State-of-the-art analysis of geochemical data for mineral exploration

E. C. Grunsky[1,2]* & P. de Caritat[3,4]

[1] Department of Earth & Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[2] China University of Geosciences, Beijing, China
[3] Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
[4] Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia
    ECG, 0000-0003-4521-163X; PdC, 0000-0002-4185-9124
* Correspondence: egrunsky@gmail.com

Abstract: Multi-element geochemical surveys of rocks, soils, stream/lake/floodplain sediments and regolith are typically carried out at continental, regional and local scales. The chemistry of these materials is defined by their primary mineral assemblages and their subsequent modification by comminution and weathering. Modern geochemical datasets represent a multi-dimensional geochemical space that can be studied using multivariate statistical methods from which patterns reflecting geochemical/geological processes are described (process discovery). These patterns form the basis from which probabilistic predictive maps are created (process validation). Processing geochemical survey data requires a systematic approach to effectively interpret the multi-dimensional data in a meaningful way. Problems that are typically associated with geochemical data include closure, missing values, censoring, merging, levelling different datasets and adequate spatial sample design. Recent developments in advanced multivariate analytics, geospatial analysis and mapping provide an effective framework to analyse and interpret geochemical datasets. Geochemical and geological processes can often be recognized through the use of data discovery procedures such as the application of principal component analysis. Classification and predictive procedures can be used to confirm lithological variability, alteration and mineralization. Geochemical survey data of lake/till sediments from Canada and of floodplain sediments from Australia show that predictive maps of bedrock and regolith processes can be generated. Upscaling a multivariate statistics-based prospectivity analysis for arc-related Cu–Au mineralization from a regional survey in the southern Thomson Orogen in Australia to the continental scale, reveals a number of regions with a similar (or stronger) multivariate response and hence potentially similar (or higher) mineral potential throughout Australia.

Keywords: geochemistry; analytical methods; compositional data; multivariate analytics; process discovery; process validation; predictive mapping; machine learning; geospatial coherence; Melville Peninsula; Nunavut; Thomson Region; New South Wales

What do geochemical data represent?

Geochemical datasets can be defined as geochemical data derived from a range of media (e.g. soil, till, regolith, lake sediments, stream sediments, bedrock) collected at a spatial scale consistent with the geological/geochemical processes being investigated. Continental, regional and local scale surveys reveal increasingly detailed processes ranging from the tectonic assemblage of continents to hydrothermal veining, for instance.

The intent of geochemical surveys is to provide a spatial geochemical description of the general geology or dominant geochemical processes as manifested in the medium being sampled. For example, the geochemistry of glacial till or regolith over an area may reflect the underlying geology, or it may reflect the source material that has been transported.

The value of geochemical datasets

Geochemical surveys generally contribute to the economy, environment and society through supporting fact-based decisions in the following applications: mineral exploration, regional geological mapping, agriculture and forestry, environmental baseline monitoring, environmental remediation, geohealth and general land use stewardship. For example, geochemical surveys can have an impact on the understanding of human health issues from natural contamination of the source rock or the effects of the urban environment through anthropogenic activities that result in local pollution.

Survey area and density

Sample density is a critical aspect of geochemical survey design and subsequent interpretation. Sample density, generally described in terms of the average area, or volume (in 3D) that each sample site represents, will have an influence on the detection and discovery of geochemical/geological processes that have acted at a specific spatial scale. Local scale or high-density surveys have sample site densities in the range of more than 100 sites per km^2 to one site per km^2. Regional scale geochemical surveys can vary from one site per km^2 to one site per 500 km^2, and continental scale surveys can vary from one site per 500 km^2 to one site per 5000 km^2 (Geological Survey of Northern Ireland 2007; Reimann *et al.* 2009, 2010, 2014; Caritat & Cooper 2011; Smith *et al.* 2011). It is evident that high-density surveys are able to detect local scale processes, which can be associated with mineral deposits. As the density of a survey decreases, the likelihood of randomly sampling a site that is associated with alteration or mineralization decreases. Conversely, studying increasing larger areas enables detection of large-scale geological processes such as continental accretion, collision, and major fault and shear zones.

Geochemical data

The following is a brief summary of the primary considerations that must be taken into account when obtaining, compiling and synthesizing geochemical data prior to statistical treatment and interpretation. This is not intended to provide all of the details that are necessary when obtaining geochemical data.

Choosing the sample material

The choice of sample media is a critical part of the strategy of any geochemical survey. Sampling bedrock may reveal 'in situ' geochemical processes pertaining to the underlying geology. Sampling regolith that has been derived by weathering of in situ bedrock may present a geochemical signature that reflects both the protolith and its weathering. Sampling transported material, such as glacial till, lake sediments, stream sediments, overbank sediments, colluvial or alluvial material may reflect varying amounts of transport and mixing of several processes, which may be desirable. It is important to recognize the nature of the sample media and abilities and limitations of what can be interpreted from the derived geochemical data.

Choosing the appropriate size fraction

In sample media that are comprised of a mix of mineral/organic matter, the size fraction of the mineral grains or particles analysed can be important in distinguishing between geological/geochemical processes. In most sample media, a distinction between a coarse-grained (typically >63 μm and <2 mm) and fine-grained (<63 μm) fraction is commonly used. Coarse-grained material can be considered, in many cases, to represent locally derived particles, or minerals that have not undergone weathering, comminution or chemical dissolution. The geochemical signature from fine-grained mineral matter may represent minerals that have undergone weathering, comminution and chemical dissolution/precipitation. The fine-grained size fraction is generally considered to reflect a greater range of geochemical processes, although this is dependent on the source material and the nature of the subsequent processes that occurred. Another consideration to be aware of relating to particle sizing is that certain sample analysis methods (e.g. fusion to prepare glass discs for XRF or LA-ICP-MS analysis) may require the sample to be ground or milled to a given specification (e.g. >X% of mass passing through a 60 μm sieve). If this is the case, the impact of breaking up mineral aggregates and/or litho-fragments at the sample preparation stage, as a fit-for-purpose strategy, must be borne in mind when interpreting the results.

Choosing the appropriate analytical method – digestion and instrumentation

The choice of analytical method, which includes the method(s) of sample digestion and subsequent instrumentation for determining elemental abundances, is critical in the interpretation of the results. The choice of sample digestion is generally the most important. Several types of acid digestion, including four-acid (HF-HCl-HNO$_3$-HClO$_4$), aqua regia (HCl-HNO$_3$) and numerous weak/partial extractions, will (preferentially) dissolve specific mineral, organic and amorphous phases, or target certain physical sites (e.g. adsorbed ions, exchangeable cations). Four-acid digestion is a 'near-complete' digestion that dissolves all but the most resistant minerals (e.g. monazite, zircon). The use of aqua regia is useful for dissolving sulphide/oxide type minerals while leaving most silicate minerals unaffected. Weak acid leaches (extractions) tend to dissolve the coatings on mineral grains and/or adsorbed species that are associated with alteration/mineralization processes. Other methods of sample preparation include the use of a total fusion (rather than digestion) whereby finely ground sample material is melted with a flux (e.g. Li$_2$B$_4$O$_7$) to form a homogeneous glass bead or disk that can then either be analysed directly or taken into solution with an acid, e.g. HNO$_3$.

The resulting acid solution is then presented to an analytical instrument after dilution as appropriate. Common technologies include inductively coupled plasma optical emission spectrometry (ICP-OES) and inductively coupled plasma mass spectrometry (ICP-MS). In these techniques, the acid digest is first aspirated into a chamber and typically admixed with argon gas before being converted at high temperature to a plasma. Subsequently an optical emission spectrum is produced where each element has a unique emission spectrum (ICP-OES). Alternatively, a mass spectrometer can be used to separate the elements or molecules based on their unique mass signature (ICP-MS). Older but still current methods of instrumentation include atomic absorption spectroscopy (AAS). Fire assay is the preferred method of preparation in ore-grade materials for the determination of Au, Pt and Pd. Methods such as X-ray fluorescence (XRF) and instrumental neutron activation analysis (INAA) have the ability to analyse a sample without a wet digestion thus delivering a true total analysis. The former is routinely used for the determination of major mineral forming oxides (e.g. Al$_2$O$_3$, SiO$_2$, etc.).

Quality assurance/quality control (QA/QC)

A critical component of geochemical analysis is the monitoring of all procedures that result in the analytical values. Quality control measures include the use of blanks, duplicates and standards to ensure that the results produced are fit for purpose. Two 'quality' parameters are commonly determined: precision, which measures the repeatability of measurements and accuracy, which quantifies how close the obtained results are to the 'real' values. Blanks allow the detection of contamination introduced at any stage of the process, from sampling to analysis. Duplicates can be separated into two types on the basis of their purpose: field duplicates (collected within a given distance from the original sample), which are used to quantify the total (sampling, preparation and analytical) precision; and laboratory duplicates (split in the laboratory under controlled conditions), which test only the analytical precision. Finally, standards also come under two guises: firstly, internal project standards (IPSs), which can track drift in the preparation and analysis steps within and between batches and secondly, certified reference materials (CRMs), which compare results to certified analytical results and are used to establish accuracy. There is a wide range of CRMs (rocks, soils, sediments, water and vegetal matters) available; those chosen should be similar to the material that is being analysed.

The compositional nature of geochemical data

Geochemical data are, by definition, compositional in nature. Elements or oxides of elements are generally expressed as parts per million (ppm), parts per billion (ppb), weight percent (wt%, or simply %) or some other form of 'proportion'. When data are expressed as proportions, there are two important limitations: first, the data are restricted to the positive number space and must sum to a constant (e.g. 1 000 000 ppm, 100%), and second, when one value (proportion) changes, one or more of the others must change too to maintain the constant sum. This problem cannot be overcome by selecting sub-compositions so that there is no constant sum. The 'constant sum' or 'closure' problem results in unreliable statistical measures. The use of ratios between elements, oxides or molecular components that define a composition is essential when making comparisons between elements in systems such as igneous fractionation (Pearce 1968). The use of logarithms of ratios, or simply logratios, is required when measuring moments such as variance/covariance (Aitchison 1986; Egozcue et al. 2003; Buccianti et al. 2006; Pawlowsky-Glahn & Buccianti 2011; Buccianti & Grunsky 2014).
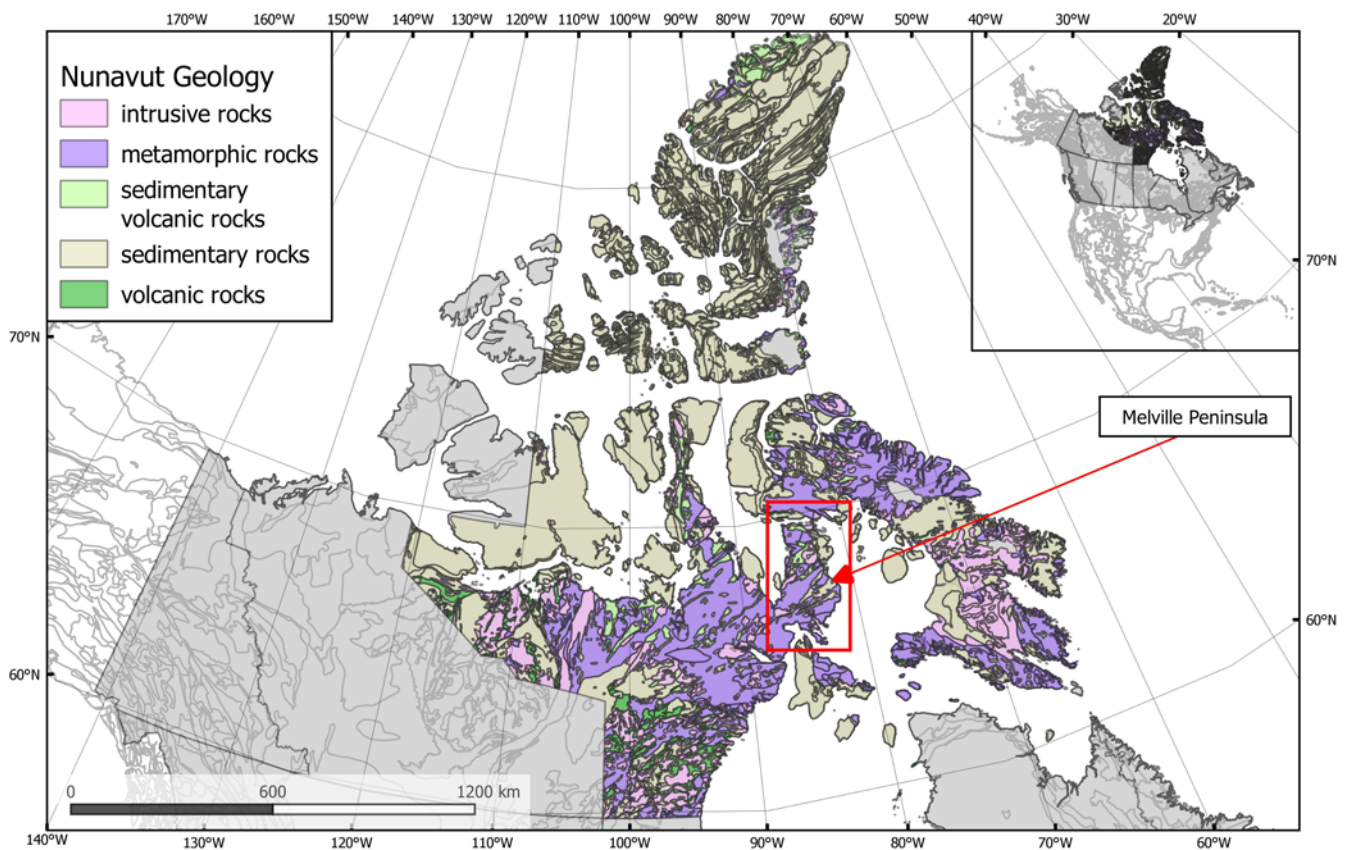
Fig. 1. Regional basement rock type map of Nunavut, Canada, showing the location of the Melville Peninsula study area.

The relationships between the elements of geochemical data are controlled by 'natural laws' (Aitchison 1999). In the case of inorganic geochemistry that law is stoichiometry, which governs how atoms are combined to form minerals, and thereby defines the structure within the data. Geochemical data are not the only type of data to exhibit structure.

Methods

Philosophy

To effectively interpret geochemical data, a two-phased approach is suggested: initial process discovery, followed by process validation. This strategy identifies geochemical/geological processes that exist in the data but may not be obvious unless robust statistical methods are utilized. The process discovery phase is most effective when carried out using a multivariate approach. Linear combinations of elements related by stoichiometry are generally expressed as strong patterns, whilst random patterns and under-sampled processes show weak or uninterpretable patterns. If the process discovery phase provides evidence that there is structure in the data, then models can be built and tested using the process validation phase. If groups of observations are associated with specific processes (bedrock, alteration, mineralization, groundwater, weathering, gravitational sorting), then the observations can be assembled into training sets in which the uniqueness of these groups can be tested.

Process discovery and process validation

A two-step approach is recommended for evaluating multi-element geochemical data. In the first 'process discovery' step patterns, trends and associations between observations (sample sites) and variables (elements) are teased out. Geospatial associations are also a significant part of process discovery. Patterns and/or processes that demonstrate geospatial coherence likely reflect an important geological/geochemical process.

Following process discovery, 'process validation' is the step in which the patterns or associations are statistically tested to determine if these features are valid or merely coincidental associations. Patterns and/or associations that reveal lithological variability in surficial sediment, for instance, can be used to develop training sets from which these lithologies can be predicted in areas where there is uncertainty in the geological mapping and/or paucity of outcrop. Patterns and associations that are associated with mineral deposit alteration and mineralization may be predicted in the same way. In low-density geochemical surveys, where processes such as those related to alteration and mineralization are generally under-sampled, it may be difficult to carry out the process validation phase relating to these processes.

A multivariate approach is an effective way to start the process discovery phase. Linear combinations of elements that are controlled by stoichiometry may emerge as strong patterns, whilst random patterns and/or under-sampled processes show weak or uninterpretable patterns. This approach was successfully used by Grunsky et al. (2012) using multi-element lake sediment geochemical data from the Melville Peninsula area, Nunavut, Canada, and by Caritat & Grunsky (2013) using continental scale multi-element catchment outlet sediment geochemical data from Australia.

Processes are recognized by a continuous range of variable responses and an associated relative increase/decrease in element concentrations. The presence of data that are reported at less than the lower limit of detection (LLD), referred to as censored data, can affect the derivation of associations in the process discovery stage. Using the detection limit, or some arbitrary replacement value (e.g. ½ LLD), as replacement values for censored data, although commonly performed, may bias any statistical (especially multivariate) calculation. Treatment for censored data has been studied within the medical epidemiology community for a long time and was recognized as a
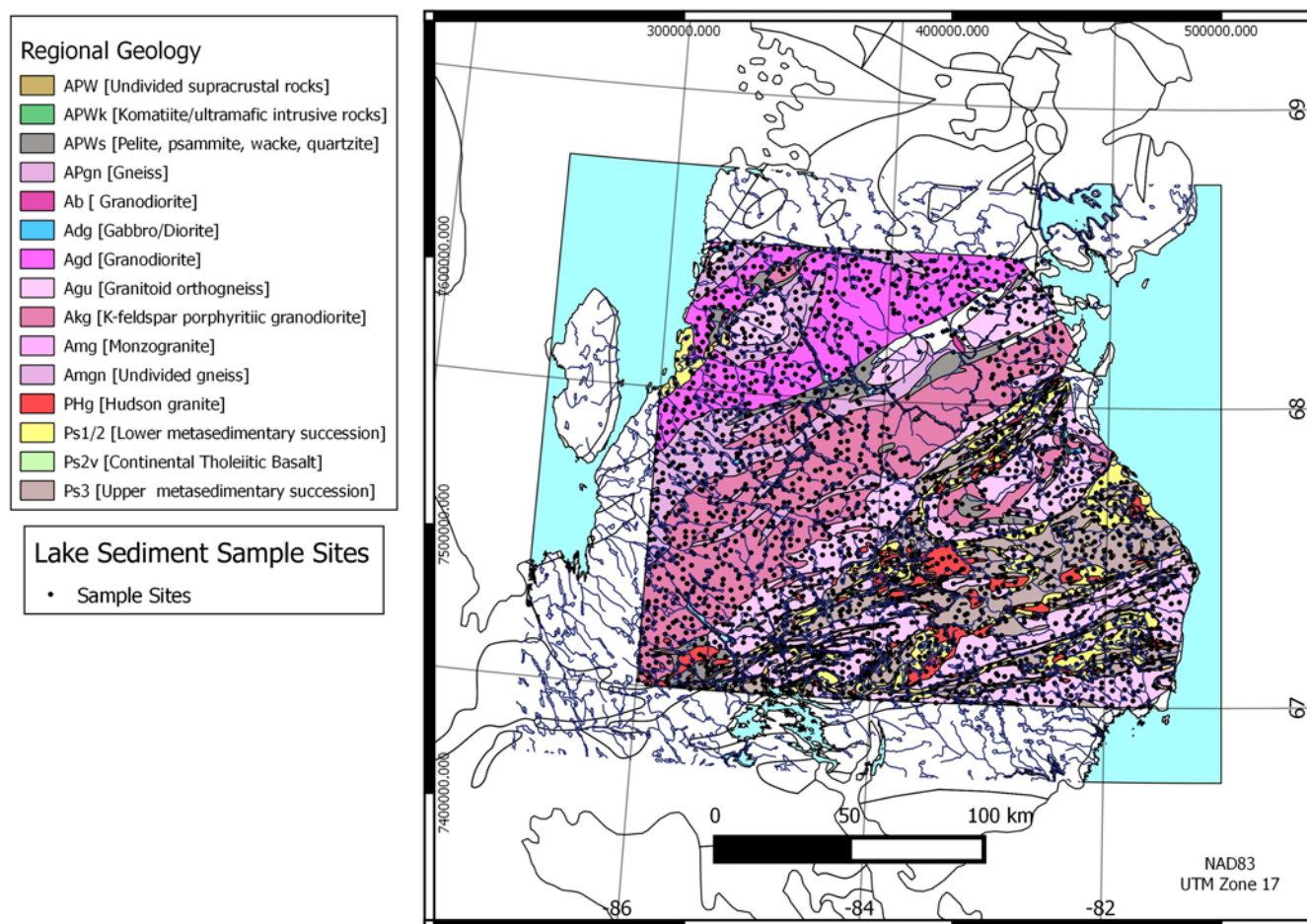
Fig. 2. Geological map of the southern part of the Melville Peninsula, Nunavut, Canada, with lake sediment sample sites (shown as black dots) of Day *et al.* (2009).

problem for geochemical data in the 1980s (Chung 1985; Campbell 1986; Sanford *et al.* 1993). Research by Martín-Fernández *et al.* (2003) and Hron *et al.* (2010) provided various methods for finding replacement values for the censored data. For instance, the R package 'zCompositions' with the function (lrEM) (Palarea-Albaladejo & Martín-Fernández 2008; Palarea-Albaladejo *et al.* 2014) can be used to determine suitable replacement values for several of the elements. Equally important is the distinction between missing values (i.e. no data) and censored data. Missing values may not be censored values, requiring a decision on how they should be replaced, or if they should be used at all (Martín-Fernández *et al.* 2003).

Advanced analytics for process discovery

Process discovery involves the use of unsupervised multivariate methods such as principal component analysis (PCA), independent component analysis (ICA), multi-dimensional scaling (MDS), or random forests (RFs), to name a few. Model-based process discovery methods can also be used, such as model-based clustering (MBC) or RFs. As described previously, statistical measures applied to geochemical data typically reveal linear relationships, which may represent the stoichiometry of rock-forming minerals and subsequent processes that modify mineral structures, including hydrothermal alteration, weathering and water–rock interaction. Physical processes such as gravitational sorting can effectively separate minerals according to the energy of the environment and mineral/grain density. Mineral chemistry is governed by stoichiometry and the relationships of the elements that make up minerals are easily described within the simplex, an *n*-dimensional

composition within the positive real number space. It has long been recognized that many geochemical processes can be clearly described using element/oxide ratios that reflect the stoichiometric balances of minerals during formation (e.g. Pearce 1968). Geochemical data derived from mineral-based materials, when expressed in elemental or oxide form, are a proxy for mineralogy. If the mineralogy of a geochemical dataset is known, then the proportions of these elements can be used to calculate normative mineral proportions (Caritat *et al.* 1994; Grunsky 2013).

An essential part of the process discovery phase is a suitable choice of coordinates to overcome the problem of closure. The centred logratio (clr) transformation (Aitchison 1986) is a useful transform for evaluating geochemical data. The principal components (PCs) of clr-transformed data are orthonormal (i.e. statistically independent) and can reflect linear processes associated with stoichiometric constraints. The PCs offer a significant advantage for subsequent process validation.

Advanced analytics for process validation

Process validation is the methodology used to verify that a geochemical composition (response) reflects one or more processes. These processes can represent lithology, mineral systems, soil development, ecosystem properties, climate or tectonic assemblages. Validation can take the form of an estimate of likelihood that a composition can be assigned membership to one of the identified processes. This is typically done through the assignment of a class identifier or a measure of probability. The prediction of class membership can be done through techniques such as linear
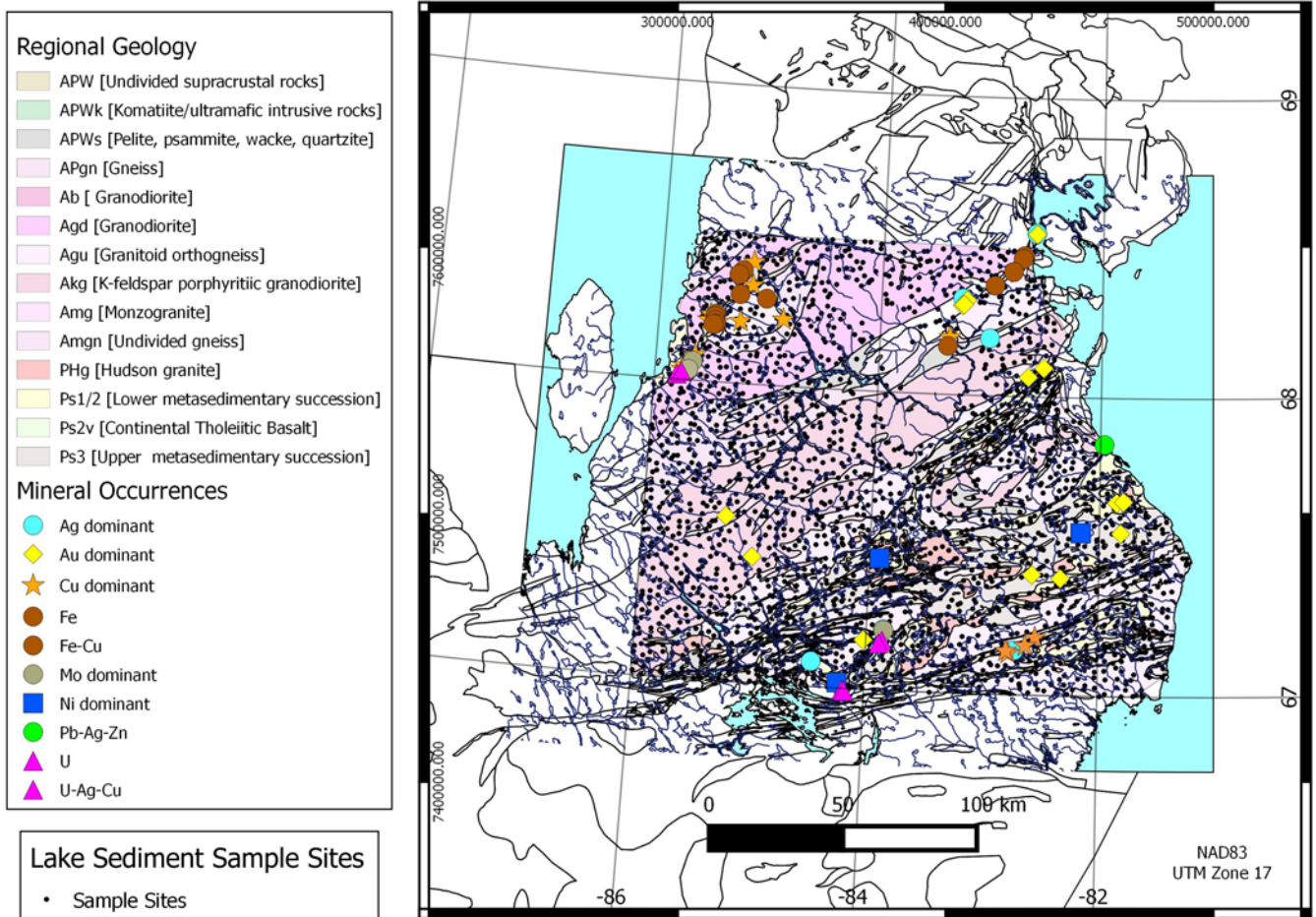
Fig. 3. Mineral occurrences obtained from NUMIN (2017).

discriminant analysis (LDA), logistic regression (LR), neural networks (NN), support vector machines (SVMs), RFs or other machine learning procedures.

A critical part of process validation is the selection of variables that produce an effective classification. This requires the selection of variables that maximize the differences between the various classes
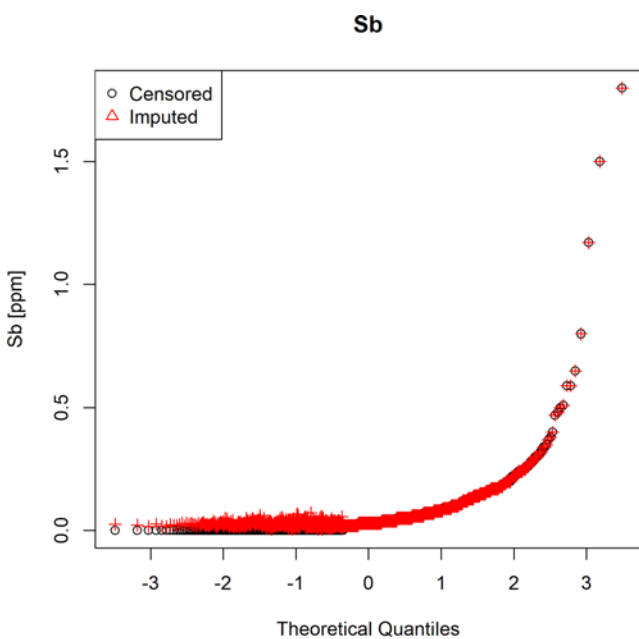

Fig. 4. Sample quantile v. theoretical quantile plot showing the effect on the data distribution of imputation for Sb in lake sediments, Melville Peninsula.
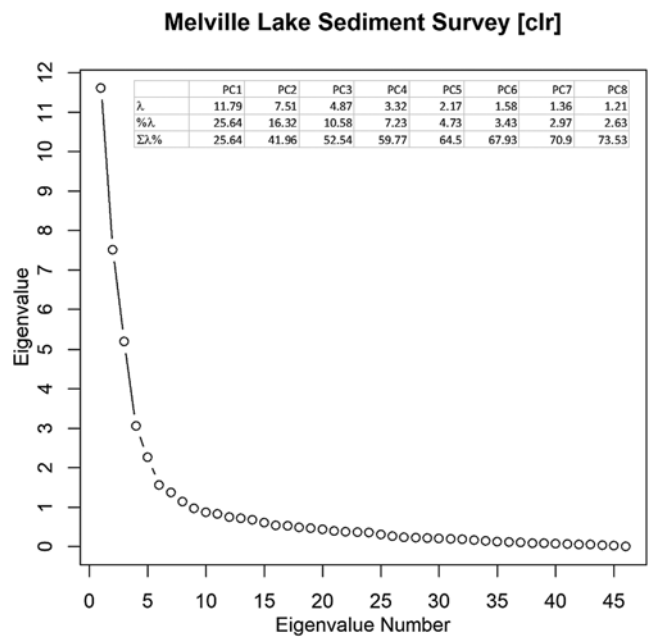
Fig. 5. Screeplot of the eigenvalues and cumulative percentage of eigenvalues derived from a PCA applied to clr-transformed lake sediment geochemical data, Melville Peninsula.
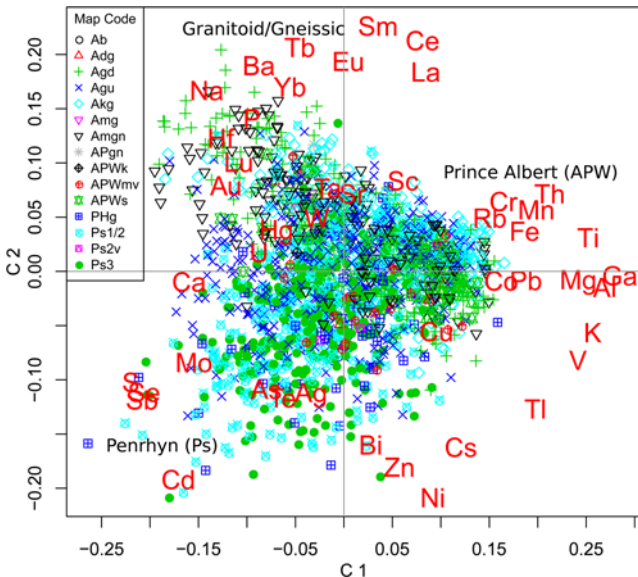
Fig. 6. Principal component biplot of clr-transformed lake sediment geochemical data, Melville Peninsula. See text for an explanation of the element associations.

and minimizes the amount of overlap due to noise, unrecognized or under-sampled processes in the data. As stated previously, because geochemical data are compositional in nature, the variables that are selected for classification require transformation to logratio coordinates. For the purposes of classification, the choice of coordinate system based on the additive logratio (alr) or the isometric logratio (ilr) yields the same results. Although the covariance matrix of clr-transformed data is singular, classification can still be undertaken using a generalized inverse and yields the same classification results as the alr and ilr transforms. Analysis of variance (ANOVA) applied to clr-transformed data enables the recognition of the compositional variables (elements) that are most effective at distinguishing between the classes. Choosing an effective alr transform (choice of suitable denominator) or balances for the ilr transform can be challenging and requires some knowledge and insight about the nature of the processes being investigated. ANOVA applied to the PCs derived from the clr transform has been shown to be highly effective at discriminating between the different classes (Grunsky *et al.* 2014 – Melville Peninsula; Grunsky *et al.* 2017 – Australia). Because the dominant PCs (PC1, PC2, …) commonly identify active processes, as discussed above, and the lesser components (PCn, PCn-1, …, where n is the number of variables) may reflect under-sampled processes or noise, the use of the dominant components can be effectively used for classification using only a few variables. Classification results can be expressed as direct class assignment or posterior probabilities (PPs) in the form of forced class allocation, or as class typicality. Forced class allocation assigns a PP based on the shortest Mahalanobis distance of a compositional observation from the compositional centroid of each class. Class typicality measures the Mahalanobis distance from each class and assigns a PP based on the F-distribution (Campbell 1984; Garrett 1990). This latter approach can result in an observation having a zero PP for all classes, indicating that its composition is not similar to any of the compositions defined by the class compositional centroids.

The application of a procedure such as LDA can make use of cross-validation procedures, whereby the classification of the data is
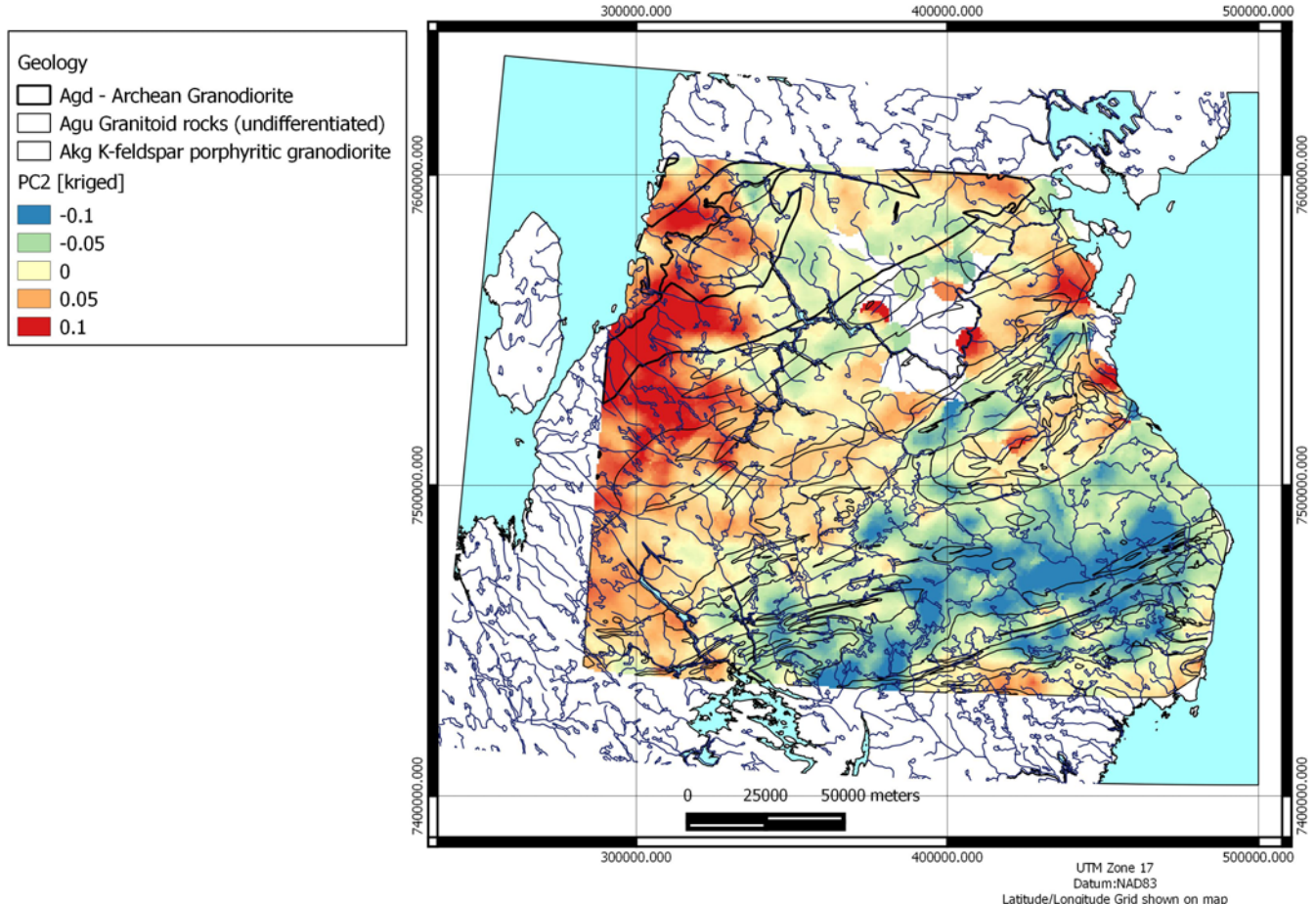

Fig. 7. Kriged map of PC2 derived from clr-transformed lake sediment geochemical data, Melville Peninsula, overlain with mineral occurrences.
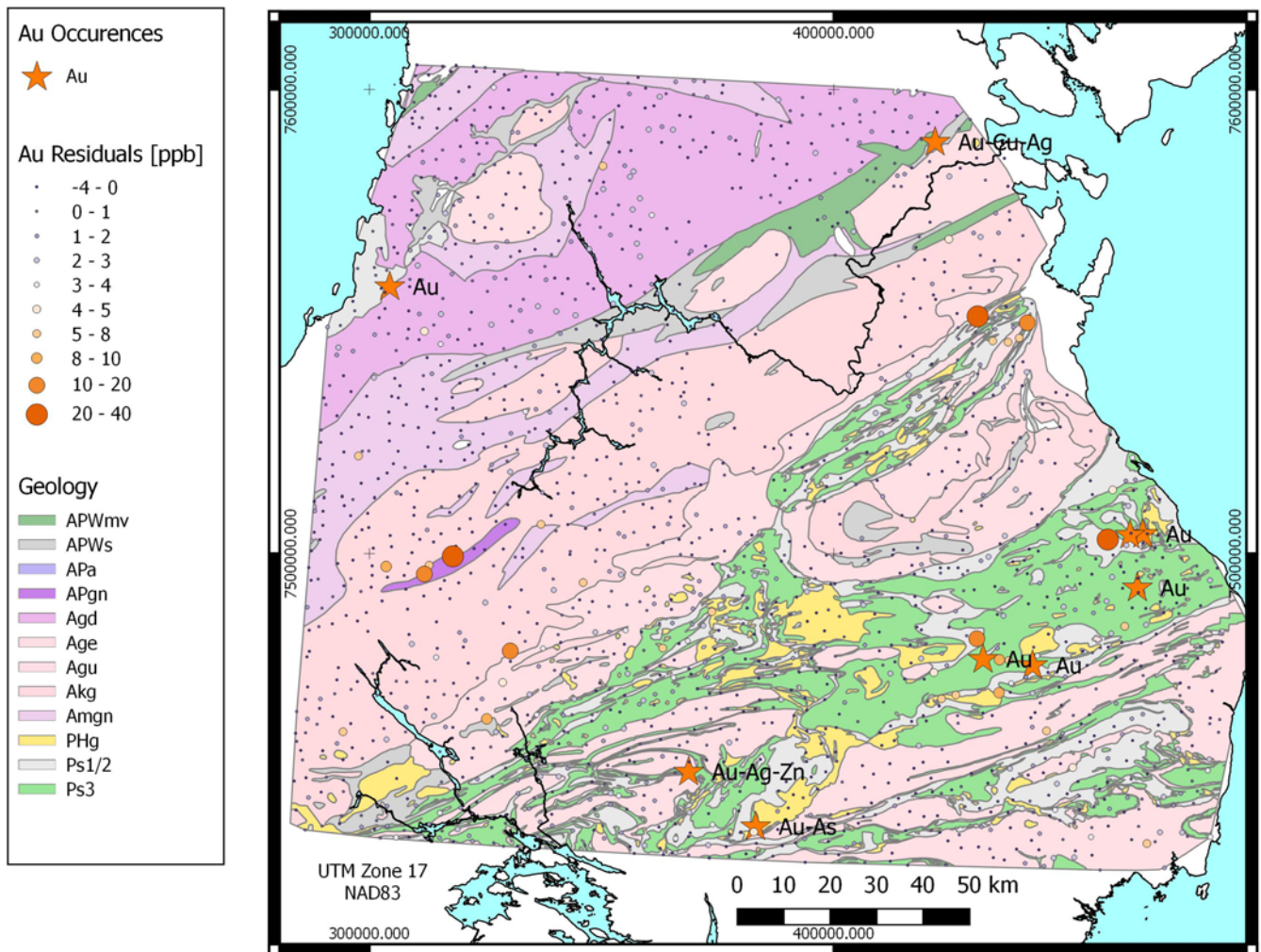
Fig. 8. Map of the residual values of Au (ppb) estimated from a robust linear regression (Au ∼ PC1 + PC2 + PC3 + PC4 + PC5) of lake sediment geochemical data, Melville Peninsula.

repeatedly run based on random partitioning of the data into a number of equal-sized subsamples. One subsample is retained for validation and the remaining subsamples are used as training set. This approach produces stable results and reduces the influence of outliers (Aitchison 1986; Tolosana-Delgado 2006; Pawlowsky-Glahn & Egozcue 2016). However, the subsequent derivation of maps displaying PPs, which are compositions in themselves, requires a suitable logratio transformation to deal with the non-negativity and the constant sum constraint of compositional data. Posterior probabilities are transformed using an alr transform followed by ordinary co-kriging after which a back transformation is carried out for geospatial rendering. It is important to note that the alr transform cannot be used to estimate kriging variance (Aitchison 1986; Tolosana-Delgado 2006). Kriging variance can be estimated by the calculation of the expected value and error variance covariance matrix by Gauss–Hermite integration (Pawlowsky-Glahn & Olea 2004) after which a back transform can be applied.

Classification accuracies can be assessed through the generation of tables that show the accuracy and errors measured from the estimated classes against the initial classes in the training sets used for the classification.

Geospatial coherence

The results from the classification of samples gathered in a geochemical survey should bear a geospatial resemblance to the area sampled. The creation of maps is part of the process validation

procedure. If a geospatial rendering of a posterior probability shows no spatial coherence (i.e. no structure, or a significant amount of 'noise'), then it is likely that the classification will be difficult to interpret within a geological context. The most effective way to test this is through the generation and modelling of semi-variograms that describe the spatial continuity of a specific class based on PPs. If meaningful semi-variograms can be created, then geospatial maps of PPs can be generated through interpolation using the kriging process. Maps of PPs may show low overall values but still be spatially coherent. This is also reflected in the classification accuracy matrix that indicates the extent of classification overlap between classes. Geospatial analysis methodology described by Bivand *et al.* (2013) and the 'gstat' package (Pebesma 2004) in R can be used to generate the geostatistical parameters and images of the PCs and PPs from kriging.

Two case studies

Melville Peninsula, Nunavut, Canada

Process discovery and validation

The Melville Peninsula region, Nunavut, has been the focus of geological mapping and lake sediment and till geochemical sampling for the past 40 years. The example presented here highlights the value of multi-element geochemical data as an aid to regional geological mapping and exploration targeting for potential base and precious metal deposits through the evaluation of regional
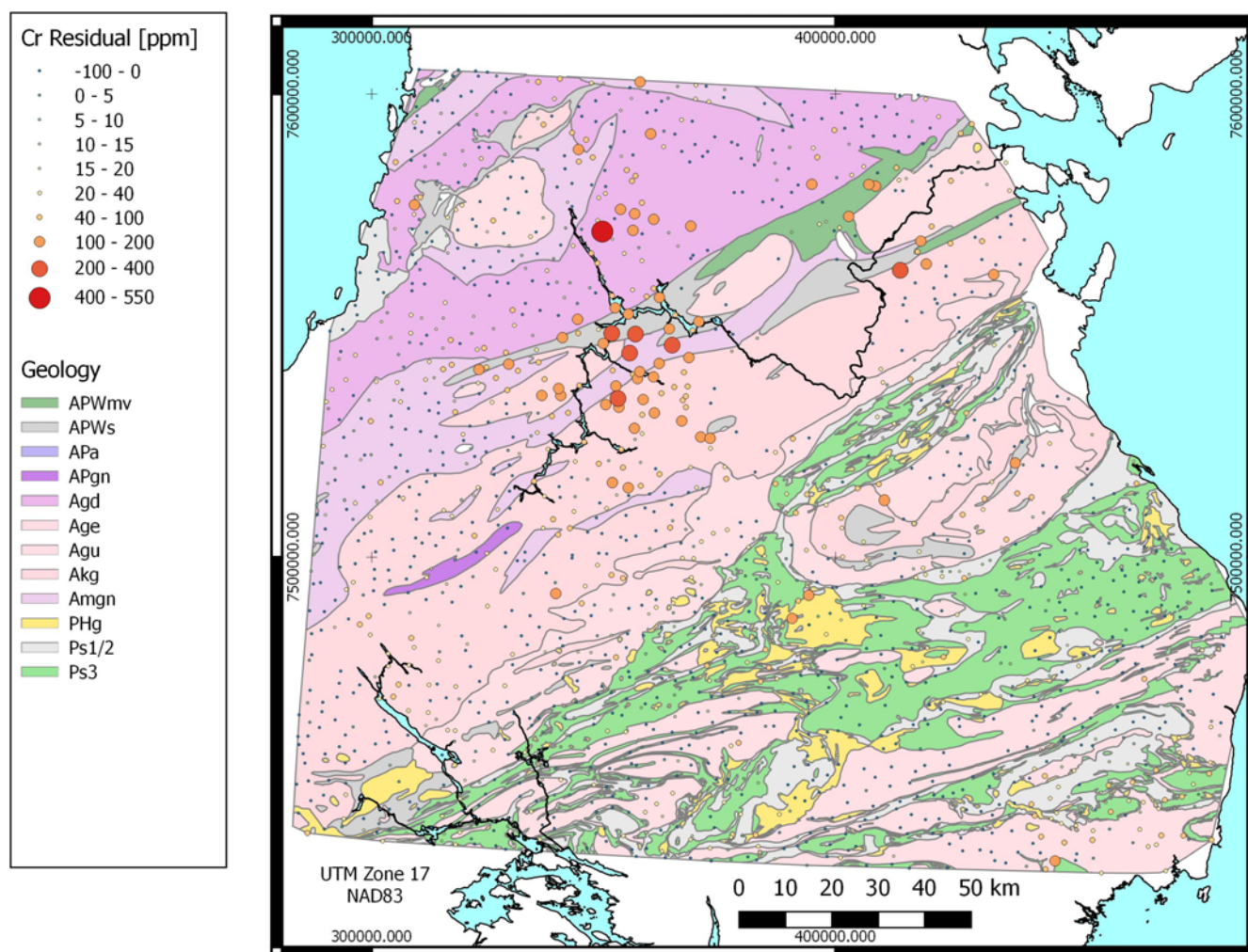
Fig. 9. Map of the residual values of Cr (ppm) estimated from a robust linear regression (Cr ∼ PC1 + PC2 + PC3 + PC4 + PC5) of lake sediment geochemical data, Melville Peninsula.

geochemical survey data in the Melville Peninsula area (Figs 1 and 2). Recent work by Grunsky et al. (2014); Harris & Grunsky, (2015) and Mueller & Grunsky, (2016) has evaluated the lake sediment and till geochemistry in the context of predictive geological mapping and mineral resource potential. Figure 2 shows a generalized geological map of the area and Figure 3 shows the principal mineral occurrences of the area. A study in the use of till geochemical data for predictive geologic mapping using multivariate spatial analysis is summarized by Mueller & Grunsky, (2016) and is not discussed here for the sake of brevity.

The geology is comprised of poly-deformed and poly-metamorphosed Archean and Paleoproterozoic assemblages (Machado et al. 2011, 2012; Corrigan et al. 2013; Grunsky et al. 2014). The area was covered by the Laurentide ice sheet during the Foxe glaciation. Sandy till covers much of the northern part of the Melville Peninsula. The central part of the area was covered by a cold-based ice cap that preserved much of the pre-glacial landscape, which is composed of weathered regolith and boulder rubble with only local glacial transport (Dredge 2009; Tremblay & Paulen 2012). According to Dredge (2009, p. 6):

Glacially scoured lake basins and classic glacial erosion forms are absent. Apart from a few scattered outcrops, the southern plateau surface consists of weathered regolith, or bouldery rubble that was glacially transported for short distances. The main glacial landforms are distinctive subglacial and ice marginal channels associated with wasting phases of the ice

sheet. The till on most of the plateau is immature, and the matrix tends to be sandy.

The lake sediments are the result of reworking and sorting of the glacial till that developed during the retreat of the ice sheet.

The lake sediment geochemical data used in the study of Grunsky et al. (2014) have been published in the Geological Survey of Canada Open File 6269 (Day et al. 2009) based on earlier studies by Hornbrook et al. (1978a, b). Details on the sampling methodology and analytical protocols are documented in Open File 6269. Sample pulps collected in the earlier field campaigns were re-analysed using aqua regia digestion and ICP-MS instrumentation. Pulps were also analysed using INAA. Where elements have been analysed using both methods, the elements were evaluated in terms of detection limit suitability and visual examination of the correlation of the element with each method. This included the evaluation of the degree of censoring. QA/QC protocols and reporting are provided in the reports by Day et al. (2009) and the data were considered adequate for statistical processing. The R statistical package (R Core Team 2014) was used to process the data.

Following the protocols described above, the data were screened for values reported < LLD. Data < LLD were imputed (estimated) using the function 'impKNNa' in the R package 'robCompositions' (Hron et al. 2010). Figure 4 shows a quantile–quantile plot of imputed Sb values that minimizes bias in calculating statistical moments.

After adjusting the censored values to minimize statistical bias, a clr transform was applied to the data. These transformed values were
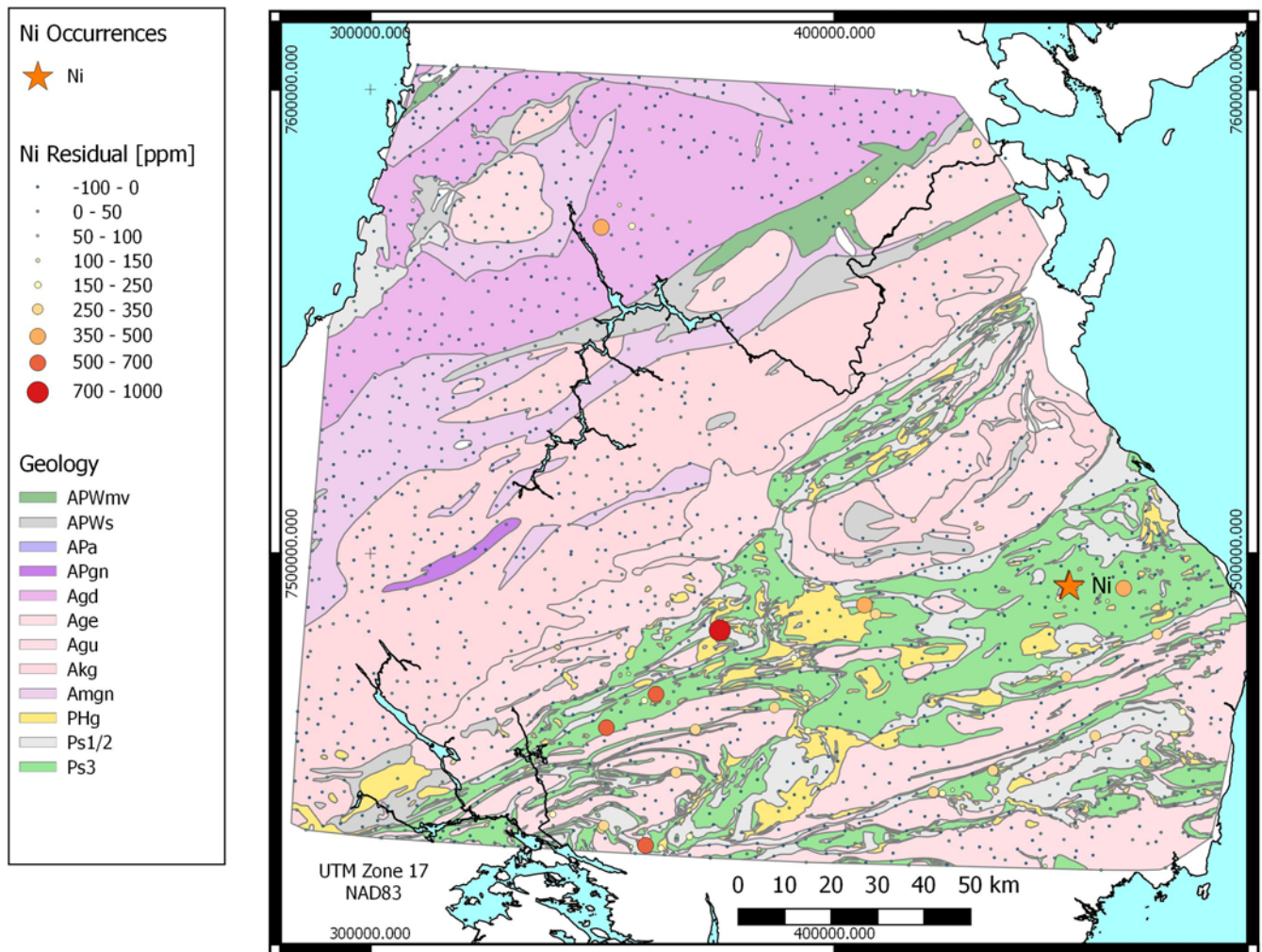
Fig. 10. Map of the residual values of Ni (ppm) estimated from a robust linear regression (Ni ~ PC1 + PC2 + PC3 + PC4 + PC5) of lake sediment geochemical data, Melville Peninsula.

then used to carry out a PCA on the data. A useful tool that is derived from PCA is the screeplot, which is shown in Figure 5. The screeplot shows the eigenvalues plotted in descending order. The figure indicates that eigenvalues decrease rapidly and that most of the variation of the data is accounted for by the first five PCs. The remaining PCs can be interpreted as under-sampled or random processes. The five largest eigenvalues indicate that there is 'structure' in the data that is controlled by mineral stoichiometry and hence, geological processes. The structure in the data can be visualized using a PC biplot (Gabriel 1971). Figure 6 shows a biplot of the first two PCs for the clr-transformed Melville Peninsula lake sediment geochemistry. Three generalized features are evident in this biplot, which accounts for 42% of the variability of the data. First, the plot indicates the relative relationships of the elements (loadings) that highlight the relative affinities of the sample sites and corresponding geological domains. Second, scores of the sample sites associated with granitoid and gneissic rocks occur along the positive PC2 axis. Third, sample sites associated with the Prince Albert Group supracrustal rocks and locally associated granitoid rocks occur along the positive PC1 axis and the sites associated with the Paleoproterozoic Penrhyn supracrustal rocks occur along the negative PC1, negative PC2 axes. Maps of the first and second PCs are shown in Grunsky et al. (2014). The map of PC1 shows a clustering of positive values that correspond with a region of granitoid material and rocks associated with the Prince Albert supracrustal and associated granitoid assemblages. The map of PC2 (Fig. 7) shows positive values associated with granitic and gneissic

rocks in the NW part of the map, and negative scores corresponding with the supracrustal assemblages in the Paleoproterozoic Penrhyn Group in the southeastern part of the map. Thus, as an initial part of process discovery, a PC biplot provides useful information on the geochemical nature and relationships of the data. Grunsky et al. (2014) provide more detail on the use of PCA in this area. There are several other ways that processes can be discovered in geochemical data as outlined previously in the process discovery section.

Mineral exploration targeting

In the example provided here with Melville Peninsula lake sediment geochemical data, the underlying geology was tagged to the sample sites, which are shown in the PC biplot of Figure 6. An analysis of the number of lake sediment sites associated with specific lithologies is summarized in Table 4 of Grunsky et al. (2014). In this case, eight dominant lithologies, derived from the revised geology of Machado et al. (2011, 2012), were tagged to the lake sediment sample sites. As part of the process discovery phase, it is reasonable to test the ability of the lake sediment geochemistry to distinguish between the dominant lithologies. This can be done by applying an ANOVA, in which the most significant PCs provide maximum distinction between the lithologies. Grunsky et al. (2014) demonstrated that PCs derived from clr-transformed geochemical data provide an effective and efficient means to demonstrate discrimination between lithologies. Since the PCs represent linear combinations of elements that are mostly controlled by
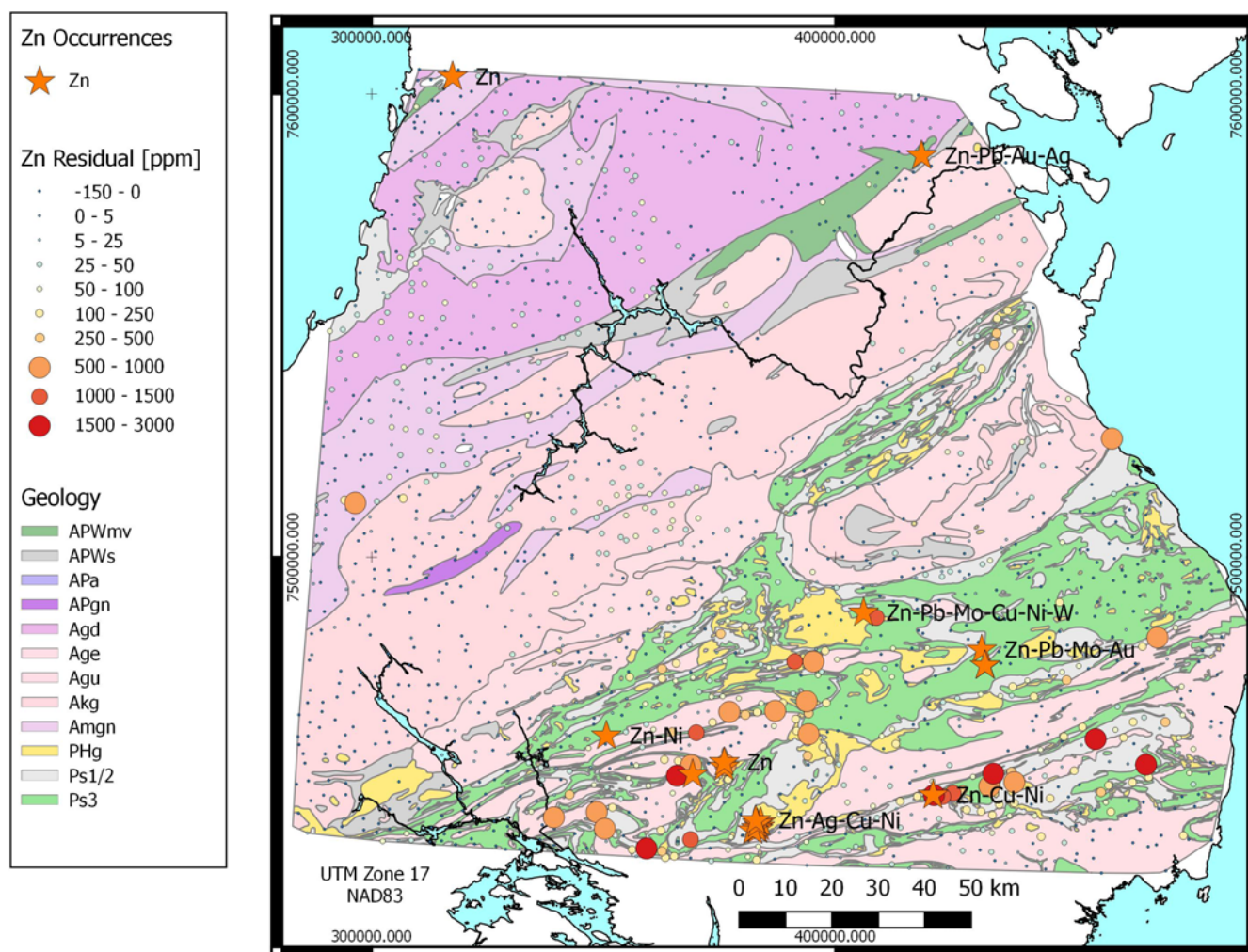
Fig. 11. Map of the residual values of Zn (ppm) estimated from a robust linear regression (Zn ∼ PC1 + PC2 + PC3 + PC4 + PC5) of lake sediment geochemical data, Melville Peninsula.

stoichiometry, more geological information is contained in fewer components; thus, this approach is more parsimonious and effective than using the elements. In the study by Grunsky *et al.* (2014), it was found that the first six PCs accounted for most of the lithological separation of the data. In contrast, almost all of the 44 elements were required to maximize differences between the lithologies.

PCA provides insight into processes controlled by mineral stoichiometry. Sampling strategies for large-scale geochemical surveys are useful for highlighting dominant processes such as the underlying bedrock, but are seldom at a sufficient spatial sampling density for detecting processes that have small spatial footprints, such as veining or mineralization associated with an ore deposit.

The following assumptions are made when considering data from a geochemical survey:

(1) The dominant PCs reflect linear combinations of elements and their relative relationships are controlled by mineral stoichiometry. Typically, rock-forming processes are highlighted in the dominant principal components. These processes were confirmed in the studies of Grunsky *et al.* (2014) and provide the confidence that the lesser components likely represent under-sampled processes.

(2) Under-sampled processes may represent processes such as alteration and mineralization associated with a specific deposit type.

(3) A regression of a commodity of interest (e.g. Zn) against the dominant PCs will reflect the association of that element with the dominant processes. The choice of using the raw values was based

on the assumption that for the sake of regression, the commodity elements are independent. In fact, there is no easy way to create a simple regression of an element in any of the logratio-transformed spaces (alr, clr, ilr). Each transform presents different problems in any analysis involving regression.

(4) High residual values derived from a regression of an element against the dominant PCs may reflect processes that are potentially associated with mineralization.

(5) These high residuals when plotted on a map may highlight areas for mineral exploration follow-up.

The screeplot of Figure 5 indicates that the first five PCs account for 65% of the overall variability in the data. The remaining 41 PCs account for under-sampled or random patterns in the data. On this basis, parametric linear modelling of specific commodity elements, Au, Cu, Ni, Cr and Zn can be applied on the first five PCs. The difference between the observed values of these elements and the estimated values derived from a linear regression define the residuals that may represent under-sampled processes that are possibly associated with mineralization.

Mineral occurrence information was obtained from the Nunavut Mineral Occurrences database (http://nugeo.ca/apps/showing/showQuery.php). Only selected commodities are presented here to demonstrate the methodology of using residual values for resource prediction.

Figures 8–11 show the residuals for linear models applied to the first five PCs derived from the clr-transformed lake sediment
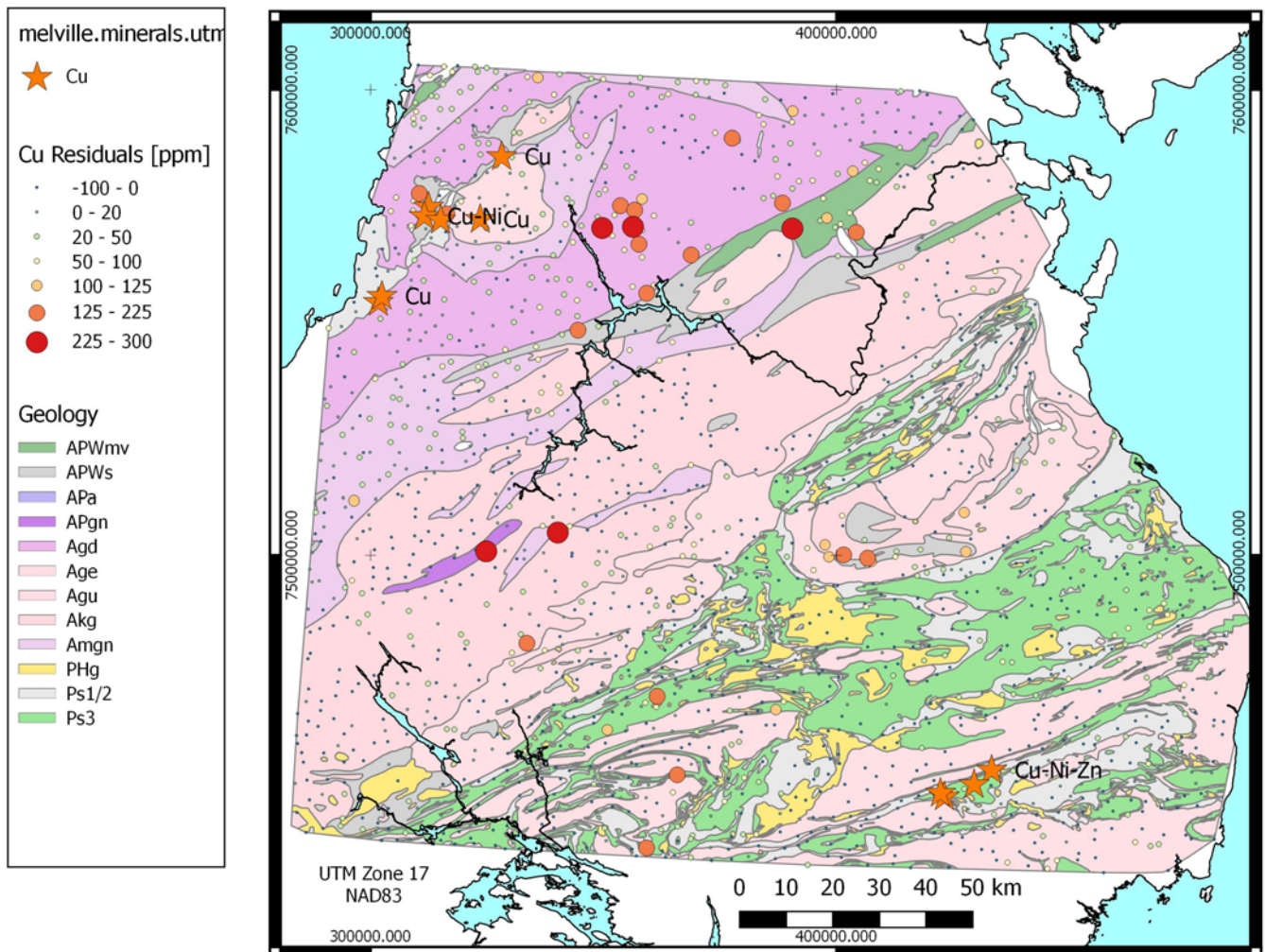
Fig. 12. Map of the residual values of Cu (ppm) estimated from a robust linear regression (Cu ~ PC1 + PC2 + PC3) of lake sediment geochemical data, Melville Peninsula.

geochemical data for untransformed values of Au, Ni, Cr and Zn. The function 'rlm' (robust fitting for linear models) from the MASS library in the R statistical environment (R Core Team 2014; Venables & Ripley 2002) was used. Figure 12 shows the results of modelling Cu against the first three PCs. The choice of three components is based on the observation that most of the variability of Cu is contained within the first three components, and the residuals generated based on linear modelling on the first three components may further highlight areas of potential mineralization.

Figure 8 shows a map of the residual values of Au. Within the Penrhyn Group (Ps1/2, Ps3) of supracrustal rocks in the southeastern part of the map area, four lake sediment sites show high values of residual Au and are associated with local Au occurrences. Other areas of high residual values are located in the northern part and Penrhyn Groups in the eastern part of the area. One isolated elevated residual Au value is located in the western part of the map area within Archean gneisses (APgn, Amgn).

Although there are no known Cr occurrences in the area, Figure 9 shows a map of residual Cr based on the first five PCs. Elevated Cr values appear to be concentrated within the north-central part of the map area possibly associated with the Prince Albert group of volcanics (ultramafic, mafic and felsic), sediments and associated mafic gneissic rocks (APwmv, APWs, APgn). It is unknown if a Cr-steel mill was used in the sample preparation process, which could contaminate the sample material. Alternatively, gravitational effects on heavy minerals may have resulted in local accumulation of Cr-rich minerals.

Figure 10 shows a map of residual Ni values, which are restricted to the Penrhyn Group supracrustal volcanic and sedimentary rocks (Ps1/2, Ps3) within the south-central part of the map area. One Ni is noted in the eastern part of the map area with a few isolated residual Ni values in the range of 150–500 ppm.

High residual values of Zn occur throughout the Penrhyn Group supracrustal assemblage (Ps1/2, Ps3) in the southern part of the map area, shown in Figure 11. These values are closely associated with known Zn occurrences. A few isolated lower residual values of Zn occur in the east-central part of the map area.

The high residual values of Cu (Fig. 12) occur within the western portion of the Penrhyn Group assemblages (Ps1/2, Ps3) in the southern part of the map area. Isolated high residual Cu values occur within the K granitoid rocks in the central part of the map area and the Prince Albert Group supracrustal assemblages (APwmv, APWs) in the east-central part of the area. High residual values also occur in the northern part of the area, associated with K granites, mixed gneiss and sediments associated with the Prince Albert Group.

Thomson region, Australia

Resolution of process discovery with regional geochemical survey data

The southern Thomson Orogen region of northern New South Wales and southern Queensland (Australia) was the subject of two
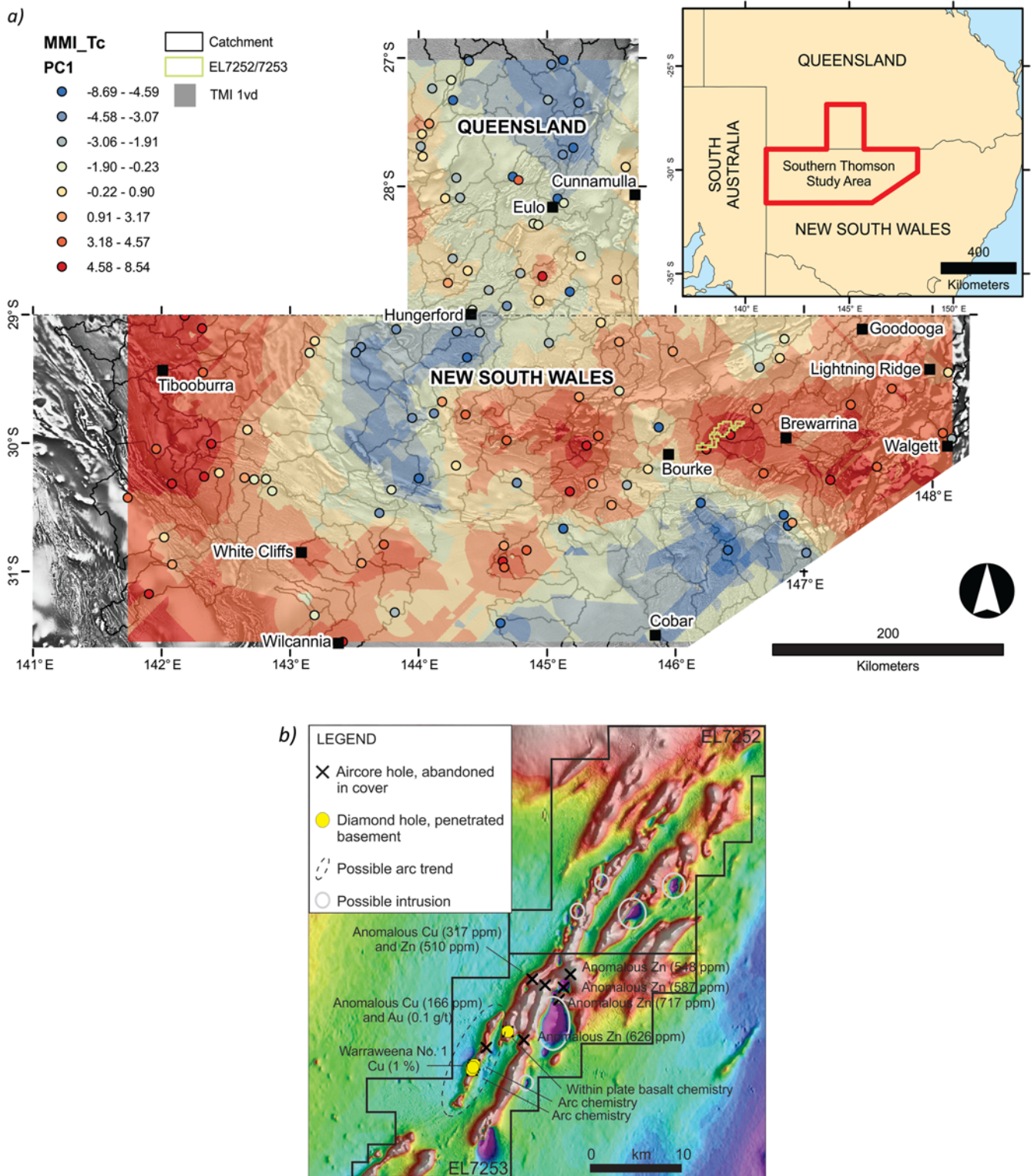
Fig. 13. Map of PC1 of the MMI extraction of coarse (<2 mm) surficial sediments from the southern Thomson survey study area in eastern Australia (**a**). Values at the sampling sites are shown by coloured circles; kriging interpolation is shown as a semi-transparent continuous raster based on the same colour scale (eight equi-percentile classes). First vertical derivative of the total magnetic intensity (TMI 1vd) surface (Nakamura 2015) is underlain (as shaded relief), as are catchment boundaries (as black/grey polygons). The green polygon between Brewarrina and Bourke shows the location of Thomson Resources Ltd tenements EL7252/7253, enlarged in (**b**), which shows the location of exploration drill holes over a TMI image (http://www. thomsonresources.com.au/projects/porphyry-copper-gold). Figure from Caritat *et al.* (2017) published with the permission of the Geological Society of Australia. Map (**b**) reproduced with the permission of the Chief Executive Officer, Thomson Resources Ltd.

adjacent regional geochemical surveys (Caritat & Lech 2007; Main & Caritat 2016), whose results were combined. Here, the thick sedimentary deposits of the Mesozoic Eromanga Basin cover the inferred boundary between the largely unknown Paleozoic

Thomson Orogen to the north and the Paleozoic Lachlan Fold Belt to the SE (which includes the Macquarie Arc, a fertile province for magmatic arc-related mineralization such as porphyry Cu deposits), and the Paleoproterozoic Broken Hill Domain to the SW,
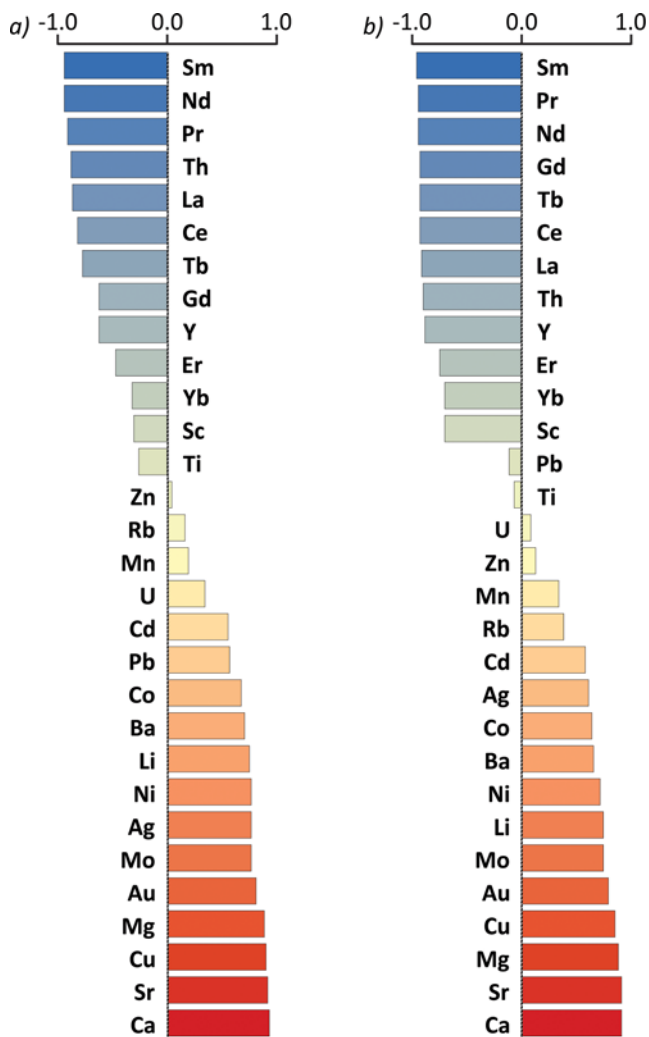
Fig. 14. Graphical representation of the scaled and ordered eigenvalues of PC1 for 30 elements analysed after MMI extraction in both (**a**) the regional (southern Thomson) and (**b**) the continental (NGSA) scale studies. Figure from Caritat *et al.* (2017) published with the permission of the Geological Society of Australia.

(which is well known for a multitude of mineral deposits but most significantly its world-class base metal endowment) (Fig. 13a).

One hundred and twenty top outlet sediments (0–25 cm), similar to floodplain or overbank sediments but with a potential contribution of aeolian material in places, were sampled by the regional surveys over a combined area of 185 000 km^2 (yielding an average site density of 1 sample per 1540 km^2) and analysed by ICP-MS for multi-element geochemistry after weak extraction using the Mobile Metal Ion® (MMI) procedure (Mann 2010; SGS 2017), among other analyses. MMI extraction was designed to target loosely bound elements adsorbed on the surfaces of soil particles, oxyhydroxide coatings and organic matter; thereby potentially representing elements that moved up in the regolith post mineralization (Mann 2010). PCA carried out after imputation of censored values and clr transformation of the data for 30 elements indicated that the variance explained by PC1 was 49.3% (Main & Caritat 2016). The composition of PC1 was dominated by positive loadings of Ca, Sr, Cu, Mg, Au and Mo, and negative loadings of the rare earth elements (REEs) and Th (Fig. 14a). This composition was interpreted to dominantly reflect regolith processes (calcrete formation: Ca, Sr, Mg; clay minerals: REEs; and accumulation of resistate minerals: REEs) with a potential overprint of mineralization expressed by the Cu, Au and Mo components.

A local exploration company also using soil MMI geochemistry had coincidentally and independently developed an empirical (not statistically derived) vector to mineralization consisting of an enrichment in Sr, Ca and Au concomitant with a depletion in REEs to successfully site drill holes targeting Cu–Au mineralization in the region (J. Macauley, pers. comm. 2013). Thus, it was postulated that the regional map of PC1 from the Thomson regional geochemical survey could serve as a first order mineral potential map for porphyry Cu–Au mineralization (Main & Caritat 2016). This could potentially be extended to other Cu–Au mineralization types as well. The implications of these findings are that several areas within the southern Thomson region were identified to potentially be of interest for exploration purposes at the regional scale: an area between Brewarrina and Bourke and its eastward continuation towards Walgett and spreading north and south; four regions in the centre of the study area (one *c.* 150 km NW of Cobar, one *c.* 50–150 km west of Bourke, one *c.* 40 km NE of Hungerford, and one *c.* 40 km NW of Eulo); a region *c.* 80 km east of White Cliffs; around Wilcannia; and in the northwestern part of the area stretching from the Koonenberry Belt to Tibooburra.

A test for this concept emerged when another local exploration company reported some historical as well as new drill hole intersections of sub-economic Pb–Zn–Cu–Au mineralization at Warraweena, between Brewarrina and Bourke in northern New South Wales (tenements EL7252/7253) (http://www.thomsonresources.com.au/projects/porphyry-copper-gold). Burton *et al.* (2008) attributed the mineralization at Warraweena to porphyry-type deposits related to volcanic arcs. The boreholes on tenements EL7252/7253 were sited not based on geochemistry but on regional and detailed airborne magnetic geophysical surveys (Fig. 13b). These tenements happened to be in the centre of a circular anomalous area of high positive PC1 values in the regional MMI survey. This was taken as providing supporting evidence for the prospectivity model developed above (Main & Caritat 2016).

Upscaling from regional to continental scale

Next, Caritat *et al.* (2017) tested whether the continental scale National Geochemical Survey of Australia (NGSA) data, which also included MMI data, would yield similar PCA results, and if so, if the patterns would be potentially useful for prioritizing mineral exploration in certain regions. The number of NGSA sites was *c.* 1200 and the area was >6 million km^2, yielding an average density of one site per 5200 km^2. The PCA conducted on the same suite of 30 elements and following the same procedure (imputation, clr transformation) yielded as PC1 (53.3% of variance) the association of Ca–Sr–Mg–Cu–Au–Mo at the positive end, and of REEs–Th at the negative end (Fig. 14b). This PC composition was noted to be strikingly similar to that obtained for the smaller southern Thomson regional survey (Caritat *et al.* 2017). The resulting national scale patterns of PC1 are shown in Figure 15. The implications of this research (Caritat *et al.* 2017) were that (1) the dominant element associations determined by multivariate statistics at the regional scale are robust enough to also emerge at much lower density continental scale surveys and vice-versa; and (2) several provinces potentially prospective for Cu–Au-base metal mineralization could be determined by the map of PC1 from the NGSA dataset (Pilbara-Capricorn-northern Yilgarn, Albany-Fraser, western Eucla, eastern Eucla, western Amadeus, Victoria, Adelaide, Eromanga, southern Georgina-Isa, Curnamona-western Murray and southern Surat geological regions). This was the first study (to our knowledge) of upscaling a methodology for mineral prospectivity analysis based on the statistical analysis of surface geochemical surveys from the regional scale to the continental scale.
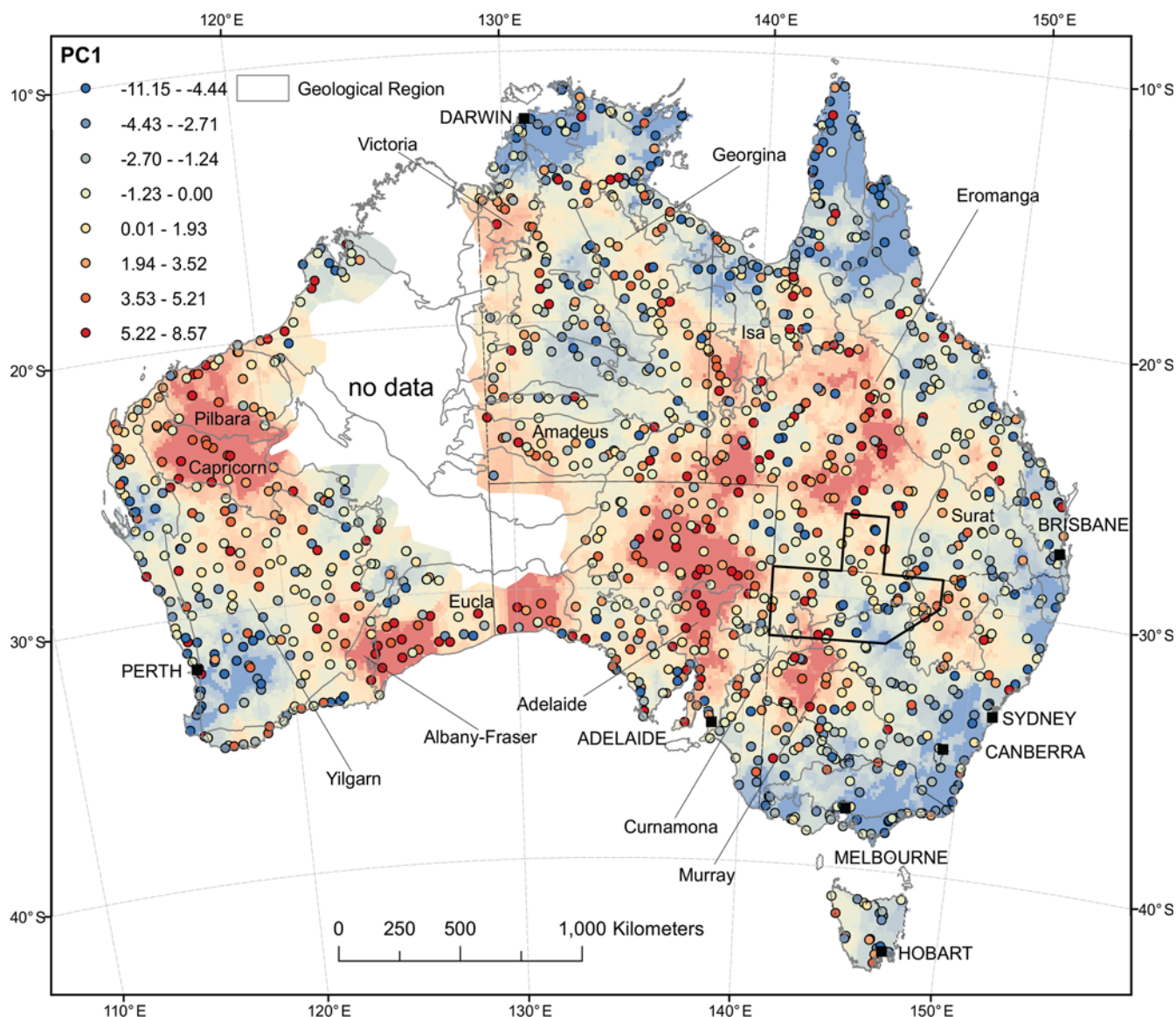
Fig. 15. Map of PC1 of the MMI extraction of coarse (<2 mm) surficial sediments from the NGSA project study area. Values at the sampling sites are shown by coloured circles; kriging interpolation is shown as a continuous raster based on the same colour scale (eight equi-percentile classes). Geological regions (Blake & Kilgour 1998) are overlain (as grey polygons); those discussed in the text are labelled. The southern Thomson project area (Fig. 13a) is shown by the black polygon. Figure from Caritat *et al.* (2017) published with the permission of the Geological Society of Australia.

Concluding remarks

This paper outlines a systematic approach by which regional and exploration scale geochemical survey data can be evaluated for the purposes of discovering geological/geochemical processes that define underlying geology and features associated with mineral systems. As a contribution to Exploration '17, this paper summarizes advances in the use of statistical and geospatial methods since the last review of geochemical methods at Exploration '07 (Grunsky 2010). The concept of process discovery facilitates the construction of geological process models that assist in identifying the dominant processes, which, in turn, assist in unmasking and 'discovering' under-sampled or 'rare-event' processes associated with mineral deposits. The methodology of applying the consecutive steps of process discovery followed by process validation provides a systematic, transparent, defensible and reproducible way for extracting useful information from a range of data that represent geological/geochemical processes. There are many methods available for both process discovery and process validation. In this paper we have highlighted only a few of these methods. Numerous

references cited in this paper provide further detail on the use and application of other methods.

The examples presented in this contribution are based on weak or partial digestions in which many of the silicate minerals are not decomposed. Despite this limitation, the results presented here demonstrate that both the underlying lithologies and processes related to alteration and mineralization can be detected using multivariate statistical methods.

Geological mapping and mineral exploration programs usually have access to a wealth of digital data that, when used effectively, provide an enormous amount of information from which mineral exploration and mapping models can be constructed. In the authors' view, the use of high quality multivariate and geospatial methods through the application of modern, advanced data analytics, at detailed and regional scales, is the next step forward in finding the greenfields mineral provinces of the future.

References

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, New York, 416.

Aitchison, J. 1999. Logratios and natural laws in compositional data analysis. *Mathematical Geology*, **31**, 563–580, https://doi.org/10.1023/A:1007568008032

Bivand, R., Pebesma, E. & Gomez-Rubio, V. 2013. *Applied Spatial Data Analysis with R*. 2nd edn. Springer, New York, 405.

Blake, D. & Kilgour, B. 1998. *Geological Regions of Australia 1:5,000,000 scale (Data Set)*. Geoscience Australia, Canberra, https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search;jsessionid=3BE16A8B235E117FC9F5448B361B954E#/metadata/32366

Buccianti, A. & Grunsky, E.C. 2014. Compositional data analysis in geochemistry: are we sure to see what really occurs during natural processes? *Journal of Geochemical Exploration*, **141**, 1–5, https://doi.org/10.1016/j.gexplo.2014.03.022

Buccianti, A., Mateu-Figueras, G. & Pawlowsky-Glahn, V. (eds) 2006. *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society, London, Special Publications, **264**, 212, https://doi.org/10.1144/GSL.SP.2006.264

Burton, G.R., Dadd, K.A. & Vickery, N.M. 2008. Volcanic arc-type rocks beneath cover 35 km to the northeast of Bourke. *Quarterly Notes of the Geological Survey of New South Wales*, **127**, 1–23, https://digsopen.minerals.nsw.gov.au/digsopen/

Campbell, N.A. 1984. Some aspects of allocation and discrimination. *In*: Howells, W.W. & van Vark, G.N. (eds) *Multivariate Statistical Methods in Physical Anthropology*. D. Reidel, Dordrecht, 177–192.

Campbell, N.A. 1986. *A General Introduction to a Suite of Multivariate Programs*. CSIRO Division of Mathematics and Statistics, unpublished report.

Caritat, P.de & Cooper, M. 2011. *National Geochemical Survey of Australia: The Geochemical Atlas of Australia*. Geoscience Australia Record, **2011/20**, 557 p (2 Volumes), http://www.ga.gov.au/metadata-gateway/metadata/record/gcat_71973

Caritat, P.de & Grunsky, E.C. 2013. Defining element associations and inferring geological processes from total element concentrations in Australian catchment outlet sediments: multivariate analysis of continental-scale geochemical data. *Applied Geochemistry*, **33**, 104–126, https://doi.org/10.1016/j.apgeochem.2013.02.005

Caritat, P.de & Lech, M.E. 2007. *Thomson Region Geochemical Survey, Northwestern New South Wales*. Cooperative Research Centre for Landscape Environments and Mineral Exploration Open File (**Report, 145**), 1021 p+CD-ROM. Available at http://crcleme.org.au/Pubs/OFRSindex.html

Caritat, P.de, Bloch, J. & Hutcheon, I. 1994. LPNORM: a linear programming normative analysis code. *Computers & Geosciences*, **20**, 313–347, https://doi.org/10.1016/0098-3004(94)90045-0

Caritat, P.de, Main, P.T., Grunsky, E.C. & Mann, A. 2017. Recognition of geochemical footprints of mineral systems in the regolith at regional to continental scales. *Australian Journal of Earth Sciences*, **64**, 1033–1043, https://doi.org/10.1080/08120099.2017.1259184

Chung, C.F. 1985. *Statistical treatment of geochemical data with observations below the detection limit*. Geological Survey of Canada, Current Research, Part B, Paper **85-1B**: 141–150.

Corrigan, D., Nadeau, L. *et al.* 2013. *Overview of the GEM Multiple Metals – Melville Peninsula project, central Melville Peninsula, Nunavut*. Geological Survey of Canada, Current Research, **2013–19**, 17.

Day, S.J.A., McCurdy, M.W., Friske, P.W.B., McNeil, R.J., Hornbrook, E.H.W. & Lynch, J.J. 2009. *Regional lake sediment and water geochemical data, Melville Peninsula, Nunavut (parts of NTS 046N, O, P, 047A and B)*. Geological Survey of Canada, Open File, **6269**, 12.

Dredge, L.A. 2009. *Till geochemistry, and results of lithological, carbonate and textural analyses, northern Melville Peninsula, Nunavut (NTS 47A,B,C,D)*. Geological Survey of Canada, Open File, **6285**, 11.

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

Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453–467, https://doi.org/10.1093/biomet/58.3.453

Garrett, R.G. 1990. A robust multivariate allocation procedure with applications to geochemical data. *In*: Agterberg, F.P. & Bonham-Carter, G.F. (eds) *Proceedings of Colloquium on Statistical Applications in the Earth Sciences*. Geological Survey of Canada Paper, Ottawa, Ontario, **89–9**, 309–318.

Geological Survey of Northern Ireland (GSNI) 2007. *Tellus Project Overview*, http://www.bgs.ac.uk/gsni/tellus/index.html

Grunsky, E.C. 2010. The interpretation of geochemical survey data. *Geochemistry, Exploration, Environment Analysis*, **10**, 27–74, https://doi.org/10.1144/1467-7873/09-210

Grunsky, E.C. 2013. Predicting Archean volcanogenic massive sulfide deposit potential from lithogeochemistry: application to the Abitibi Greenstone Belt. *Geochemistry: Exploration, Environment, Analysis*, **13**, 317–336, https://doi.org/10.1144/geochem2012-176

Grunsky, E.C. & Caritat, P.de 2017. *Advances in the use of geochemical data for mineral exploration*. Exploration 2017, Proceedings, October, 2017, Toronto.

Grunsky, E.C., Corrigan, D., Mueller, U. & Bonham-Carter, G.-F. 2012. *Predictive geologic mapping using lake sediment geochemistry in the Melville Peninsula*. Geological Survey of Canada, Open File, 7171, 1 sheet.

Grunsky, E.C., Mueller, U.A. & Corrigan, D. 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: applications for predictive geological mapping. *Journal of Geochemical Exploration*, **141**, 15–41, https://doi.org/10.1016/j.gexplo.2013.07.013

Grunsky, E.C., Caritat, P.de & Mueller, U.A. 2017. Using surface regolith geochemistry to map the major crustal blocks of the Australian continent. *Gondwana Research*, **46**, 227–239, https://doi.org/10.1016/j.gr.2017.02.011

Harris, J.R. & Grunsky, E.C. 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers & Geosciences*, **80**, 9–25, https://doi.org/10.1016/j.cageo.2015.03.013

Hornbrook, E.H.W. & Lynch, J.J. 1978*a*. National geochemical reconnaissance release NGR 33A-1977, regional lake sediment and water geochemical reconnaissance data, Melville Peninsula, Northwest Territories, Geological Survey of Canada, Open File **522A**, 1978.

Hornbrook, E.H.W. & Lynch, J.J. 1978*b*. National geochemical reconnaissance release NGR 33B-1977, regional lake sediment and water geochemical reconnaissance data, Melville Peninsula, Northwest Territories, Geological Survey of Canada, Open File **522B**, 1978; 14 sheets.

Hron, K., Templ, M. & Filzmoser, P. 2010. Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics and Data Analysis*, **54**, 3095–3107, https://doi.org/10.1016/j.csda.2009.11.023

Machado, G., Houlé, M.G., Richan, L., Rigg, J. & Corrigan, D. 2011. *Geology, southern part of Prince Albert Greenstone Belt, Prince Albert Hills, Melville Peninsula, Nunavut*. Geological Survey of Canada, Canadian Geoscience Map, **44** (preliminary), 1 sheet, 1 CD-ROM.

Machado, G., Corrigan, D., Nadeau, L. & Rigg, J. 2012. *Geology, northern part of Prince Albert Greenstone Belt, Prince Albert Hills, Melville Peninsula, Nunavut*. Geological Survey of Canada, Canadian Geoscience Map, **81** (preliminary), scale 1:25 000.

Main, P. & Caritat, P.de 2016. *Geochemical Survey of the Southern Thomson Orogen, Southwestern Queensland and Northwestern New South Wales – The Chemical Composition of Surface and Near-Surface Catchment Outlet Sediments*. Geoscience Australia Record, **2016/11**, 136. Available at http://www.ga.gov.au/metadata-gateway/metadata/record/83230

Mann, A.W. 2010. Strong versus weak digestions: ligand-based soil extraction geochemistry. *Geochemistry: Exploration, Environment, Analysis*, **10**, 17–26, https://doi.org/10.1144/1467-7873/09-216

Martín-Fernández, J.A., Barceló-Vidal, C. & Pawlowsky-Glahn, V. 2003. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, **35**, 253–278, https://doi.org/10.1023/A:1023866030544

Mueller, U.A. & Grunsky, E.C. 2016. Multivariate spatial analysis of lake sediment geochemical data; Melville Peninsula, Nunavut, Canada. *Applied Geochemistry*, **75**, 247–262, https://doi.org/10.1016/j.apgeochem.2016.02.007

Nakamura, A. 2015. *Magnetic Anomaly Map of Australia (Data Set)*. Geoscience Australia, Canberra. Available at http://www.geoscience.gov.au/gadds

NUMIN 2017. *Nunavut Minerals*, http://www.nunavutgeoscience.ca/pages/en/numin.html

Palarea-Albaladejo, J. & Martín-Fernández, J.A. 2008. A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. *Computers & Geosciences*, **34**, 902–917, https://doi.org/10.1016/j.cageo.2007.09.015

Palarea-Albaladejo, J., Martín-Fernández, J.A. & Buccianti, A. 2014. Compositional methods for estimating elemental concentrations below the limit of detection in practice using R. *Journal of Geochemical Exploration*, **141**, 71–77, https://doi.org/10.1016/j.gexplo.2013.09.003

Pawlowsky-Glahn, V. & Buccianti, A. (eds) 2011. *Compositional Data Analysis, Theory and Applications*. Wiley & Sons, New York, 378.

Pawlowsky-Glahn, V. & Egozcue, J.J. 2016. Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*, **164**, 28–32, https://doi.org/10.1016/j.gexplo.2015.12.010

Pawlowsky-Glahn, V. & Olea, R.A. 2004. *Geostatistical Analysis of Compositional Data*. Studies in Mathematical Geology. Oxford University Press, New York, **7**, 181.

Pearce, T.H. 1968. A contribution to the theory of variation diagrams. *Contributions to Mineralogy and Petrology*, **19**, 142–157, https://doi.org/10.1007/BF00635485

Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, **30**, 683–691, https://doi.org/10.1016/j.cageo.2004.03.012

R Core Team 2014. *The R Project for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/

Reimann, C., Matschullat, J., Birke, M. & Salminen, R. 2009. Arsenic distribution in the environment: the effects of scale. *Applied Geochemistry*, **24**, 1147–1167, https://doi.org/10.1016/j.apgeochem.2009.03.013

Reimann, C., Matschullat, J., Birke, M. & Salminen, R. 2010. Antimony in the environment: lessons from geochemical mapping. *Applied Geochemistry*, **25**, 175–198, https://doi.org/10.1016/j.apgeochem.2009.11.011

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds) 2014. Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. Geologische Jahrbuch, **B102**, 528 p, 1 DVD.

Sanford, R.F., Pierson, C.T. & Crovelli, R.A. 1993. An objective replacement method for censored geochemical data. *Mathematical Geology*, **25**, 59–80, https://doi.org/10.1007/BF00890676

SGS 2017. *Mobile Metal Ions (MMI)*, http://www.sgs.com/en/mining/exploration-services/geochemistry/mobile-metal-ions-mmi

Smith, D.B., Cannon, W.F. & Woodruff, L.G. 2011. A national-scale geochemical and mineralogical survey of soils of the conterminous United States. *Applied Geochemistry*, **26**, 250–255, https://doi.org/10.1016/j.apgeochem.2011.03.116

Tolosana-Delgado, R. 2006. *Geostatistics for constrained variables: positive data, compositions and probabilities. Applications to environmental hazard monitoring*. PhD thesis, University of Girona, 215.

Tremblay, T. & Paulen, R. 2012. *Glacial geomorphology and till geochemistry of central Melville Peninsula, Nunavut*. Geological Survey of Canada, Open File, **7115**.

Venables, W.N. & Ripley, B.D. 2002. *Modern Applied Statistics with S*. 4th edn. Springer, Berlin, 495.