Mineral-Resource Prediction Using Advanced Data Analytics and Machine Learning of the QUEST-South Stream-sediment Geochemical Data, Southwestern British Columbia (Parts of NTS 082, 092)

E.C. Grunsky, Dept. Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada, egrunsky@gmail.com

D.C. Arne, Telemark Geosciences, Yackandandah, Victoria, Australia, dennis.arne@telemarkgeosciences.com

Geoscience BC Report 2020-06

April 30, 2020

Table of Contents

Summary

Geoscience BC conducted an additional stream-sediment sampling program in the QUEST-South project area in 2009 and reanalysed older archived regional geochemical survey (RGS) samples using ICP-MS in 2010. Catchments were later determined for these samples and a preliminary interpretation of the geochemical data undertaken using the dominant rock type in the catchments to level the data for the effects of variable background. In this Geoscience BC project, we apply multivariate statistical methods, including the random forests classification method, to interpret the data from 8545 stream-sediment samples. Data for 35 elements were levelled for laboratory analytical effects and values below the lower limit of detection were replaced with imputed values prior to a centred logratio transformation to moderate the effects of geochemical closure, a unique feature of geochemical data that requires all data to sum to 100%. Principal components were derived from the data using correlations between elements to reduce the number of variables required to enhance the geochemical signals associated with a variety of mineral deposits. Each sample was attributed with the closest MINFILE occurrence provided that it was within 2.5 km of the sample site. MINFILE occurrences were grouped based on similarities in British Columbia Geological Survey (BCGS) mineral deposit models (GroupModels) and geochemical signatures for statistical training purposes. A training data set of 474 samples, including 100 samples not attributed with a MINFILE occurrence and the most significant principal components, was used to generate random forests prediction models from which posterior probabilities were estimated for the remaining 8071 samples.

Two different dimensionality reduction approaches prior to the random forests procedure have been tested: principal component analysis (PCA) and t-distributed stochastic neighbour embedding using 9 dimensions (t-SNE9). The predicted posterior probabilities for the GroupModels have been used to generate kriged images to test for geospatial coherence within the predictions. These have been compared to kriged images for porphyry Cu-Au-Mo deposits generated using raw data and data corrected for the effects of catchment bedrock type using more conventional data analysis approaches. In the case of the random forests prediction of porphyry Cu-Au-Mo deposits, there are many more predicted sites than MINFILE records. The t-SNE9 posterior probabilities provide a better visual fit to the distribution of known mineral occurrences and have a slightly higher level of accuracy compared to the posterior probabilities obtained using PCA. Catchment polygons were thematically coded using the t-SNE9 posterior probabilities to create maps of exploration potential for 13 of the GroupModels.

The use of random forests provides predictions of mineral occurrences ranging from 0 to 60%. The number of training sites for each GroupModel has an influence on the prediction accuracy. Additionally, several of the GroupModels have similar geochemical compositions, which results in significant overlap in the prediction accuracy. The geospatial continuity of the GroupModel predictions provides evidence of regions potentially endowed with one or more GroupModels.

Several assumptions have been made for the prediction of the mineral deposit types based on stream-sediment geochemistry and are discussed. The results presented here indicate that various types of mineral-deposit can be predicted with a confidence similar to more conventional geochemical interpretative methods involving catchment analysis and the use of expert knowledge-based models.

Introduction

The QUEST-South project area in southern British Columbia (BC) was a focus for geochemical and geophysical research by Geoscience BC in 2009 and 2010 (Figure 1). New geochemical data were obtained for the -80 mesh (<177 μm) grain size fraction of 8536 stream-sediment samples. These samples were originally collected near stream outlets under the Regional Geochemical Survey (RGS) program between 1976 and 1981 from within the QUEST-South project area. The samples were originally analyzed using atomic absorption (AA) and instrumental neutron activation analysis (INAA), but the data suffer from limited elements and high lower limits of detection compared to analytical methods currently available. Available archived material from these samples was reanalyzed in 2009 (Jackaman, 2010a) using an aqua-regia digestion followed by a combination of inductively coupled plasma atomic–emission spectrometry (ICP-AES) and inductively coupled plasma–mass spectrometry (ICP-MS) at ALS Global (North Vancouver, BC; method code ME-MS41L). A new stream-sediment survey was undertaken in 2009, adding 785 new samples that were analyzed using the same grain size with a similar acid digestion and instrumental finishes at Eco Tech Laboratories Ltd. (Kamloops, BC), as well as by INAA (Jackaman, 2010b). The use of two different laboratories for analyses from the QUEST-South project area raises some issues in terms of data quality, as will be discussed in the following section.

The newly acquired stream-sediment data for the samples were interpreted by Arne and Bluemel (2011) using a catchment-analysis approach. The locations of the RGS samples were manually transcribed from hard copy 1:50 000 scale topographic maps in NAD27 and subsequently transformed to NAD83. Global positioning satellite (GPS) receivers were used to locate only the 785 new stream-sediment samples. The historical sample locations are known to be inconsistent with the 1:20 000 scale provincial Terrain Resource Information Management Program (TRIM I) hydrology data (Cui, 2010). As a result, considerable effort was expended by Arne and Bluemel (2011) to validate the recorded sample locations using scanned images of the archived topographic maps that had been used in the original sampling programs. Sample locations were adjusted where they were inconsistent with the original survey maps, and each sample location was given a confidence ranking. Catchment polygons for each adjusted sample were delineated by the BCGS for the adjusted sample locations using the approach described by Cui et al. (2009), which involves calculating the total drainage area for an individual sample from the nearest downstream junction.

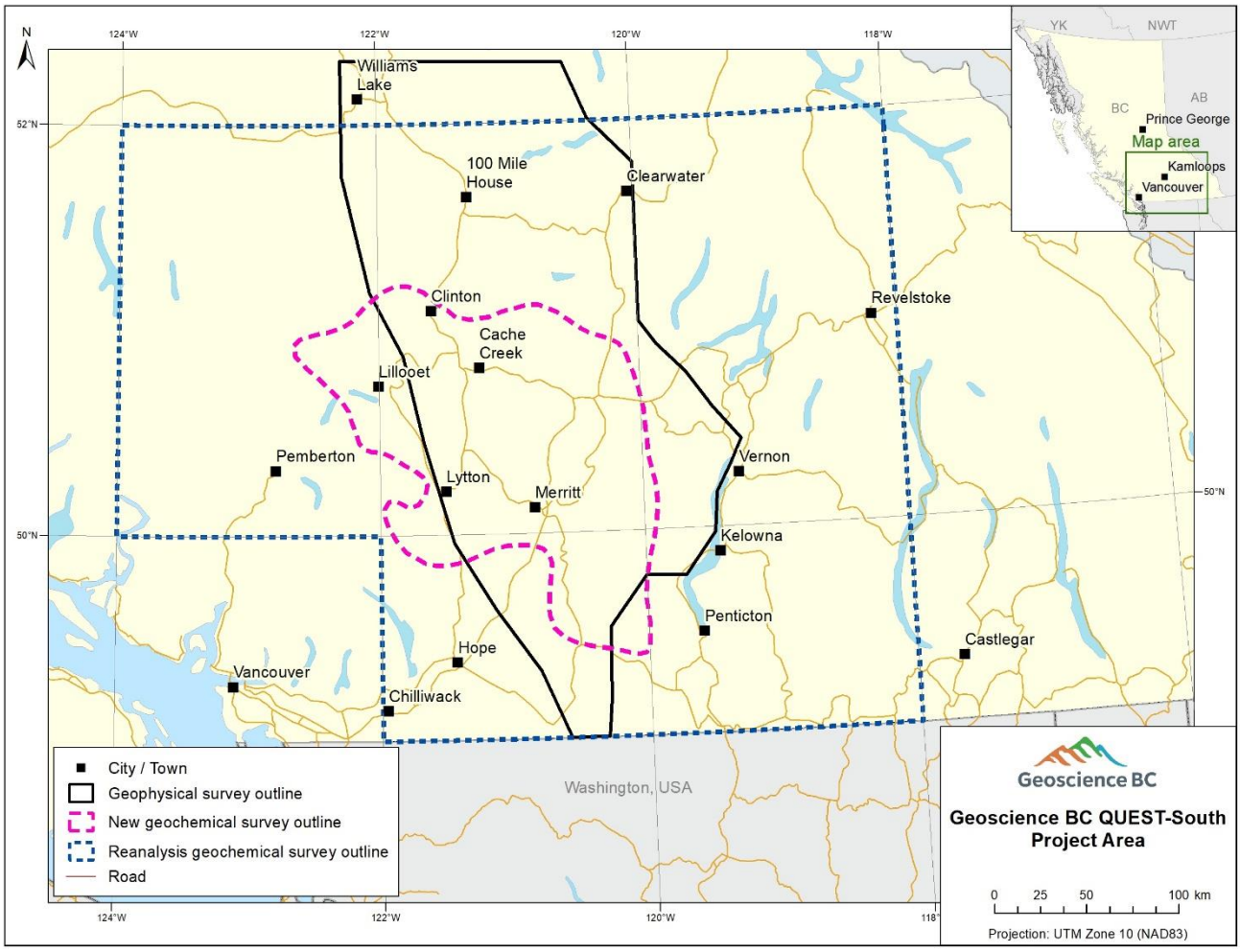


Figure 1. Location of QUEST-South project activities, including the areas from which archived stream-sediment samples were obtained (outlined in blue dashes), the area of infill stream-sediment sampling (outlined in pink dashes) and the area of geophysical surveys (outlined in solid black).

Arne and Bluemel (2011) assessed the stream-sediment geochemical data through correction of the data for variable catchment geology. The catchment polygons were used to determine the dominant rock type for each catchment area, information that was then applied to the QUEST-South stream-sediment samples. It has previously been established that the dominant control on regional stream-sediment geochemistry are catchment rock types (Bonham-Carter and Goodfellow, 1986; Bonham-Carter et al., 1987; Carranza and Hale, 1997). Dominant catchment rock types were therefore used to normalize the stream-sediment geochemistry data for the effects of variable background influence on their geochemistry.

Exploratory data analysis (EDA) of the geochemical data also indicated that there were positive correlations between some elements with Fe and/or Mn, suggestive of scavenging of metals by secondary hydroxides. High metal values compared to those expected from a correlation with Fe, known as positive residuals from linear

regression analysis, were also used to identify areas of anomalous metal concentrations. Additive indices for several common mineral-deposit types from the QUEST-South project area were then calculated using residuals and/or data normalized for catchment geology. For example, an additive index of Cu+Au+Mo was used to model the presence of porphyry Cu-Au-Mo deposits.

The approach used by Arne and Bluemel (2011), as well as by other previous studies (see references therein), relies on the use of drainage catchments for constraining bedrock type to define background values. Several assumptions are implied by the catchment-analysis approach:

1. The samples are accurately located and thus can be attributed to the correct catchment area.
2. The bedrock geology of the area is well known and accurately represented by the available geological mapping.
3. All areas of the catchment, and thus all rock types, contribute equally to the sediment load of the stream draining past the sample location.
4. The influence of transported materials such as till or glaciofluvial sediments is minimal.

Previous methods and reviews for assessing mineral resource potential were carried out by Kilby (2004), Grunsky et al. (1994), MacIntrye et al. (2004) and Mihalsky et al. 2013). The previous approaches were based on the use of defined geological tracts, previous work and investment from assessment file report, and expert knowledge from the mineral exploration industry and the provincial and federal governments. The approach taken in this study is defined solely by the regional stream sediment geochemistry and is empirically based.

Grunsky et al. (2010), de Caritat and Grunsky (2013), and Grunsky et al. (2014) demonstrated that the lithological controls on the geochemistry of regional drainage-sediment samples can be extracted from the data, particularly using PCA. Arne et al. (2018a) used regression analysis of key pathfinder or target-commodity elements against those principal components which strongly represent lithological control to calculate residual values for those elements that were elevated above what would be expected. Geological processes, including responses related to exposed mineralization, are inherent in the data, e.g., de Caritat et al. (2016), Harris et al. (2015) and Arne et al. (2018b) and therefore demonstrate that the use of machine-learning algorithms can provide useful predictions of where mineralization is likely to be found using publicly available regional geochemical data. These predictions can then be applied to catchment polygons in the case of stream-sediment surveys to generate predictive maps for mineral exploration. This project extends the work of Harris et al. (2015) and Arne et al. (2018b) and applies it to the QUEST-South project area using advanced data analytics and machine learning. The use of geochemical data corrected for the influences of catchment geology as input into machine learning algorithms produced a similar outcome in predicting the locations of known mineral occurrences as using uncorrected data in an analysis of stream-sediment geochemical data in northwestern BC (Arne et al., 2018b). Therefore, random forests analysis in

this study uses data uncorrected for the effects of lithology, metal scavenging or dilution and the results are compared to a more conventional catchment analysis approach for porphyry Cu-Au-Mo deposits.

In a study of this complexity, many assumptions and compromises must be made:

1) The geochemical composition of the stream-sediment associated with individual mineral-deposit models is uniquely distinct.
2) The stream-sediment samples represent a suitable medium from which the geochemical characteristics of mineral systems can be identified.
3) The MINFILE model identification is accurate.
4) The MINFILE site is accurately located, although the associated stream-sediment site may not be within the same catchment area.

Further details on these assumptions are provided in the discussion.

Data Quality

Geochemical data require quality-assurance and quality-control (QA-QC) screening prior to the application of statistical methods and subsequent interpretation. The main data quality aspect assessed in this study is whether there is a systematic bias for data from the two different laboratories used to generate the data used in this report.

The analyses from standard reference materials (SRMs) submitted with the samples during the original RGS survey were not provided in Jackaman (2010a, b); therefore, only a perfunctory review of data quality could be made by Arne and Bluemel (2011) using the available field duplicate data. Arne and Bluemel (2011) did note, however, that there was poor correlation (Spearman Rank correlation coefficient of 0.44) between reanalyzed ICP-MS and historical INAA data for Au. Digestions for the ICP-MS data used 0.5 g of –177 μm sediment, whereas the historical INAA samples averaged 23 g. The INAA Au data were preferred for data interpretation by Arne and Bluemel (2011), given the larger sample mass. Despite this preference, the precision of the INAA Au analyses is also poor given the nugget effect of Au distribution using -80 mesh (e.g. Arne and MacFarlane, 2014).

Subsequently, reanalyzed SRM data from the original RGS surveys and data from the SRMs submitted with the additional samples were made available by Jackaman (2018), including the RGS SRMs Red Dog (84) and SQ (22), and a small number of samples of certified reference material (CRM) Canmet STSD-2 (7). A larger number of Geological Survey of Canada (GSC) SRMs were also reanalyzed with samples from the original RGS surveys but these SRMs were not available for analysis of the infill survey samples to provide overlapping SRM data sets for comparison.

A comparison of SRM data for the Red Dog and SQ for selected elements indicates systematic relative biases for several elements of significance for mineral deposits in the QUEST-South region (Figure 2), including As, Ag, Mo

and Sb, in part due to differences in lower limits of detection. The elements Ba and La also show significant relative biases. Those elements with significant relative biases (i.e., >±5%) have been adjusted using the RGS Red Dog median data for those elements prior to a centred logratio transformation of the data. This correction was validated on the stream-sediment data from samples located in an area of overlap sampling but was found to make only a slight difference in gridded images of the data.

A comparison was also made of the three analytical methods used on the stream-sediment samples: ICP-MS/AES, INAA and AA. The AA results were not considered further for this study due to the limited number of elements analyzed, as those elements were already present in the reanalyzed ICP-MS/AES results which have a lower detection limit compared to the AA analyses. The ICP-MS/AES data are derived from an aqua-regia digestion, which is a partial extraction for many elements, whereas the INAA data represent a complete analysis but for fewer elements. Previous studies have shown that material dissolved with aqua regia provides a multi-element signature that reflects silicate-bearing assemblages, most likely through partial dissolution of sheet silicates (Grunsky et al., 2014). The decision was made to use only the ICP-MS data in this study for the sake of consistency, in spite of the better precision of the INAA Au analyses, as both the INAA and ICP-MS Au data generate similar spatial trends when the data are gridded. Data from the following 35 elements were therefore used: Au, Ag, Al, As, Ba, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, La, Mg, Mn, Mo, Na, Ni, P, Pb, S, Sb, Sc, Se, Sr, Th, Ti, Tl, U, V, W and Zn. In total, data from 8545 stream-sediment sites for which ICP-MS/AES data are available were used in the study. Both Ag and Au were converted from ppb to ppm prior to any data analysis.

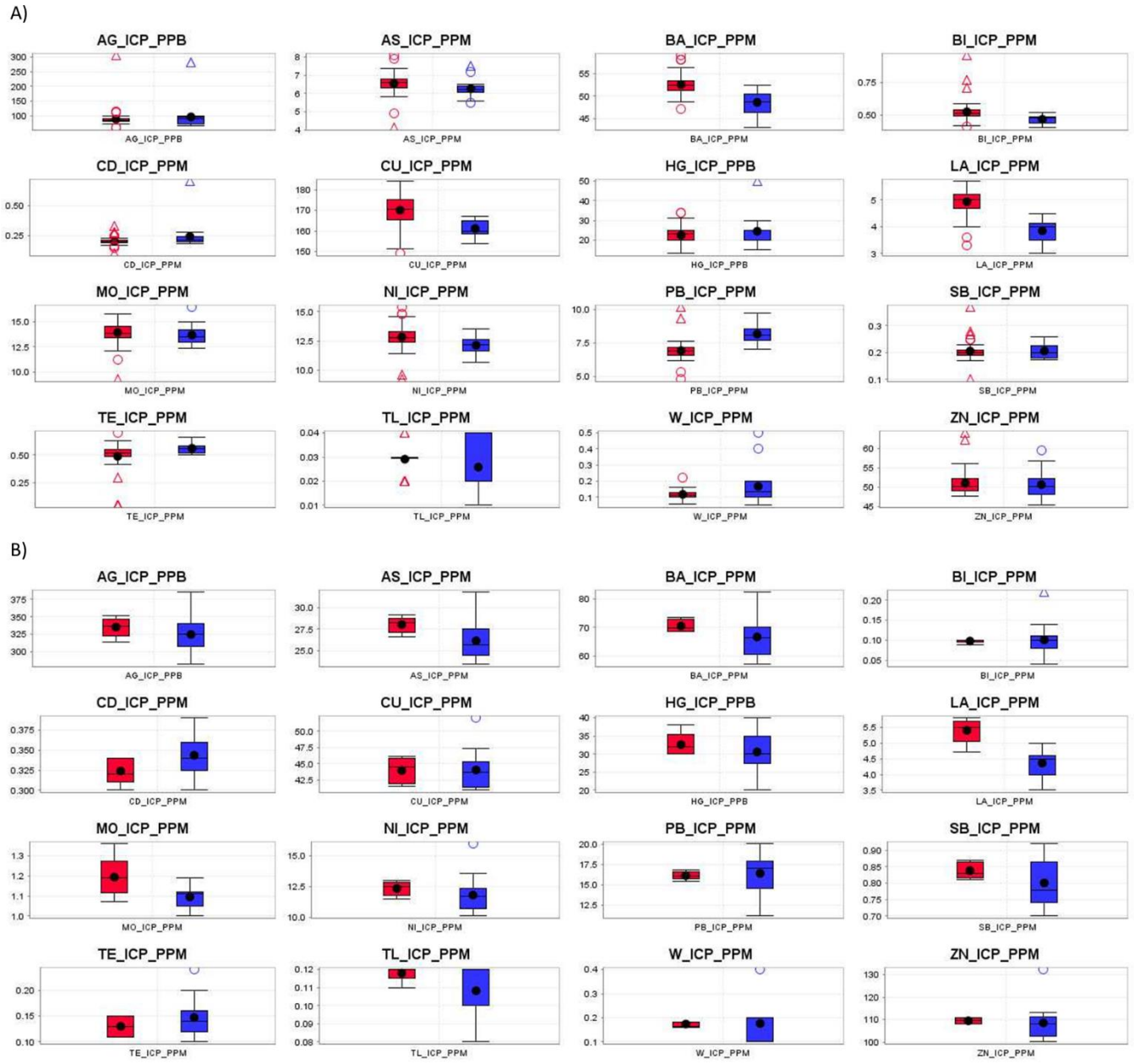


Figure 2. Box and whisker plots of selected elements from ICP-MS analysis of 84 Red Dog (66 from ALS; 18 from EcoTech) **(a)** and 22 SQ (5 from ALS; 17 from EcoTech) **(b)** standard reference materials used to correct the stream-sediment data for systematic laboratory biases. Data in red are from ALS Global and data in blue are from EcoTech.

Methods

Data Screening and the Compositional Nature of Geochemical Data

All data processing was carried out using the R programming and statistical environment (R Core Team, 2019), and geospatial rendering was carried out using the Quantum Geographic Information System (QGIS Development

Team, 2019). Of the 9321 geochemical analyses assembled by Arne and Bluemel (2011), 485 were field duplicates, archived material from 280 RGS stream-sediment samples were unavailable for reanalysis, and catchments were not derived for 11 samples. The duplicate analyses were removed to provide 8545 analyses for evaluation.

Major-element concentrations, reported as percentages, were converted to parts per million (ppm) to allow for centred logratio (clr) transformation. Geochemical data reported at less than the lower limit of detection (censored data) can bias the estimates of mean and variance; therefore, a replacement value that more accurately reflects an estimate of the true mean is preferred. Replacement values for censored geochemical data can be determined using several methods (Grunsky, 2010; Hron et al., 2010; Palarea-Albaladejo et al., 2014). In this study, the lrEM function from the zCompositions package (Palarea-Albaladejo et al., 2014) was used to estimate replacement values. Values greater than the upper limit of detection were interpreted using the 'maximum' value reported by the laboratory.

After QA-QC, the geochemical data were subjected to an empirical investigation in which the assumptions about the data were minimal. Because geochemical data are compositions, the issue of closure also becomes important. As compositional data sum to a constant (i.e., 100%, 1 000 000 ppm), when one value changes, all others must change to maintain the constant sum. Thus, the data are 'closed' and the variables are not independent, but standard statistical methods are based on variables that are independent. For geochemical data, this lack of independence can result in meaningless statistical results. To deal with the effect of closure, data for the 35 selected elements were centred logratio transformed (Aitchison, 1986).

Integration of Geology and MINFILE Attributes with the Stream-Sediment Geochemistry

Various data sources were integrated and displayed in QGIS in projection NAD 83 UTM Zone 10. The PDF maps included in Appendix 4 of this report were prepared in NAD83 latitude/longitude to allow integration with other geographical layers.

Digital files of the bedrock geology (Cui et al., 2017), regional terranes (Nelson et al., 2013) and MINFILE data were obtained from the BCGS (https://www2.gov.bc.ca/gov/content/industry/mineral-exploration-mining/british-columbia-geological-survey) in May 2019. An initial selection from the MINFILE database yielded 4877 records for the QUEST-South study area. However, as the focus of this study is on metallic mineral deposits, MINFILE data that were classified as industrial minerals were removed from further consideration, resulting in a total of 4108 MINFILE records. It was found that Polymetallic Ag-Pb-Zn veins (deposit type I05) are by far the most common mineral occurrence in the QUEST-South area (31.5% of all MINFILE occurrences) but have geochemical characteristics that overlap with several mineral-deposit types that are more economically significant.

The QGIS plug-in function 'NNJoin' was used to find the closest MINFILE point to each stream-sediment sampling site. and each sample site was tagged with the nearest distance to a MINFILE site. These distances range from 0.7 to 42848 m. Table 2 summarizes the number of stream-sediments sites associated with each MINFILE mineral

deposit type. A histogram of distance values is shown in Figure 3. Figure 4 shows a map of the stream-sediment sites and a summary of the distances between a stream-sediment site and the closest MINFILE site. Based on the empirical observations from Figure 4 and some experimentation with selecting different distance thresholds (500m, 1000m, 2500m, 5000m), a threshold distance of 2500m was selected.

There are two issues associated with choosing a suitable threshold distance. The first; if the threshold distance is too low, then there will be too few sites for the creation of a training set that represents the mineral deposit models. If the threshold is too large, then too many sites will be selected for the training set and that will result in mineral deposit models that overlap each other in terms of geography and stream-sediment geochemical composition. The second issue is somewhat more subjective; what is a reasonable distance for a stream-sediment geochemical signature from a mineral deposit? Depending on the size of the catchment, the geology of the terrain and the characteristics of the mineral deposit, the distance threshold may be quite variable. Given these uncertainties, a distance threshold of 2500m was considered reasonable. Although not pursued in this study, different distance threshold might be applied to different mineral deposit models.

It is important to note that the location of a MINFILE site and the associated stream-sediment site may not be within the same catchment area. MINFILE sites were captured by catchment polygons in a previous study using a catchment analysis approach (Arne et al., 2018b). However, this approach also resulted in several ambiguities. Large catchments often contain multiple MINFILE occurrences for several different mineral deposit types, raising the question of which MINFILE occurrence should be attributed to the sample site. Further, the locations of the MINFILE sites may be incorrect, as was the case with many of the historical RGS sample sites. The MINFILE location may also not represent the areal extent of any geochemical surface expression of mineralization, and it is possible that the geochemical signature may extend across catchment divides. This is certainly the second author's experience with significant mineral deposits. Finally, if there is a requirement for the location of a MINFILE site and associated stream-sediment site to be in the same catchment, the number of sites for the training set would be significantly reduced.

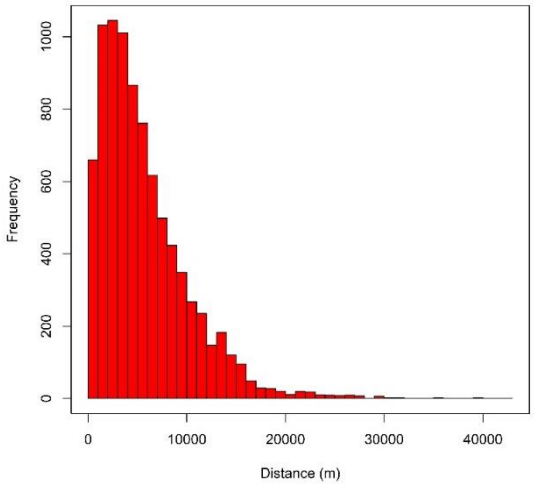


Figure 3. Histogram of the distances between stream-sediment sites and MINFILE sites, based on the QGIS function 'NNJoin'.

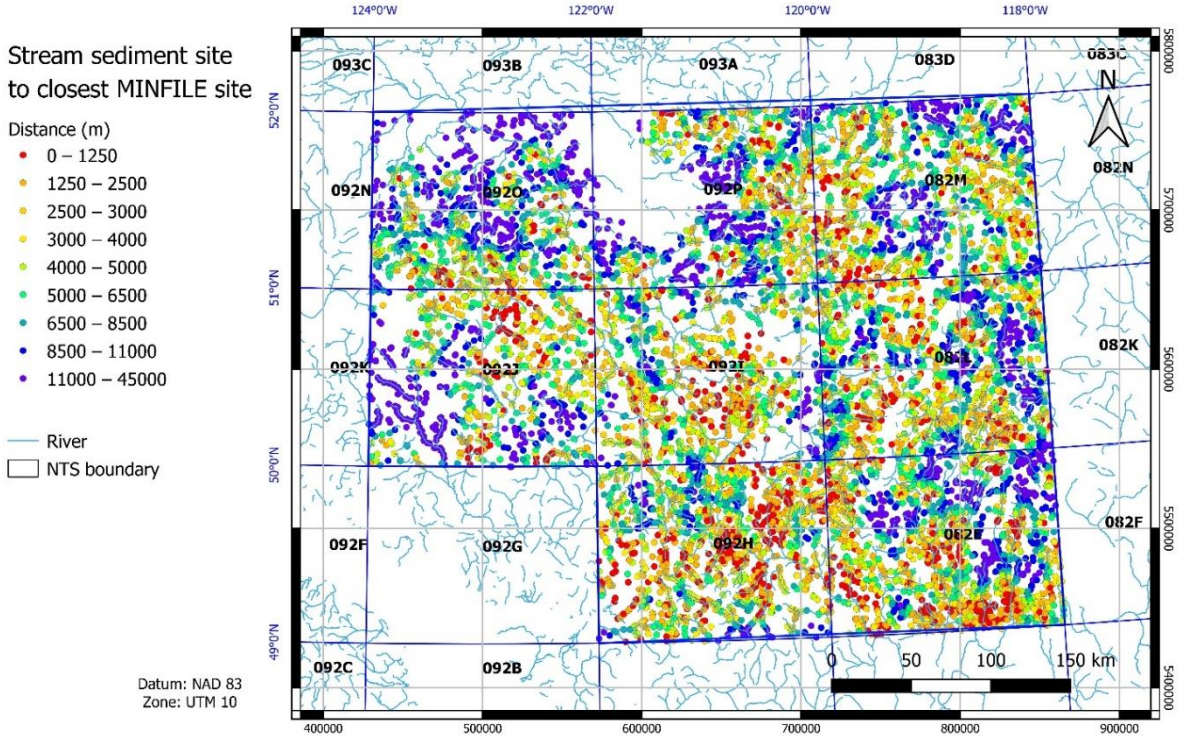


Figure 4. Geographic distribution of the distance measures between a stream-sediment site and the closest MINFILE site, based on the QGIS function 'NNJoin'.

The QGIS function ‘Intersect’ was used to merge the bedrock geology and geological terrane designation with the stream-sediment geochemical data and the closest MINFILE point. The tagging of a MINFILE site with a stream-sediment site is based on the closest distance between the two sites, regardless of the MINFILE ‘Status’ designation and catchment delineation. Thus, MINFILE sites with the status of Producer or Past Producer may not be tagged with the closest stream-sediment site if another MINFILE site with the status of Developed Prospect, Prospect, Showing or Anomaly is closer. Some MINFILE sites may not be tagged if there is no stream-sediment site nearby and the likelihood of a geochemical expression of the mineralization is difficult to estimate. If the measured distance between a stream-sediment site and a MINFILE site was greater than 2500 m, the stream-sediment site MINFILE Model designation was tagged as ‘Unknown’. Table 1 lists the number of stream-sediment sample sites associated with each MINFILE Status attribute.

Table 1. MINFILE Status for the tagged Quest-South stream-sediment data.

MINFILE Status	Number of MINFILE/Stream-sediment Sites
Anomaly	116
Developed Prospect	341
Past Producer	682
Producer	38
Prospect	1144
Showing	6224
Total	**8545**

Interpolation of principal-component scores and random forests posterior probabilities was carried out using a geostatistical framework. Posterior probabilities are estimated by calculated averages of class assignments over all trees that defines a probability vector (the number of tree votes for each class), which is interpreted as a posterior probability. The gstat package for R (Pebesma, 2004) was used to generate and model semi-variograms with sufficient parameters to produce interpolated raster images through kriging. Variogram analysis was based on the stream-sediment site with the assigned posterior probability. Catchment boundaries were not considered in the kriging interpolation. The cell size used for image interpolation was chosen as 2.5 km for the images generated by PCA and t-SNE9 (van Maaten and Hinton, 2008) for the random forests predictions. Each of these methods provides different coordinate systems that can reveal features and patterns related to geochemical processes.

PCA reduces the number of variables necessary to describe the observed variation within a dataset. This is achieved by forming linear combinations of the variables (components) that describe the distribution of the data. These linear combinations are derived from some measure of association (i.e. correlation or covariance matrix). Mineral stoichiometry typically controls the observed linear combinations and variability in geochemical datasets. t-SNE is

a machine learning algorithm for dimensionality reduction. It is a nonlinear dimensionality reduction technique that is particularly well-suited for embedding high dimensional data into a space of two or more dimensions, which can then be visualized in a scatter plot. Specifically, it models each high-dimensional object by a two- or three-dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points.

Characterizing Mineral Occurrence Information

Each MINFILE record lists a mineral-deposit model derived from the BCGS Mineral Deposit Profiles (British Columbian Geological Survey, 1996). The number of MINFILE sites associated with each model in the QUEST-South area is shown in Table 2. The large number of mineral-deposit types for which there are only a few sites creates difficulty in a statistical assessment of the data because these techniques require multivariate input, therefore subgroups with sparse data are not easily classified. Consequently, the models were merged as shown in Table 3 based on geochemical similarities between the individual deposit types. These merged models, termed 'GroupModels', were the basis for assessing the multivariate geochemical patterns. Figure 5a shows the GroupModel designation for each of the tagged stream-sediment sites and Figure 5b shows the Status of the MINFILE sites, labelled with the BCGS Mineral Deposit Profile that is listed in the MINFILE record variable 'Deposit Type'. Figure 6 shows a graphical legend for lithology and the GroupModel classes that are used in the subsequent figures of this report, with the latter as mnemonic symbols with a brief description of the GroupModels.

Error! Reference source not found. shows the frequency of the GroupModel class for all stream-sediment sites that met the criteria of being less than 2500 m from a MINFILE site with the Status class, as described above. Table 2 shows that, for the 61 Model deposit types that were identified, many are associated with less than 10 sites. As a result, the Deposit Types were merged into the GroupModels, as shown in Table 3, with the corresponding number of sites shown in **Error! Reference source not found.**.

Table 2. Number of stream-sediment sites associated with a MINFILE model.

Model[1]	Frequency	Model	Frequency	Model	Frequency	Model	Frequency
C01	118	G03	1	I11	7	L05	52
D03	115	G04	34	I12	4	L07	1
D04	13	G05	6	I14	3	L08	14
D06	6	G06	102	J01	11	M01	1
E01	2	G07	2	J04	2	M02	26
E03	2	H02	13	K01	151	M03	21
E04	3	H03	2	K02	28	M04	2
E05	2	H05	44	K03	20	M05	26
E12	12	H08	10	K04	54	N01	15
E13	5	I01	311	K05	25	N03	1
E14	92	I02	30	K07	5	O01	8
E15	1	I05	988	K09	6	O02	24
E16	2	I06	82	L01	61	S01	7
F01	7	I07	3	L02	7	Unknown*	901
G01	5	I08	4	L03	198		
G02	2	I09	24	L04	384		

[1] MINFILE 'Model' designation (see 'Deposit' section on 'Mineral Occurrence' tab of MINFILE Search Page at <http://MINFILE.ca/>.
* Unknown means no mineral deposit model was assigned to the MINFILE record.

Table 3. Merged mineral-deposit models (GroupModels) for statistical processing of the QUEST-South stream-sediment data.

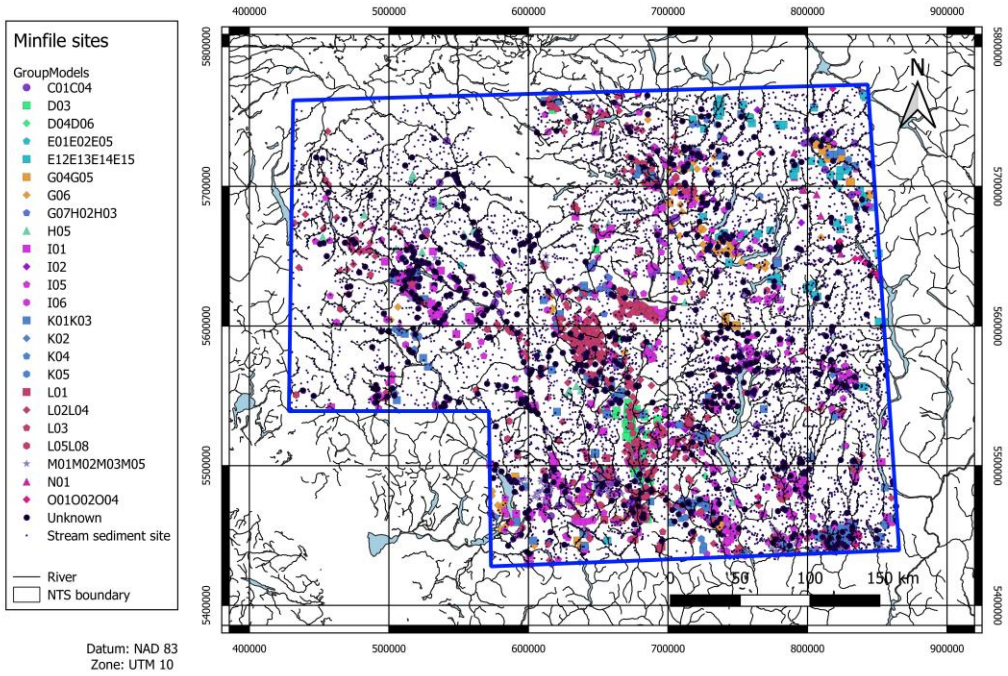
Models[1]	GroupModels	GroupModel Description
C01, C04	C01C04	Surficial & buried placer Au
D03	D03	Volcanic redbed Cu
D04, D06	D04D06	Basal U; volcanic-hosted U
E01, E04, E05	E01E04E05	Sediment-hosted Hg-Cu-Pb
E12, E13, E14	E12E13E14	MVT Pb-Zn; Irish-type Pb-Zn; SEDEX Pb-Zn
G04, G05	G04G05	Massive sulphide Cu-Zn
G06	G06	Volcanogenic Cu-Pb-Zn
G07, H02, H03	G07H02H03	Hot spring Au-Ag-Hg
H05	H05	Low-sulphidation epithermal Ag-Au
I01	I01	Au-quartz veins
I02	I02	Intrusion-related Au
I05	I05	Polymetallic vein Ag-Pb-Zn-Au
I06	I06	Cu +/-Ag quartz veins
K01, K03	K01K03	Cu-Fe skarn
K02	K02	Pb-Zn skarn
K04	K04	Au skarn
K05	K05	W skarn
L01	L01	Sub-volcanic Cu-Au-Mo
L02, L04	L02L04	Porphyry Cu-Au-Mo
L03	L03	Alkalic porphyry Cu-Au
L05, L08	L05L08	Porphyry Mo
M01, M02, M03, M05	M01M02M03M05	Mafic-hosted Ni-Cu-Cr
N01	N01	Carbonatite
O01, O02, O04	O01O02O04	Rare earth element pegmatites

[1] MINFILE 'Model' designation (see 'Deposit' section on 'Mineral Occurrence' tab of MINFILE Search Page at http://MINFILE.ca/ See Table 2 for the Model descriptions.

Table 4. Merged Mineral Deposit Models (GroupModels) tagged at the stream-sediment sites. Note that 100 Unknown sites were used with the training set for the application of Random Forest classification/prediction. The remaining 8071 sites were used to classify the 'Unknown' GroupModels.

GroupModel	Frequency
C01C04	83
D03	5
E01E04E05	3
E12E13E14E15	22
G04G05	11
G06	27
G07H02H03	6
H05	18
I01	41
I02	1
I06	5
K01K03	23
K02	2
K04	9
K05	6
L01	6
L02L04	61
L03	15
L05L08	13
M01M02M03M05	17
Unknown - test	8071
Unknown - train	100
Total	**8545**

a)



b)

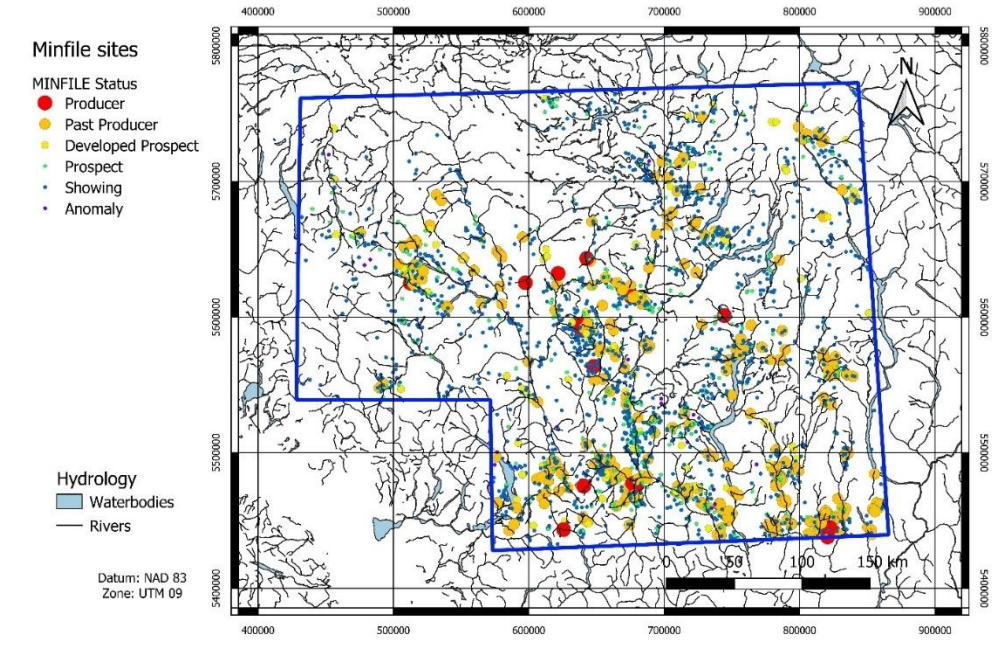


Figure 5. Geographic distribution of **a)** MINFILE sites classified by GroupModel; and **b)** MINFILE sites classified by Status and labelled by BCGS Mineral Deposit Model. See Table 2 for a description of the Mineral Deposit Model Mnemonics (e.g. L03).

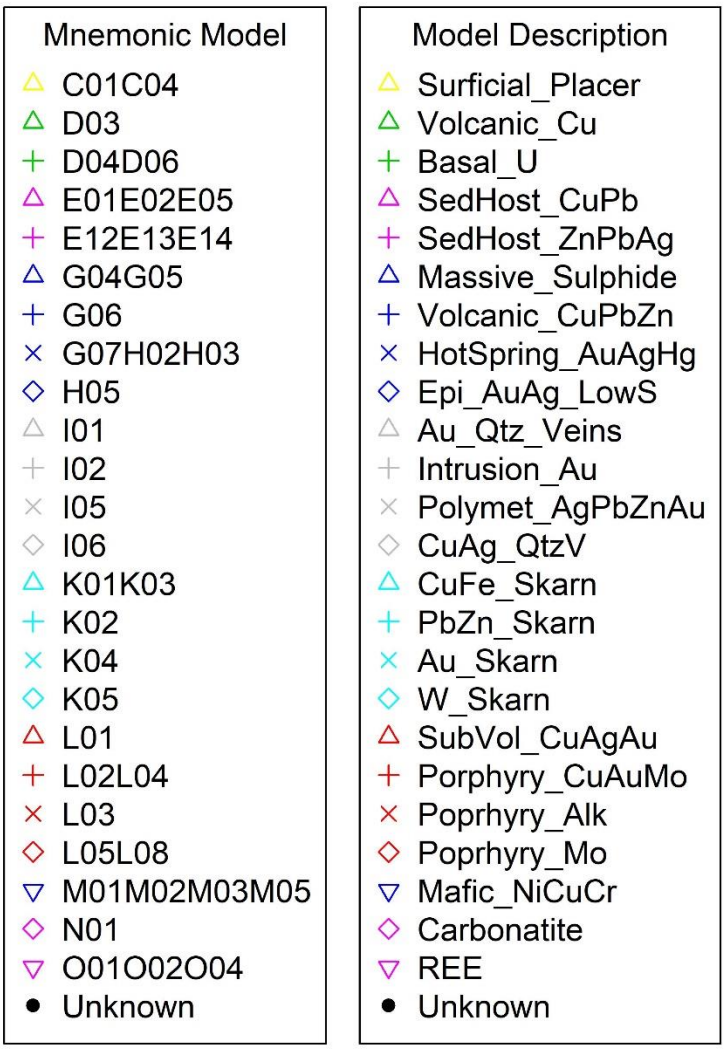


Figure 6. Legends showing colours and symbols for lithology (left) mineral deposit types as BCGS Mineral Deposit Model mnemonics (centre) and short descriptions of the respective BCGS Mineral Deposit Models (right).

Selecting the Training and Test Datasets

In this study, mineral-deposit prediction is based on the selection of a training set of stream-sediment sites that are tagged with the nearest MINFILE site. A stream-sediment site that is more than 2500 m from a MINFILE site is also classed as Unknown for the associated MINFILE Status and Model classes. MINFILE Status designations of Anomaly or Occurrence were classed as Unknown regardless of the distance between the two sites. Thus, the training set was comprised of stream-sediment sites with a MINFILE GroupModel designation, classed as Producer, Past Producer, Developed Prospect, Prospect that is less than 2500m in distance.

After some experimentation, it was decided that the Mineral Deposit Model I05 (polymetallic veins) created a significant amount of confusion in the prediction of the other mineral-deposit types. This issue was also noted in a previous study (Arne et al., 2018b). Consequently, stream-sediment sites that were labelled as I05 were relabeled as Unknown. The test set contains all the stream-sediment sites where the GroupModel class is Unknown. It is unrealistic to consider that every stream-sediment site must have a MINFILE Model or GroupModel designation. Thus, a random selection of 100 stream-sediment sites with a GroupModel of Unknown was made. In this way, sites that do not have a geochemical signature that reflects a form of mineralization may have the possibility of being assigned as belonging to an Unknown GroupModel class. This resulted in a training set of 474 sites and a test set of 8071 sites. **Error! Reference source not found.** summarizes the GroupModel classes that are part of the training dataset.

Process Discovery – Empirical Investigation of Geochemistry

Multivariate methods were applied to the clr-transformed data for the purposes of discovering patterns and features that potentially describe relationships amongst geochemical, geological and geophysical parameters, as well as the effects of gravitational processes (Grunsky et al., 2010). These methods included PCA and t-SNE (van Maaten and Hinton, 2008). Each of these methods provides different axial coordinate systems that can reveal features and patterns related to geochemical processes when viewed in the geospatial domain.

The coordinates resulting from the application of PCA and t-SNE were used to discover patterns and features in the data. The method of PCA used in this study is based on the methodology of Zhou et al. (1983) and Grunsky (2001). The geochemistry of the stream-sediments was evaluated using a simultaneous R- and Q-mode extraction of eigenvalues/eigenvectors. The R-package, tsne, (R Core Team, 2019) was used to generate the t-SNE coordinates.

Process Validation – Modelled Investigation of Geochemistry

A training set comprised of the stream-sediments associated with a GroupModel, as shown in Table 4, was established. This training set has both principal component scores (PCA coordinates) and t-SNE coordinates (9 dimensions). The training dataset included 100 sites that had a GroupModel class of “Unknown”. The set of data that was tested contained stream-sediment sites with both PCA and t-SNE coordinates with the GroupModel set as “Unknown”. A GroupModel class was predicted for the test dataset from the training dataset using the method of random forests (Breiman, 2001).

Random forests was previously employed by Harris and Grunsky (2015), Arne et al. (2018b) and Grunsky et al. (2018) and used as part of a remote predictive-mapping strategy (Harris et al., 2008). The method of random forests is based on the construction of classification trees (Venables and Ripley, 2002, Chapter 9) in which nodes (splits in classes) are based on continuous variables from which a series of branches in the tree correctly classify all of the data into categorical variables. A more detailed description of how the random forests classification method was

used with soil-geochemical data is provided in Harris et al. (2015). It should be noted that cross-validation is built into the method of random forests as it repeatedly samples the data population to grow trees from which votes are cast to determine class membership.

Interpolated maps of the posterior probabilities derived from the classification method of random forests can be created using geostatistical methods such as kriging. However, since the posterior probabilities are compositions and sum to 1.0, these values should be logratio transformed, followed by subsequent co-kriging, and then back-transformed for subsequent geographic rendering (Pawlowsky-Glahn and Egozcue, 2015; Mueller and Grunsky, 2016). This approach is potentially problematic because, in cases where posterior probabilities are very low or zero, the results from kriging may be unreliable. It can be argued that the posterior probabilities for each predicted class are independent, since there is no intention, or value, of assessing the variables of probabilities in terms of any interactions. Additionally, maps of the posterior probabilities for each of the classes can be created by posting the sample sites with points and colours. An alternative to this would be to consider the un-normalized (raw) votes as independent and carry out kriging on these estimations. For this study, the posterior probabilities were interpolated using the gstat function “krige” with the assumption of independence between the estimated classes.

Note that kriged images based on point data have been used for validation purposes to test the sensitivity of various model input parameters and that thematically coded catchment maps have been generated with predictive results for a number of mineral-deposit types using the preferred modelling inputs.

Process Validation – Comparison with Conventional Approaches

Arne and Bluemel (2011) undertook a limited interpretation of the QUEST-South data using the dominant catchment bedrock types to level raw Cu data for the effects of variable background. Data for some elements were also regressed against Fe given the possibility that those elements were scavenged by secondary Fe hydroxides within the samples. Levelled data and residuals from regression analysis were then used to generate a series of additive indices for a variety of deposit types. However, as pointed out by Bonham-Carter and Goodfellow (1986), use of data for all rock types in the catchment areas is preferable to using only the dominant rock type for levelling purposes. One common approach to achieve this outcome is multiple regression analysis of an element against the proportion of rock types found within all catchments (Bonham-Carter and Goodfellow, 1986; Bonham-Carter et al., 1987; Carranza and Hale, 1997). An alternative approach involving regression of key commodity and pathfinder elements against principal components in which they define lithological controls, was used by Arne et al. (2018a) to interpret stream-sediment geochemical data from the Yukon, followed by combining the residuals in weighted sums models following the approach described by Garrett and Grunsky (2001).

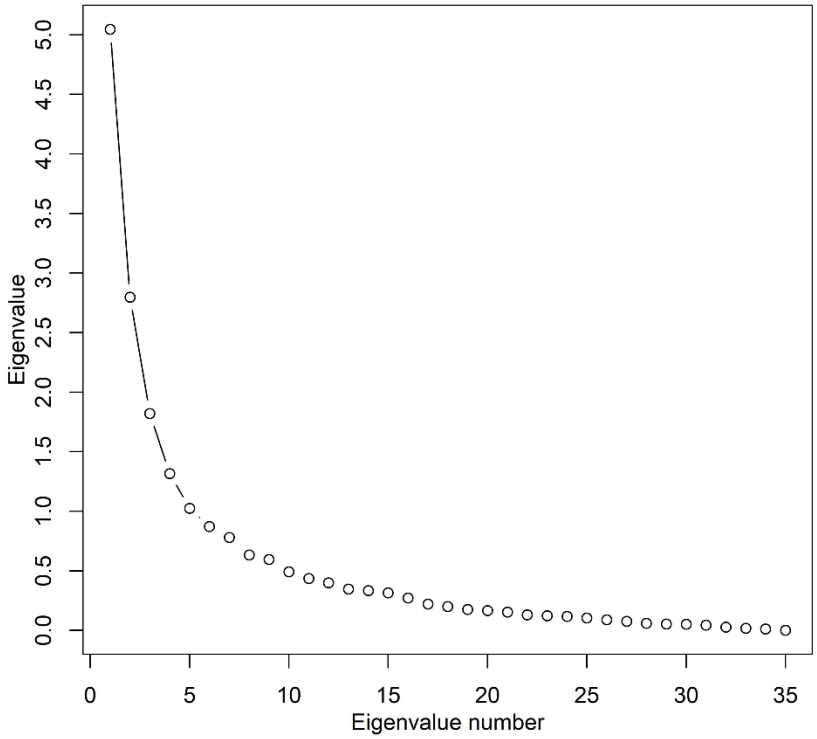
Accordingly, predictive models using these more conventional interpretive approaches were generated for comparison to random forests posterior probabilities for porphyry Cu-Au-Mo deposits for 8534 of the samples.

Catchments were not determined for 11 samples by Arne and Bluemel (2011) given uncertainties in their locations and so no catchment rock types could be determined for these samples. Linear multiple regression analysis of Cu and Mo was undertaken against the proportions of the main catchment rock types defined by Arne and Bluemel (2011). The results of this analysis were compared with regression analysis of Cu and Mo against PC1 and PC2, respectively, but the results from the multiple regression analysis of rock type were found to provide a better visual fit to MINFILE porphyry Cu-Au-Mo occurrences. Given the possible scavenging of both Cu and Mo by secondary Fe and Mn hydroxides identified by Arne and Bluemel (2011), the residuals from multiple regression analysis were in turn regressed against Fe. These final residuals were then used to generate a porphyry Cu-Mo additive index. The multiple regression residuals, without further regression against Fe, were also used to generate a weighted sums model for comparison.

Results

PCA Process Discovery

Principal components derived from the clr-transformed data are shown in a screeplot in

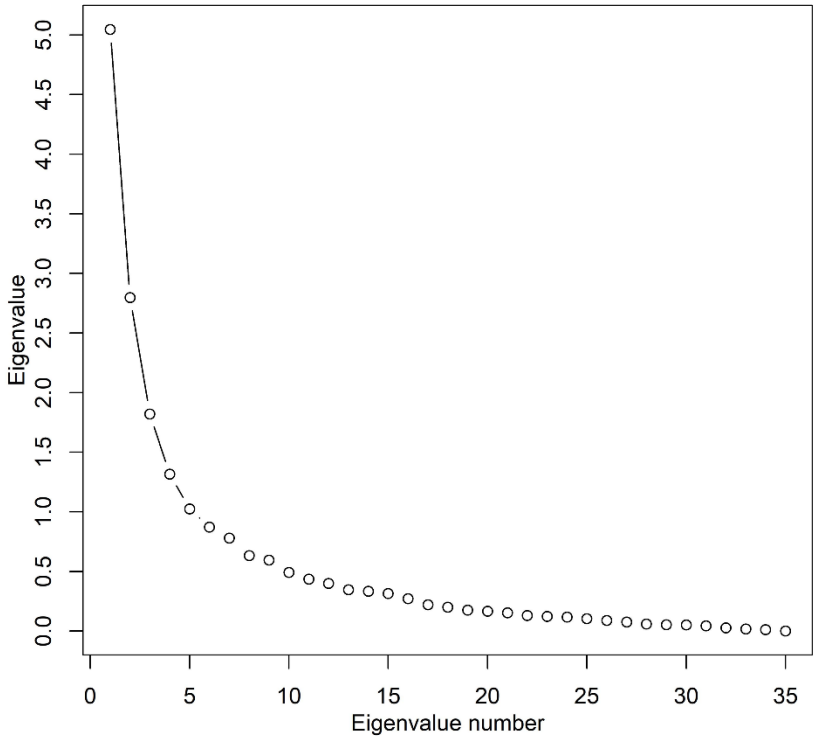


Eigenvalues	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
λ	5.05	2.8	1.82	1.32	1.02	0.87	0.78	0.63	0.59	0.49	0.43	0.4	0.35	0.33	0.32
λ%	26.1929	14.5228	9.4398	6.8465	5.2905	4.5124	4.0456	3.2676	3.0602	2.5415	2.2303	2.0747	1.8154	1.7116	1.6598
Σλ%	26.1929	40.7158	50.1556	57.0021	62.2925	66.805	70.8506	74.1183	77.1784	79.7199	81.9502	84.0249	85.8402	87.5519	89.2116

Figure **7**. The screeplot shows a steep decay for the first six eigenvalues, after which the curve flattens. The first six principal components can be interpreted as containing the 'structure' of the data that reflect the relationships between the variables (e.g., mineral stoichiometry) and the observations (scores of dominant processes). The remaining eigenvalues (7–35) may represent under-sampled geochemical or random processes. Typically, in regional geochemical surveys, elements associated with mineral deposits are under-sampled and the relationships of the elements associated with mineralization do not appear in the dominant principal components (Grunsky et al., 2014).

A full display of PCA biplots is not feasible in this report, so only the biplots of selected principal components are shown in order to illustrate the associations between the stream-sediment sites and the elements. Table 5 shows the relative contributions of the PCA results. The contribution of variability for each element is shown across the first 15 principal components.

The amount of variability in the data is shown in the table that accompanies

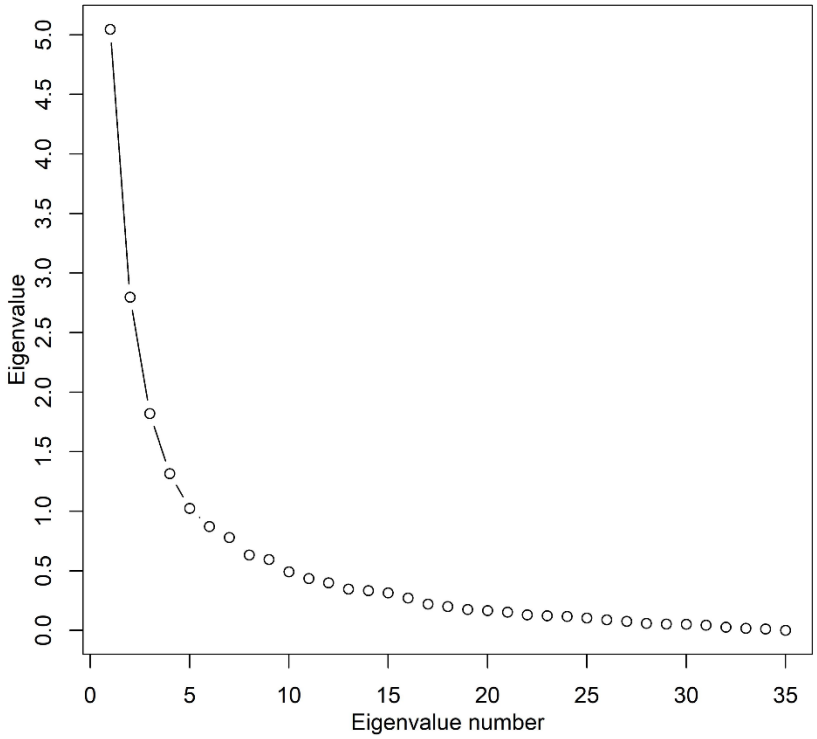


Eigenvalues	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
λ	5.05	2.8	1.82	1.32	1.02	0.87	0.78	0.63	0.59	0.49	0.43	0.4	0.35	0.33	0.32
λ%	26.1929	14.5228	9.4398	6.8465	5.2905	4.5124	4.0456	3.2676	3.0602	2.5415	2.2303	2.0747	1.8154	1.7116	1.6598
Σλ%	26.1929	40.7158	50.1556	57.0021	62.2925	66.805	70.8506	74.1183	77.1784	79.7199	81.9502	84.0249	85.8402	87.5519	89.2116

Figure **7**. Table 5 lists the coefficients of the principal component loadings and can be examined to determine which elements have the most variability across the principal components. Figure 7 shows that the first two principal components account for 40.7% of the overall variability. Table 5 shows the relative contributions that each element makes across the principal components. It can be seen that As, Bi, Cd, Co, Cu, K, La, Mg, Ni, P, Sb, Se, Th, Tl, U and W account for most of the variability in PC1 and Ag, Al, Bi, Cd, Co, Cr, Fe, Ga, K, Mg, Mo, Na, Ni, P, Pb, Sc, Se, Ti and V account for most of the variability in PC2. Principal component 4 (PC4) accounts for 86% of the variability of Au in the data and PCs 1, 12 and 14 account for most of the variability of Cu. Based on the contents

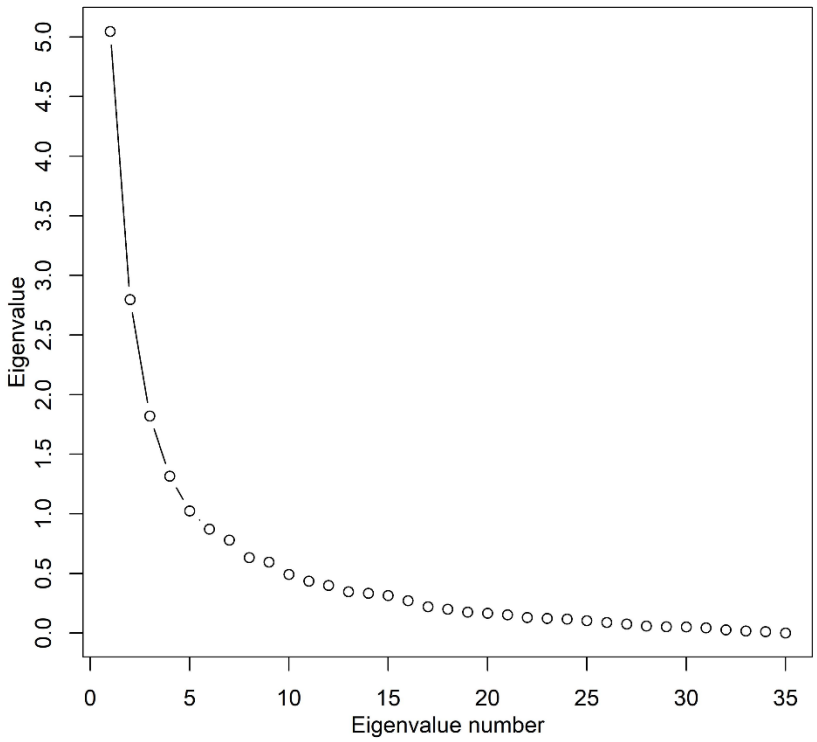
of Table 5, biplots of PC1 vs. PC2, PC3 vs. PC4 and PC12 vs. PC14 are shown in Figures 8 and 9 and reveal information on the processes that are reflected by the relative relationships of the principal component loadings (elements) and scores (stream-sediment sites).

The biplots of Figure 8a,,c, e show the principal component scores for PC1 and PC2. The scores are coded by their tectonic terrane (Figure 8a), regional rock type (Figure 8c) and GroupModel (Figure 8e). Figure 8b, d and f show the mean values of PC1-PC2 for the tectonic terranes, regional rock types and the GroupModels. The principal component loadings of the elements are plotted in each of the four figures. The associations of the principal component loadings reflect the dominant processes as expressed in



Eigenvalues	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
λ	5.05	2.8	1.82	1.32	1.02	0.87	0.78	0.63	0.59	0.49	0.43	0.4	0.35	0.33	0.32
λ%	26.1929	14.5228	9.4398	6.8465	5.2905	4.5124	4.0456	3.2676	3.0602	2.5415	2.2303	2.0747	1.8154	1.7116	1.6598
Σλ%	26.1929	40.7158	50.1556	57.0021	62.2925	66.805	70.8506	74.1183	77.1784	79.7199	81.9502	84.0249	85.8402	87.5519	89.2116

Figure **7**.



Eigenvalues	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
λ	5.05	2.8	1.82	1.32	1.02	0.87	0.78	0.63	0.59	0.49	0.43	0.4	0.35	0.33	0.32
λ%	26.1929	14.5228	9.4398	6.8465	5.2905	4.5124	4.0456	3.2676	3.0602	2.5415	2.2303	2.0747	1.8154	1.7116	1.6598
Σλ%	26.1929	40.7158	50.1556	57.0021	62.2925	66.805	70.8506	74.1183	77.1784	79.7199	81.9502	84.0249	85.8402	87.5519	89.2116

Figure 7. Screeplot of the eigenvalues derived from PCA applied to the clr-transformed data from the QUEST-South stream-sediment geochemistry results. A table of the first 15 eigenvalues and their contribution to the overall variance is given below the figure.

Table 5. Relative contributions of the elements over the first 15 principal components. Relative values >10 are highlighted in bold.

Element	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
Au	0.699	9.269	0.161	**86.551**	0.777	0.060	1.794	0.019	0.076	0.003	0.407	0.041	0.021	0.006	0.006
Ag	6.679	**41.483**	1.146	0.236	0.038	8.722	0.143	2.539	0.004	2.155	6.757	**13.581**	2.560	0.645	0.258
Al	0.007	**19.932**	0.783	1.314	0.296	3.167	**17.916**	0.133	0.248	0.084	**13.243**	0.646	0.552	0.300	2.676
As	**52.771**	9.123	**14.705**	0.544	4.063	5.209	0.427	0.464	1.172	4.188	0.472	0.000	0.417	0.207	3.505
Ba	0.925	5.799	8.726	3.621	4.514	0.295	**12.024**	3.842	0.742	2.607	1.406	0.719	**15.010**	0.023	1.273
Bi	**19.295**	**25.379**	8.163	0.013	2.793	2.764	3.213	0.036	5.571	0.109	6.865	2.677	1.405	5.610	0.398
Ca	8.657	8.023	**39.455**	0.352	0.119	4.366	**20.845**	4.729	0.003	0.028	0.560	0.010	1.009	1.876	0.124
Cd	**11.991**	**37.800**	1.125	2.241	2.089	**10.698**	0.858	**11.799**	1.237	2.422	1.913	0.163	1.407	3.829	0.835
Co	**21.876**	**30.746**	**12.366**	0.001	2.134	3.833	0.357	6.443	0.233	2.941	0.061	1.074	0.175	1.322	0.870
Cr	3.953	**32.265**	**14.462**	0.750	3.533	1.635	2.437	3.575	**19.949**	0.075	0.034	2.099	0.010	0.447	0.015
Cu	**28.365**	0.436	1.169	0.370	6.919	0.455	0.024	0.069	6.315	0.000	5.101	**18.405**	0.521	**14.485**	2.326
Fe	1.239	**26.304**	**12.216**	0.485	6.532	0.383	0.277	7.676	5.241	**15.327**	0.055	0.199	1.038	0.003	0.039
Ga	9.557	**26.774**	3.565	1.444	4.354	0.061	**12.659**	0.097	0.943	0.347	6.089	0.024	0.246	1.573	0.463
Hg	7.683	2.797	5.598	2.590	1.921	2.704	**11.999**	1.270	8.979	2.366	7.176	**14.719**	2.829	4.662	**14.526**
K	**28.210**	**10.953**	0.342	0.017	7.164	1.216	1.622	3.981	**10.468**	**20.583**	1.824	0.804	1.567	0.101	0.028
La	**65.982**	0.362	0.505	0.003	**10.900**	0.996	2.977	1.060	0.610	0.480	0.014	0.533	0.371	2.142	0.392
Mg	**17.842**	**38.923**	7.042	0.432	6.636	1.455	1.517	0.236	0.866	0.164	0.000	0.439	0.038	0.500	1.556
Mn	8.236	0.095	0.381	4.934	8.069	0.822	4.106	1.450	0.556	**15.597**	1.976	**13.076**	4.844	0.474	**10.573**
Mo	0.044	**24.814**	0.591	5.882	0.983	0.029	5.726	0.395	0.556	9.635	**22.227**	0.004	5.342	**13.474**	0.850
Na	0.873	**36.465**	**12.131**	1.757	0.493	4.340	2.170	5.055	0.820	0.063	0.941	0.185	**21.694**	0.339	0.101
Ni	**17.135**	**17.705**	8.265	0.476	9.298	**10.775**	6.020	2.194	**19.147**	1.207	0.133	0.450	0.121	0.284	0.416
P	**24.338**	**10.649**	2.863	0.169	1.969	1.994	0.773	1.346	2.442	1.028	4.595	0.744	6.028	1.495	0.037
Pb	5.807	**27.125**	0.126	0.243	2.407	7.311	5.371	0.293	1.979	0.313	**11.323**	5.350	1.147	1.877	0.551
S	6.814	5.777	**37.175**	0.112	**14.612**	0.818	0.046	**26.469**	3.859	0.119	0.085	0.019	0.071	2.921	0.048
Sb	**61.604**	7.451	2.914	0.891	7.597	4.594	0.777	0.247	0.477	1.576	0.002	0.183	0.037	0.336	4.034
Sc	8.436	**29.235**	6.357	0.157	1.633	1.003	4.780	0.056	0.090	0.505	5.586	0.372	0.004	0.049	0.447
Se	**13.210**	**10.469**	**20.761**	3.939	3.386	0.420	0.015	0.125	5.598	0.428	5.801	1.352	0.038	0.438	0.970
Sr	1.402	9.901	**46.979**	0.169	5.225	3.665	**10.377**	6.245	0.240	0.174	0.213	0.106	0.004	1.562	2.192
Th	**68.977**	0.163	5.591	0.360	3.150	0.035	**11.752**	1.089	1.409	0.066	0.742	0.014	0.024	0.000	2.922
Ti	3.844	**50.598**	2.211	0.004	0.001	0.469	8.984	0.726	0.016	0.257	0.000	0.679	1.465	4.721	0.784
Tl	**33.895**	1.355	0.059	1.690	3.930	9.210	7.339	3.204	0.663	**15.743**	6.174	0.417	0.811	0.561	0.002
U	**63.609**	9.613	2.941	1.662	2.817	0.645	1.234	1.528	6.683	1.070	0.079	2.435	1.120	0.631	1.872
V	0.980	**33.614**	6.045	0.257	7.199	6.568	6.882	0.247	0.607	**12.219**	0.001	8.147	0.090	0.847	0.687
W	**25.992**	8.842	**11.581**	0.060	**20.987**	**24.991**	0.039	2.520	0.917	1.310	0.251	0.319	1.194	0.424	0.094
Zn	7.849	1.345	1.482	4.434	0.259	**17.452**	0.000	4.889	3.796	3.315	0.000	3.765	2.208	2.035	0.000

Figure 8a shows a clear distinction between the three dominant terranes in the region. The Omineca and Coast terranes are dominated by felsic intrusive rocks which occur along the negative portion of the PC1 axis and have an association of U-Th-La-Tl-K-P. The Intermontane region is dominated by rocks that have relative enrichment in Cr-Ni-Mg-Co-Fe-Ca-V and likely reflect volcanic rocks. The Coast terrane also shows relative enrichment in chalcophile elements (Sb-As-Ag-Cd-S-Se-Hg) that represent sedimentary rocks and sites that have relative enrichment in Fe-Cr-Ni-Mg that reflect ultramafic rocks. This is also evident in Figure 8b where the mean values of the four terranes are shown.

In Figure 8c, the negative part of PC1 shows a relative association of U-Th-La-W-Tl-K-Bi-P that reflects felsic, mostly granitoid and metamorphic rocks. Figure 8d shows the mean values of the different regional rock types. The positive PC1–negative PC2 quadrant scores shows an association of Sb-As-Ag-Cd-S-Se-Hg-Au and reflects a dominantly chalcophile assemblage of elements. This portion of the biplot is dominated by clastic sedimentary rocks. The positive PC1–positive PC2 quadrant shows an association of siderophile elements (V-Cr-Ni-Co-Cu-Fe-Sc) and are coded as volcanic rocks. The Coast terrane ultramafic rocks and intermediate intrusive rocks are also displayed in this quadrant.

Figure 8e and Figure 8f display the PC scores that are coded according to the GroupModel designation. Sites with the GroupModel status as "Unknown" are not included in the figure. It is difficult to see specific trends in the dense cloud of points in Figure 8e. However Figure 8f shows the mean PC1-PC2 values for each of the GroupModels. GroupModels that have relative enrichment in U-Th-La-W-Tl-Bi-K-P are: carbonatite (N01), REE (O01O02O04), basal U (D04D06), W skarn (K05) and sediment-hosted Zn-Pb-Ag (E12E13E14). GroupModels that are associated with the siderophile elements (Fe-Ni-Cr-Co-Mg-Cu) include: hot spring-associated Au-Ag-Hg (G07H02H03), low-sulphidation epithermal Au-Ag (H05), mafic Ni-Cu-Cr (M01M02M03M05), sediment-hosted Cu-Pb (E01E04E05) and alkalic porphyry Cu-Au (L03). GroupModels that have an affinity with the chalcophile group of elements in the positive PC1 – negative PC2 quadrant include: porphyry Cu-Au-Mo (L02L04), subvolcanic Cu-Ag-Au (L01), Pb-Zn skarn (K02), Au quartz veins (I01), polymetallic Ag-Pb-Zn-Au (I05), massive sulphide (G04G05), volcanic-hosted Cu-Pb-Zn (G06), Au skarn (K04) and Mo porphyry (L05L08). GroupModels that plot near the origin of the biplot include Cu-Fe skarn (K01K03) and Cu-Ag quartz veins (I06). The position of the mean values of the GroupModels across the PC1-PC2 biplot suggest that there is reasonable contrast between the GroupModels that is reflected by the relative enrichment/depletion of the elements and provides a framework from which GroupModels can be predicted from the multi-element geochemistry.

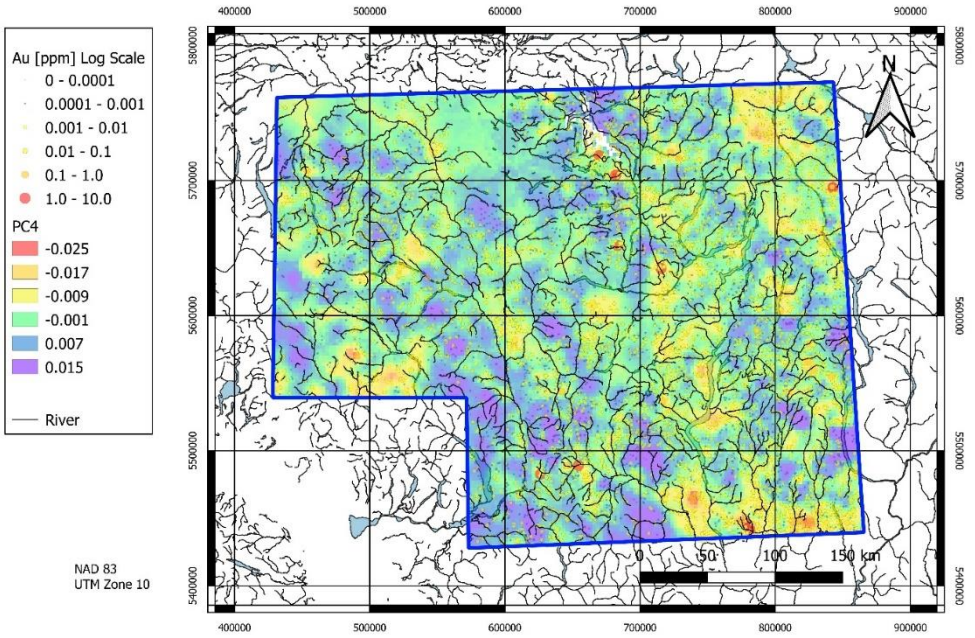
Figure 9 shows principal component biplots for PC3-PC4 and PC12-PC14. Figure 9a shows the principal component scores for PC3-PC4 coded with GroupModel symbols/colours (see Figure 6 for legend). Along the positive portion of the PC3 axis there is relative enrichment of both chalcophile (As-W-Bi-Sb), siderophile (Fe-Cr-Ni-V-Co-Mg) and lithophile (Th-K). The negative portion of the PC3 axis shows relative enrichment of lithophile

(Sr-Ca-Na-U-Ba) and chalcophile (S-Se-Hg-Cd) elements. The most significant association is along the negative axis of the PC4 which shows extreme relative enrichment of Au. A number of GroupModel types plot along the negative PC4 axis, including massive sulphides (G04G05), polymetallic veins (I05), Cu-Au-Mo porphyry (L02L04), Mo porphyry (L05L08), Au quartz veins (I01) and surficial placers (C01C04).

Figure 9b shows the mean values of PC3-PC4 for the GroupModels of which REE (O01O02O04), carbonatite (N01), Au quartz veins (I01), sediment-hosted Zn-Pb-Ag (E12E13E14) and Cu-Fe skarns (K01K03) dominate. Note that the scaling of the mean values has been changed to enhance the separation

The biplot of PC12-PC14 (Figure 9c) shows the relative enrichment of Cu along the positive PC14 axis and the negative PC12 axis. The sites identified with relative Cu enrichment are associated with L02L04 and L05L08 (porphyry Cu-Au-Mo and porphyry Mo) MINFILE designations. A biplot of the mean values of the PC12-PC14 scores for each of the GroupModel classes is given in Figure 9d for clarity. The position of the GroupModel mean symbols indicates relative enrichment and depletion of the elements with the GroupModels. Note that the scaling of the mean values has been changed to enhance the separation. The relative positions of the GroupModel icons do not match the scales of the biplot axes.

Kriged images, along with individual point scores for PC4 and PC12, are shown in



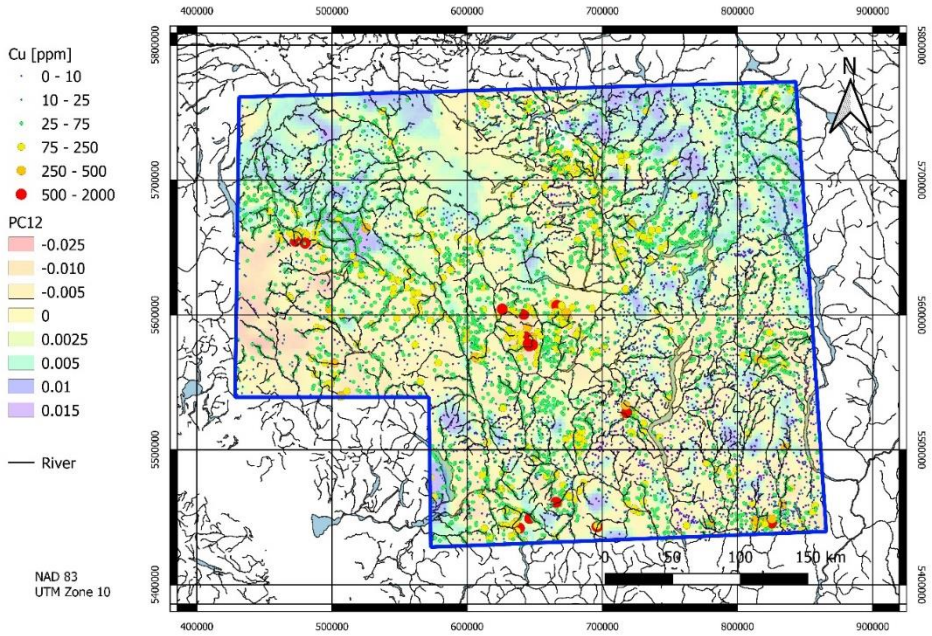
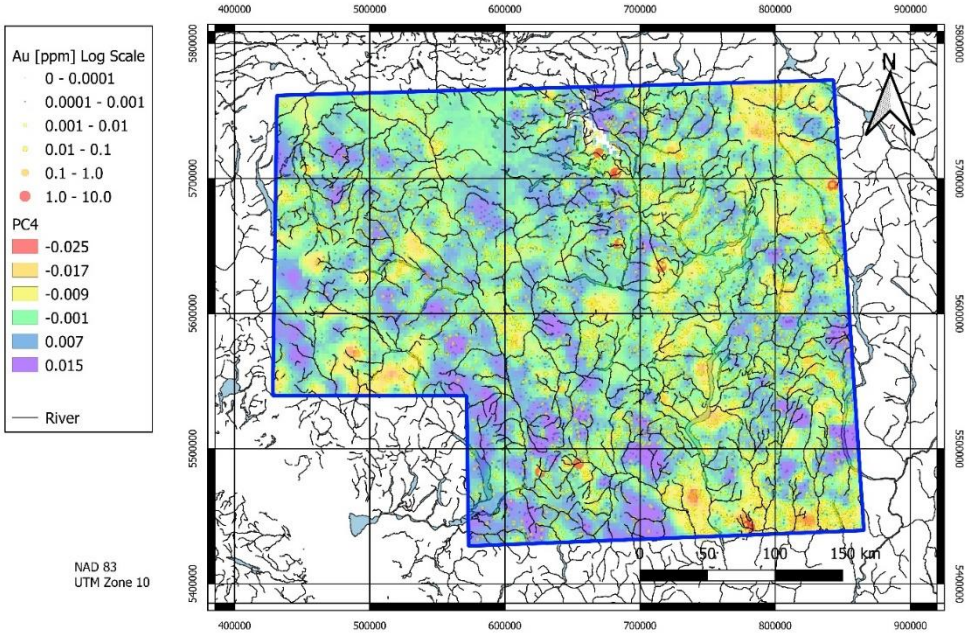


Figure 10. Regions of relative Au enrichment (



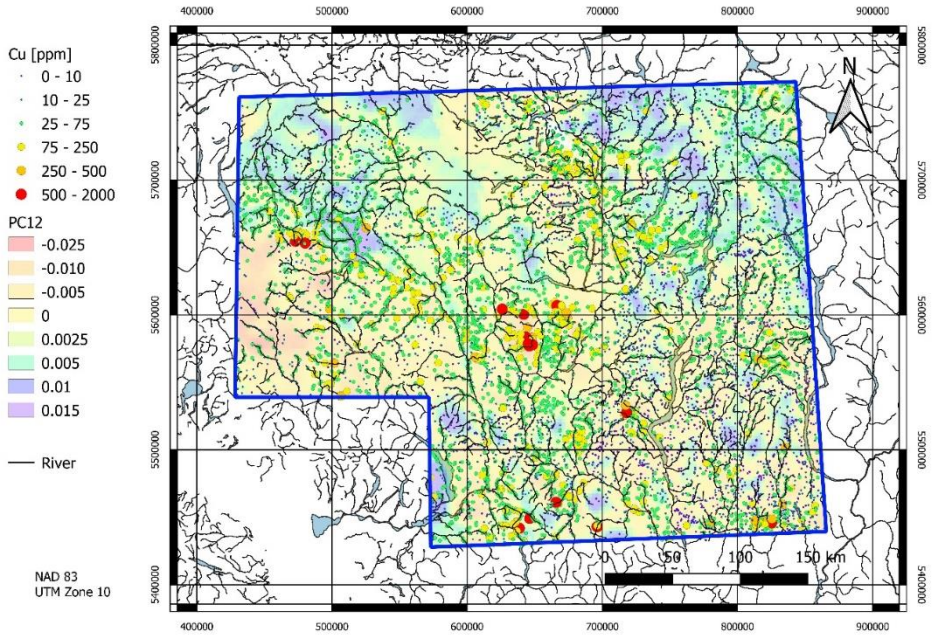
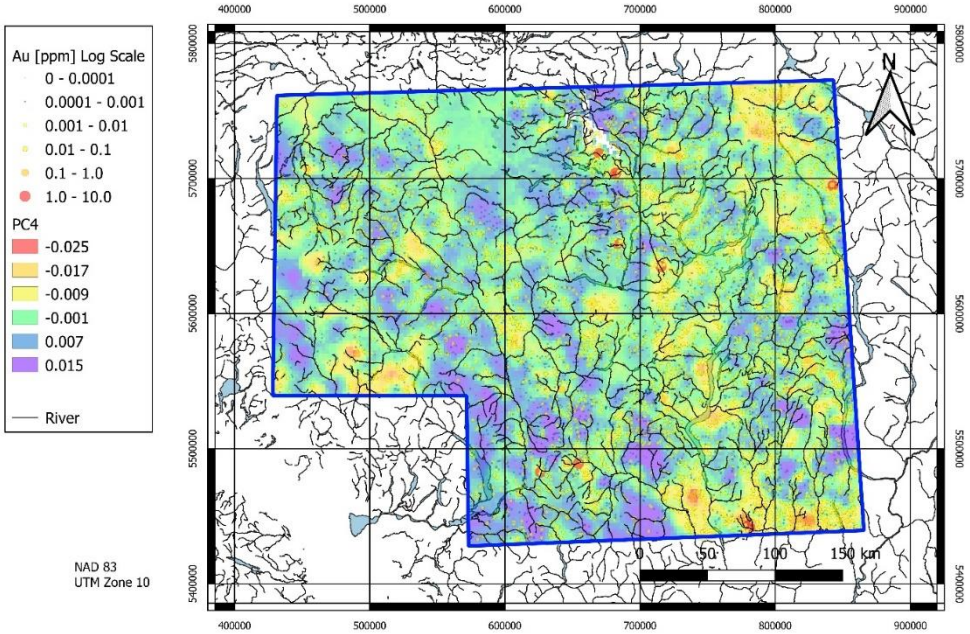


Figure 10a) and relative Cu enrichment (



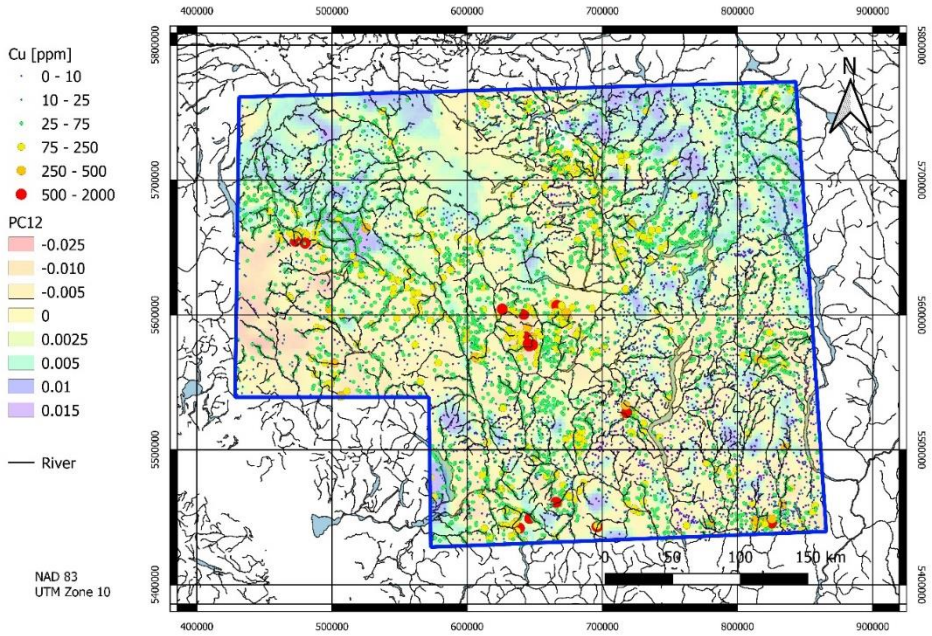
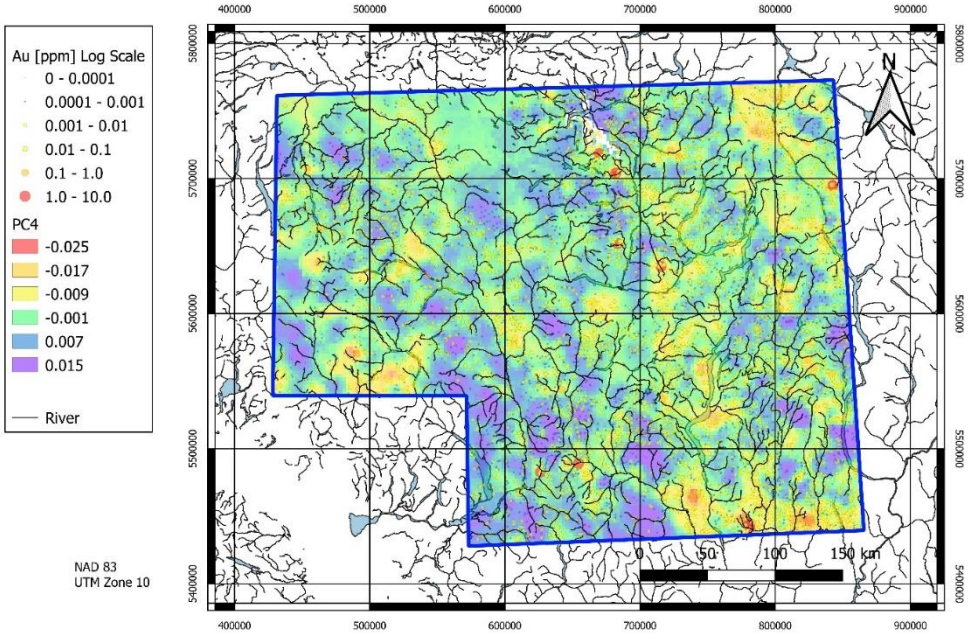


Figure 10b) are clearly visible on these maps. In



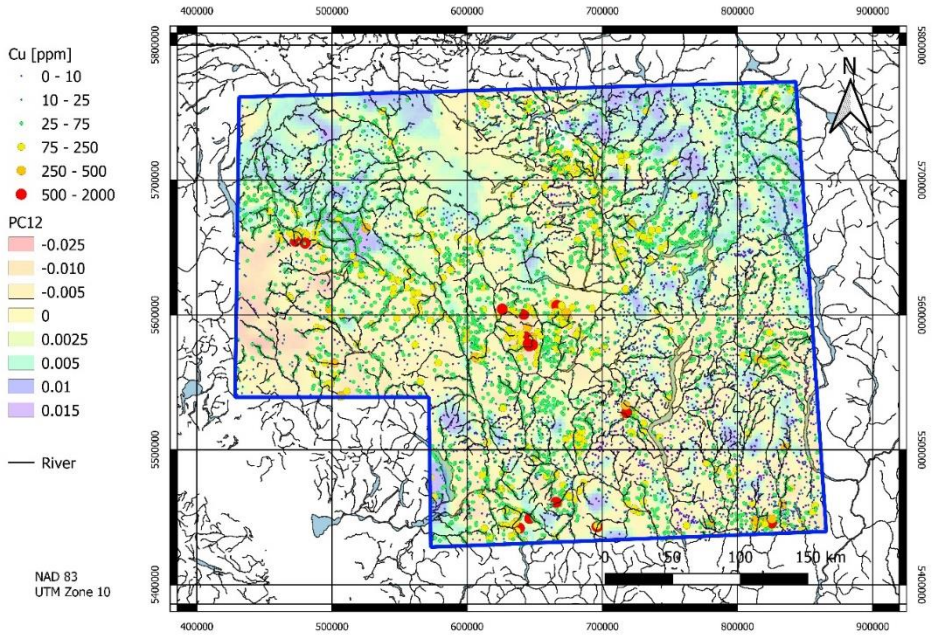
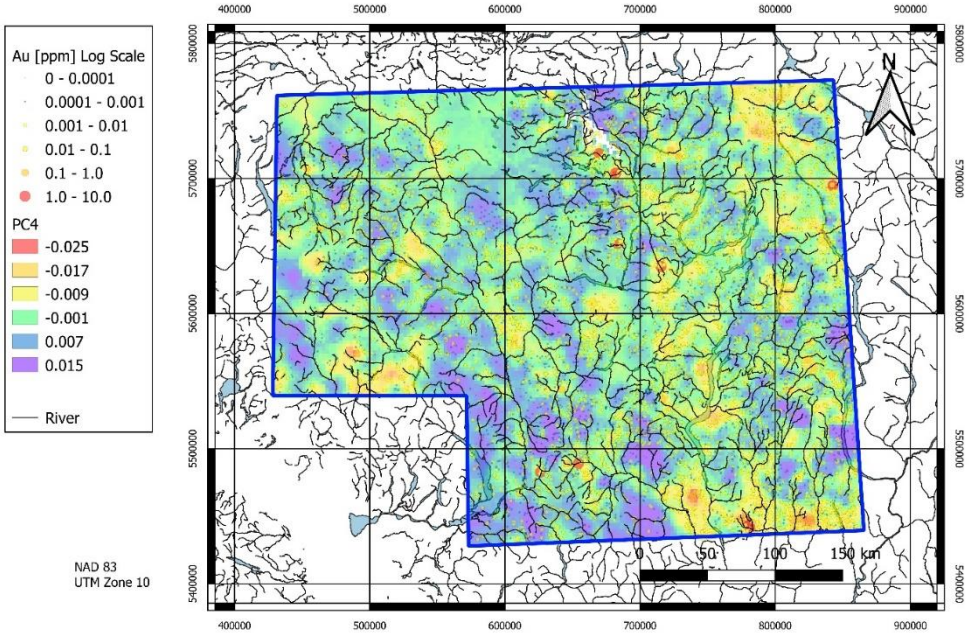


Figure 10a, PC4 indicates relative Au enrichment associated with negative (red) values. In



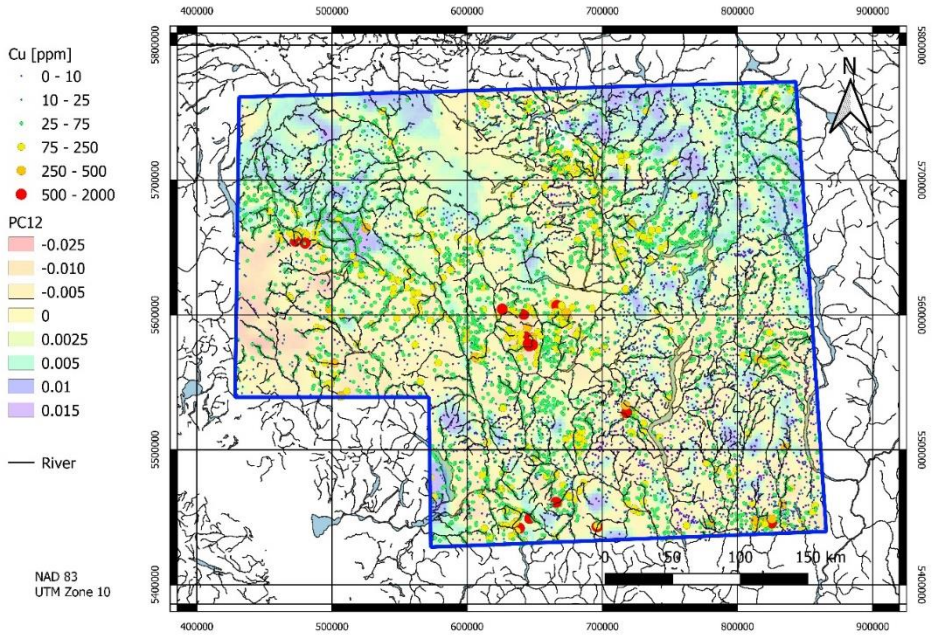


Figure 10b, PC12 indicates relative Cu enrichment associated with negative (red) values. In both figures, the trend of relative enrichment is not necessarily an indicator of mineralization, however the patterns reveal trends in the elements that may be associated with regions of higher mineralization potential.

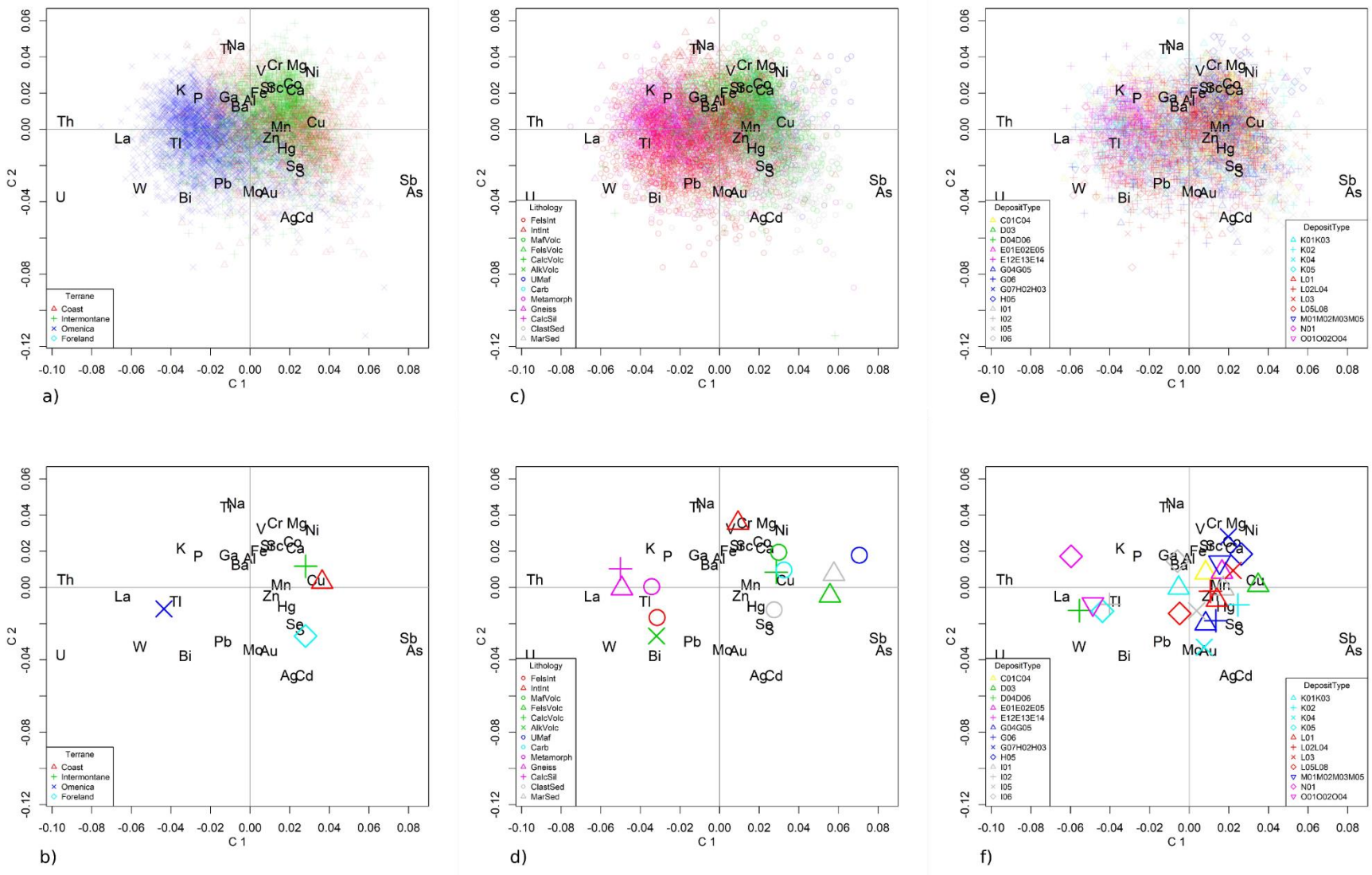

a)
b)
c)
d)
e)
f)
C 1
C 2
Terrane
Coast
Intermontane
Omenica
Foreland
Lithology
FelsInt
IntInt
MafVolc
FelsVolc
CalcVolc
AlkVolc
UMaf
Carb
Metamorph
Gneiss
CalcSil
ClastSed
MarSed
DepositType
C01C04
D03
D04D06
E01E02E05
E12E13E14
G04G05
G06
G07H02H03
H05
I01
I02
I05
I06
K01K03
K02
K04
K05
L01
L02L04
L03
L05L08
M01M02M03M05
N01
O01O02O04

Figure 8. a) Biplot of PC1-PC2 showing the relative relationships between the principal terranes in the QUEST-South region. The Omineca terrane shows relative enrichment in Th-U-La while the Coast terrane shows relative enrichment of Sb-As indicating provenance with chalcophile-rich rocks and relative enrichment in Th-U-La. The Intermontane terrane shows a mixture of lithophile, siderophile and chalcophile elements; **b)** Biplot of PC1-PC2 showing the mean values of the 2 components for each of the tectonic terranes. See text for detailed description. **c)** Biplot of PC1-PC2 showing the relative relationships between the generalized lithologies of the region. Felsic intrusive rocks show relative enrichment in Th-U-La and sedimentary rocks show relative enrichment in Sb-As. Volcanic rocks show a mixture of lithophile and siderophile elements; d) Biplot of PC1-PC2 showing the mean values of the two principal components for each of the regional rock types. **e)** Biplot of PC1-PC2 showing the relative relationships between the GroupModels as defined from the proximity of stream-sediment sites and MINFILE sites; **f)** Biplot of PC1-PC2 showing the mean values of the PC1-PC2 scores for each of the GroupModel classes. Relative relationships between the GroupModels are defined from the proximity of stream-sediment sites and MINFILE sites. Epithermal Au-Ag deposits show relative enrichment with siderophile elements. Porphyry deposits show relative enrichment with chalcophile elements. Carbonatite, REE, basal U, W skarn and sediment-hosted deposits show relative enrichment with U-Th-La-W-Tl. Note that the scaling of the mean values has been changed to enhance the separation. The relative positions of the GroupModel icons do not match the scales of the biplot axes.

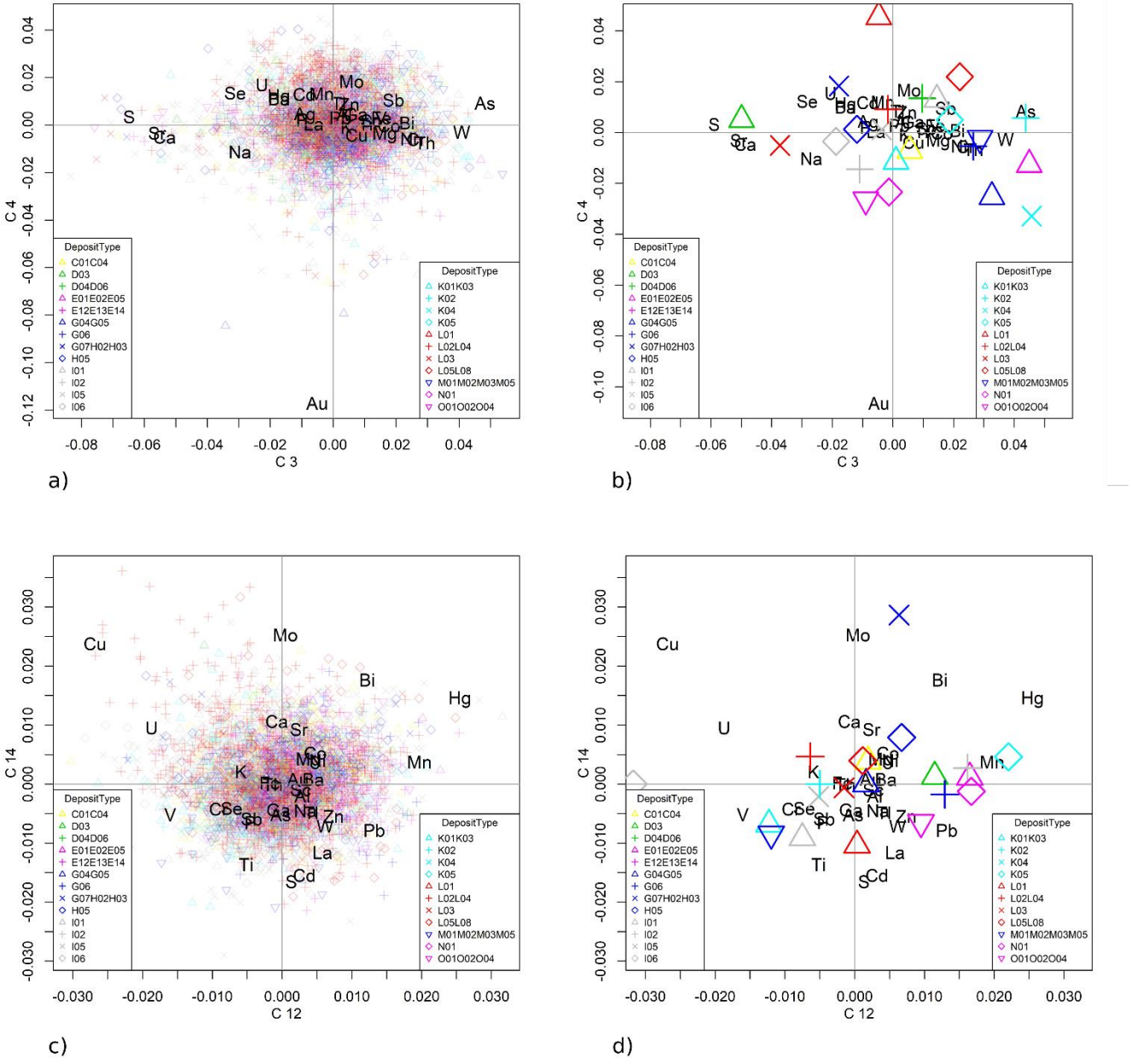
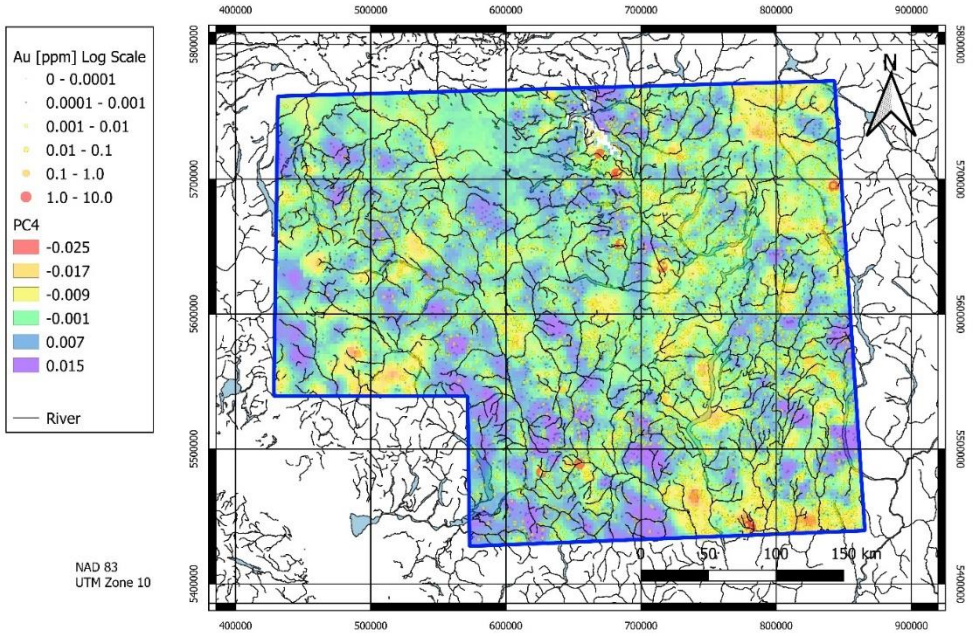


Figure 9. a) Biplot of PC3-PC4 showing the relative enrichment of Au along the negative PC4 axis. **b)** Biplot of PC1-PC2 showing the mean values of the PC1-PC2 scores for each of the GroupModel classes. Relative relationships between the GroupModels are defined from the proximity of stream-sediment sites and MINFILE sites. GroupModels that show a relative increase in Au include: REE, Carbonatite, Massive sulphides, Au skarn, Au quartz veins, Cu-Fe skarn and sediment-hosted deposits; **c)** Biplot of PC12-PC14 showing the relative enrichment of Cu along the positive PC14 axis and the negative PC12 axis. The sites identified with relative Cu enrichment are associated with L02L04 and L05L08 (porphyry Cu, Mo) MINFILE designations; **d)** Biplot of PC12-PC14 showing the mean values of the PC12-PC14 scores for each of the GroupModel classes. The position of the GroupModel mean symbols indicates relative enrichment and depletion of the elements with the GroupModels. Note that the scaling of the mean values has been changed to enhance the separation. The relative positions of the GroupModel icons do not match the scales of the biplot axes.

a)



b)

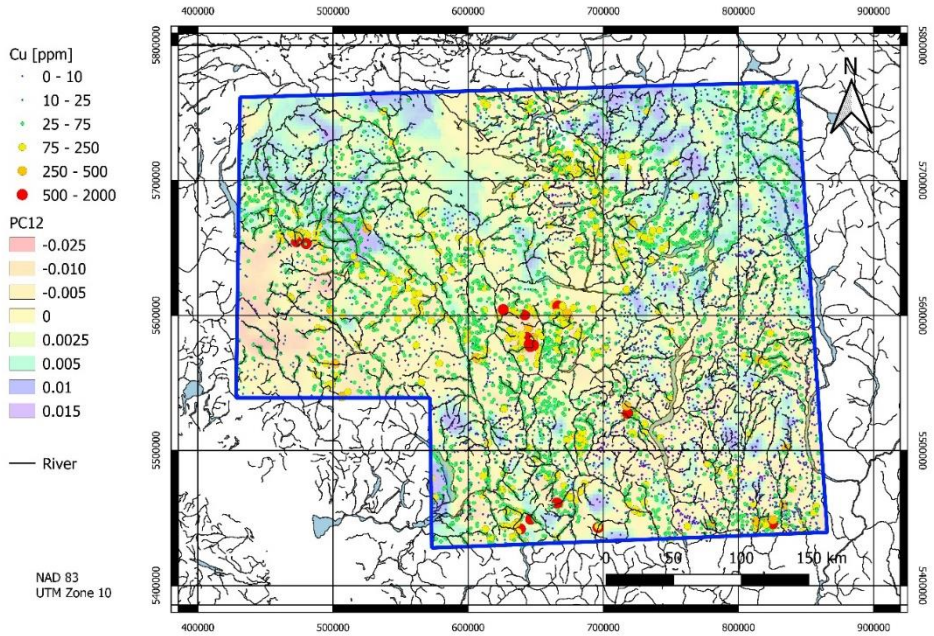


Figure 10. Geographic distribution of **a**) Au plotted over a kriged image map of principal component 4. The relative increase in Au shown on the map corresponds with the relative increase in Au along the negative PC4 axis in Figure 9a; and b) of Cu plotted over a kriged image map of principal component 12. The relative increase in Cu shown on the map corresponds with the relative increase in Cu along the negative PC12 axis in Figure 9b.

t-SNE Process Discovery

The use of t-SNE coordinates requires some careful investigation in determining the optimum parameters that best describe the distribution of the data in a reduced dimensional space derived from the initial 35 multi-element space. Based on the previous experience with the use of the principal component metric, a 9-dimenional t-SNE space was used with a perplexity constant of 0.75. Perplexity is a value that is used to balance the effects of local versus global features in data. It is a tuning parameter that requires some experimentation. The value used in this study was determined after testing several measures of perplexity. Figure 11a shows a scatter plot of the stream-sediment sites tagged with the coordinates for t-SNE3 vs. t-SNE9. These two coordinates were used based on the application of the random forests prediction methodology, which is explained below. The points are coded with coloured symbols that represent the unique rock types over the area. Figure 11a reveals that the rock types segregate into different groups across the plot. Volcanic and sedimentary rocks dominate the upper portion of the plot; intrusive rocks occur along the positive t-SNE3 axis and metamorphic rocks occur along the negative t-SNE9 axis. Carbonate and carbonate-bearing sedimentary rocks occur long the negative t-SNE3 axis. The plot also shows significant overlap between the classes. The positive t-SNE3 – positive t-SNE9 quadrant shows a cluster of carbonate rocks that occur between the domain of volcanic/sedimentary rocks and intrusive rocks. These carbonate rocks may represent a mix of rock types but were labelled as carbonate. The mean values of t-SNE3 and t-SNE9, for each of the lithologies, are also plotted on the figure and show the distinct compositional differences between the different rock types.

Figure 11b shows the t-SNE3 vs. t-SNE9 plot with symbols labelled according to the GroupModel designation for the stream-sediment sites. The sites labelled as “Unknown” are plotted at a reduced symbol size. The plot shows many clusters of nearly unique GroupModels. The mean values of the GroupModels are shown in Figure 11c. The mean values show that the GroupModels occur in three clusters. The groups: carbonatite (N01), REE (O01O02O04), sediment-hosted Zn-Pb-Ag (E12E13E14) and Intrusion-related Au(I02) plot along the positive t-SNE3 axis. Mean values for porphyry deposits (L01L02L04L05L08) plot along the positive t-SNE9 axis along with Au, W, Cu-Fe and Pb-Zn skarn deposits (K03K05K01K03K02). Volcanic-hosted Cu deposits (D03) plot along the negative t-SNE3 axis. Hot spring-associated Au-Ag-Hg (G07H02H03), low sulphidation epithermal Au-Ag (H05), alkalic porphyry Cu-Au (L03), massive sulphide (G04G05), and sediment-hosted Pb-Zn-Ag (E01E04E05) define a trend along the negative t-SNE3-negative t-SNE9 quadrant. The trend of mean values across the t-SNE3 vs. t-SNE9 scatterplot provides evidence that there is a reasonable distinction between the different GroupModels that will permit a predictive classification scheme to validate the existing GroupModels and predict the stream-sediment sites that are tagged as “Unknown”.

Maps of the two dominant t-SNE coordinates for t-SNE3 and t-SNE9 are shown in Figure 12. Both plots show broad spatially coherent regions that reflect the regional geology of the area. There is no evidence of specific

GroupModel regions in these plots, although the broad patterns likely reflect significant geochemical differences between these regions.

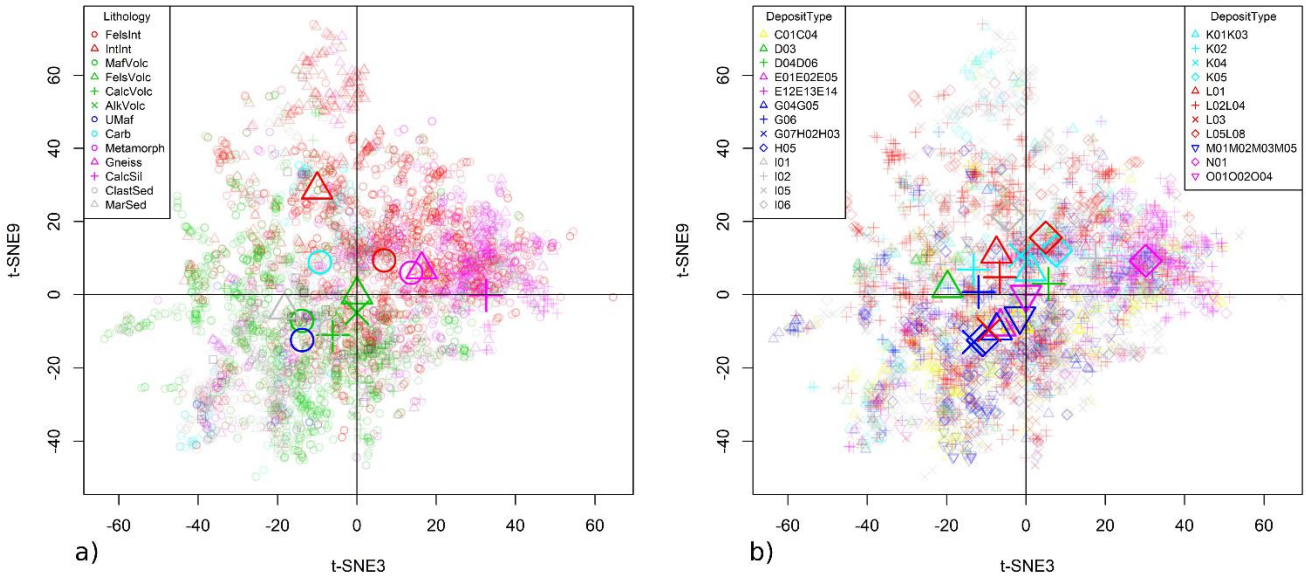
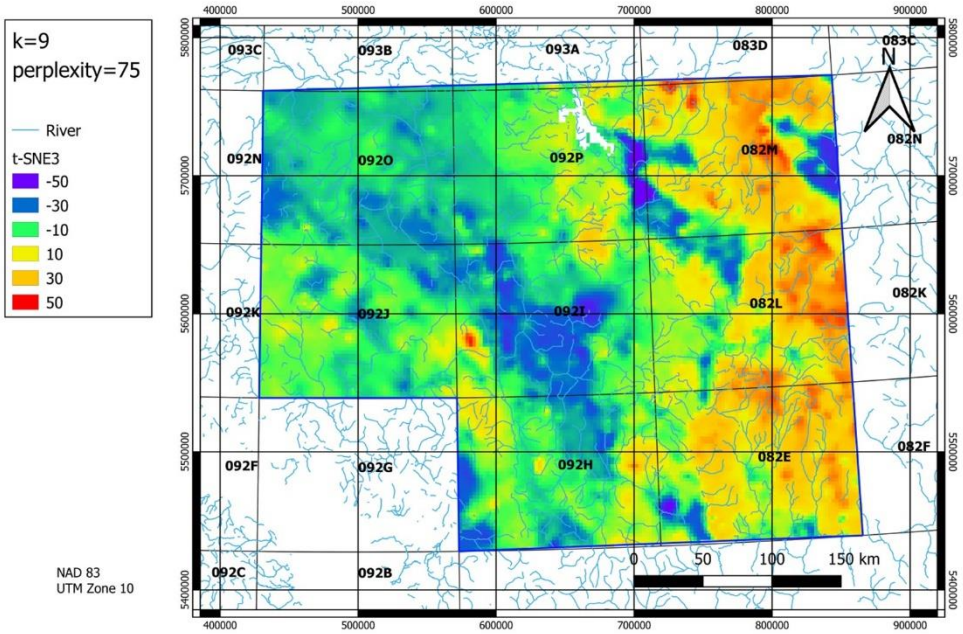


Figure 11. a) Scatter plot of t-SNE3 vs. t-SNE9 coded by rock type. The plot shows distinct groupings of the rock types. The mean values of the lithologies are also plotted and show the distinctive compositional differences between the lithologies. No scaling factor was applied; **b)** Scatter plot of t-SNE3 vs. t-SNE9 coded by GroupModel. The plot shows some distinct groupings of the GroupModels but also considerable overlap between many of them. Sites that have a GroupModel class as “Unknown” are not shown; Scatter plot of the mean values of t-SNE3 vs. t-SNE9 coded by GroupModel are also shown. The symbols show a distinct trend of the GroupModel means. Porphyry and skarn deposits tend to cluster together. These two t-SNE coordinates provide the best discrimination between GroupModels using the random forests classification method. See the text for more details.

a)



b)

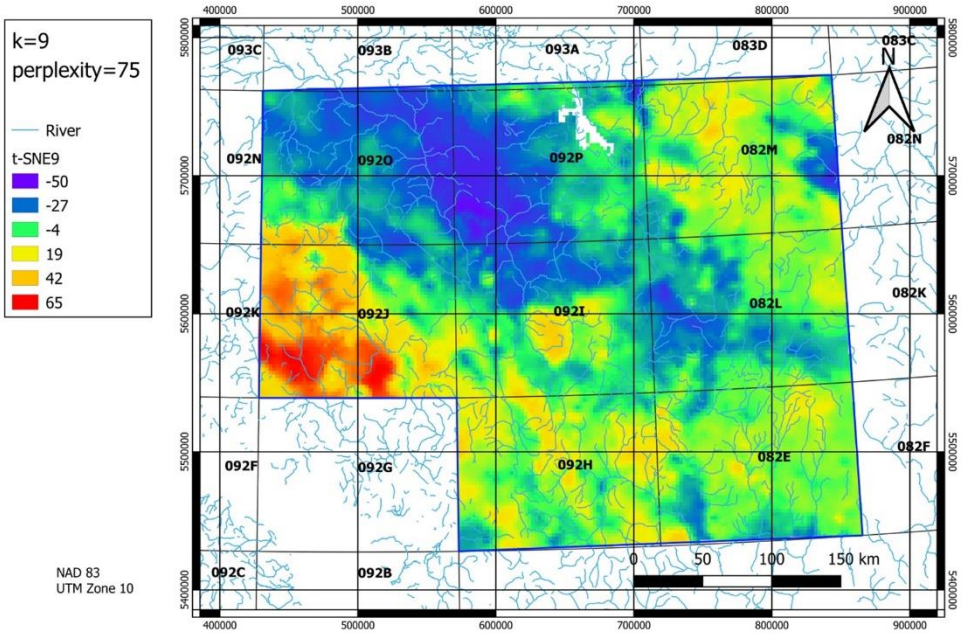


Figure 12. Kriged images of the two dominant t-SNE coordinates computed in a 9 dimensional t-SNE space. **a)** of t-SNE3; and **b)** Kriged image of t-SNE9.

Process Validation

Random Forests GroupModel Prediction Based on a Principal Component Metric

The random forests function 'randomForest' (package randomForest for R; Breiman, 2001) was used to predict a GroupModel classification based on the training set of 474 sites using the distance threshold of 2500 m. Unlike other methods of classification (e.g. linear discriminant analysis), one advantage of the random forests process is that *a priori* selection of variables is not required. The procedure starts with all the variables (PC1–PC35) and then reduces the number of variables to those that provide the best node separation in the trees that are generated. Figure 13 shows the significance of the variables derived from the random forest procedure. The significance is measured by the 'Mean Decrease in Gini'. This measure of variable importance is based on the 'node impurity' (i.e., the rate of misclassification). Lower rates of misclassification correspond to higher values of the Gini index. The figure indicates that PC1 is by far the most significant variable, followed by PC6 and PC11. The remaining variables show a monotonic decrease in significance until the seventeenth place, where an inflection point occurs, and the remaining values are essentially insignificant. It should be noted that the method of random forests does not require any cross-validation with the data as this is implicitly done during the construction of the trees.

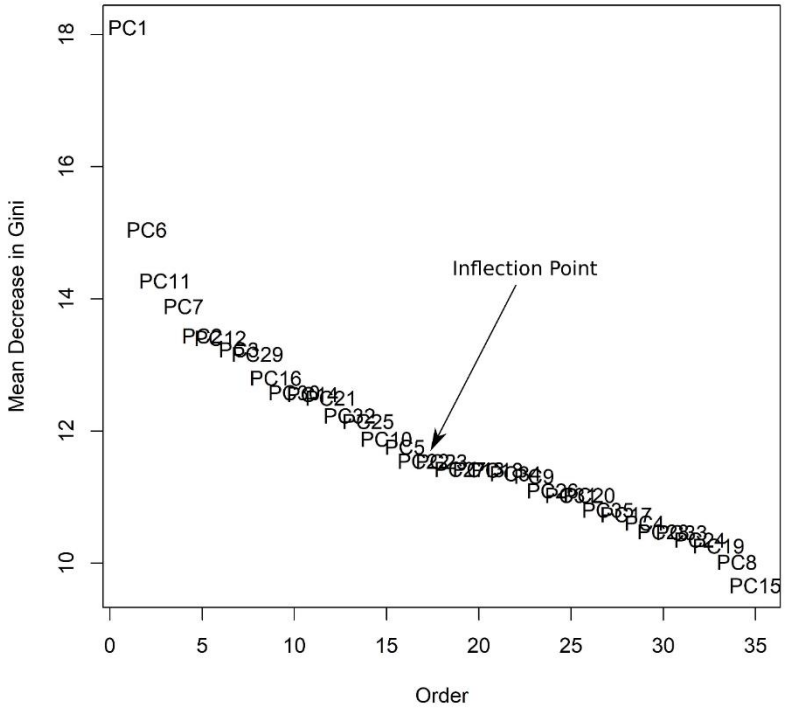


Figure 13. Plot of 'Mean Decrease in Gini' for the principal components used in the application of random forests prediction based on the training data for mineral occurrences within 2500 m.

Table 6 shows the accuracy of classification in terms of percentage, based on the training data set only. The overall classification accuracy is 37.55% when Model I05 (polymetallic veins) is excluded from the modelling runs. Several of the GroupModel classes show a classification accuracy of zero, including volcanic-hosted Cu (D03), sediment-hosted Cu-Pb (E01E04E05), massive sulphide (G04G05), hot spring-associated Au-Ag-Hg (G07H02H03), Pb-Zn skarn (K02), and W skarn (K05). The confusion among these GroupModels is likely due to significant overlap of their geochemical signatures with those of other GroupModels and too few sites for each of the classes. As discussed previously, the choice of the distance threshold between a stream-sediment site and a MINFILE site also has an influence in the outcome of the predictions. Despite the low accuracies shown in Table 6, the kriged images of the posterior probabilities may display geospatial continuity and provide some insight into the prospectivity of deposits defined by the GroupModel classes.

The random forests procedure estimates posterior probabilities for each GroupModel at each stream-sediment site. The assigned class is selected from the GroupModel with the highest posterior probability. A predictive map of the posterior probabilities can therefore be created for each GroupModel class. Areas of contiguous elevated posterior probabilities for a given class define the ‘geospatial coherence’ of a GroupModel. It is expected that the maps of posterior probability will show overlap because of compositional overlap between the classes. This also results in lower values of posterior probabilities for all the GroupModels due to the overlap. Also, because of compositional overlap, the posterior probabilities for many GroupModels can be very low. However, geospatial coherence in the interpolated image for a given GroupModel increases the potential that the area is associated with that GroupModel. A given stream-sediment site could have nearly equal posterior probabilities for several GroupModels. This increases the confusion and resulting overlap in the classification and, in the cases where there is geospatial coherence for several GroupModels in the same area(s), further investigation is required to determine which GroupModel is most feasible.

Validation of Predictions Against GroupModels

In this section, kriged images of posterior probabilities from random forests using PCA are compared to GroupModel occurrences in order to visually validate the predictions. For the sake of clarity and brevity, only one GroupModel, L02L04 (Porphyry Cu-Au-Mo) is presented. Maps of the predicted GroupModels are shown in Appendices 2 and 3. Appendices 2 and 3 show maps for kriged posterior probabilities for each GroupModel based on the PCA and t-SNE metric, respectively, based on the locations of the stream-sediment sites. Appendix 4 shows maps where the catchment areas are assigned the posterior probability of the stream-sediment site located within each catchment for the t-SNE metrics, only for those GroupModels with measurable probabilities.

Table 6. Accuracy matrix for the GroupModels training set, derived from the application of random forests classification using PCA. Values in bold type are the percent accuracies of predictions for a particular GroupModel using the training data set.

% Accuracy																						
	C01C04	D03	E01E04E05	E12E13E14E15	G04G05	G06	G07H02H03	H05	I01	I02	I06	K01K03	K02	K04	K05	L01	L02L04	L03	L05L08	M01M02M03M05	Unknown	class.error
C01C04	61.45	0	0	0	0	2.41	0	0	0	0	0	0	0	0	0	0	6.02	0	0	1.20	28.92	0.39
D03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60.00	0	0	0	40.00	1
E01E04E05	33.33	0	0	0	0	0	0	0	33.33	0	0	0	0	0	0	0	0	0	0	0	33.33	1
E12E13E14E15	9.09	0	0	22.73	0	4.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63.64	0.77
G04G05	36.36	0	0	0	9.09	0	0	0	9.09	0	0	9.09	0	0	0	0	9.09	0	0	0	27.27	0.91
G06	37.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25.93	0	0	0	37.04	1
G07H02H03	0	0	0	0	0	0	0	0	16.67	0	0	0	0	0	0	0	0	0	0	0	83.33	1
H05	11.11	0	0	0	0	0	0	33.33	5.56	0	0	0	0	0	0	0	5.56	0	0	0	44.44	0.67
I01	34.15	0	0	0	0	2.44	0	0	26.83	0	0	0	0	0	0	0	4.88	0	0	0	31.71	0.73
I02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00	1
I06	40.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40.00	0	0	0	20.00	1
K01K03	21.74	0	0	0	0	0	0	4.35	4.35	0	0	8.70	0	0	0	0	13.04	4.35	0	0	43.48	0.91
K02	50.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50.00	1
K04	22.22	0	0	0	0	0	0	0	11.11	0	0	0	0	22.22	0	0	0	0	0	0	44.44	0.78
K05	0	0	0	0	0	0	0	0	16.67	0	0	0	0	0	0	0	33.33	0	0	0	50.00	1
L01	16.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66.67	0	0	0	16.67	1
L02L04	13.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54.10	0	0	0	32.79	0.46
L03	26.67	0	0	0	0	0	0	0	6.67	0	0	13.33	0	0	0	0	26.67	6.67	0	0	20.00	0.93
L05L08	7.69	0	0	0	0	0	0	0	7.69	0	0	0	0	0	0	0	30.77	0	0	0	53.85	1
M01M02M03M05	58.82	0	0	0	0	0	0	0	5.88	0	0	0	0	0	0	0	0	0	0	0	35.29	1
Unknown	21.00	0	0	0	0	1.00	0	0	2.00	0	0	0	0	0	0	0	10.00	0	0	0	66.00	0.34

Figure 14 shows a predictive map of the combined porphyry deposit models for Cu-Au-Mo porphyry (L02L04). The map of the posterior probabilities is less than 0.5 with many regions showing elevated values in the range of 0.25. There are clusters of L02L04 sites in the vicinity of currently producing mines in the centre of NTS area 092I. Additional sites that are identified by MINFILE sites and classed as L02L04 by random forests are shown in NTS areas 092H 082E, 092P and 092O. The kriged image of the posterior probabilities coincides with both the MINFILE sites and the predicted classes. The elevated zone of posterior probabilities and class predictions is broad across the western and southern part of the map area. There are many more L02L04 class predictions than MINFILE sites.

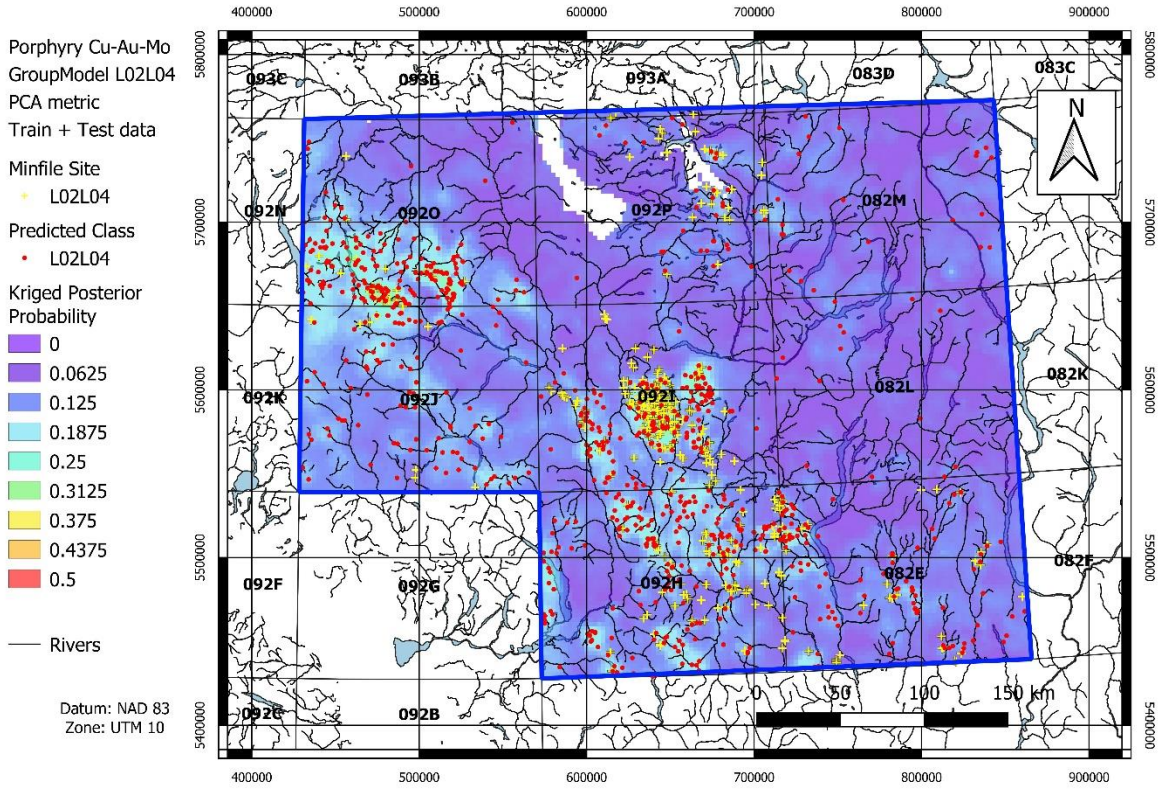


Figure 14. Geographic distribution of individual sites for GroupModel porphyry Cu-Au-Mo (L02L04) overlain on a kriged image of the posterior probabilities for porphyry Cu-Au-Mo (L02L04) prediction using random forests and the PCA metric, based on the test data and a distance threshold of 2500 m. MINFILE sites tagged as L02L04 are shown as yellow crosses. Stream-sediment sites identified as class L02L04 by random forests are shown as red dots. Areas of increased potential for L02L04 deposits are shown by colour shading of the kriged image.

Random Forest GroupModel Prediction Based on a t-SNE 9-Dimensional Metric

The results of the application of random forests to the t-SNE 9-dimensional dataset are shown in Figure 15 and Table 7. Figure 15 shows that the most significant variables are t-SNE3, t-SNE9 and t-SNE4. The diagonal of Table 7 shows the percentage of prediction accuracy for each of the GroupModel classes. The off-diagonal elements of the matrix show where there is uncertainty/confusion in the class assignment. The overall accuracy is shown as 43.25%, which is slightly better than the overall accuracy based on PCA of 37.55%. The two metrics show very different rates of class confusion that is more clearly expressed in Figure 16, which plots the accuracy of prediction for each GroupModel for both the PCA- and t-SNE-based metrics. The figure shows that overall, the t-SNE metric results in higher prediction accuracy for all of the GroupModel classes. The increased prediction accuracy is also reflected in the kriged images of the posterior probabilities for several of the GroupModels.

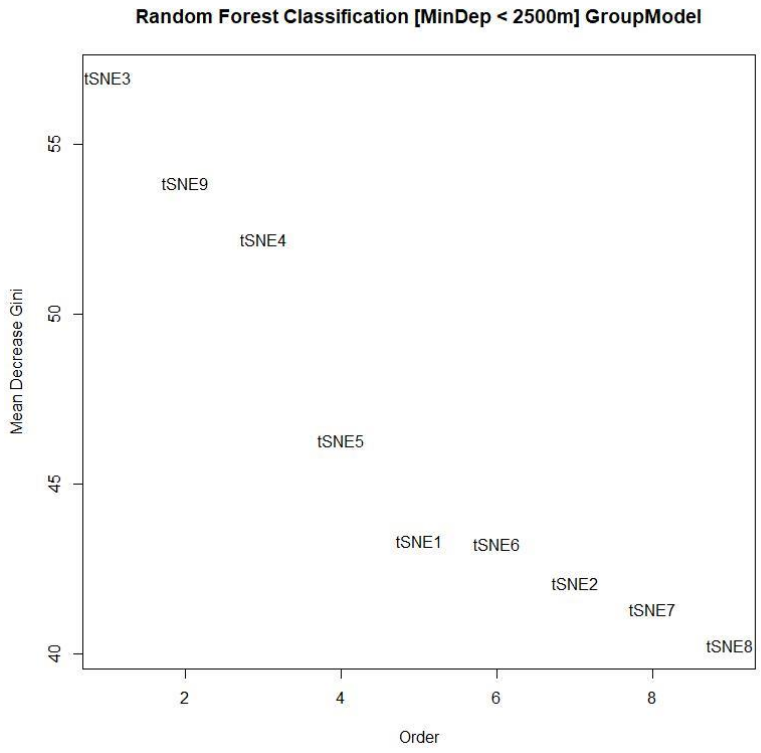


Figure 15. Plot of Mean Decrease in GINI for the t-SNE coordinates used in the application of random forests prediction; based on the training data. t-SNE3 is the most significant variable for classification. t-SNE8 is the least significant variable.

Table 7. Accuracy matrix for the GroupModels training set, derived from the application of random forests classification using t-SNE9. Values in bold type are the percent accuracies of predictions for a particular GroupModel using the training data set.

Accuracy %																					
	C01C04	D03	E01E04E05	E12E13E14E15	G04G05	G06	G07H02H03	H05	I01	I02	I06	K01K03	K02	K04	K05	L01	L02L04	L03	L05L08	M01M02M03M05	Unknown	class.error
C01C04	54.22	0	0	1.20	0	2.41	0	2.41	10.84	0	0	0	0	0	0	0	6.02	1.20	0	3.61	18.07	0.46
D03	0	0	0	0	0	0	0	0	0	0	0	20.00	0	0	0	0	60.00	0	0	0	20.00	1.00
E01E04E05	0	0	0	0	0	0	0	0	66.67	0	0	0	0	0	0	0	0	0	0	0	33.33	1.00
E12E13E14E15	4.55	0	4.55	54.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36.36	0.45
G04G05	9.09	0	0	0	45.45	0	0	9.09	9.09	0	0	0	0	0	0	0	0	0	0	0	27.27	0.55
G06	14.81	0	0	0	0	33.33	0	0	3.70	0	0	3.70	3.70	0	0	0	11.11	0	0	0	29.63	0.67
G07H02H03	0	0	0	0	0	0	0	0	16.67	0	0	0	0	0	0	0	0	0	0	0	83.33	1.00
H05	11.11	0	0	5.56	0	0	0	61.11	11.11	0	0	0	0	0	0	0	5.56	0	0	0	5.56	0.39
I01	17.07	0	2.44	0	2.44	2.44	2.44	2.44	39.02	0	0	0	2.44	0	0	0	2.44	2.44	0	2.44	21.95	0.61
I02	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
I06	0	0	0	0	0	0	0	0	0	0	20.00	0	0	40.00	0	0	0	0	0	0	40.00	0.80
K01K03	17.39	0	0	0	0	0	0	0	0	0	0	43.48	0	0	0	0	0	17.39	0	4.35	17.39	0.57
K02	0	0	0	0	0	50.00	0	0	50.00	0	0	0	0	0	0	0	0	0	0	0	0	1.00
K04	11.11	0	0	0	0	0	0	0	11.11	0	0	0	0	55.56	0	0	11.11	0	0	0	11.11	0.44
K05	0	0	0	0	0	0	0	0	16.67	0	0	0	0	0	0	0	16.67	16.67	0	0	50.00	1.00
L01	16.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66.67	0	0	16.67	0	1.00
L02L04	4.92	0	0	0	0	1.64	0	0	0	0	0	1.64	0	0	0	4.92	62.30	1.64	4.92	0	18.03	0.38
L03	0	0	0	0	0	0	0	0	6.67	0	0	33.33	0	0	13.33	0	13.33	20.00	0	0	13.33	0.80
L05L08	15.38	0	0	0	0	0	0	0	15.38	0	0	0	0	0	7.69	0	0	7.69	46.15	0	7.69	0.54
M01M02M03M05	23.53	0	0	0	0	0	0	0	0	0	0	5.88	0	0	0	0	5.88	0	0	47.06	17.65	0.53
Unknown	15.00	0	1.00	6.00	2.00	7.00	2.00	1.00	6.00	0	1.00	0	0	1.00	3.00	0	17.00	1.00	1.00	0	36.00	0.64

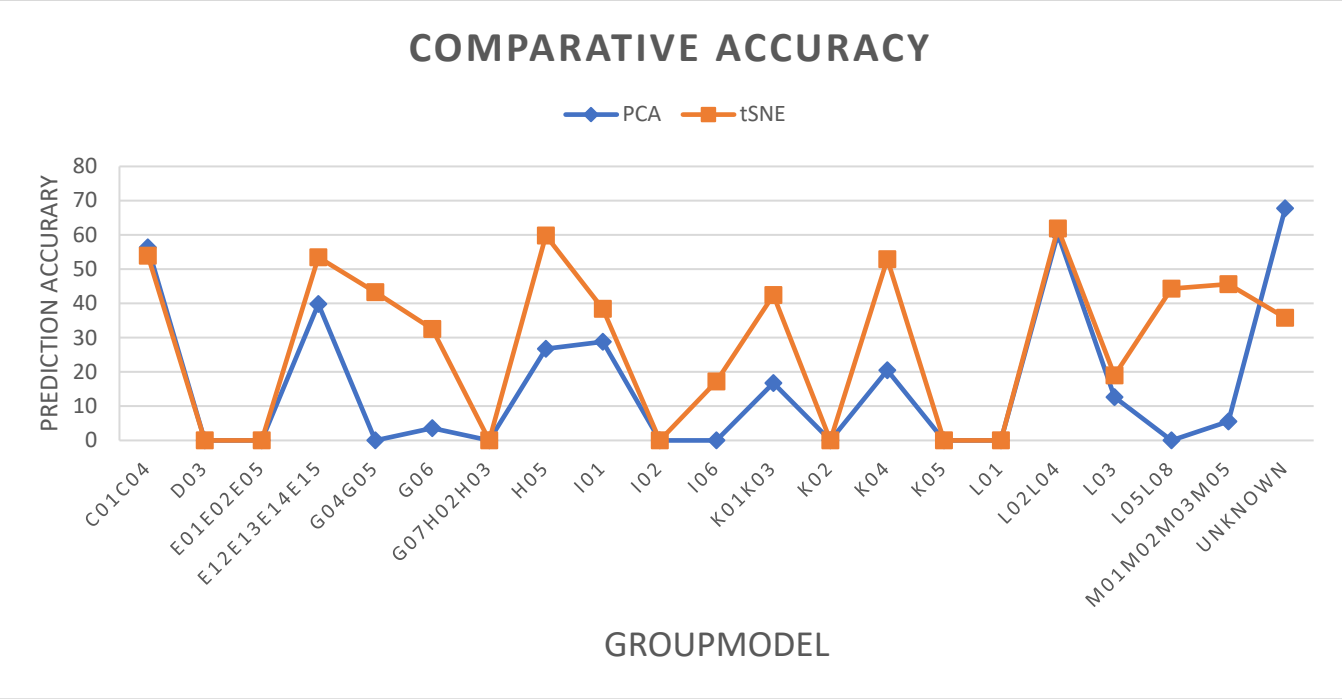


Figure 16. Plot of the accuracy of prediction for each GroupModel for both the PCA- and tSNE-based metrics. The tSNE metric outperforms the PCA metric.

Figure 17 show a map of the kriged posterior probabilities predicted for GroupModel L02L04 (porphyry Cu-Au-Mo) using the t-SNE metric. The MINFILE sites are shown as yellow crosses and the predicted classes for each of the stream-sediment sites are shown as red dots. There is a broad region of elevated posterior probabilities that coincides with the L02L04 class predictions and numerous MINFILE sites. The highest posterior probabilities and class prediction occur in the central part of NTS map sheet 092I, which is well known for its porphyry-style deposits.

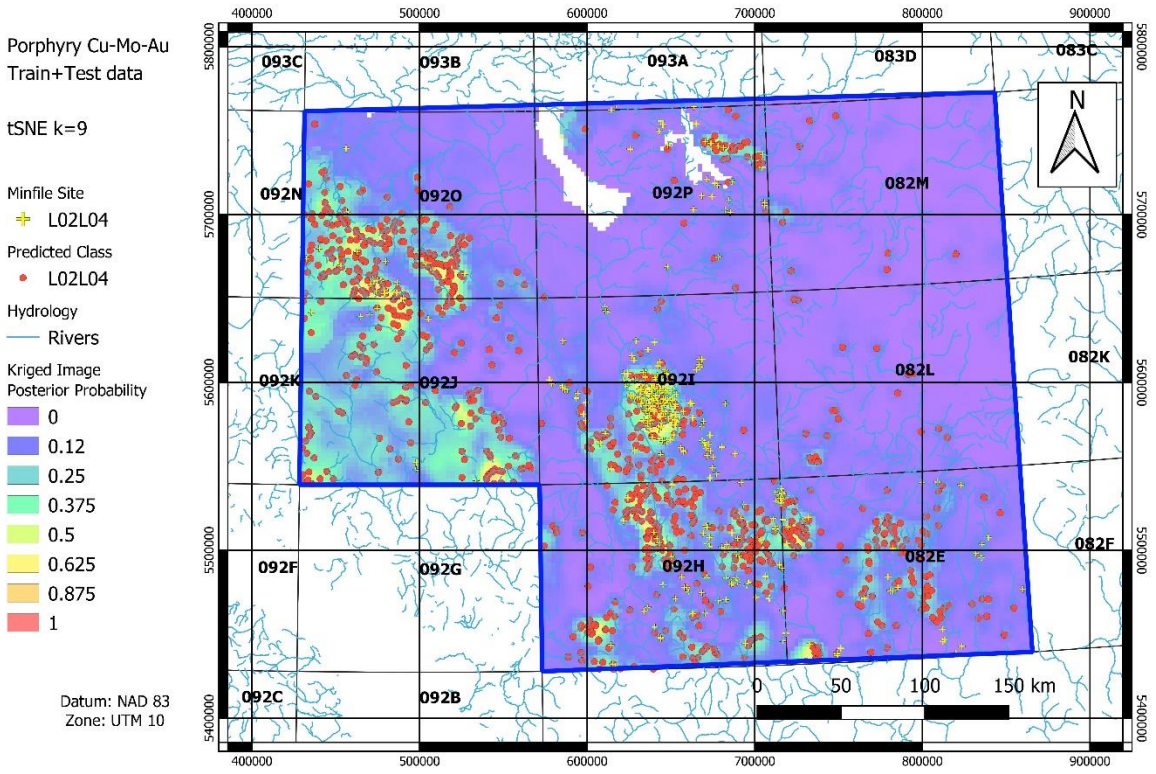


Figure 17. Geographic distribution of individual sites for GroupModel porphyry Cu-Mo-Au (L02L04) overlain on a kriged image of the posterior probabilities for porphyry Cu-Mo-Au (L02L04) prediction using random forests and the t-SNE9 metric, based on the train + test data and a distance threshold of 2500 m. MINFILE sites tagged as L02L04 are shown in yellow crosses. Stream-sediment sites identified as class L02L04 by random forests are shown in red dots. Areas of increased potential for L02L04 deposits are shown by map colour shading of the kriged image.

Comparisons with Conventional Approaches

Residuals from multiple regression analysis of Cu and Mo as a function of the most common rock types found within the catchments were calculated for the samples. In addition, residuals following regression of Cu and Mo against PC1 and PC2, respectively, both of which are dominated by rock type controls, were also calculated. Both approaches generated residuals that highlighted areas of data within the upper 98th percentiles of the respective data sets that are similar to those evident in the raw data (Figure 18 and Figure 19), but the residuals from multiple regression analysis against common rock types provide a clearer definition of trends and so are preferred over the residuals from regression against principal components. Given that there is also a positive correlation of both Cu and Mo with both Fe and Mn, the residuals from multiple regression analysis were also regressed against Fe to correct for possible metal scavenging by secondary hydroxide minerals. This approach is similar to the levelling of residuals employed by Arne and Bluemel (2011) and results in a tighter spatial association between the highest residuals and the locations of known porphyry Cu-Au-Mo deposits, in the case of Cu.

The multiple regression residuals corrected for a positive correlation with Fe were used to generate an additive index based on 1xMo residual plus 2xCu residuals, reflecting the relative importance of Cu in most porphyry deposits (Figure 20). As an alternative model, the multiple regression residuals were used in a weighted sums model with the following elements (followed by importance, or ”weights”, in parentheses): Cu (2), Mo (1), Fe (-2). The inclusion of Fe with a negative importance is designed to minimize the potential effects of metal scavenging. Note that Au was not included in either model given the imprecision in the ICP-MS Au data, as well as the absence of any clear lithological control on its distribution. Both models produce similar results and are generally consistent with many of the known porphyry Cu-Au-Mo mineral occurrences in the central and western portions of the project area. Both models produce what initially appear to be spurious elevated scores in the eastern portion of the project area. There could be several reasons for this difference, including inaccurate rock type information, poorly represented rock types that are not captured by the multiple regression process, and Fe distributions not necessarily controlled by the presence of secondary hydroxide minerals in all samples. However, when the random forests-predicted GroupModels are overlain on the kriged weighted sums and additive models, many of the elevated model scores correspond with random forests predictions for porphyry Cu-Au-Mo. This correspondence provides confidence in the random forests predictions, at least for this particular deposit type.

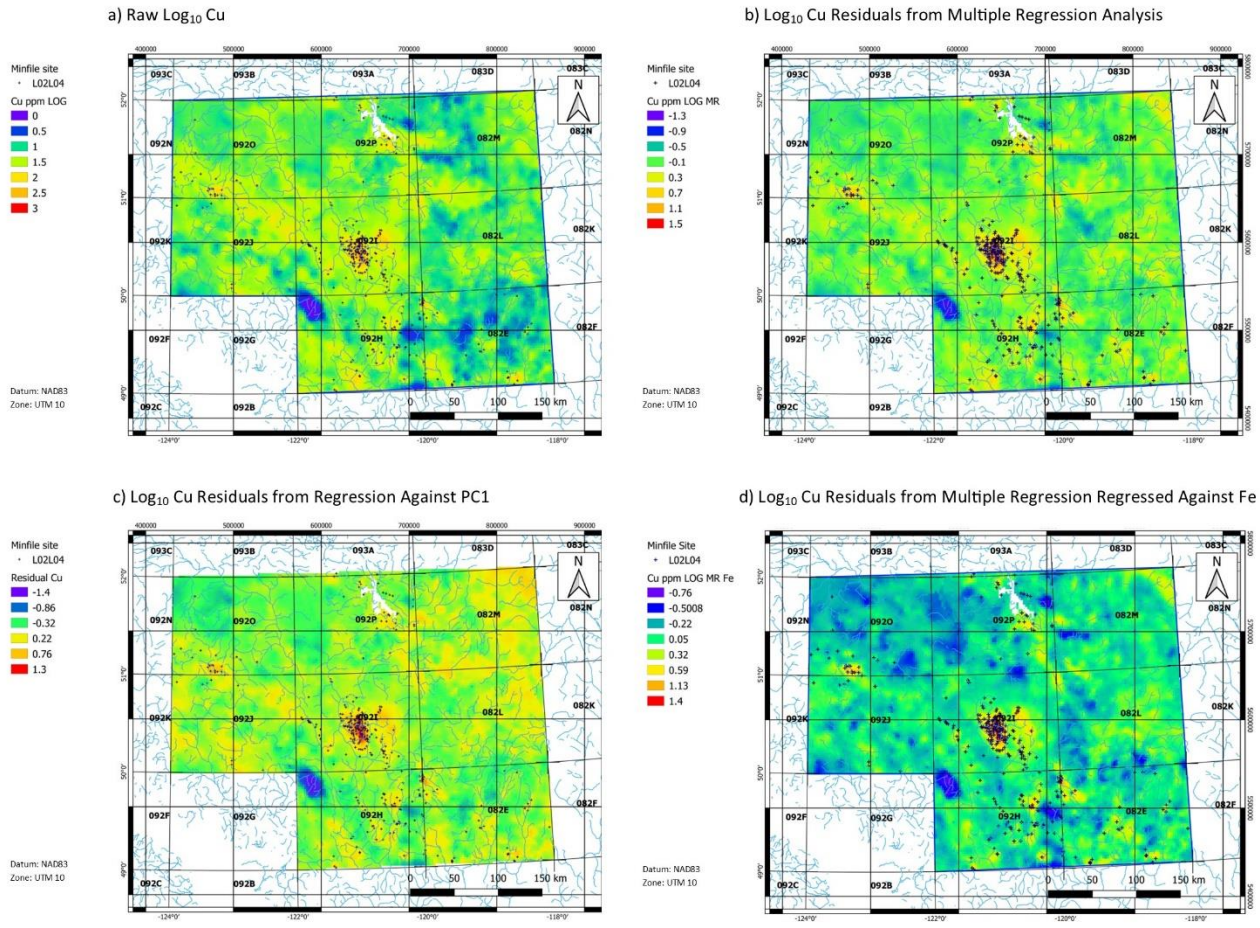


Figure 18. Geographic distribution of **a)** raw Log_{10} Cu; **b)** Log_{10} Cu residuals following multiple regression against the proportions of the most common bedrock types in the catchment areas; **c)** Log_{10} Cu residuals following regression against PC1; **d)** Log_{10} Cu residuals following multiple regression against the proportions of the most common bedrock types in the catchment areas, regressed against Fe in the samples to account for potential effects of metal scavenging.

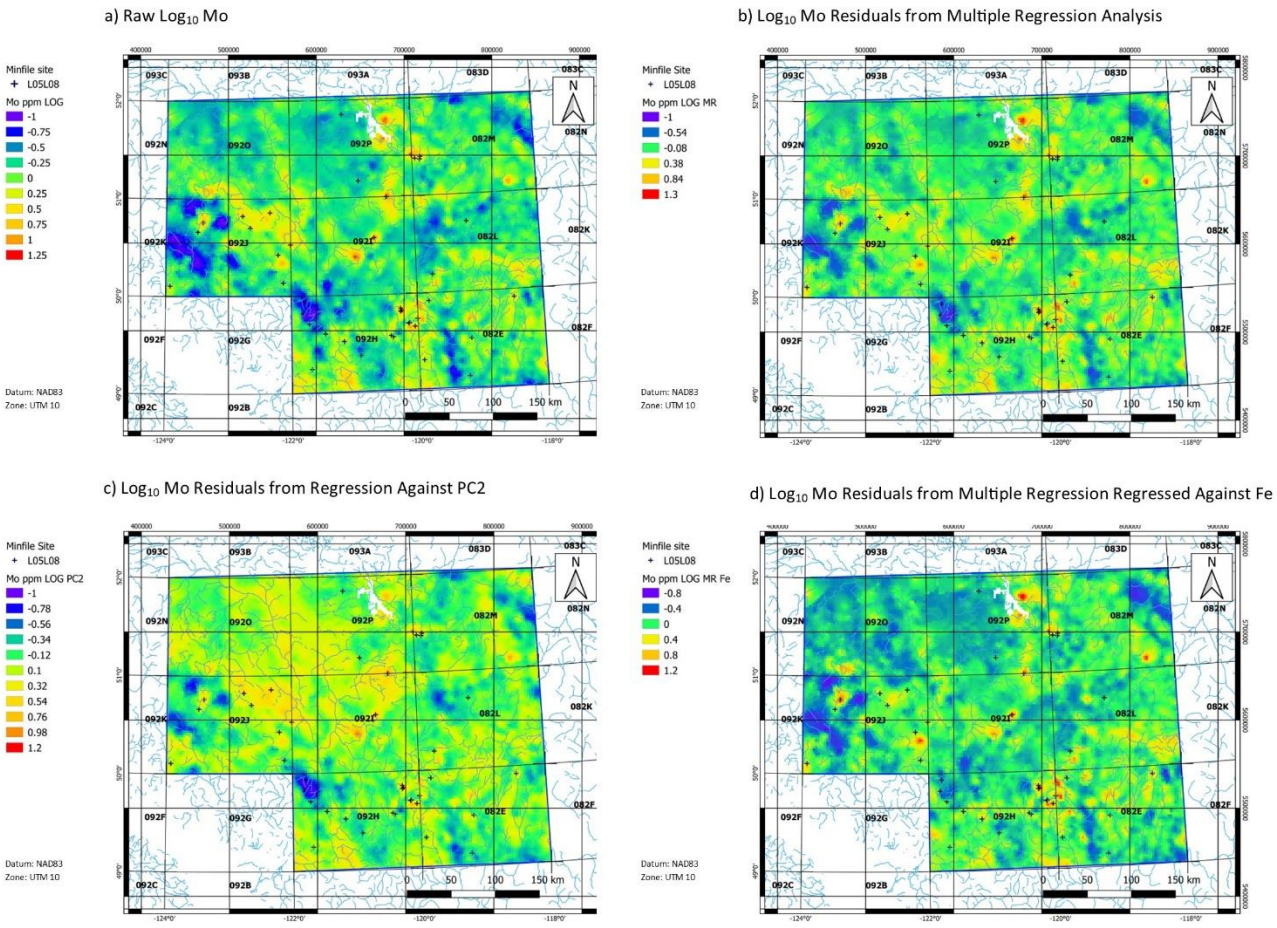


Figure 19. Geographic distribution of **a)** raw Log_{10} Mo; **b)** Log_{10} Mo residuals following multiple regression against the proportions of the most common bedrock types in the catchment areas; **c)** Log_{10} Cu residuals following regression against PC2; **d)** Log_{10} Mo residuals following multiple regression against the proportions of the most common bedrock types in the catchment areas, regressed against Fe in the samples to account for potential effects of metal scavenging.

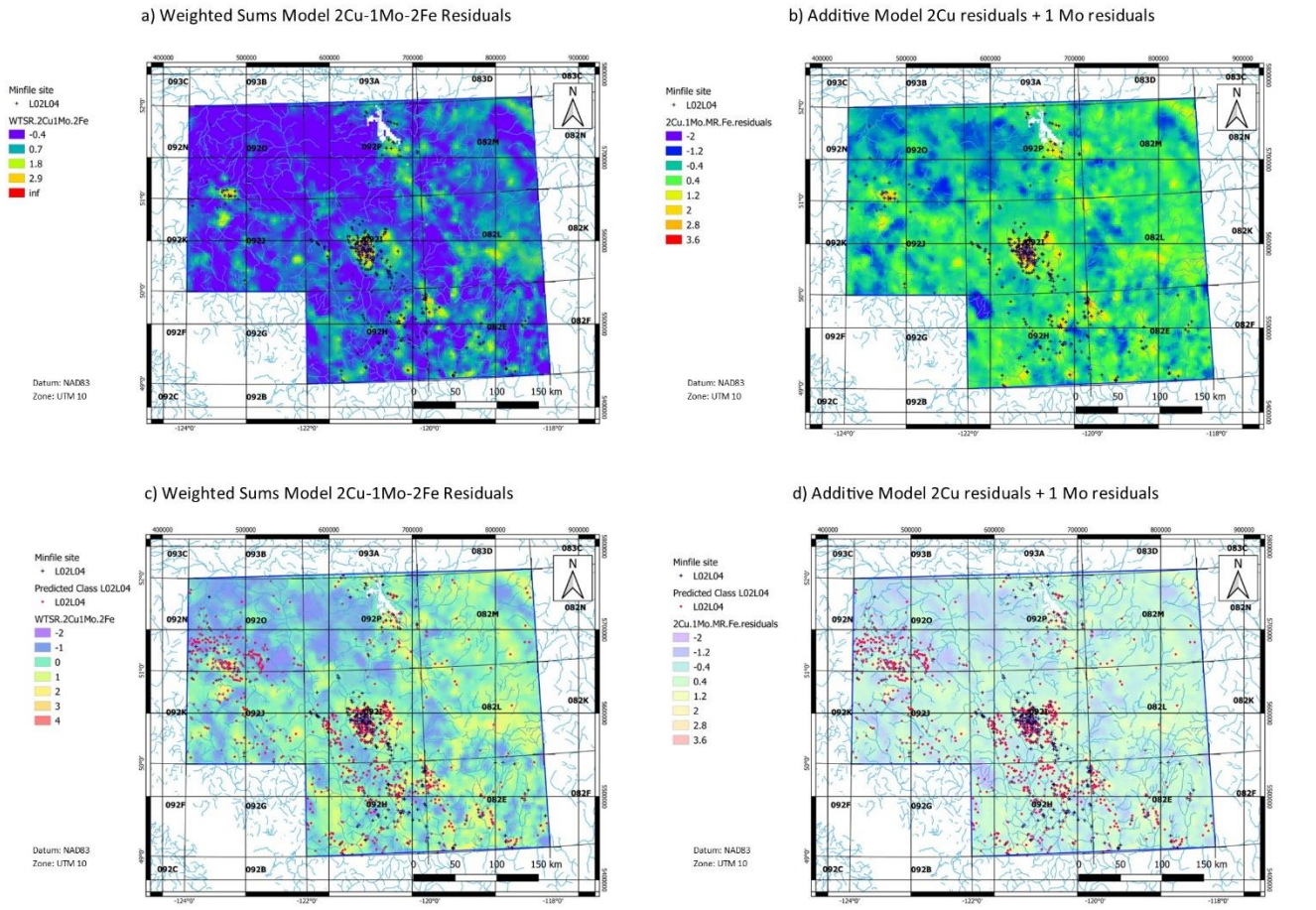


Figure 20. Geographic distribution of **a)** weighted sums model consisting of Log_{10} Cu multiple regression residuals (2), Log_{10} Mo multiple regression residuals (1) and Log_{10} Fe (-2) shown with MINFILE porphyry Cu-Au-Mo occurrences; **b)** An additive model of 2xLog_{10} Cu multiple regression residuals regressed against Log_{10} Fe plus 1xLog_{10} Mo multiple regression residuals regressed against Log_{10} Fe shown with MINFILE porphyry Cu-Au-Mo occurrences; **c)** weighted sums model consisting of Log_{10} Cu multiple regression residuals (2), Log_{10} Mo multiple regression residuals (1) and Log_{10} Fe (-2) shown with MINFILE porphyry Cu-Au-Mo occurrences and random forests t-SNE9 class predictions for porphyry Cu-Au-Mo (in red symbols); **d)** An additive model of 2xLog_{10} Cu multiple regression residuals regressed against Log_{10} Fe plus 1xLog_{10} Mo multiple regression residuals regressed against Log_{10} Fe shown with MINFILE porphyry Cu-Au-Mo occurrences and random forests t-SNE9 class predictions for porphyry Cu-Au-Mo (in red symbols); note that the intervals and colour schemes have been modified slightly in the figures showing random forests class predictions in order to display subtle regional differences and to provide better contrast with the class prediction sample locations.

Practical Considerations

Data summarizing the random forests class predictions and posterior probabilities using both PCA and t-SNE9 have been provided as digital files to accompany this report in Appendix 1. The class predictions show which GroupModel class is most likely to be associated with the geochemical data from a given stream-sediment sample. The posterior probabilities lie between 0 and 1. They give an indication for a particular GroupModel what the

probability is that geochemical data from a given sample are consistent with that GroupModel. The posterior probabilities are therefore more sensitive indicators of prospectivity than the class predictions and, as numerical values, have been used to generate a series of larger scale percentile-based thematic catchment maps to show the catchments having the highest probability of being prospective for a particular GroupModel deposit type.

The probabilities have not been adjusted for catchment area following the methodology described by Hawkes (1976) for raw element data as it is not clear what effect dilution has on the posterior probabilities, although it is noted that the probabilities for the various GroupModel deposit predictions tend towards zero in the largest catchments. In addition, there are many zero probabilities that obviate a simple multiplication of probability by catchment area, although it could be argued that the background probability for a given deposit type should be zero.

The percentile probability catchment maps indicate areas that may be prospective for particular deposit types but should be used with caution. They have been identified based on the assumptions that have been made in calculating the posterior probabilities, as discussed in the following section. They are also subject to the uncertainties in sample locations originally note by Cui (2010) and evaluated by Arne and Bluemel (2011). The authors have not re-checked adjusted sample locations from Arne and Bluemel (2011) and have accepted these as the best available. Re-sampling of the stream-sediments in prospective catchments is recommended to confirm reanalyzed geochemical data from the original samples and the locations of anomalous catchments. This is strongly recommended in the larger catchments for which sample locations may be incorrect and, if correct, in which any anomalous geochemical signature is likely to be strongly diluted. Detailed follow-up sampling would be recommended in these circumstances.

Deliverables to accompany this report include files containing centred logratio transformed values of the NAD 83 UTM Zone 10 coordinates of the stream-sediments, MINFILE attributes, elements, PCA scores, t-SNE9 scores, random forests votes, normalized votes, posterior probabilities and class predictions. The files containing this information are in either ESRI shapefile format or Microsoft Excel™. The predictive maps of the GroupModels are presented as interpolated maps for the PCA and t-SNE metrics, in Appendices 2 and 3, respectively. Predictive catchment maps using posterior probabilities from t-SNE9 at A3 scale are included in Appendix 4 of this report for GroupModels that have >0% accuracy in class prediction.

Discussion

The results presented here do not represent the entire range of mineral-deposit types or additional results that were determined by changing the selection of the GroupModels or the distance threshold. For some mineral-deposit types, changing the distance threshold to 1000 or 5000 m yielded different and reasonable predictions. Changing the distance threshold changes the size of the training set and can significantly affect the number of sites available

for each GroupModel. For some types of mineral deposits and different sample media (e.g. lake sediments), the choice of threshold distance may be varied to optimize the prediction of some mineral deposit types.

The predictions of the GroupModels presented in this report demonstrate the use of the two different metrics (PCA and t-SNE) and the application of random forests classification for defining zones of specific classes of mineral deposits. The methodology applied in the study shows the relationship between the predictions and the actual MINFILE sites, which provides a guideline for the accuracy of the method. In addition to the confirmation of known GroupModels from the MINFILE sites, areas of elevated posterior probabilities highlight the potential for additional deposits.

There are also several MINFILE sites that are not associated with a geospatially continuous region or individual class predictions. There are also some class predictions with elevated posterior probabilities in the central part of the map with no known MINFILE sites. In the former case, it may be that stream-sediment geochemistry does not contain the signature of the mineralization at the MINFILE site. Other media (soil, talus, bedrock) might be more appropriate in defining the signature of the mineralization. In the latter case, areas with elevated posterior probabilities may reflect previously unknown sites of mineralization

For some GroupModels, such as those in the K group (K01K03, K02, K04, K05), skarn mineralization, and G group (G04G05, G06, G07H02H03), volcanic-hosted and hot spring mineralization, the primary geospatial footprints are very limited, which is expected for these types of deposits.

It is worth noting, as shown in Figure 16, that the prediction accuracy based on the t-SNE metric outperforms the prediction accuracy based on the PCA metric. From Table 6 and Table 7 it can be seen that some GroupModels show more confusion with other models. For both the PCA and t-SNE metrics, the GroupModels C01C04, I01 and L02L04 have overlap with many of the other GroupModels. The reason for the overlap may be due to compositional similarities between the mineral deposit types based solely on the geochemistry of stream-sediments, or the geochemical composition of the GroupModel may be very simple (i.e. Au or Cu) and thus be like several other GroupModels.

An important concept in predictions based on spatially-based geochemical data is the geospatial coherence of the geochemical data and the subsequent predictions (see Grunsky and Kjarsgaard, 2016; Grunsky and Caritat, 2019). In a regional survey context, geospatial coherence of geochemical data for process discovery or process prediction provides an objective way in which geochemical processes are identified and validated. The results shown in this study demonstrate that, even with low posterior probabilities, the prediction of GroupModels can be validated using the location of the MINFILE sites that were used to build the GroupModels

Within the current scope and context of this study, some fundamental assumptions have been made:

5) The geochemical composition of the stream-sediment associated with individual mineral-deposit models is uniquely distinct. In some cases, this assumption is not warranted. For example, the mineral-deposit model I05 (polymetallic veins) has characteristics that overlap with many other mineral-deposit types, resulting in confusion of prediction. As a result, this model was re-assigned as "Unknown" for the GroupModel classes. Consequently, polymetallic vein mineral potential was not determined in this study.
6) The stream-sediment samples represent a suitable medium from which the geochemical characteristics of mineral systems can be identified. Not all mineral-deposit types can be best represented in stream-sediments. The size fraction and the analytical methods used may not extract unique information to distinguish a sub-cropping mineral deposit or distinguish between different mineral-deposit types, as in the case for the polymetallic vein class (I05). The method of dissolution using aqua regia is useful for sheet silicates and sulphide minerals, but aqua-regia digestion does not dissolve many silicate minerals. Thus, some unique geochemical aspects of specific mineral-deposit types based on silicate mineral assemblages may not be recognized.
7) The MINFILE model identification is accurate. This may not be the case for some types of mineral systems and, as a result, there will be an increase in confusion of prediction. The identification of the BCGS Mineral Deposit Profiles, as specified in the MINFILE field 'Deposit Type', may be incorrect or inconclusive, or the locations may be inaccurate. This can lead to misclassification errors in the subsequent application of machine-learning prediction methods.
8) As stated previously, the location of a MINFILE site and the associated stream-sediment site may not be within the same catchment area. Thus, the assumption was made that the effect of catchment is not significant. If there is a requirement for the location of a MINFILE site and associated stream-sediment site to be in the same catchment, the number of sites for the training set would be significantly reduced. It can also be argued that the placement of the stream-sediment site and the MINFILE site in separate catchments may not be an issue if the geological environment (lithology, alteration) is similar.
9) The fact that the corresponding MINFILE site and the stream-sediment sample site are not co-located means that there is always the likelihood that the stream-sediment composition does not reflect the observed mineralization at the MINFILE site. The distance threshold is 2.5 km appears to work for some mineral deposit types, but not necessarily for others. Changing the threshold distance for different mineral deposit models may help in a more refined estimate of mineral resource prediction.
10) Although the issue of spatial autocorrelation may be an issue in data that are spatially associated, this potential issue was not deemed as relevant and not addressed in the spatial interpolation of the images through kriging.

Conclusions

This report summarizes the rationale and methodology for the prediction of mineral-deposit types based on the mineral deposit model framework we have developed. The use of centred logratio transforms to overcome the effect of closure, and the application of multivariate methods to the stream-sediment geochemistry establish an objective framework for characterizing the data, termed 'process discovery'. The application of a tree-based method (random forests) for predicting potential mineral-deposit sites offers a repeatable, consistent and defensible methodology, termed 'process prediction', that offers the ability to identify prospective terrains and mineral systems. Together, they will enhance and encourage exploration strategies in the province of British Columbia.

The results presented here indicate that various types of mineral-deposit can be predicted with a confidence similar to more conventional geochemical interpretative methods involving catchment analysis and the use of expert knowledge-based models. Random forests posterior probabilities based on t-SNE with 9 dimensions provide slightly more accurate predictions than those made using PCA. Although many of the predictions have low values of posterior probability, the geospatial coherence of many of these sites provide evidence that the region is potentially prospective. In cases where isolated sites are identified in regions not previously known to be prospective, these can be considered either as 'new' prospective sites or as representing an overlap with other types of mineral deposits.

Acknowledgments

The authors thank Geoscience BC for funding this project under funding agreement 2018-016. Yao Cui of the BCGS is thanked for originally providing the catchments used in this project and for reviewing an early draft of this report. We would also like to acknowledge the thorough and thoughtful reviews provided by three anonymous reviewers for Geoscience BC. Geoscience BC is an independent, non-profit organization that generates earth science in collaboration with First Nations, local communities, government, academia and the resource sector. Their independent earth science enables informed resource management decisions and attracts investment and jobs. Geoscience BC gratefully acknowledges the financial support of the Province of British Columbia.

References

Aitchison, J. (1986): The Statistical Analysis of Compositional Data; Chapman and Hall, New York, New York, 416 p.

Arne, D.C. and Bluemel, E.B. (2011): Catchment analysis and interpretation of stream-sediment data from QUEST South, British Columbia; Geoscience BC Report 2011-5, 24 p., URL <http://www.geosciencebc.com/reports/gbcr-2011-05/> [November 2019].

Arne, D. and MacFarlane, B. (2014): Reproducibility of gold analyses in stream-sediment samples from the White Gold District and Dawson Range, Yukon Territory, Canada. (Newsletter for the Association of Applied Geochemists), no. 164, 1-10. URL < https://www.appliedgeochemists.org/images/Explore/Explore_Number_164_Sept_2014.pdf> [November 2019].

Arne, D., Mackie, R. and Pennimpede, C. (2018a): Catchment analysis of re-analyzed regional stream-sediment geochemical data from the Yukon; Explore (Newsletter for the Association of Applied Geochemists), no. 179, URL <https://www.appliedgeochemists.org/sites/default/files/documents/Explore%20issues/Explore179-June2018-website.pdf> [November 2019].

Arne, D., Mackie, R., Pennimpede, C., Grunsky, E. and Bodnar, M. (2018b): Integrated assessment of regional stream-sediment geochemistry for metallic deposits in northwestern British Columbia (parts of NTS 093, 094, 103, 104), Canada; Geoscience BC, Report 2018-14, 87 p., URL <http://www.geosciencebc.com/i/project_data/GBCR2018-14/GBCReport2018-14_Report.pdf> [November 2019].

BC Geological Survey (1996): British Columbia mineral deposit profiles; BC Geological Survey, URL <http://cmscontent.nrs.gov.bc.ca/geoscience/PublicationCatalogue/Miscellaneous/BCGS_MP-86.pdf> [November 2019].

BC Geological Survey (2019): MINFILE BC mineral deposits database; BC Ministry of Energy, Mines and Petroleum Resources, URL <http://MINFILE.ca> [November 2019].

Bonham-Carter, G.F and Goodfellow, W.D. (1986): Background corrections to stream geochemical data using digitized drainage and geological maps: application to Selwyn Basin, Yukon and Northwest Territories; Journal of Geochemical Exploration, v. 25, p. 139–155.

Bonham-Carter, G.F., Rogers, P.J. and Ellwood, D.J. (1987): Catchment basin analysis applied to surficial geochemical data, Cobequid Highlands, Nova Scotia: Journal of Geochemical Exploration, v. 29, p. 259–278.

Breiman, L. (2001): Random Forests; Machine Learning, v. 45, p. 5–32.

Carranza, E.J.M. and Hale, M. (1997): A catchment basin approach to the analysis of reconnaissance geochemical-geological data from Albay Province, Philippines; Journal of Geochemical Exploration, v. 60, p. 157–171.

Comon, P. (1994): Independent component analysis: a new concept? Signal Processing, v. 36, p. 287–314.

Cui, Y. (2010): Regional geochemical survey: validation and refitting of stream sample locations; *in* Geological Fieldwork 2010, BC Ministry of Energy, Mines and Petroleum Resources, BC Geological Survey, Paper 2011-1, p. 169–179, URL <http://cmscontent.nrs.gov.bc.ca/geoscience/PublicationCatalogue/Paper/BCGS_P2011-01-12_Cui.pdf> [November 2019].

Cui, Y., Eckstrand, H. and Lett, R.E. (2009): Regional geochemical survey: delineation of catchment basins for sample sites in British Columbia; *in* Geological Fieldwork 2008, BC Ministry of Energy, Mines and Petroleum Resources, BC Geological Survey, Paper 2009-1, p. 231–238, URL <http://cmscontent.nrs.gov.bc.ca/geoscience/PublicationCatalogue/Paper/BCGS_P2009-01-19_Cui.pdf> [November 2019].

Cui, Y., Miller, D., Schiarizza, P. and Diakow, L.J. (2017): British Columbia digital geology; BC Ministry of Energy, Mines and Petroleum Resources, BC Geological Survey, Open File Report 2017-8, 14 p., URL <http://cmscontent.nrs.gov.bc.ca/geoscience/PublicationCatalogue/OpenFile/BCGS_OF2017-08.pdf> [November 2019].

de Caritat, P. and 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, v. 33, p. 104–126.

de Caritat, P., Main, P.T., Grunsky, E.C. and Mann, A.W. (2016): Recognition of geochemical footprints of mineral systems in the regolith at regional to continental scales; Australian Journal of Earth Sciences, v. 64, p. 1033–1043.

Garrett, R.G. & Grunsky, E.C. (2001): Weighted sums – knowledge based empirical indices for use in exploration geochemistry. Geochemistry: Exploration, Environment, Analysis, v. 1, p. 135–141.

Grunsky, E.C. (2001): A program for computing rq-mode principal components analysis for S-Plus and R; Computers and Geosciences, v. 27, p. 229–235.

Grunsky, E.C. (2010): The interpretation of geochemical survey data; Geochemistry, Exploration, Environment Analysis, v. 10, p. 27–74.

Grunsky, E.C., Drew, L.J. and Sutphin, D.M. (2010): Process recognition in multi-element soil and stream-sediment geochemical data; Applied Geochemistry, v. 24, p. 1602–1616.

Grunsky, E.C., Mueller, U.A. and 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, v. 141, p. 15–41.

Grunsky, E.C. and Kjarsgaard, B.A. (2016). Recognizing and Validating Structural Processes in Geochemical Data. In Compositional Data Analysis, J.A. Martin-Fernandez and S. Thio-Henestrosa (eds.), Springer Proceedings in Mathematics and Statistics, 187. 85-116, 209pp., doi: 10.1007/978-3-319-44811-4_7, ISBN 978-3-319-44811-4.

Grunsky, E.C., Drew, L.J. and Smith, D.B. (2018): Analysis of the United States portion of the North American Soil Geochemical Landscapes Project – a compositional framework approach; *in* Handbook on Mathematical Geosciences: Fifty Years of IAMG, Springer, p. 313–346, URL <https://www.springer.com/gp/book/9783319789989> [November 2019].

Grunsky, E.C. and de Caritat, P. (2019): State-of-the-Art Analysis of Geochemical Data for Mineral Exploration, Geochemistry, Exploration Environment Analysis, Special Issue from Exploration 17, October, 2017, Toronto, Canada, DOI 10.1144/geochem2019-031

Grunsky, E.C., Massey, N.W.D., Kilby, W.E. (1994): Mineral resource assessment in British Columbia, The Mineral Potential Project; Nonrenewable Resources, v. 3, p. 271-283.

Hawkes, H.E. (1976): The downstream dilution of stream-sediment anomalies. Journal of Geochemical Exploration, v. 6, p. 345-358.

Harris, J.R., and Grunsky, E.C. (2015): Predictive lithological mapping of Canada's north using Random Forest classification applied to geophysical and geochemical data; Computers & Geosciences, v. 80, p. 9–25.

Harris, J.R., Grunsky, E., Behnia, P. and Corrigan, D. (2015): Data- and knowledge-driven mineral prospectivity maps for Canada's north; Ore Geology Reviews, v. 71, p. 788–803.

Harris, J.R., Schetselaar, E.M., Lynds, T. and deKemp, E.A. (2008). Remote predictive mapping: a strategy for geological mapping of Canada's north; *in* Remote Predictive Mapping: An Aid for Northern Mapping; J.R. Harris (ed.), Geological Survey of Canada, Open File 5643 p. 5–27.

Hron, K., Templ, M. and Filzmoser, P. (2010): Imputation of missing values for compositional data using classical and robust methods; Computational Statistics and Data Analysis, v. 54, no. 12, p. 3095–3107.

Jackaman, W. (2010a): QUEST-South Project sample reanalysis; Geoscience BC, Report 2010-4, 4 p., URL <http://www.geosciencebc.com/reports/gbcr-2010-04/> [November 2019].

Jackaman, W. (2010b): QUEST-South regional geochemical data, southern British Columbia; Geoscience BC, Report 2010-13, 152 p., URL<http://www.geosciencebc.com/reports/gbcr-2010-13/> [November 2019].

Jackaman, W. (2018): A compilation of quality control data from Geoscience BC RGS initiatives; Geoscience BC, Report 2018-15, 9 p. URL <http://www.geosciencebc.com/projects/2016-018/> [September 2019].

Kilby, W.E. (2004): The British Columbia Mineral Potential Project, 1992-1997, Methodology and Results, British Columbia Geological Survey Geofile, p. 2004-2330.

MacIntyre, D.G., Massey, N.W.D. and Kilby, W.E. (2004). The BC Mineral Potential Project - New Level 2 Mineral Resource Assessment Methodology and Results. In: Geological Fieldwork 2003. British Columbia

Ministry of Energy and Mines, British Columbia Geological Survey Paper 2004-01, p. 125-140., URL < http://cmscontent.nrs.gov.bc.ca/geoscience/PublicationCatalogue/Paper/BCGS_P2004-01-10_MacIntyre.pdf> [April 2020].

Mihalasky, M.J., Bookstrom, A.A., Frost, T.P., and Ludington, S., with contributions from Logan, J.M., Panteleyev, A., and Abbot, G. (2011, revised 2013): Porphyry copper assessment of British Columbia and Yukon Territory, Canada: U.S. Geological Survey Scientific Investigations Report 2010-5090-C, v. 1.1, 128 p., URL > https://pubs.usgs.gov/sir/2010/5090/c/< [April 2020].

Mueller, U.A. and Grunsky, E.C. (2016): Multivariate spatial analysis of lake sediment geochemical data; Melville Peninsula, Nunavut, Canada; Applied Geochemistry, v. 75, p. 247–262.

Nelson, J.L., Colpron, M. and Israel, S. (2013). The Cordillera of British Columbia, Yukon and Alaska: Tectonic and Metallogeny; *in* Tectonics, Metallogeny, and Discovery: The North American Cordillera and Similar Accretionary Settings, M. Colpron, T. Bissig, B.G. Rusk and J.F.H. Thompson (ed.), Society of Economic Geologists, Special Publication 17, p. 53–109.

Palarea-Albaladejo, J., Martín-Fernández, J.A. and Buccianti, A. (2014). Compositional methods for estimating elemental concentrations below the limit of detection in practice using R; Journal of Geochemical Exploration, v. 141, p. 71–77.

Pawlowsky-Glahn, V. and Egozcue, J.-J. (2015): Spatial analysis of compositional data: a historical review; Journal of Geochemical Exploration, v. 164, p. 28–32.

Pebesma, E.J. (2004): Multivariable geostatistics in S: the gstat package; Computers & Geosciences, v. 30, p. 683–691.

QGIS Development Team (2019): QGIS Geographic Information System; Open Source Geospatial Foundation Project, URL <http://qgis.osgeo.org> [October 2019].

R Core Team (2019): R: a language and environment for statistical computing: R Foundation for Statistical Computing; Vienna, Austria, URL < http://www.r-project.org> [November 2019].

van der Maaten, L.J.P. and Hinton, G.E. (2008): Visualizing data using t-SNE; Journal of Machine Learning Research, v. 9, p. 2579–2605.

Venables, W.N. and Ripley, B.D. (2002): Modern Applied Statistics with S (Fourth Edition); Springer, Berlin, 504 p., URL <http://www.bagualu.net/wordpress/wp-content/uploads/2015/10/Modern_Applied_Statistics_With_S.pdf> [November 2019].

Zhou, D., Chang, T. and Davis, J.C. (1983): Dual extraction of R-Mode and Q-Mode factor solutions; Mathematical Geology, v. 15, p. 581–606.

Appendix 1

ESRI shapefiles

Catchment_Basins_Predicted_Class_PCA.shp

Catchment_Basins_Predicted_Class_tSNE9.shp

Stream sediment sites with associated random forests PCA prediction posterior probabilities

[qss_pca_predicted_classes.shp]

Stream sediment sites with associated random forests t-SNE prediction posterior probabilities

[qss_tsne9_predicted_classes.shp]

Microsoft Excel™ spreadsheets

GBC 2020-06 SRM data

qss_pca_predicted_classes

qss_tsne9_predicted_classes

Appendix 2

Geotiff raster images of PCA metric posterior probability predictions for 20 GroupModel categories

Jpeg raster images of PCA metric posterior probability predictions for 20 GroupModel categories

Appendix 3

Geotiff raster images of t-SNE metric posterior probability predictions for 20 GroupModel categories

Jpeg raster images of t-SNE metric posterior probability predictions for 20 GroupModel categories

Appendix 4

PDF maps at A3 scale of:

Geological terranes of the QUEST-South project area

Catchments and stream-sediment sample locations for the QUEST-South project area

Catchments thematically coded with t-SNE9 posterior probabilities for C01C04 (Placer Au)

Catchments thematically coded with t-SNE9 posterior probabilities for E12E13E14 (Sediment-hosted Pb-Zn-Ag)

Catchments thematically coded with t-SNE9 posterior probabilities for G04G05 (Volcanic-hosted Cu-Zn)

Catchments thematically coded with t-SNE9 posterior probabilities for G06 (Volcanic-hosted Cu-Pb-Zn)

Catchments thematically coded with t-SNE9 posterior probabilities for H05 (Low-sulphidation Epithermal Ag-Au)

Catchments thematically coded with t-SNE9 posterior probabilities for I01 (Au Quartz Veins)

Catchments thematically coded with t-SNE9 posterior probabilities for I06 (Cu +/- Ag Quartz Veins)

Catchments thematically coded with t-SNE9 posterior probabilities for K01K03 (Cu-Fe Skarns)

Catchments thematically coded with t-SNE9 posterior probabilities for K04 (Au Skarns)

Catchments thematically coded with t-SNE9 posterior probabilities for L02L04 (Porphyry Cu-Au-Mo)

Catchments thematically coded with t-SNE9 posterior probabilities for L03 (Alkalic Porphyry Cu-Au)

Catchments thematically coded with t-SNE9 posterior probabilities for L05L08 (Porphyry Mo)

Catchments thematically coded with t-SNE9 posterior probabilities for M01M02M03M05 (Mafic-hosted Ni-Cu-Cr)