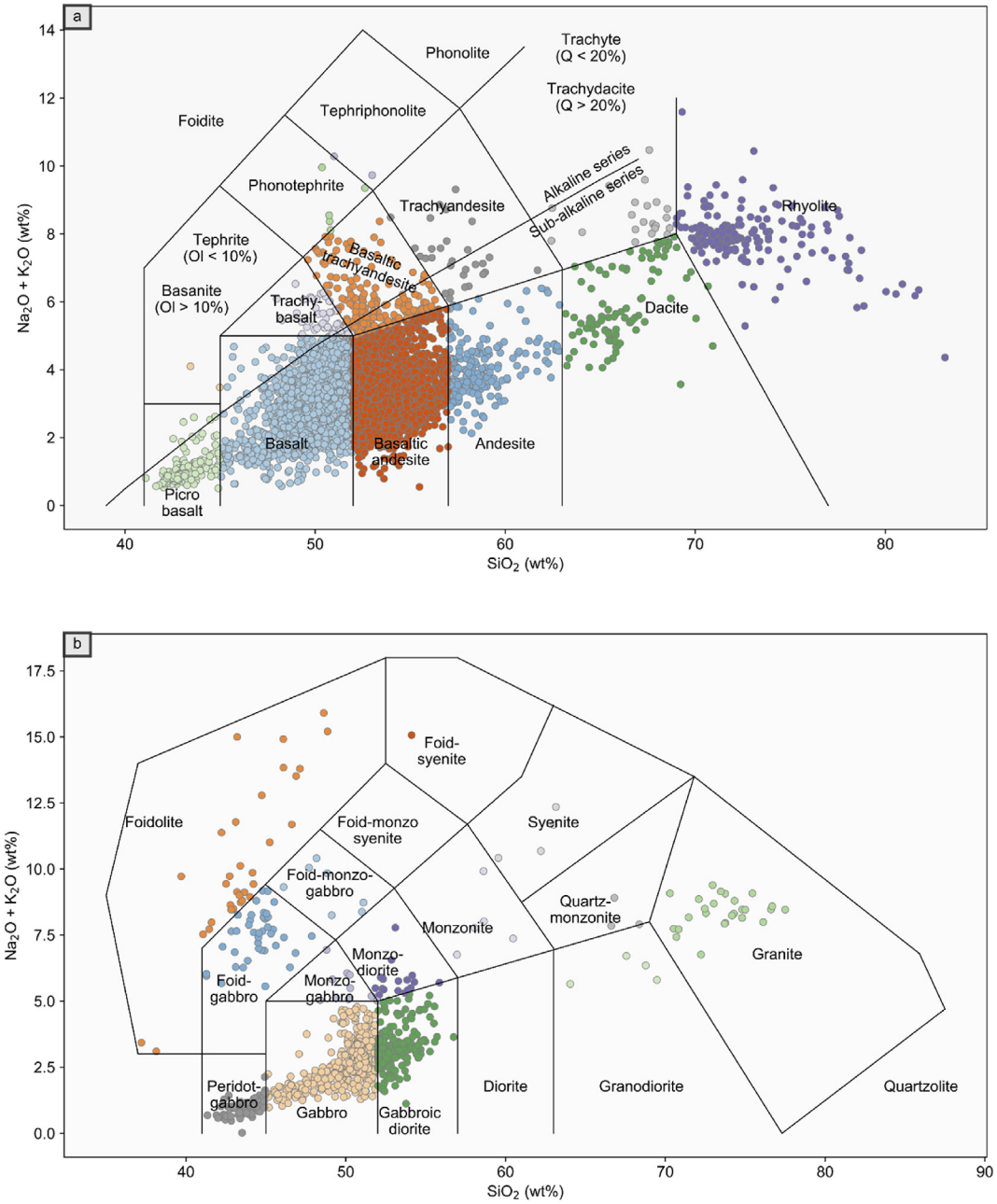

Fig. 2. Total alkalies vs. silica (TAS) diagram of the Karoo (a) volcanic and (b) plutonic units showing a compilation of whole-rock major and minor element analyses of samples taken from South Africa, Lesotho, Eswatini (formerly Swaziland), Zimbabwe, Botswana, Mozambique, Namibia, Antarctica, Australia, Tasmania and the Falkland Islands. Alkaline-subalkaline series divisions are from Irvine and Baragar (1971).

minimize cross-contamination between samples. Major and minor element concentrations were determined by X-ray fluorescence (XRF) spectrometry on fusion glass disks. Loss on ignition was determined via thermogravimetry. The analytical error for most of the elements is <0.10% (relative), except for MgO (0.50%). Trace element concentration was determined using either XRF, inductively coupled plasma-mass spectrometry (ICP-MS) or inductively coupled plasma-optical emission spectrometry (ICP-OES) depending on the available instruments during the time of the analysis, e.g. in the 1980s, trace elements were analysed using XRF. Various certified reference materials were used during the major, minor and trace element analyses. For the dataset, only the major and minor elements were normalised to unity. Some trace elements feature a substantial portion of the data near the instrumental lower detection limits, and as such, their concentration data is heavily quantised (e.g. low U, Th and radiogenic Pb concentrations in mafic rocks).

4.2. Rock classification and lithogeochemistry

For this study, the major, minor and trace element analyses must cover a range of rock types. As the data source is secondary, the type of rocks is not documented consistently across the primary data sources. To ensure that the rock types are consistent, we standardise the classification of rock types in the database using the TAS (total alkali vs. silica) classification scheme for volcanic rocks, with coordinates as given in Le Bas et al. (1986) (Fig. 2a), and similarly, we classify plutonic rocks using coordinates and labels as given in Middlemost (1994) (Fig. 2b). During rock classification, highly altered rocks were removed from the database by comparing their composition with the bulk rock compositions within the database using the interquartile range. Both the Karoo volcanic and plutonic suites cover a wide compositional range from ultramafic to felsic.

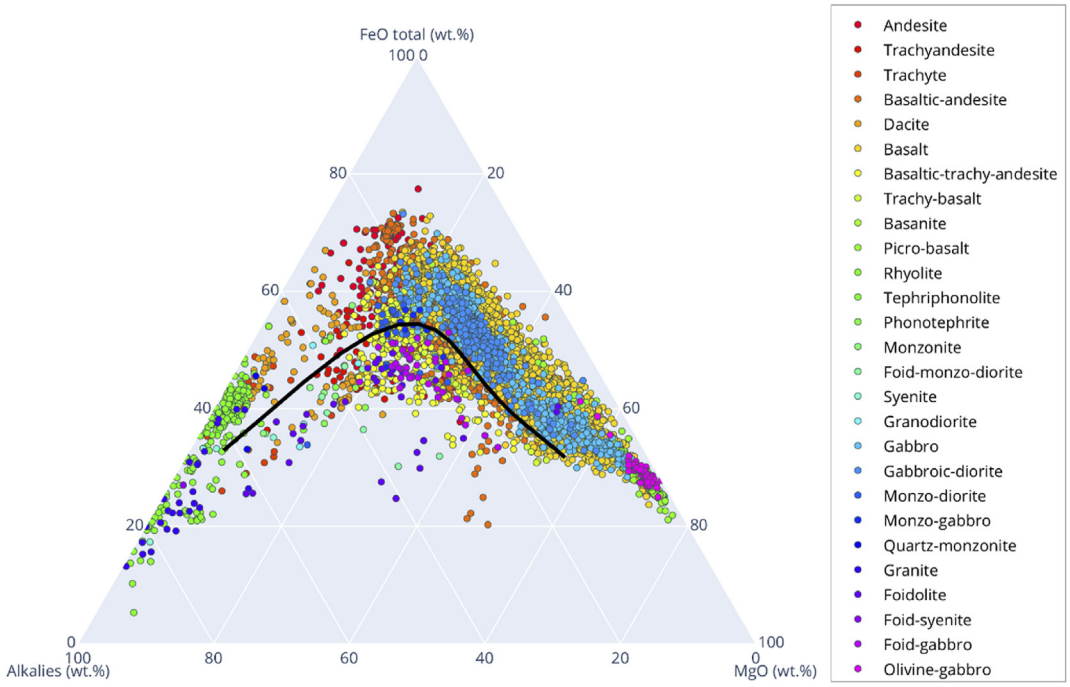

Fig. 3. An AFM ternary diagram showing the relative proportions of the oxides. The black line represents a division between tholeiitic (above black curve) and calc-alkaline (below black curve) series for various extents of magma evolution. The magma series division (black curve) is from Irvine and Baragar (1971).

Binary and ternary discrimination diagrams that use only a few elements are useful for the geochemical discrimination of magmatism (Wood, 1980; Agrawal et al., 2008; Vermeesch, 2013). Rocks in the tholeiitic magma series are distinguished from those in the calc-alkaline magma series by the redox state of the parent magmas (tholeiitic magmas are reduced; calc-alkaline magmas are oxidised) (Sisson and Grove, 1993; Chin et al., 2018). The difference between these two magma series can be seen on a $Na_2O + K_2O$ (A for alkalis), $FeO + Fe_2O_3$ (F for total Fe content), and MgO (M) ternary, known as the AFM diagram (Fig. 3). The dataset contains both tholeiitic (above the black curve) and calc-alkaline (below the black curve) rocks. This diversity of rocks is important in understanding the generalisability of the predictive models and ideally, the predictive modelling should incorporate data belonging to both trends (Gong et al., 2019).

4.3. General data pre-processing

Systematic differences between analyses from different laboratories for all analysed elements are impossible to determine with the available data, as original analytical and field duplicates are unavailable and sampling methods were varied. However, this is an intended challenge in our application, since a key objective of this study is to understand the ability of some common machine learning algorithms to tolerate low-quality legacy data that is highly variable and therefore infer the feasibilty of our method to other similar types of data (e.g. Karpatne et al., 2018). For machine learning algorithms, the best model performance is obtained by using training data that matches the variability of the testing data and in some applications, noise is purposefully injected into clean data to improve model performance (Goodfellow et al., 2016). Predictive modelling performance obtained using high quality data cannot be generalised to applications using legacy data. In addition, by using a low quality dataset, it would be possible to interpret the models' performance as a conservative estimate of the capability of our approach. For elemental concentrations that are significantly higher than the instrumental detection limits, data levelling discrepancies are likely to be small relative to the measurements themselves. As such, and for predictive modelling, the accuracy of most preparation and analytical methods are typically more than sufficient to compare major and minor (rock-forming) elements between laboratories. Therefore, we consider data levelling discrepancies unimportant for major and minor elements. This may be true for high concentrations of trace elements. Data levelling discrepancies that are differential across the elements analysed (e.g. an enrichment of some trace elements due to contamination or matrix effects) cannot generally be removed by using ratios of compositions (e.g. a log-ratio transformation). However, several elements that are close to the detection limit (e.g. U and Th) may be more affected by differences in data levelling, especially due to discrepancies in instrumental lower detection limits and preparation methods. In general, we have no evidence that our dataset is uniformly levelled, which is ideal for creating a methodology that would generalise well to similar types of datasets.

All oxides have been recalculated on an anhydrous basis. Missing data is sparse in the major and minor element data. Our approach requires complete major and minor elements (training features) but not complete trace elements (prediction targets). Therefore, missing compositions within the major and minor elements ($n = 32$; 0.4% of the whole dataset) are imputed using the k-nearest neighbors algorithm (Hron et al., 2010). Given the small fraction of data points requiring imputation and the nature of major and minor elements (in that they are rock-forming elements), the resulting imputation were satisfactory as the largest deviation in closure after imputation was less than 2%. The check for closure after imputation is not to enforce closure, but to ensure that the results are realistic. This should not affect our workflow, as we do not use all the elements together in a multivariate sense (e.g. through principal component analysis).

4.4. Feature engineering, machine learning algorithms and workflow

All rock-forming elements (major and minor elements) are used as machine learning features. Ratios of elements or any other mathematical manipulations of features as part of data-preprocessing are known as feature engineering, whose purpose is to measurably improve the performance of the algorithms (Hastie et al., 2009; Domingos, 2012). Feature space is a vector space (Hastie et al., 2009) and in the case of log-ratio-transformed compositional data, the feature space is unaffected

by properties of data closure (e.g. the range restriction is removed). However, in general, machine learning algorithms do not assume specific geometries of the feature space, nor feature variable distributions (e.g. normality). The choice of the geometry of the embedding vector space (and its associated properties, such as linear transformations and metrics) is made based on the structure of the data and verified through algorithm performance profiling. Embedding data into a geometry that is more suitable for the data's structure often, but not always results in better algorithmic behaviour, and a variety of geometries are available, such as the Euclidean, hyperbolic and spherical (Gu et al., 2018).

4.4.1. CLR-transform vs untransformed (raw) data

For compositional data, the Aitchinson geometry (Aitchison, 1982) is the native vector space geometry. The use of compositional data usually occurs outside of its native vector space through a choice of log-ratio transformations on the data, such that the resulting vector space is Euclidean. Outside of machine learning, this transformation traditionally facilitates the use of multivariate analysis techniques, much of which was originally built for the Euclidean geometry (e.g. Aitchison, 1982; Grunsky and de Caritat, 2019). For machine learning algorithms, if the algorithms do not explicitly require properties of any specific geometry (e.g. linear transformations), then leaving the data in its native vector space is unproblematic. To use a consistent set of training data for machine learning algorithms, we chose to transform the feature portion of the data (i.e. major and minor elements) using the CLR transformation, which removes the range restriction of the Aitchison geometry (Aitchison, 1982). The CLR transformation has been used with success when followed by traditional multivariate geochemical data analysis routines. In the case of Grunsky and de Caritat (2019) and other similar studies (e.g. Grunsky, 2013; Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018), the process discovery routine contains an application of principal component analysis (PCA), which is an algorithm that also belongs to machine learning (depending on its application and purpose). For these reasons, we follow this established data-preprocessing convention (Grunsky, 2013; Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018; Grunsky and de Caritat, 2019) and also make use of the CLR transformation for a portion of this study. In the CLR transformation, all compositions $x_i \ldots x_D$ are divided by the geometric mean of the vector, i.e. $g(x) = \sqrt[D]{x_i \ldots x_D}$ and subsequently, a logarithm is taken of the ratios, i.e. for sample x_j, $CLR(x_j) = \left[ln\frac{x_{ij}}{g(x_j)}; \ldots; ln\frac{x_{Dj}}{g(x_j)}\right]$. In the machine learning context, whether this type of data preprocessing (or any other form of feature engineering) is warranted depends on its impacts on the predictive modelling performance (Karpatne et al., 2018). In this study, we asses the performance of machine learning algorithms using two parallel workflows, one that employs the CLR transformation to engineer features suitable for Euclidean-geometry aware algorithms, and one without it (i.e. using raw data).

4.4.2. Machine learning algorithms and workflow

We employ a range of algorithms, some of which are geometry aware and employ Euclidean metrics, while others are completely unaware of the feature space geometry. The prediction targets (i.e. trace elements to be predicted) are not transformed, as they are not part of the feature space. If the features encode characteristics of rocks, which also differ by rock type, then the machine can identify these differences and their relationships that can be used for predictive modelling. Whether the choice of the embedding vector space (and therefore the log-ratio transformation) is appropriate should be evaluated empirically using performance metrics (Gu et al., 2018). In this study, we adopt two performance metrics, the coefficient of determination (CoD or R^2) to assess model fitting, and the median absolute prediction error (MAPE) to assess typical prediction performance. MAPE is more robust to outliers compared to the CoD metric.

Two major types of machine-learning algorithms are suitable for predictive modelling using geochemical data: (a) supervised and (b) unsupervised learning. Semi-supervised machine learning as a hybrid of (a) and (b) is also possible. In unsupervised learning, the class labels (e.g. a concentration or rock type) are unknown and the machine attempts to deduce categorisations within the data to create a classification scheme, which can then be used to classify new data (Hastie et al., 2009). In supervised learning, the data is labelled and the algorithms automatically deduce relationships between features and labels (Russell and Norvig, 2010). The algorithm's hyperparameters are tuned through cross-validation (Hastie et al., 2009). The resulting models are then used to inductively predict either continuous (e.g. the concentration of an element) or discrete (e.g. types of rocks) labels. The prediction results can be assessed for accuracy through yet another dataset that does not overlap with the training and cross-validation datasets. For our purpose, supervised machine learning algorithms that are capable of regression are appropriate.

There are many supervised regression algorithms. Algorithm selection is based on many factors, such as (a) computation time, (b) data density including feature space density, (c) bias-variance trade-off, (d) function complexity, (e) feature space dimensionality, (f) input and prediction noise, and (g) feature interactions. An experimental approach is often used to maximise some combination of factors. Prediction error can be interpreted through the bias, variance and noise model. The bias is the tendency of the model to default to some class label. The variance gauges the relative change in the model output given a change in the model's input. The noise is the portion of prediction error that is neither bias nor variance. The total prediction error is the quadrature sum of the three sources of errors. Algorithms generally exhibit different behaviours in terms of their bias and variance, and it might be desirable to trade some bias for a greater reduction in variance in a particular context.

The k-nearest neighbors algorithm (KNN; Cover and Hart, 1967; Fix and Hodges, 1951) for regression is a non-parametric method that uses k neighboring training samples in feature space (a hyperparameter of the model) to construct a consensus (by averaging) that is used to estimate unknown targets that fall within the neighborhood (Kotsiantis et al., 2007; Witten and Frank, 2005). The KNN algorithm utilises a sequence of steps: (a) evaluate the feature space distance between a target and each training sample (e.g. Euclidean distance $= \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$, although other metrics are also applicable); (b) the closest k data points are selected and; (c) the average of these data points is the prediction for the target. Large values of k may lead to model overfitting (Hastie et al., 2009), which increases the prediction variance.

Elastic-Net is a regularised regression method that linearly combines the L_1 and L_2 penalties of the lasso (least absolute shrinkage and selection operator; Santosa and William, 1986; Tibshirani, 1996) and ridge (Tikhonov regularisation; Tikhonov, 1943) algorithms, respectively (Zou and Hastie, 2005). Elastic-Net's objective function is:

$$min_\omega {}^1/_{2N}\left\{X\omega - Y_2^2 + \alpha\rho\omega_1 + \alpha(1-\rho)/2\ \alpha\omega_2^2\right\},$$

where X is the input features, Y is the regression output, ω is the coefficients, N is the number of samples, α is a regularisation parameter and ρ is the mixture ratio of the ridge and lasso contributions. Elastic-Net integrates ridge regression's ability to uniquely determine useful features while retaining lasso regression's ability to completely remove useless features. The use of the L_2 metric both in a regularisation term and in measuring the distance of the data points to the model implies that the native feature space geometry is Euclidean.

The support vector machines (SVM; Vapnik, 1998) algorithm is similar to other Euclidean-metric regression algorithms and is typically used to define nonlinear decision boundaries or regression models in high-dimensional feature space (Hsu and Lin, 2002; Karatzoglou et al., 2006). SVM maximizes the number of training samples (support vectors)

that are closest to the regression hyperplane within a boundary region defined by ε. The objective function measures the L_2 norm of the model coefficients and the margin, which is the sum of the Euclidean distances between the hyperplane and data points outside of the boundary, multiplied by a parameter C. Increasing C promotes an increasingly more complex model. As SVM only uses support vectors, it can automatically ignore some outliers. The kernel of SVMs is another algorithm hyperparameter and many parametric functions can be used, including the highly flexible radial basis functions (RBFs).

Decision trees are flowchart-like hierarchical structures that partition the trees recursively. Internal nodes represent features, branches represent decision rules, and each leaf represents an outcome. This algorithm learns to partition the data based on feature values. The flowchart-like structure is easy to interpret and visualise. The decision rule to split a node is based on a metric to maximise some notion of the difference between the resulting leaves. For regression, the mean squared error metric can be used to measure the goodness of fit of the leaf to its surrounding samples. The depth of the tree is a hyperparameter. Decision trees are weak regressors in the sense that they regress above chance, but not substantially. It is possible to convert a weak regressor into a strong one by using various statistical approaches, such as ensemble methods (Ho, 1995; Breiman, 1996a, 1996b; Kotsiantis, 2014; Freund and Schapire, 1995; Sagi and Rokach, 2018), if each weak regressor is better than a random guess. Random forest is a type of bagged decision tree that mitigates the noise sensitivity of individual trees by constructing an ensemble of trees and averaging the output, as long as the trees are not correlated (Ho, 1995). The removal of correlation between individual trees is via sampling of random subsets of features (e.g. bootstrap sampling). This leads to better model performance than decision trees in general because the model variance is reduced but without introducing additional bias. The maximum amount of features per tree, the number of trees, and the minimum number of samples per split are model hyperparameters in addition to the tree depth parameter that is inherited from decision trees. AdaBoost can use decision trees as a base weak regressor and in this form, is a boosted decision tree that uses adaptive boosting (Freund and Schapire, 1995), which combines the output of weak regressors into a weighted sum that represents the final output. Adaptation occurs by adjusting weights of subsequent regressors according to the error of the current prediction to focus on cases that are more difficult. The rate of adaptation and the number of trees are model hyperparameters. Tree-based methods do not assume feature space geometry.

Multilayer perceptron classifier (MLP) is a class of feedforward artificial neural network, which is a collection of connected nodes (artificial neurons) that loosely resemble biological brains (Hastie et al., 2009). Connections between the neurons transmit real numbers to other neurons and the output of each neuron is a nonlinear function of the sum of its inputs (similar to the activation potential in biological neurons). The connections and the neuron outputs are typically weighted and the weights are adjusted through experience. Neurons activate according to some function, which may exhibit a threshold or maybe linear (Hastie et al., 2009). Neurons are usually connected layer-wise and each layer performs a different transformation on their inputs. It is possible for signals to recurrently travel the same network multiple times in other artificial neural network designs, although the feedforward designs are single-pass. Artificial neural networks are universal function approximators and are extremely useful algorithms in data-rich applications such as image classification and natural language processing. To date and in an increasing number of applications, artificial neural networks are capable of surpassing human performance in several tasks (e.g. He et al., 2015; Lundervold and Lundervold, 2019). An MLP contains a minimum of three layers of neurons: an input stage; a hidden layer; and an output layer, and because of its simplicity, it is a trivial example of artificial neural networks. Input nodes are linearly activated, while the subsequent layers are nonlinear. The supervised learning technique uses an objective function and backpropagation for model training. The objective function is any metric that evaluates the desirability of the output (e.g. mean squared error for regression). Backpropagation computes the gradient of the objective function concerning the weights of the network for each training example by using the chain rule, iterating over each layer at a time. It allows the weights to be updated following a gradient descent approach to minimize the objective function (Curry, 1944; Lemaréchal, 2012; Rosenblatt, 1961; Rumelhart et al., 1986). MLPs are capable of distinguishing data that are not linearly separable (Cybenko, 1989). There are several hyperparameters including the following: activation, which is the type of mathematical function used to activate the hidden and final layers and could include the identity ($f(x) = x$), the logistic sigmoid function ($f(x) = 1/(1 + e^{-x})$), the hyperbolic tangent function ($f(x) = tanh(x)$) and the rectified linear unit function (relu, $f(x) = max(0, x)$); the L_2-norm-based regularisation parameter α, which can be tuned to balance the model bias and variance; and the learning rate parameter, which can be constant, decreased over each time step using a power function (invscaling), and adaptive, which keeps the learning rate at the initial constant rate until the loss function ceases to decrease, at which point, the learning rate is decreased five-fold. MLP does not assume any feature space geometry.

Table 1
Model parameters used in the grid search.

Algorithm	Parameter Grid
KNN	k = {1, 2, 4, 6}
SVM	C = {10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 750, 1000}, ε = {0.00001, 0.0001, 0.001, 0.01, 0.1, 0.5, 1.0}, kernel = {linear, RBF}
Elastic-Net	L_1 ratio = {.1, .5, .7, .9, .95, .99, 1}
Random Forest	Ensemble size = 500; maximum depth = {5, 4, 3, 2, 1, unlimited}; maximum number of features is = {1, 2, 3, 4, 5}; minimum number of samples for a split = {2, 3, 4}; minimum number of samples for a leaf = {1, 2, 3}
AdaBoost	Learning rate = 1, number of classifiers = 100, base classifier = decision tree with maximum depth = {0, 2, 4, 6}
MLP	α = {0.001, 0.01, 0.1, 1.0}, activation = {identity, logistic, tanh, relu}, learning rate = {constant, inverse scaling, adaptive}

Model selection and tuning for supervised machine-learning algorithms are usually accomplished through cross-validation, which is an out-of-sample testing technique. In cross-validation, the dataset is split into several non-overlapping sets, the larger of which is the training dataset that is used to train the models. Then the remainder validation dataset is used to profile the prediction performance of the models and the model hyperparameters are adjusted. Subsequently, the models are re-trained and re-validated to optimize the hyperparameters. Issues such as excessive model variance and selection bias are minimised through this process. In most machine learning applications, either 5- or 10-fold cross-validation is recommended. The 5-fold cross-validation is preferred as it minimised computation time, especially where prediction accuracy is within acceptable level depending on the objectives of the study or application (Zhang et al., 1999; An et al., 2007). We use a grid search (Table 1) combined with 5-fold cross-validation to determine the best algorithms.

5. Results

5.1. Prediction of trace element concentrations

A total of 32 trace elements (Sc, V, Cr, Co, Ni, Cu, Zn, Rb, Sr, Y, Zr, Nb, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, Pb, Th and U) were predicted using a variety of machine learning algorithms (Fig. 4). The entire dataset is divided into a training and testing dataset, whereby the training dataset is further divided into cross-validation and training subsets using 5-fold cross-validation. The best algorithm for each element is determined using the CoD. The best algorithms are then used for performance assessment. The final model testing is performed on the testing dataset to obtain a stable result. The final testing uses 5-fold cross-validation that is repeated 25 times (125 total results per element) with

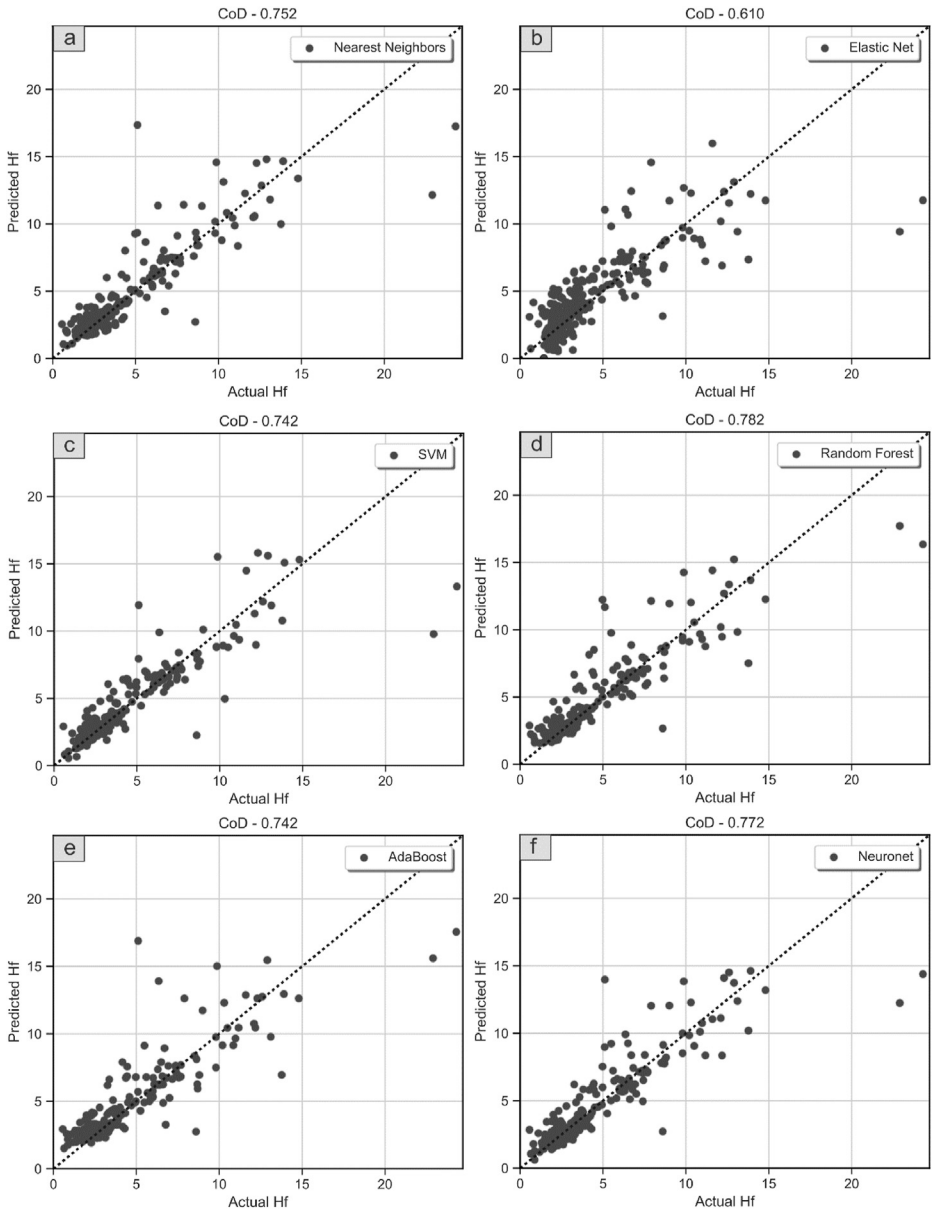

Fig. 4. Regression scatter figures of a single cross-validation run of Hf using (a) KNN, (b) Elastic-Net, (c) SVM, (d) random forest, (e) AdaBoost, and (f) neural network regressors.

randomly sampled subsets of data for training and testing, and the results for each iteration CoD and MAPE are averaged; subsequently, the results from the 25 runs are further averaged to produce a single measure of CoD and MAPE per element. The MAPE metric is difficult to interpret as the MAPE generally scales with the mean elemental concentration. Therefore, to compare between different elements, the MAPE metric is converted to a dimensionless quantity by dividing it by the mean elemental concentration. The results for algorithm selection and final testing (Fig. 5) shows that the best algorithm generally varies depending on the element being predicted. AdaBoost and random forest are two of the best performing algorithms across all elements. In the past, for a similar task that consisted of the prediction of Au using geochemical data, Rodriguez-Galiano et al. (2014) showed that the random forest algorithm is very powerful for mineral potential modelling and outperforms logistic regression when used in classification tasks. A benefit of the tree-based approaches (such as random forest and AdaBoost) is that they can incorporate non-numerical and categorical data as features and therefore predictive methods based on them can be easily extended to include descriptive geology, mineralogy and a range of other types of evidence. Indeed, Rodriguez-Galiano et al. (2014) used multi-sourced datasets (gravimetric, magnetic, lithogeochemistry, lithology and satellite data) and did not consider a large range of algorithms (only random forest and logistic regression were evaluated). In any case, for this study and the dataset used, we are unable to replicate that random forest is the best algorithm all around. Instead, other methods such as AdaBoost, k-nearest neighbors and SVM are highly competitive and are often more accurate than the random forest on a per-element basis. The most accurate algorithm by the number of elements that it predicts the most accurately is AdaBoost. It is possible that robust neural networks beyond MLPs can outperform tree-based methods for prediction of elemental concentrations using compositional geochemical data, but this conjecture remains to be tested (Chen et al., 2020). The testing CoDs of elements such as Pb, Cu, U, Th, Sc and Zn are lower than 0.5. However, for Th and Zn, the CoDs are heavily affected by outliers, which results in low CoDs with low ratios of MAPEs to elemental mean concentrations (Fig. 5). In the case of Pb, U and Th, it is possible to observe data quantisation due to data being close to lower detection limits. This negatively affects the quality of the training data and therefore the performance of all algorithms. Adding

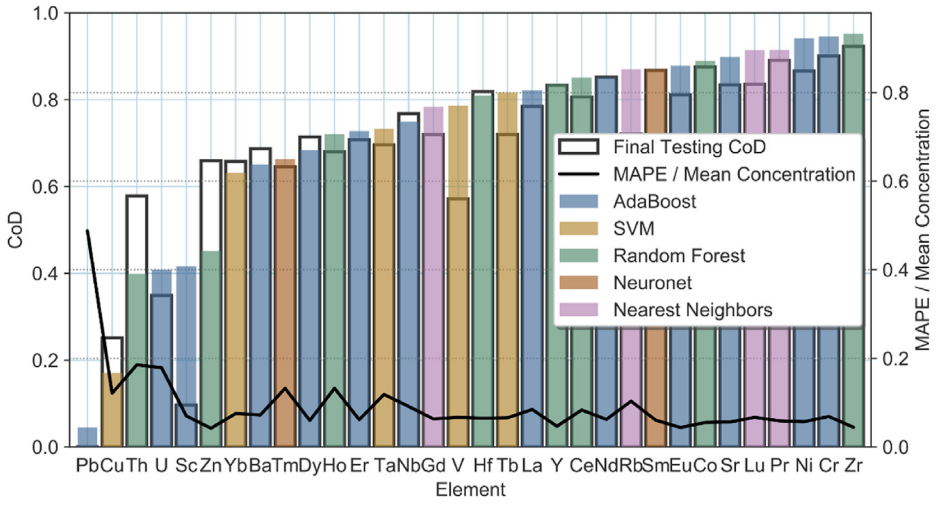

Fig. 5. Results of the algorithm selection and performance assessment, as measured using the coefficient of determination (CoD), of all trace elements. The best machine learning algorithms are shown for each element.

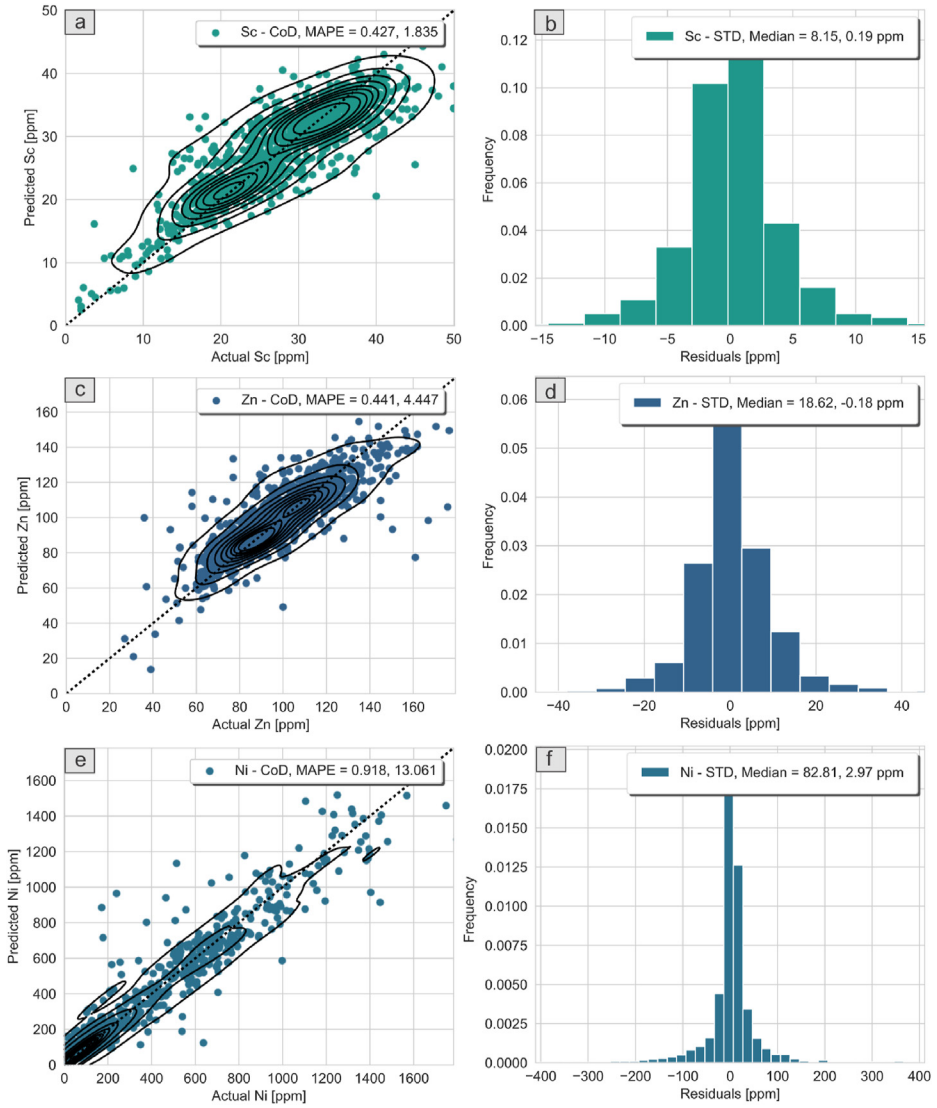

Fig. 6. Final results of prediction performance for select elements and their prediction residuals. (a) (c), and (e) shows that the predicted values are very close to the actual values. (b), (d), and (f) illustrates the distribution of residuals. The corresponding coefficient of determination (CoD) and median absolute error (MAPE) are shown in the scatter plots. For the histograms, distribution median and standard deviation (STD) are shown.

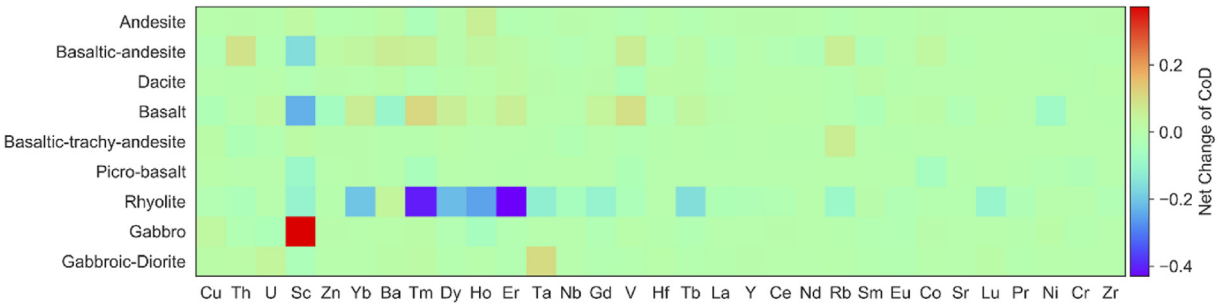

Fig. 7. Changes in prediction performance relative to the all rock-types baseline (as measured by the coefficient of determination (CoD)) with the removal of each type of rock. Pb is removed due to its substantial change with rock-type removal, which obscures the trend for all other elements.

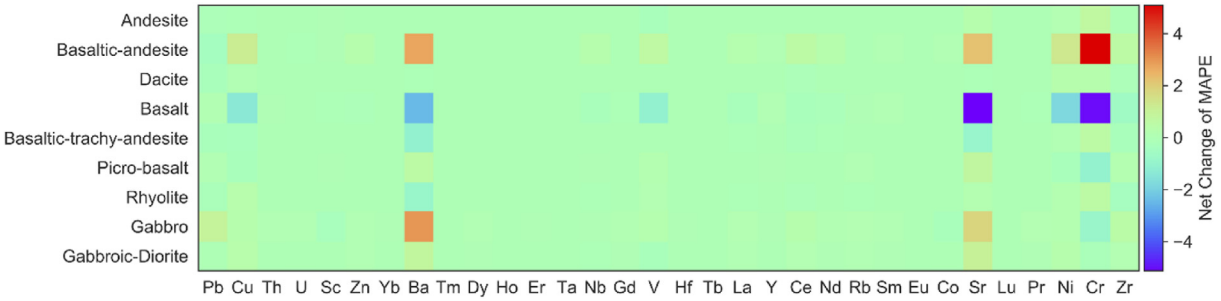

Fig. 8. Changes in prediction performance relative to the all rock-types baseline (as measured by the median absolute error (MAPE)) with the removal of each type of rock.

some noise to quantised data may be effective to increase the prediction performance. However, this remains to be tested.

Prediction performance as assessed through the CoD alone does not reveal the quality of prediction residuals. Ideal prediction residuals should feature a minimum standard deviation, and where there are no anomalies expected, the distribution should be symmetric. As prediction residuals approach this ideal, the prediction results are more useful for regional mapping or large-scaled block modelling, as the baseline for geochemical anomalies is increased. The histograms of prediction residuals most often show the asymmetric distribution and are (approximately) independently distributed with a relatively uniform variance or standard deviation (Fig. 6). Deviation from the normal distribution represents geochemical variation that is unexplained by the model. The scatter clouds for each predicted element do not exhibit strong bias (e.g. no significant rotation relative to the perfect prediction line). The standard deviation generally varies on a per-element basis (e.g. Zn and Sc prediction residuals are similar, but the concentration of Zn is on average substantially higher than Sc; Fig. 6).

5.2. *Influence of rock type on prediction performance*

The influence of rock type on prediction performance is important to understand whether particular rock types (e.g. volcanic rocks) are detrimental to the overall prediction performance. This information provides a basis to determine the generalisability of the method proposed in this study to new environments. A combination of MAPE and CoD were used to monitor prediction performance changes relative to the removal of various rocks in the studied magmatic suites, again using 5-fold cross-

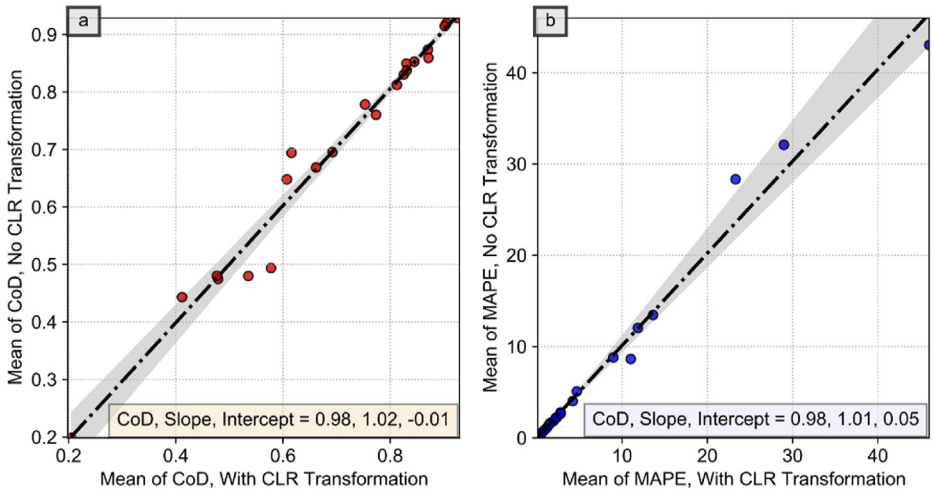

Fig. 9. (a) Mean of the coefficient of determination (CoD) compared between training data that employed the CLR transformation and without CLR transformation. (b) Mean of the median absolute error (MAPE) compared between training data that employed CLR transformation and without CLR transformation.

validation and with 50 randomly generated train-test splits of the dataset. The changes in MAPE and CoD were measured by using the net change of these performance metrics against a baseline using all rock types (i.e. CoD per rock type removed – CoD of all rocks). A positive change indicates an improvement with the removal of a rock type and vice versa (Figs. 8 and 9). Elements such as Sc show improved prediction performance when gabbro is removed from the magmatic suites, while Tm shows a decline in prediction when rhyolite is removed. Some elements show similar compatibility in various rocks based on their geochemical classification, that is, the chalcophile (e.g. Cu and Zn) and lithophile elements in mafic rocks (e.g. Ba). The lithophile elements include: Ba, Hf, Nb, Rb, Sc, Sr, Th, U, V, Y, Zr, and the lanthanides or rare earth elements. The highest prediction performance variability occurs in felsic rocks (e.g. rhyolite). The removal of rhyolite substantially decreases the ability to predict some elements, primarily in the lanthanides (Yb, Tm, Dy, Ho, Er, Gd, Tb and Lu, see Fig. 7) and Sc. The MAPE is not as heavily affected as the CoD with changes in the rock types. For the MAPE, several elements (Cu, Ba, Sr and Cr) show a consistent change in performance depending on the rock types removed (Fig. 8).

5.3. Impact of CLR-transformation on predictive modelling

The use of log-ratio transformations is standard practice in multivariate compositional data analysis, to remove the data closure effect that introduces spurious negative correlations between components (Aitchison, 1982; Grunsky and de Caritat, 2019). In this case, multivariate analyses of compositional data (e.g. correlation matrices) are unreliable with untransformed data exhibiting effects of closure, because multivariate analysis techniques such as principal component analysis assume a Euclidean geometry for the feature space. However, this may not be the case for many machine learning algorithms both theoretically and empirically. For example, neural networks and tree-based methods do not assume Euclidean geometry, or the existence of geometry altogether, in general. Transformation of data for data pre-processing or feature engineering for these algorithms should strictly serve to increase predictive modelling performance objectively. However, SVM, KNN and Elastic-Net algorithms can use the Euclidean distance metric for fitting of data points (although other distance metrics are possible, e.g. Lin and Ye, 2020). The empirical effect of data closure is unclear for any of the machine learning algorithms in our application. In this study, we took an experimental approach to understand the effects of the CLR transformation on the performance of machine learning algorithms. We generated two sets of testing results, one with CLR transformation and one without using two parallel workflows (each with its own algorithm selection, model fitting and cross-validation, independent from the other), and subsequently compared them on a per-element basis (Fig. 9). It is clear that for the machine learning algorithms that were applied in this study, there is no statistically relevant difference between the results as measured using both the CoD and MAPE metrics. In addition, the mean and standard deviations of the averaged prediction residuals across 50 runs are essentially indistinguishable. Potential deleterious effects of closed feature space, such as negative predictions were not observed. However, for the application of the proposed technique for geochemical anomaly detection (e.g. for prospectivity mapping), even negative prediction results are suitable and appropriate, as we would only make use of prediction residuals (predicted minus actual), not the predicted quantities themselves, which are not useful. In the case of spatially unaware algorithms (e.g. most tree-based methods), there are no known benefits to using log-ratio transformations as a data pre-processing or feature engineering method. Instead, as log-ratio transformations are unable to handle missing data, using log-ratio transformations necessitate uses of imputation that may not be warranted in all contexts. For spatially aware algorithms, the validity of log-ratio transformations, such as the CLR transformation as a data pre-processor or feature engineering method requires a per-application investigation.

6. Discussion

6.1. Predictive modelling results

In general, the prediction performance of trace elements that are investigated in this study is satisfactory using machine learning across both the CoD and MAPE metrics. The prediction performance is very good (CoD>0.8) for many elements (Tb, La, Y, Ce, Nd, Rb, Sm, Eu, Co, Sr, Lu, Pr, Ni, Cr and Zn; Fig. 5). It is well known that trace elements play a role in fingerprinting igneous differentiation. Since major, minor and trace elements vary throughout the differentiation process, high-dimensional relationships may likely exist between major, minor and trace elements in igneous rocks. However, creating conventional prospective models requires substantial subject matter expertise in ore geology, petrography and exploration geochemistry. Since igneous systems are dynamically complex with intrinsic spatial and temporal variability, data-driven approaches are likely to be successful, since they intrinsically leverage high-dimensional elemental associations (Chen et al., 2020). As such, data-driven approaches are increasingly becoming a core staple of exploration (Grunsky, 2013; Grunsky et al., 2014; Chen et al., 2018; Grunsky and de Caritat, 2019; Chen et al., 2020; Grunsky and Arne, 2020; Lawley et al., 2021).

Using our proposed approach, most (27 out of 32) of the predicted elements exhibited CoDs higher than 0.6 (Fig. 5). Only Pb, Cu, Th, U and Sc exhibited poor prediction performance. In the case of Cu, Sc, and to some extent Ba, the prediction is poor as gauged through the CoD metric, because the CoDs are deflated through the presence of sufficient outliers. In the case of Th and U, much of the data within the dataset is heavily quantised, since their concentrations are very close to the lower detection limits of the analytical instruments. For Pb, there is a combination of sufficient outliers and some prediction bias. In general, Pb and U are highly mobile elements and are amenable to redistribution via secondary processes such as alteration and weathering. However, it is not possible to further substantiate the causes for these issues from a data perspective. To further understand the variability in prediction performance across trace elements, it would be necessary to investigate this dataset's accuracy and precision, spatial variability and elemental partitioning effects, such as the role of trace elements in mineral and glass chemistry and structures, which is out of the scope of this study. It is also clear that depending on the predictive modelling task using geochemical data, tree-based algorithms are powerful. However, the most optimal algorithm must be determined through trial and error in a data-driven manner, which is the essence of algorithm selection in machine learning. Algorithm selection is a common and essential task in data-driven predictive modelling and any geochemical data-driven predictive modelling workflow should incorporate algorithm selection to determine the most suitable algorithm.

In many geochemical studies, the data coverage is usually biased in terms of sampling density, analytical quality and rock type. This is due to multiple factors, such as differential exposure that limits sample availability and hypothesis-driven sampling strategies. Our dataset exhibits strong coverage bias towards basalt (n = 3039), basaltic-andesite (n = 2643), gabbro (n = 680) and gabbroic-diorite (n = 359). Therefore, when these types of rocks are removed from the prediction database, we expected that there might be an overall performance change across nearly all elements. However, this is not observed except for the loss of performance in the prediction of the lanthanides with the removal of rhyolite (Figs. 8 and 9). As rhyolite and basalt represent the most felsic and mafic rock types in the dataset (for volcanic rocks), they are known to exhibit a contrasting suite of trace elements. More compatible elements would tend to be found in basalt samples, while relatively more incompatible elements would tend to be more concentrated in felsic magmas. In this case, rhyolite samples are highly valuable for the prediction of lanthanides, while basalt seems to be generally useful. The behaviour of Sc as a function of the removal of either gabbro or basalt is contrasting and it is not clear why this occurs in a geological sense. The

prediction performance as measured by the MAPE exhibits a somewhat different trend than that of the CoD as a function of rock-types removed (Fig. 8). In this case, the MAPE metric is more indicative of changes in prediction performance in a manner that is more robust to outliers than the CoD metric. No single type of rock seems to significantly affect the MAPE across a range of elements. However, several elements such as Ba, Sr and Cr are more affected by the removal of rock types, particularly basaltic rocks. This may be due to the large proportion of mafic rocks in the database and the existence of significant relationships between Ba, Sr and Cr, and major and minor elements inside mafic rocks. Such relationships are lost with the subtraction of mafic rock types. The elements that were predicted with the lowest CoD typically have a low concentration (e.g. Th, U and Pb), which suggests that instrumental detection limits play a major role in the prediction performance at low concentration levels. In addition, the mobility of Pb may also play a role in its poor prediction performance. Despite the many sources of data resulting from multiple sampling and analytical methods that were followed over several decades, the legacy data is useful for predictive modelling and method development, in addition to its primary value for understanding igneous systems. The variability introduced by this type of data does not appear to hinder the machine learning algorithms in their ability to extract high-dimensional patterns (e.g. Fig. 6). This is significant as previous methodologies rely on high-quality data that are fully documented and vetted for mapping and modelling (e.g. Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018; Grunsky and de Caritat, 2019). Machine learning algorithms are different from traditional approaches in this regard, that the variability of the training data may be an asset, as the subsequent predictive models are more generalizable (Therrien and Doyle, 2018; Hyontai, 2018; Chen et al., 2020). A high performance (e.g. accuracy) alone is not indicative of a predictive model or even an algorithm's performance in a different setting.

6.2. Implications for regional resource exploration

There are many knowledge- and data-driven approaches to creating prospectivity maps, all of which invariably try to contextualise elemental distributions as either part of normal regional variability or something that requires further investigation, such as a mineral deposit. One established method of generating prospectivity maps using compositional data and multivariate analysis relies on three key sequential procedures: (a) a PCA-based method to extract linear combinations of elements that co-vary; (b) a regression-based analysis of residuals of elements of interest (e.g. an indicator or deposit-vectoring element) against the dominant principal components to determine residuals; and (c) creating a map of regression residuals (Grunsky, 2013; Grunsky et al., 2014; Grunsky and de Caritat, 2019). The fundamental logic for this approach and other similar approaches relies on the idea that main variations of stoichiometry are related to lithological variability and secondary processes, and if these processes do not adequately explain the occurrence of a particular element of interest, then it is possible that other processes that are under-sampled (except noise) may be responsible for potentially interesting targets for further exploration (Grunsky, 2013; Harris et al., 2015; Chen et al., 2016; Grunsky and de Caritat, 2019). If high residuals of the regression of these elements against the principal components form spatially coherent patterns, then the patterns are more likely to be associated with a physical process instead of sampling noise. In this approach, PCA serves several overlapping roles that include dimensionality reduction of the high-dimensional geochemical feature space (into key linear combinations of features) and highlighting key stoichiometric variability in the data (that are interpreted to be associated with sample mineralogy changes). However, in this approach, the quality of the prospectivity map is entirely limited by the amount of geochemical variability that the selected principal components can explain. This is because the extent to which regression residuals represent anomalies highly depends on the explanatory power of the principal components, which provides a regional geochemical baseline model. In this sense, increases in the quality of the models (higher explanatory power models, e.g. a higher CoD) increases the baseline of geochemical anomalies. In effect, a higher quality (more predictive) model renders the geochemical anomalies more selective and less noisy. Therefore, to be able to predict any element of interest as accurately as possible is the main consideration to increasing the targeting selectivity of prospectivity maps. A higher-quality model is capable of filtering out some of the regional variability of an element of interest that are not associated with potentially interesting geochemical anomalies.

From the results, it is clear that the proposed methodology can predict a range of trace elements with excellent accuracy, such as the lanthanides as measured by both the CoD and MAPE (Figs. 4 and 5). Tree-based models in comparison to the PCA-based approach have several advantages, which aside from being able to incorporate categorical features (e.g. Harris et al., 2015), are also highly capable of capturing feature interactions, especially nonlinear feature interactions, while not assuming any geometry of the feature space. In this sense, the tree-based approach in comparison to the PCA-based approach is more general, more capable and should generally exhibit a higher explanatory power. This distinction should be more prominent where the feature interactions are increasingly nonlinear, such as in geological samples that may be highly non-stoichiometric (volcanic glasses, highly altered clays, etc.). By using a variety of machine learning algorithms over a parameter grid, the proposed methodology can automatically produce models that do not assume linear or even parametric elemental interactions, and the prediction residuals can be used instead of linear regression residuals against principal components.

A few key improvements that the proposed approach offers compared to the traditional approach (i.e. Grunsky, 2013; Grunsky et al., 2014; Harris et al., 2015; Chen et al., 2018; Grunsky and de Caritat, 2019) are:

(a) minimal data preprocessing and in particular log-ratio transforms were unnecessary for our workflow as used in this study;
(b) PCA is not a prerequisite, hence the new approach does not rely on the variability of well-constrained stoichiometries; and
(c) linear or even parametric elemental concentration-interactions are not required and is not a prior assumption.

Since we do not use PCA or other dimensionality reduction techniques, the approach is better suited to using solely the major and minor elements content as features and because we do not assume the existence of minerals in the samples, and therefore, stoichiometry in the data. This distinction should be more prominent where the feature interactions are increasingly nonlinear, such as in geological samples that may be highly non-stoichiometric (volcanic glasses, highly altered clays, etc.).

In this study, we demonstrated that volcanic rocks can be readily incorporated into the new approach and that a combination of mineralised and volcanic rocks does not significantly impact the prediction performance. The prediction residuals resulting from the predictions can also be combined in a post-processing manner to be used for multivariate prospectivity mapping, for example, through a knowledge-based approach (weights assigned to different elemental maps) or a data-driven approach (autoencoders, linear or kernel PCA, or self-organizing maps).

In this sense, the machine learning-based approach in comparison to the PCA-based approach is more general, more capable and should generally produce models of higher explanatory powers where sample mineralogy is either not present, and/or is ineffective in the modelling of geological processes. This greater flexibility implies that our method can serve either as a replacement for the traditional workflow (i.e. Grunsky, 2013; Harris et al., 2015; Grunsky and de Caritat, 2019) or as a pre-processor to increase the geochemical baseline for anomaly detection. Since the prediction residuals do not suffer from closure and are not strictly positive, they do not require log-ratio transformations and should facilitate further modelling and interpretations. Another key benefit of the proposed approach is its high degree of automatability. In the

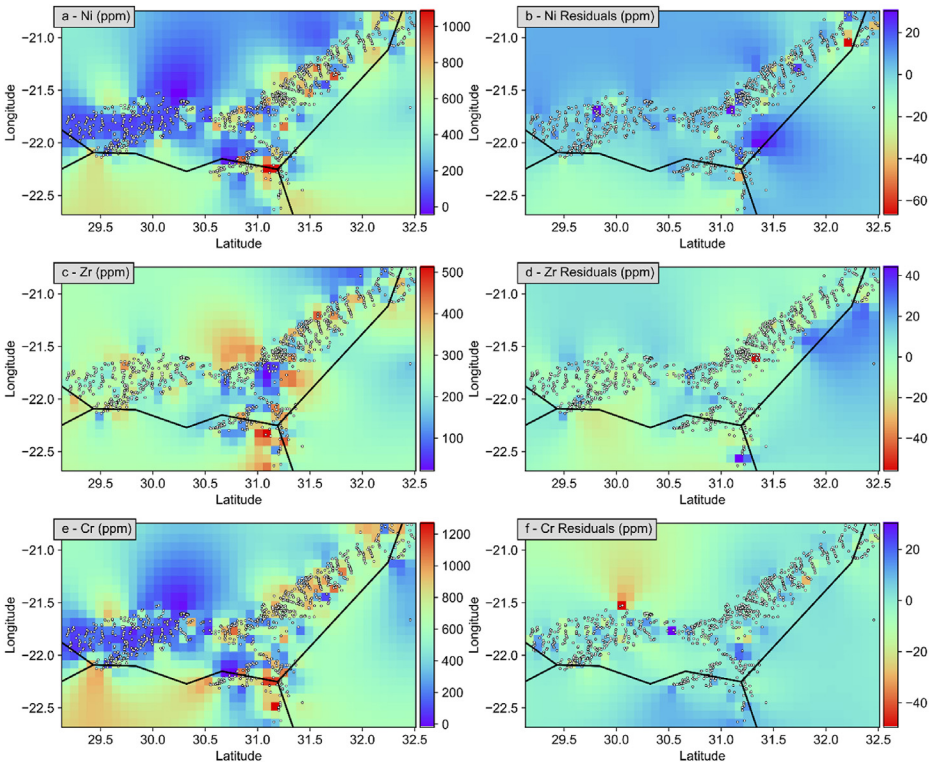

Fig. 10. Elemental concentration and the prediction residual maps for Zr, (a) and (b), Ce, (c) and (d) and Ni, (e) and (f), respectively. The black lines are national borders (Mozambique to the east, Zimbabwe to the north, Botswana to the west and South Africa to the south). This region is dominated by the volumetrically minor high-Ti suite of Karoo magmatic rocks.

implementation, all of the algorithm selection, model tuning and prediction performance profiling, to mapping are automated within a single workflow. In addition, the objectivity of the resulting prospectivity maps is increased, because there is no need to interpret intermediary results (i.e. the stoichiometry and meaning of principal components). However, this does result in a loss of scientific insight into the geological processes that may be operating in the sampled region if the prospectivity maps are not further interpreted after their generation.

Since the prediction accuracy as measured by the CoD is satisfactory for a range of trace elements (Fig. 5) and the prediction residuals are symmetrical for most elements (Fig. 6), it is possible to use the new approach to generate regional prospectivity maps. However, in the context of the Karoo dataset, the data was not originally acquired for regional mapping or prospectivity purposes and as such, much of it does not follow a systematic grid-based sampling and mapping-associated quality control and quality assurance. Unlike some of the large igneous provinces such as the platinum group elements- (PGE) rich rocks of the Bushveld Igneous Complex, there are no known economic mineralisation in the Karoo large igneous province. Ni-sulfide mineralisation is found at the Insizwa (or Mount Ayliff) Complex (Maier et al., 2002; Marsh et al., 2003). However, at present, the Insizwa Complex mineralisation is sub-economic. Therefore, all trace elements were given equal attention and no economic mineralisation hotspots were detected in this current study. This does not mean that there is no probability of mineralisation potential in magmatic rocks of the Karoo since the data used in this study is sparse and mostly limited to outcrops. It is possible to visualise subsets of the results for a qualitative interpretation. To demonstrate the use of the method for mapping, we created maps of element distributions and of the prediction residuals for basaltic samples in an area with good coverage using an RBF interpolator at a grid size of ~8.5 km to create maps (Fig. 10). RBF interpolation is similar to dual kriging (Horowitz et al., 1996) and is suitable where the sample coverage is highly variable (Stewart et al., 2014) and can provide reliable results (Yamamoto, 2002). However, the results should be taken qualitatively given the limitations of the data. The maps suggest that the prediction residuals, compared to original elemental distribution maps, exhibit less spatial variability (Fig. 10). This is precisely the result expected if, for a trace element, some portion of its variability is explained by rock-forming processes. In other words, the remainder of the variability seen on the residual maps is more likely to be true geochemical anomalies. In the residual maps, it is much easier to attribute localised variability to one or a few specific data points compared to that of the elemental distribution maps (Fig. 10). For example, in the Zr residuals map and compared to the Zr concentration map (Fig. 10), the variability near the centre of the Zr concentration map near the tri-national border is substantially attenuated in the Zr residuals map. This illustrates that the observed local variation in Zr concentration can be largely explained by changes in the rock-forming elements of the samples and the amount of anomalous concentration of Zr in this region is substantially smaller than observed changes in Zr concentration. Similarly, within the coverage of the data, for the entire Zr maps, much of the concentration variability are non-anomalous. For the Ce and Ni maps (Fig. 10), the detected amount of geochemical anomaly is far smaller than suggested by the concentration maps. This suggests that relative to the areal mean of elemental concentrations, the machine learning models' prediction of trace element concentrations provide a much higher geochemical baseline, against which the strengths of geochemical anomalies can be measured. Such detected geochemical anomalies can be mapped or used as evidence layers to assist with targeting for exploration purposes.

7. Conclusion

High-dimensional relationships between major, minor and trace elements are useful for the prediction of trace elements. In this study, we showed that a range of trace elements can be adequately predicted using machine learning algorithms by using both major and minor elemental data. This discovery has several implications; first, for process discovery in geochemical prospectivity mapping, it is not necessary to manually

discover elemental associations to determine a geochemical baseline for anomaly detection. We have demonstrated that this process can be more data-driven and automated. The second finding is that the proposed method is capable of using raw data, instead of log-ratio transformed data, both for algorithms that are aware and unaware of the feature space geometry. Indeed, performing such a transformation as part of data preprocessing leads to no statistically significant change in the performance of the method across two performance metrics. By corollary, this implies that the new workflow can be easily integrated with a traditional geochemical prospectivity mapping workflow, and either as a preprocessor to increase the baseline for anomalies before multivariate modelling (data transformation, principal component analysis, component interpretation, regression of elements of interest against components and mapping) or as a replacement for multivariate modelling if the target can be expressed as a combination of predictable elements (e.g. a trace element that is an indicator or vector to deposit). Used as a data preprocessor, the resulting prediction residuals, unlike compositional data, are not strictly positive and do not exhibit closure, and therefore may be used directly for additional multivariate modelling if necessary (to explore regional patterns). A third implication is that the approach does not assume the existence of stoichiometry in the data (minerals in the samples) and therefore, works equally well for volcanic rocks as we have demonstrated here. The proposed approach works well considering that the original data was unintended for predictive modelling and mapping and that we do not restrict the data to specific rock types. Qualitative maps using the prediction residuals compared to those of the elemental distributions show an increase in the baseline of geochemical anomalies. The last benefit of the proposed approach is that machine learning algorithms benefit from training using data that exhibits some variability, such as variability introduced by sampling and analytical methods. Given that the proposed method works with legacy data that is multiply sourced, applying this new method to high-quality prospectivity mapping should meet or exceed the performance observed in this study.

Funding

Glen Nwaila acknowledges funding from the National Research Foundation (NRF) Thuthuka Grant and from DSI-NRF CIMERA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank an anonymous reviewer and Pierre-Marc Godbout (Geological Survey of Canada) for their enlightening comments which have greatly improved this paper. Daniel Frederick Wright (Geological Survey of Canada) is also thanked. We thank Hua Wang for editorial handling.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiig.2021.11.002.

References

Adcock, S.W., Spirito, W.A., Garrett, R.G., 2013. Geochemical data management – issues and solutions. Geochem. Explor. Environ. Anal. 13, 337–348. https://doi.org/10.1144/geochem2011-084.

Agrawal, S., Guevara, M., Verma, S.P., 2008. Tectonic discrimination of basic and ultrabasic volcanic rocks through log-transformed ratios of immobile trace elements. Int. Geol. Rev. 50 (12), 1057–1079. https://doi.org/10.2747/0020-6814.50.12.1057.

Aitchison, J., 1982. The statistical analysis of compositional data. J. R. Stat. Soc. Series B. 44 (2), 139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.

An, S., Liu, W., Venkatesh, S., 2007. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. Pattern Recogn. 40 (8), 2154–2162. https://doi.org/10.1016/j.patcog.2006.12.015.

Ashwal, L.D., 2021. Sub-lithospheric mantle sources for overlapping southern african large igneous provinces. S. Afr. J. Geol. 124 (2), 421–442. https://doi.org/10.25131/sajg.124.0023.

Ashwal, L.D., Ziegler, A., Truebody, T., Bolhar, R., 2021. Origin of Sr-enriched glassy picrites from the Karoo large igneous province. Geochemistry, Geophysics, Geosystems (G-cubed). ESSOAr. https://doi.org/10.1002/essoar.10503697.1.

Breiman, L., 1996a. Bagging predictors. Mach. Learn. 24 (2), 123–140.

Breiman, L., 1996b. Stacked regressions. Mach. Learn. 24 (1), 49–64.

Buiter, S.J.H., Torsvik, T.H., 2014. A review of Wilson Cycle plate margins: a role for mantle plumes in continental break-up along sutures? Gondwana Res. 26, 627–653. https://doi.org/10.1016/j.gr.2014.02.007.

Burkov, A., 2020. Machine Learning Engineering. True Positive Inc, ISBN 978-1777005467.

Carranza, E.J.M., 2009. Fractal analysis of geochemical anomalies. In: Carranza, E.J.M. (Ed.), Handbook or Exploration and Environmental Geochemistry, vol. 11. Elsevier Science, pp. 85–114. https://doi.org/10.1016/S1874-2734(09)70008-7.

Catuneanu, O., Wopfner, H., Eriksson, P.G., Cairncross, B., Rubidge, B.S., Smith, R.M.H., Hancox, P.J., 2005. The Karoo basins of south-central Africa. J. Afr. Earth Sci. 43 (1–3), 211–253. https://doi.org/10.1016/j.jafrearsci.2005.07.007.

Chen, L., Wang, L., Miao, J., Gao, H., Zhang, Y., Yao, Y., Bai, M., Mei, L., He, J., 2020. Review of the application of big data and artificial intelligence in Geology. J. Phys. Conf. Ser. 1684 (1), 012007. https://doi.org/10.1088/1742-6596/1684/1/012007.

Chen, S., Hattory, K., Grunsky, E.C., 2016. Multivariate statistical analysis of the REE-mineralization of the maw zone, athabasca basin, Canada. J. Geochem. Explor. 161, 98–111. https://doi.org/10.1016/j.gexplo.2015.11.009.

Chen, S., Hattori, K., Grunsky, E.C., 2018. Identification of sandstones above blind uranium deposits using multivariate statistical assessment of compositional data, Athabasca Basin, Canada. J. Geochem. Explor. 188, 229–239. https://doi.org/10.1016/j.gexplo.2018.01.026.

Chen, Z., Chen, J., Tian, S., Xu, B., 2017. Application of fractal content-gradient method for delineating geochemical anomalies associated with copper occurrences in the Yangla ore field, China. Geosci. Front. 8 (1), 189–197. https://doi.org/10.1016/j.gsf.2015.11.010.

Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. J. Geochem. Explor. 51 (2), 109–130. https://doi.org/10.1016/0375-6742(94)90013-2.

Cheng, Q., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. Nat. Resour. Res. 9, 43–52. https://doi.org/10.1023/A:1010109829861.

Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol. Rev. 32 (1-2), 314–324. https://doi.org/10.1016/j.oregeorev.2006.10.002.

Chin, E.J., Shimizu, K., Bybee, G.M., Monica, E., Erdman, M.E., 2018. On the development of the calc-alkaline and tholeiitic magma series: a deep crustal cumulate perspective. Earth Planet Sci. Lett. 482, 277–287. https://doi.org/10.1016/j.epsl.2017.11.016.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theor. 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964.

Curry, H.B., 1944. The method of steepest descent for non-linear Minimisation problems. Q. Appl. Math. 2, 258–261. https://doi.org/10.1090/qam/10667.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control. Signals, Syst. 2 (4), 303–314. https://doi.org/10.1007/BF02551274.

Deutsch, C.V., Journel, A.G., 1992. Geostatistical Software Library and User's Guide, vol. 119. New York, 147.

Deutsch, C.V., Journel, A.G., 1997. In: GSLIB Geostatistical Software Library and User's Guide, second ed. Oxford University Press, New York.

Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55, 78–87. https://doi.org/10.1145/2347736.2347755.

Duncan, A.R., Erlank, A.J., Marsh, J.S., 1984. Regional geochemistry of the Karoo igneous province. In: Erlank, A.J. (Ed.), Petrogenesis of the Volcanic Rocks of the Karoo Province, Special Publication 13. Geological Society of South Africa, Johannesburg, pp. 355–388.

Du Toit, A.L., 1954. In: The Geology of South Africa, third ed. Oliver and Boyd, Edinburgh.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Bardeló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35, 279–300. https://doi.org/10.1023/A:1023818214614.

Ellefsen, K.J., Van Gosen, B.S., 2020. Bayesian Modeling of Non-stationary, Univariate, Spatial Data for the Earth Sciences. U.S. Geological Survey Techniques and Methods. https://doi.org/10.3133/tm7C24 book 7, chap. C24.

Flzmoser, P., Hron, K., 2009. Correlation analysis for compositional data. Math. Geosci. 41, 905. https://doi.org/10.1007/s11004-008-9196-y.

Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. Environmetrics 20 (6), 621–632. https://doi.org/10.1002/env.966.

Fix, E., Hodges, J.L., 1951. An important contribution to nonparametric discriminant analysis and density estimation. Int. Stat. Inst. 57, 233–238. https://doi.org/10.2307/1403796.

Fletcher, W.K., 1981. Analytical Methods in Geochemical Prospecting. Handbook of Exploration Geochemistry, vol. 1. Elsevier, Amsterdam.

Freund, Y., Schapire, R.E., 1995. A decision-theoretic generalization of online learning and an application to boosting. In: Vitányi, P. (Ed.), Second European Conference on Computational Learning Theory. Springer.
Gómez-Hernández, J.J., 1991. A Stochastic Approach to the Simulation of Block Conductivity Fields Conditioned upon Data Measured at a Smaller Scale. PhD thesis. Stanford University, California, United States of America.
Gong, Z., Zhong, P., Hu, W., 2019. Diversity in machine learning. IEEE Access 7, 64323–64350. https://doi.org/10.1109/ACCESS.2019.2917620.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Goovaerts, P., 1994. Comparative performance of indicator algorithms for modelling conditional probability distribution functions. Math. Geol. 26 (3), 385–410. https://doi.org/10.1007/BF02089230.
Grunsky, E.C., 2013. Predicting Archaean volcanogenic massive sulphide deposit potential from lithogeochemistry: application to the Abitibi Greenstone Belt. Geochem. Explor. Environ. Anal. 13, 317–336. https://doi.org/10.1144/geochem2012-176.
Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: application for predictive geological mapping. J. Geochem. Explor. 141, 15–41. https://doi.org/10.1016/j.gexplo.2013.07.013.
Grunsky, E.C., de Caritat, P., 2019. State-of-the-art analysis of geochemical data for mineral exploration. Geochem. Explor. Environ. Anal. 20, 217–232. https://doi.org/10.1144/geochem2019-031.
Grunsky, E.C., Arne, D., 2020. Mineral-resource prediction using advanced data analytics and machine learning of the QUEST-South stream-sediment geochemical data, Southwestern British Columbia, Canada. Geochem. Explor. Environ. Anal. 23 (1). https://doi.org/10.1144/geochem2020-054 geochem2020-g2054.
Gu, A., Sala, F., Gunel, B., Ré, C., 2018. Learning mixed-curvature representations in product spaces. In: International Conference on Learning Representations.
Harris, J.R., Grunsky, E., Behnia, P., Corrigan, D., 2015. Data-and knowledge-driven mineral prospectivity maps for Canada's North. Ore Geol. Rev. 71, 788–803. https://doi.org/10.1016/j.oregeorev.2015.01.004.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. IEEE International Conference on Computer Vision (ICCV) 1026–1034. https://doi.org/10.1109/ICCV.2015.123, 2015.
Ho, T.K., 1995. Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994. Montréal, Canada.
Hyontai, S., 2018. Performance of machine learning algorithms and diversity in data. In: MATEC Web Conf, vol. 210, 04019. https://doi.org/10.1051/matecconf/201821004019.
Horowitz, F.G., Hornby, P., Bone, D., Craig, M., 1996. Fast multidimensional interpolations. In: Ramani, R.V. (Ed.), Proceedings of the Application of Computers and Operations Research in the Mineral Industry (APCOM 26). Society Mining Metallurgy and Exploration (SME), Littleton, Colorado.
Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal. 54 (12), 3095–3107. https://doi.org/10.1016/j.csda.2009.11.023.
Hsu, C.W., Lin, C.J., 2002. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Network. 13, 415–425. https://doi.org/10.1109/72.991427.
Irvine, T.N., Baragar, W.R.A., 1971. A guide to the chemical classification of the common volcanic rocks. Can. J. Earth Sci. 8, 523–548. https://doi.org/10.1139/e71-055.
Isaaks, E., Srivastava, R., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.
Johnson, M.R., 1994. Lexicon of South African Stratigraphy. Part 1: Phanerozoic Units. South African Committee for Stratigraphy, Council for Geoscience Pretoria, South Africa.
Johnson, M.R., van Vuuren, C.J., Hegenberger, W.F., Key, R., Shoko, U., 1996. Stratigraphy of the Karoo Supergroup in southern Africa: an overview. J. Afr. Earth Sci. 23 (1), 3–15. https://doi.org/10.1016/S0899-5362(96)00048-6.
Johnson, M.R., van Vuuren, C.J., Visser, J.N.J., Cole, D.J., Wickens, H.deV., Christie, A.D.M., Roberts, D.L., 1997. The foreland Karoo basin, South Africa. In: Selley, R.C. (Ed.), African Basins – Sedimentary Basins of the World. Elsevier, Amsterdam.
Journel, A., 1974. Geostatistics for conditional simulation of orebodies. Econ. Geol. 69 (5), 673–687. https://doi.org/10.2113/gsecongeo.69.5.673.
Journel, A., 1980. The lognormal approach to predicting local distributions of selective mining unit grades. J. Int. A. Math. Geol. 12, 285–303. https://doi.org/10.1007/BF01029417.
Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support vector machines in R. J. Stat. Software 15, 1–28. https://doi.org/10.18637/jss.v015.i09.
Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2018. Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. 31 (8), 1544–1554. https://doi.org/10.1109/TKDE.2018.2861006.
Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: a review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering 160 (1), 3–24.
Kotsiantis, S.B., 2014. Bagging and boosting variants for handling classifications problems: a survey. Knowl. Eng. Rev. 29, 78–100. https://doi.org/10.1017/S0269888913000313.
Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. J. Chem. Metall. Min. Soc. S. Afr. 119–139. https://hdl.handle.net/10520/AJA0038223X_4792.
Krige, D.G., 1952. A statistical analysis of some of the borehole values in the Orange Free State goldfield. J. Chem. Metall. Min. Soc. S. Afr. 47–64. https://hdl.handle.net/10520/AJA0038223X_5358.
Krige, D.G., 1955. Travaux de M.D.G. KRIGE sur l'evaluation des gisements dans les mines d'or sud-africaines. Ann. Mine. 12, 3–39.
Lawley, C.J.M., Tschirhart, V., Smith, J.W., Pehrsson, S.J., Schetselaar, E.M., Schaeffer, A.J., Houlé, M.G., Eglington, B.M., 2021. Prospectivity modelling of Canadian magmatic Ni (±Cu ± Co ± PGE) sulphide mineral systems. Ore Geol. Rev. 132, 103985. https://doi.org/10.1016/j.oregeorev.2021.103985.
Le Bas, M., Le Maitre, R., Streckeisen, A., Zanettin, B., IUGS Subcommission on the Systematics of Igneous Rocks, 1986. A chemical classification of volcanic rocks based on the total alkali-silica diagram. J. Petrol. 27 (3), 745–750. https://doi.org/10.1093/petrology/27.3.745.
Lin, Y., Ye, Q., 2020. Support vector machine classifiers by non-Euclidean margins. Math. Found. Comput. 3 (4), 279–300. https://doi.org/10.3934/mfc.2020018.
Lemaréchal, C., 2012. Cauchy and the gradient method. Doc. Math Extra 251 (254), 10.
Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. 29, 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002.
Maier, W.D., Marsh, J.S., Barnes, S.-J., Dodd, D.C., 2002. The distribution of platinum group elements in the Insizwa lobe, Mount Ayliff complex, South Africa: implications for Ni-Cu-PGE sulphide exploration in the Karoo large igneous province. Econ. Geol. 97, 1–14. https://doi.org/10.2113/gsecongeo.97.6.1293.
Marsh, J.S., Allen, P., Fenner, N., 2003. The geochemical structure of the Insizwa lobe of the Mount Ayliff complex with implications for the emplacement and evolution of the complex and its Ni-sulphide potential. S. Afr. J. Geol. 106, 409–428. https://doi.org/10.2113/106.4.409.
Matheron, G., 1962. Traité de géostatistique appliquée, vol. 1. Editions Technip, Paris.
McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., de Caritat, P., Filzmoser, P., van den Boogaart, K.G., Tolosana-Delgado, R., 2016. The single-component geochemical map: fact or fiction? J. Geochem. Explor. 162, 16–28. https://doi.org/10.1016/j.gexplo.2015.12.005.
Middlemost, E.A., 1994. Naming materials in the magma/igneous rock system. Earth Sci. Rev. 37 (3-4), 215–224. https://doi.org/10.1016/0012-8252(94)90029-9.
Nwaila, G.T., Zhang, S.E., Frimmel, H.E., Manzi, M.S.D., Dohm, C., Durrheim, R.J., Burnett, M., Tolmay, L., 2020. Local and target exploration of conglomerate-hosted gold deposits using machine learning algorithms: a case study of the Witwatersrand gold ores, South Africa. Nat. Resour. Res. 29, 135–159. https://doi.org/10.1007/s11053-019-09498-1.
Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015. Modeling and Analysis of Compositional Data. John Wiley & Sons.
Pyrcz, M.J., Deutsch, C.V., 2014. Geostatistical Reservoir Modeling. Oxford University Press, New York.
Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: death of a myth: consequences for the statistical treatment of geochemical and environmental data. Environ. Geol. 39, 1001–1014. https://doi.org/10.1007/s002549900081.
Rodriguez-Galiano, V.F., Chica-Olmo, M., Chica-Rivas, M., 2014. Predictive modelling of gold potential with the integration of multisource information based on Random Forest: a case study on the Rodalquilar area, Southern Spain. Int. J. Geogr. Inf. Sci. 28 (7), 1336–1354. https://doi.org/10.1080/13658816.2014.885527.
Rosenblatt, F., 1961. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC. https://doi.org/10.1007/978-3-642-70911-1_20.
Rubidge, B.S., Hancox, P.J., Catuneanu, O., 2000. Sequence analysis of the Ecca-Beaufort contact in the southern Karoo of South Africa. S. Afr. J. Geol. 103 (1), 81–96. https://doi.org/10.2113/103.1.81.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., PDP research group, 1986. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. Foundation. MIT Press.
Russell, S.J., Norvig, P., 2010. In: Artificial Intelligence: A Modern Approach, third ed. Prentice-Hall, ISBN 9780136042594.
SACS (South African Committee for Stratigraphy), 1980. Stratigraphy of South Africa, Part 1. Handbook Geological Survey South Africa, Pretoria, p. 8.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 8 (4), e1249. https://doi.org/10.1002/widm.1249.
Samuel, A.L., 1959. Some studies in Machine Learning using the game checkers. IBM J. Res. Dev. 3 (3), 210–229. https://doi.org/10.1147/rd.33.0210.
Santosa, F., William, W.S., 1986. Linear inversion of band-limited reflection seismograms. J. Sci. Stat. Comput. 7, 1307–1330. https://doi.org/10.1137/0907087.
Sisson, T.W., Grove, T.L., 1993. Experimental investigations of the role of H_2O in calc-alkaline differentiation and subduction zone magmatism. Contrib. Mineral. Petrol. 113, 143–166. https://doi.org/10.1007/BF00283225.
Smith, R.M.H., Eriksson, P.G., Botha, W.J., 1993. A review of the stratigraphy and sedimentary environments of the Karoo-aged basins of Southern Africa. J. Afr. Earth Sci. 16 (1-2), 143–169. https://doi.org/10.1016/0899-5362(93)90164-L.
Stanley, C.R., 2003. THPLOT.M: a MATLAB function to implement generalized Thompson-Howarth error analysis using replicate data. Comput. Geosci. 29 (2), 225–237.
Stanley, C.R., Sinclair, A.J., 1986. Relative error analysis of replicate geochemical data: advantages and applications. In: GeoExpo – 1986. Exploration in the North American

Cordillera. Programs and Abstracts. Association of Exploration Geochemists, Vancouver, Canada, pp. 77–78.
Stewart, M., de Lacey, J., Hodkewicz, P.F., Lane, R., 2014. Grade estimation from radial basis functions – how does it compare with conventional geostatistical estimation?. In: Proceedings of the Ninth International Mine Geology Conference. The AusIMM, Melbourne, pp. 129–142.
Storey, B.C., 1995. The role of mantle plumes in continental breakup: case histories from Gondwanaland. Nature 377, 301–308. https://doi.org/10.1038/377301a0.
Storey, B.C., Kyle, P.R., 1997. An active mantle mechanism for Gondwana breakup. S. Afr. J. Geol. 100, 283–290. https://hdl.handle.net/10520/EJC-929a29b77.
Svensen, H., Corfu, F., Polteau, S., Hammer, Ø., Planke, S., 2012. Rapid magma emplacement in the Karoo large igneous province. EPSL 325–326, 1–9. https://doi.org/10.1016/j.epsl.2012.01.015.
Templ, M., 2021. Artificial neural networks to impute rounded zeros in compositional data. In: Filzmoser, P., Hron, K., Martín-Fernández, J.N., Palarea-Albaladejo, J. (Eds.), Advances in Compositional Data Analysis. Springer International Publishing, pp. 163–187.
Therrien, R., Doyle, S., 2018. Role of training data variability on classifier performance and generalizability. In: Proceedings SPIE 10581, Medical Imaging 2018: Digital Pathology, p. 1058109. https://doi.org/10.1117/12.2293919.
Thompson, M., 1973. DUPAN 3, A subroutine for the interpretation of duplicated data in geochemical analysis. Comput. Geosci. 4, 333–340. https://doi.org/10.1016/0098-3004(78)90096-1.
Thompson, M., 1982. Regression methods and the comparison of accuracy. Analyst 107 (1279), 1169–1180.
Thompson, M., Howarth, R.J., 1973. The rapid estimation and control of precision by duplicate determinations. Analyst 98 (1164), 153–160.
Thompson, M., Howarth, R.J., 1976a. Duplicate analysis in geochemical practice – Part 1. Theoretical approach and estimation of analytical reproducibility. Analyst 101 (1206), 690–698.
Thompson, M., Howarth, R.J., 1976b. Duplicate analysis in geochemical practice – Part 2. Examination of proposed methods and examples of its use. Analyst 101 (1206), 699–709.
Thompson, M., Howarth, R.J., 1978. A new approach to the estimation of analytical precision. J. Geochem. Explor. 9 (1), 23–30. https://doi.org/10.1016/0375-6742(78)90035-3.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Series B Methodol. 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
Tikhonov, A.N., 1943. On the stability of inverse problems. Dokl. Akad. Nauk SSSR 39, 195–198.
Vapnik, V., 1998. Statistical Learning Theory. Springer, New York.
Vermeesch, P., 2013. Tectonic discrimination diagrams revisited. G-cubed 74, 466–480. https://doi.org/10.1029/2005GC001092.
Visser, J.N.J., 1991. The paleoclimatic setting of the late paleozoic marine ice sheet in the Karoo basin of southern Africa. In: Anderson, J.B., Ashley, G.M. (Eds.), Glacial Marine Sedimentation: Paleoclimatic Significance, vol. 261. Geological Society of America Special Paper, pp. 181–189.
Witten, I.H., Frank, E., 2005. In: Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufman, San Francisco.
Wood, D.A., 1980. The application of a Th, Hf, Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary volcanic province. EPSL 50, 11–30. https://doi.org/10.1016/0012-821X(80)90116-8.
Yamamoto, J.K., 2002. Ore reserve estimation using radial basis functions. Rev. Institut. Geol. 23 (1), 25–38. https://doi.org/10.5935/0100-929X.20020003.
Yu, H., Li, R., Wang, G., Wang, Q., 2019. Current development of landscape geochemistry with support of geospatial technologies: a review. Crit. Rev. Environ. Sci. Technol. 49 (9), 745–790. https://doi.org/10.1080/10643389.2018.1558890.
Zhang, G., Hu, M.Y., Patuwo, B.E., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. Eur. J. Oper. Res. 116 (1), 16–32.
Zhang, S.E., Nwaila, G.T., Tolmay, L., Frimmel, H.E., Bourdeau, J.E., 2021. Integration of machine learning algorithms with Gompertz curves and kriging to estimate resources in gold deposits. Nat. Resour. Res. 30, 39–56. https://doi.org/10.1007/s11053-020-09750-z.
Zou, H., Hastie, T., 2005. Regularisation and variable selection via the elastic net. J. R. Stat. Soc. Series B 67 (2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.