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, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, egrunsky@gmail.com

D.C. Arne, Telemark Geosciences, Yackandandah, Victoria, Australia

Grunsky, E.C. and Arne, D.C. (2020): 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); *in* Geoscience BC Summary of Activities 2019: Minerals, Geoscience BC, Report 2020-01, p. 55–76.

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). Regional stream-sediment samples, originally collected under the Regional Geochemical Survey (RGS) program between 1976 and 1979 from the QUEST-South project area, were reanalyzed in 2009. This was done using an aqua-regia digestion followed by a combination of inductively coupled plasma–emission spectrometry (ICP-ES) and inductively coupled plasma–mass spectrometry (ICP-MS; Jackaman, 2010a). These samples were analyzed by ALS Global (North Vancouver, BC) using method ME-MS41L. An infill stream-sediment survey was also undertaken in 2009 and the samples analyzed using similar digestions and instrumental finishes at Eco Tech Laboratories Ltd. (Kamloops, BC; 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 9321 samples were interpreted by Arne and Bluemel (2011) using a catchment-analysis approach. Of these samples, 8536 were originally collected between 1976 and 1979 and the locations transcribed from hard copy 1:50 000 scale topographic maps. Global positioning satellite (GPS) receivers were used to locate only the 785 new infill 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 sample were delineated by the British Columbia Geological Survey (BCGS) for the adjusted sample locations using the approach described by Cui et al. (2009), which involves computing the total drainage area for an individual sample from the nearest downstream stream junction for the adjusted sample location.

The catchment polygons thus obtained were used by Arne and Bluemel (2011) to query the bedrock types of the catchments and to determine the dominant rock type for each catchment area. It has previously been established that the dominant control on regional stream-sediment geochemistry is catchment rock type (Bonham-Carter and Goodfellow, 1986; Bonham-Carter et al., 1987; Carranza and Hale, 1997). Dominant catchment rock types were therefore used to level the geochemical data for the effects of variable background influence on stream-sediment geochemistry. Exploratory data analysis indicated that there were positive correlations between some elements and Fe and/or Mn, suggestive of scavenging of metals by secondary hydroxides. Positive residuals from linear-regression analysis of these metals against Fe were also used to identify areas of anomalous metal concentrations. Additive indices for several common ore-deposit types from the QUEST-South project area were then calculated using residuals and/or levelled Z-scores.

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.

This publication is also available, free of charge, as colour digital files in Adobe Acrobat® PDF format from the Geoscience BC website: http://www.geosciencebc.com/updates/summary-of-activities/.

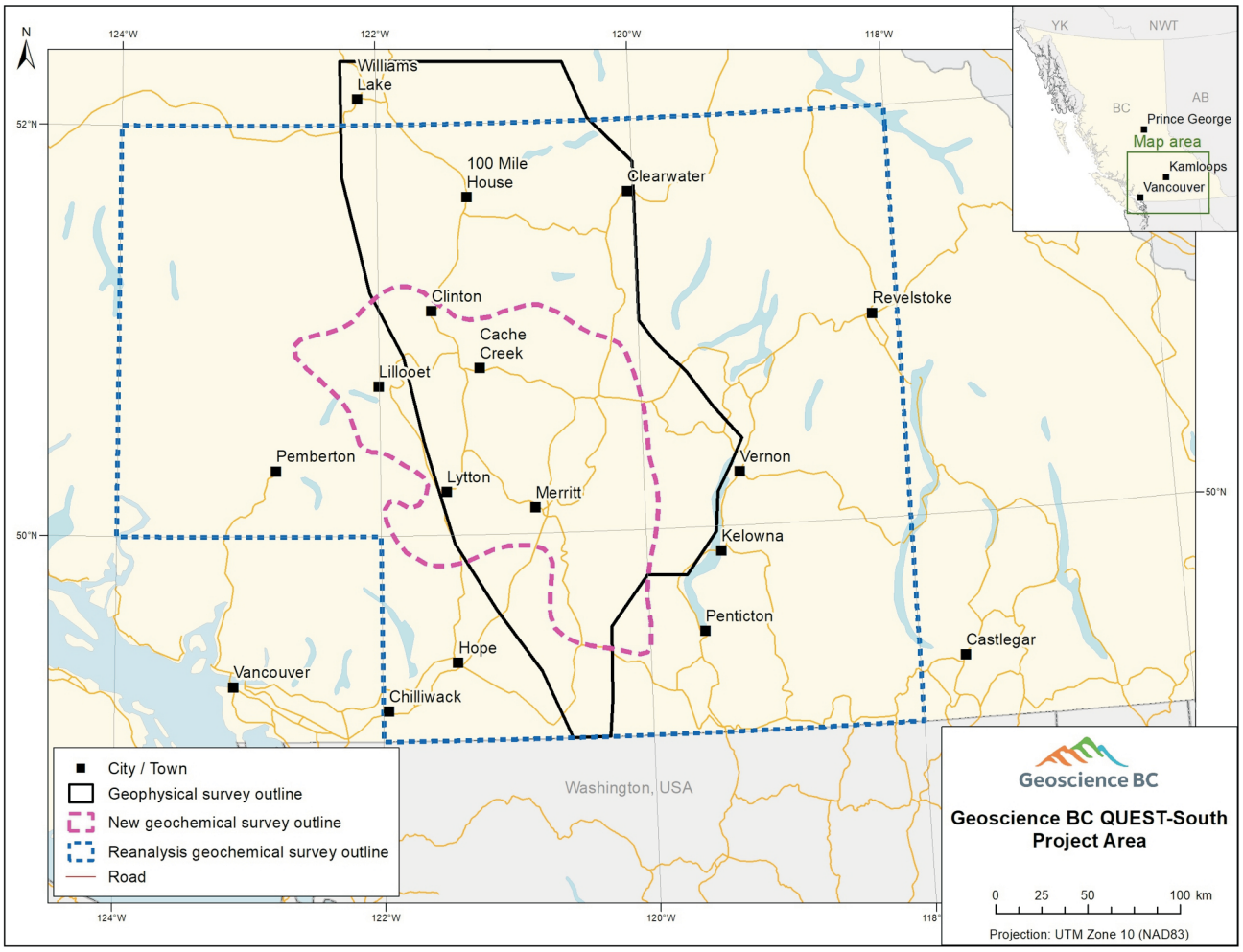


Figure 1. Location of the QUEST-South project area, showing the map areas from which archived stream-sediment samples were obtained (outlined in blue dashes), the area of infill stream-sediment sampling (outlined in pink dashes) and the area of geophysical surveys (outlined in solid black).

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.

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 principal-component analysis. Arne et al. (2018a) used regression analysis of key pathfinder or target-commodity elements against those principal components in which they were strongly represented to calculate residual values for those elements that were elevated above what would be expected, given lithological and other geochemical controls. Given that 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) demonstrated that the use of machine-learning algorithms could 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 that work and applies it to the QUEST-South project area.

Data Quality

The analyses from standard reference materials (SRMs) submitted with the samples during the original survey and analyses were not provided in Jackaman (2010a, b). Therefore, only a perfunctory review of data accuracy could be made by Arne and Bluemel (2011) using the available field duplicates and blind (pulp) duplicate data. They did note, however, that there was poor correlation (Spearman Rank correlation coefficient of 0.44) between ICP-MS Au and the historical instrumental neutron activation (INAA) data for Au. Digestions for the ICP-MS data involved 0.5 g of –177 µm sediment, whereas the historical INAA samples averaged 23 g. The INAA Au data were preferred for data

interpretation, given the larger sample mass. Despite this preference, the precision of the INAA Au analyses is considered to be poor.

Reanalyzed SRM data from the original survey and data from the SRMs submitted with the infill samples were made available by Jackaman (2018), including the Regional Geochemical Survey (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 from the original surveys but were generally not available for inclusion with the infill survey samples to provide overlapping SRM data sets for comparison.

A comparison of SRM data for the Red Dog and SQ SRMs 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. The elements Ba and La also show significant relative biases. Those elements with significant biases (i.e., ±5%) have been adjusted using the RGS Red Dog median data, so that the data remain in the unit of measurement to allow a centred-log transformation of the data. This correction was validated on the stream-sediment data from samples located in the area of infill sampling and was found to make only a slight difference.

A comparison was made of the three analytical methods used on the stream-sediment samples: ICP-MS/ES, INAA and atomic absorption (AA). The AA results were not considered further for this study due to the limited number of elements analyzed, which were already present in the reanalyzed ICP results and have higher detection limits for AA compared to the more recent ICP-MS results. The ICP data are based on an aqua-regia digestion, which is a partial extraction for many elements. The INAA results in a 'complete' analysis. A comparison of the detection limits and ranges of the analytical results generated by the two methods indicates that, although INAA data reflect a 'complete' composition, there are fewer elements available and the detection limits for some of the elements are higher than for the same elements analyzed by ICP-MS. Previous studies have shown that the material dissolved with aqua regia provides a multi-element signature that reflects silicate-bearing assemblages, most likely through partial dissolution of the silicates (Grunsky et al., 2014). The decision was made to use only the ICP-MS data in this study for the sake of consistency. Data from the following 35 elements were 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 were used in the study.

Methods

Data Screening and the Compositional Nature of Geochemical Data

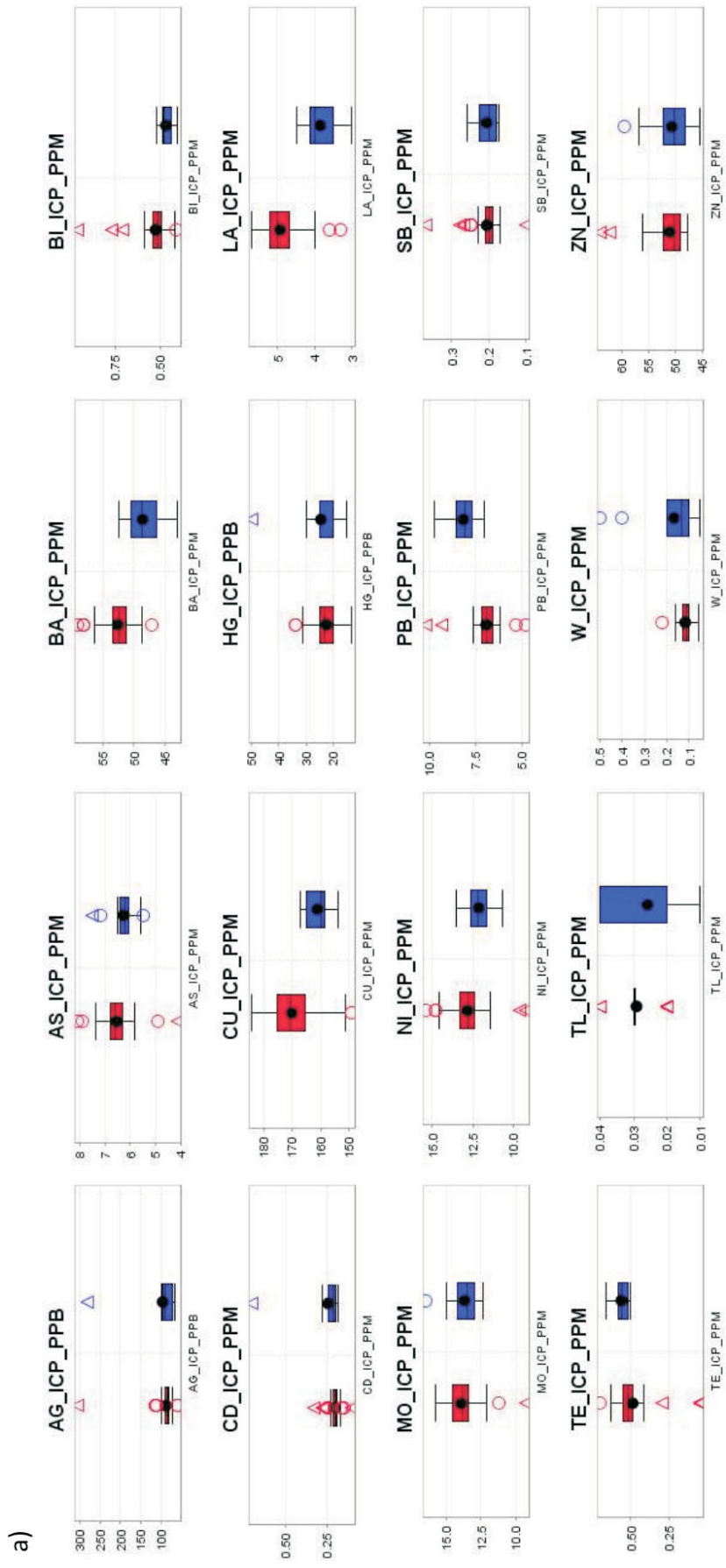
Geochemical data require quality-assurance and quality-control (QA-QC) screening prior to the application of statistical methods and subsequent interpretation. A sequence of data QA-QC strategies (Grunsky, 2010) was applied to the data. All data processing was carried out using the R programming and 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, 496 were blind (pulp) duplicates and 280 were field duplicates. The duplicate analyses were removed to provide 8545 analyses for evaluation.

Major-element concentrations, reported as percentages, were converted to parts per million (ppm). 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 reported at greater than the upper limit of detection were not addressed in this study. The 'maximum' value reported by the laboratory was used.

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

The QGIS software was used for the integration of various data sources and the geospatial rendering of the results. The projection used to manage and display the data is based on the North American Datum of 1983 (NAD 83) and the Universal Transverse Mercator (UTM) Zone 10.

Digital files of the bedrock geology (Cui et al., 2017), regional terranes (Nelson et al., 2013) and MINFILE data were obtained from the website of the BC Geological Survey (https://www2.gov.bc.ca/gov/content/industry/mineral-exploration-mining/british-columbia-geological-survey) in May 2019. The focus of this study was on metallic mineral deposits. An initial selection from the MINFILE database yielded 4877 records. MINFILE data that were classified as industrial minerals were dropped from further consideration, resulting in a total of 4108 records. 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 geo-


a)
AG_ICP_PPB
AS_ICP_PPM
BA_ICP_PPM
BI_ICP_PPM
CD_ICP_PPM
CU_ICP_PPM
HG_ICP_PPB
LA_ICP_PPM
MO_ICP_PPM
NI_ICP_PPM
PB_ICP_PPM
SB_ICP_PPM
TE_ICP_PPM
TL_ICP_PPM
W_ICP_PPM
ZN_ICP_PPM

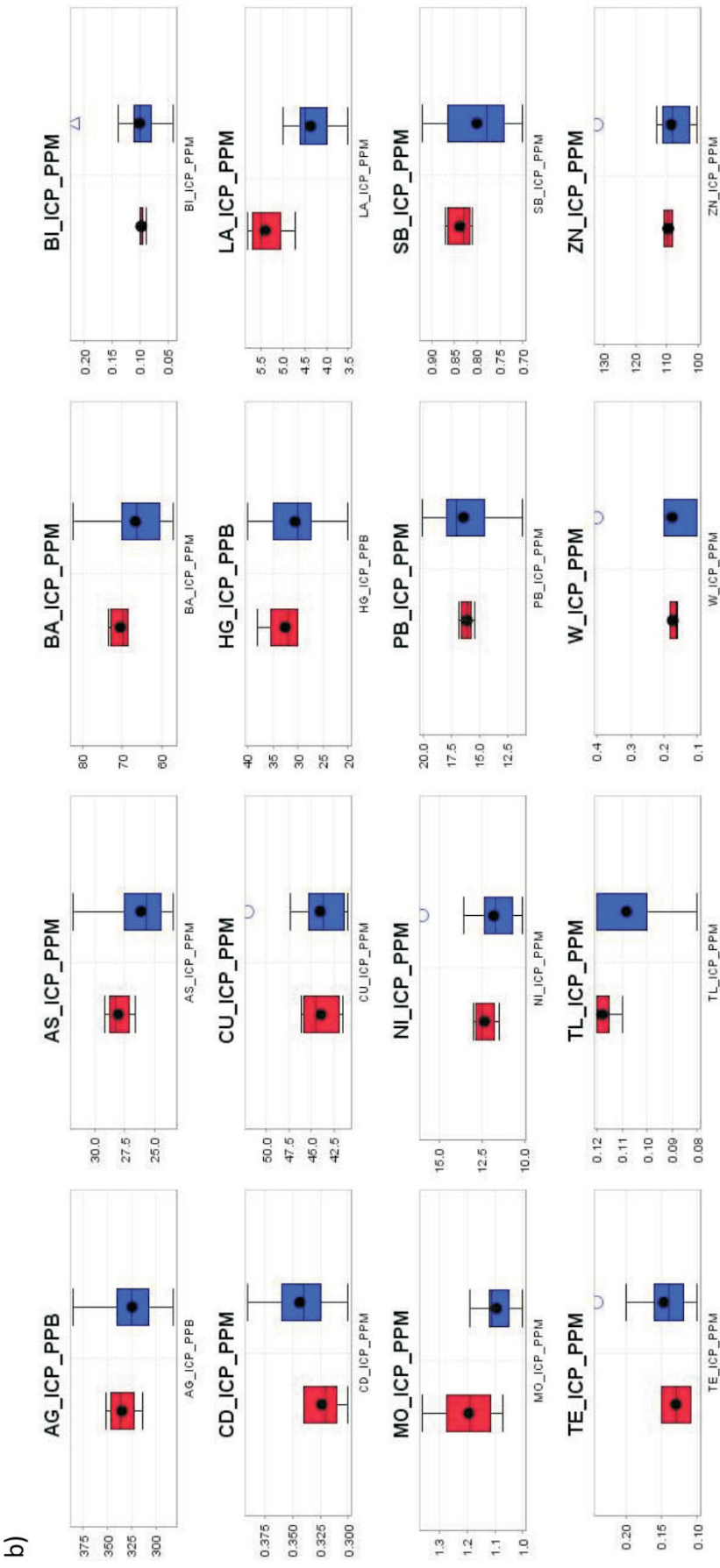


Figure 2. Box and whisker plots of selected elements from ICP-MS analysis of the Red Dog **(a)** and SRM SQ **(b)** standard reference materials. Data in red are from ALS Global and data in blue are from EcoTech.

chemical characteristics that overlap with several mineral-deposit types that are more economically significant.

The QGIS function 'NNJoin' was used to find the closest MINFILE point to each stream-sediment sampling site. Each RGS site was tagged with the nearest distance to a MINFILE site. These distances range from 0.7 to 42 848 m. 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.

The QGIS function 'Intersect' was used to merge the bedrock geology and 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. Because some MINFILE sites (Producer, Past Producer, Developed Prospect) may not be tagged if there is no stream-sediment site nearby, the likelihood of a geochemical expression of the mineralization may be difficult to estimate. Table 1 lists the number of stream-sediment sample sites associated with each MINFILE Status attribute. 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'.

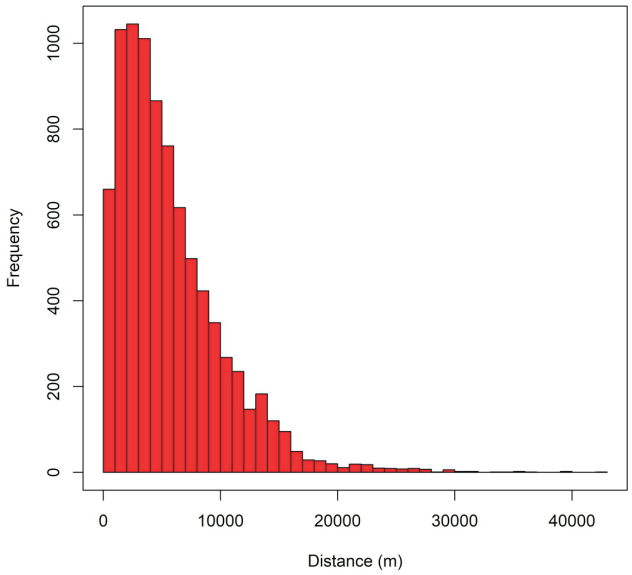


Figure 3. Histogram of the distances between stream-sediment sites and MINFILE sites, based on the QGIS function 'NNJoin'.

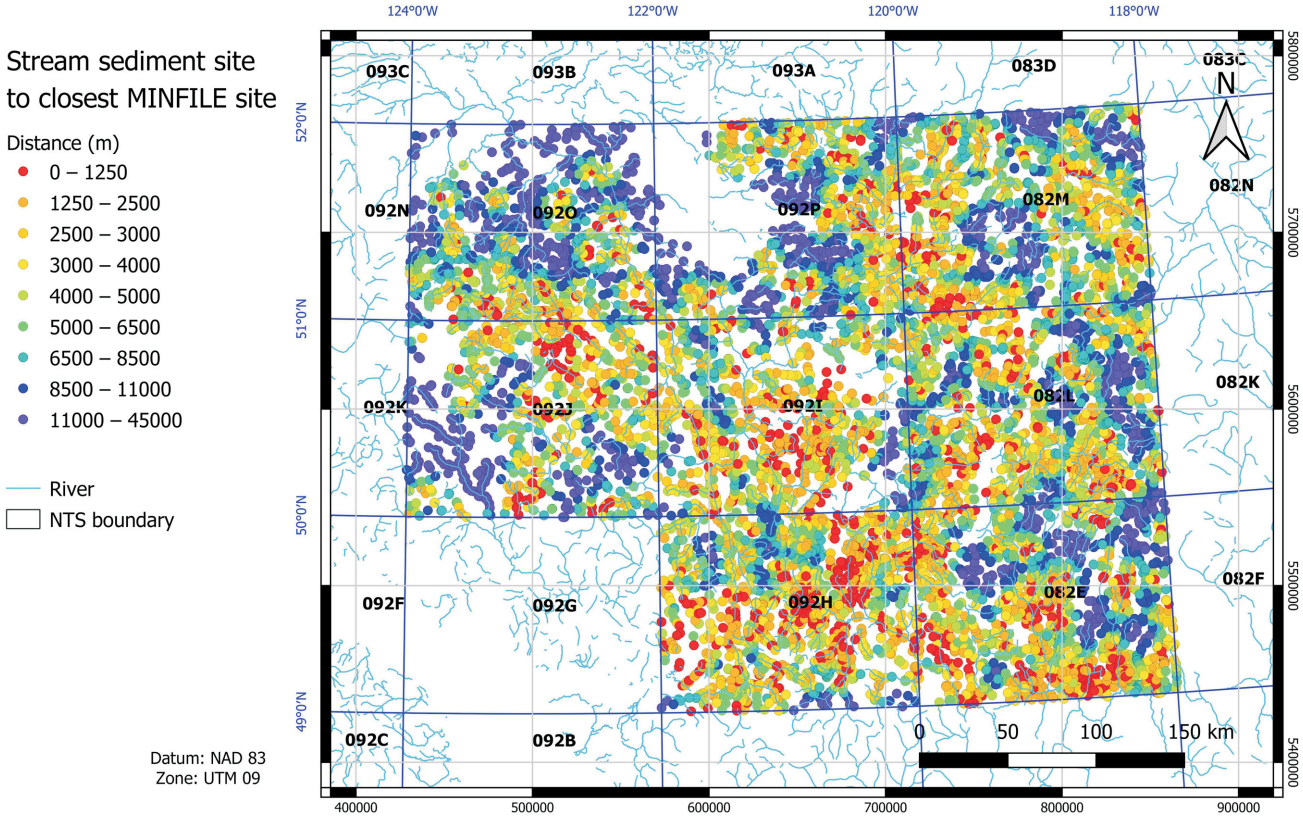


Figure 4. Geographic distribution of the distance measures between a stream-sediment site and the closest MINFILE site, based on the QGIS function 'NNJoin'.

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

MINFILE status	Frequency
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. The gstat package for R (Pebesma, 2004) was used to generate and model semi-variograms with sufficient parameters to produce interpolated images through kriging. The cell size used for image interpolation was chosen as 5.0 km for the images generated by principal-component analysis (PCA) and a cell size of 2.5 km was used for the images generated by the Random Forest predictions. This paper contains only the results of the application of PCA and the posterior probabilities of the mineral-deposit prediction derived from the application of random forests.

Characterizing Mineral Occurrence Information

Each MINFILE record lists a mineral-deposit model derived from the BCGS Mineral Deposit Profiles (BC Geological Survey, 1996). The number of MINFILE sites associated with each model is shown in Table 2.

The large number of mineral-deposit types for which there are only a few sites was considered to create difficulty in a statistical assessment of the data. Consequently, the models were merged as shown in Table 3. 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 the GroupModel classes that are used in the subsequent figures of this paper, where the left legend is the mnemonic category of the BCGS Mineral Deposit Profile and the right legend provides a description of the respective BCGS Mineral Deposit Profile.

Mineral-Deposit Models and Their Geospatial Footprint

An important consideration in the use of machine-learning methods for resource-potential prediction is the geospatial extent of the footprint of the mineral-deposit model. For many types of mineral deposits, namely vein (I01–I06), skarn (K01–K05), carbonatite (N01) and rare-earth elements (REE; O01–O04), the geospatial extent is quite limited, typically less than 200 m. As a result, the geochemistry

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

Model[1]	Frequency	Model[1]	Frequency	Model[1]	Frequency	Model[1]	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[2]	901
G01	5	I08	4	L03	198		
G02	2	I09	24	L04	384		

[1]MINFILE 'Model' designation (see 'Deposit' section of 'Mineral Occurrence' tab of MINFILE Search page at <http://minfile.ca/>.

[2]Unknown means no mineral deposit model was assigned to the MINFILE record.

Note: . For many types of mineral deposits, namely, vein (I01 – I06), skarn (K-01 – K05), carbonatite (N01) and REE (O01 – O04), the geospatial extent is quite limited, typically less than 200 m. As a result, the geochemistry of stream sediment sites that are tagged with MINFILE sites with these models may not reflect the geochemical signature of the mineral deposit. Other deposit types such as placer (C01 – C04), volcanic-hosted Cu (D03), basal U (D04- D06), sediment-hosted massive sulphides (E01, E02, E05, E12, E13, E14, E15), volcanic-hosted massive sulphides (G04, G05, G06), porphyry systems (L01 – L08) and mafic volcanic/ intrusive-hosted Ni, Cu, Cr (M01 – M05) may have broader geospatial signatures (> 200 m).

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

Models[1]	GroupModel
C01, C04	C01C04
D03	D03
D04, D06	D04D06
E01, E02, E05	E01E02E05
E12, E13, E14, E15, S01	E12E13E14E15
G04, G05	G04G05
G06	G06
G07, H02, H03	G07H02H03
H05	H05
I01	I01
I02	I02
I05	I05
I06	I06
K01, K03	K01K03
K02	K02
K04	K04
K05	K05
L01	L01
L02, L04	L02L04
L03	L03
L05, L08	L05L08
M01, M02, M03, M05	M01M02M03M05
N01	N01
O01, O02, O04	O01O02O04

[1]MINFILE 'Model' designation (see 'Deposit' section of 'Mineral Occurrence' tab of MINFILE Search page at <http://minfile.ca/>.

of stream-sediment sites that are tagged with MINFILE sites with these models may not reflect the geochemical signature of the mineral deposit. Other deposit types, such as placer (C01–C04), volcanic-hosted Cu (D03), basal U (D04- D06), sediment-hosted massive sulphides (E01, E02, E05, E12, E13, E14, E15), volcanic-hosted massive sulphides (G04, G05, G06), porphyry systems (L01–L08) and mafic volcanic/intrusive–hosted Ni, Cu and Cr (M01–M05), may have broader geospatial signatures (>200 m). Consequently, the ability to predict the various types of mineral deposits will depend on the proximity of the stream-sediment sample site to the MINFILE site.

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 (as described previously). Additionally, MINFILE Status designations of Anomaly or Occurrence were classed as Unknown. 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.

Table 4 shows the frequency of the MINFILE 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 Model classes were merged into the GroupModels, as shown in Table 3, with the corresponding number of sites shown in Table 4.

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 noted in a previous study (Arne et al., 2018b). Consequently, stream-sediment sites that were labelled as I05 were re-labelled as Unknown. The training set contains all stream-sediment sites where the GroupModel is not classed as Unknown. The test set contains all of 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. Table 4 summarizes the GroupModel classes that are part of the training dataset.

Process Discovery – Empirical Investigation of Geochemistry

After QA-QC, the geochemical data were subjected to an empirical investigation in which the assumptions about the data were minimal. Because geochemical data, by definition, are compositions, the issue of closure also becomes important. As compositional data sum to a constant (i.e., 100%, 1 000 000 ppm), then by definition, 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 log-centred (clr) transformed (Aitchison, 1986).

Multivariate methods were applied to the clr-transformed data for the purposes of discovering patterns and features that potentially describe relationships among geochemical, geological and geophysical parameters, as well as the effects of gravitational processes (Grunsky et al., 2010). These methods included principal-component analysis (PCA), independent-component analysis (ICA; Comon, 1994) and t-distributed stochastic neighbour embedding (t-SNE; van Maaten and Hinton, 2008). Each of these methods provides different co-ordinate systems that can reveal features and patterns related to geochemical processes. Only the PCA results are presented in this paper.

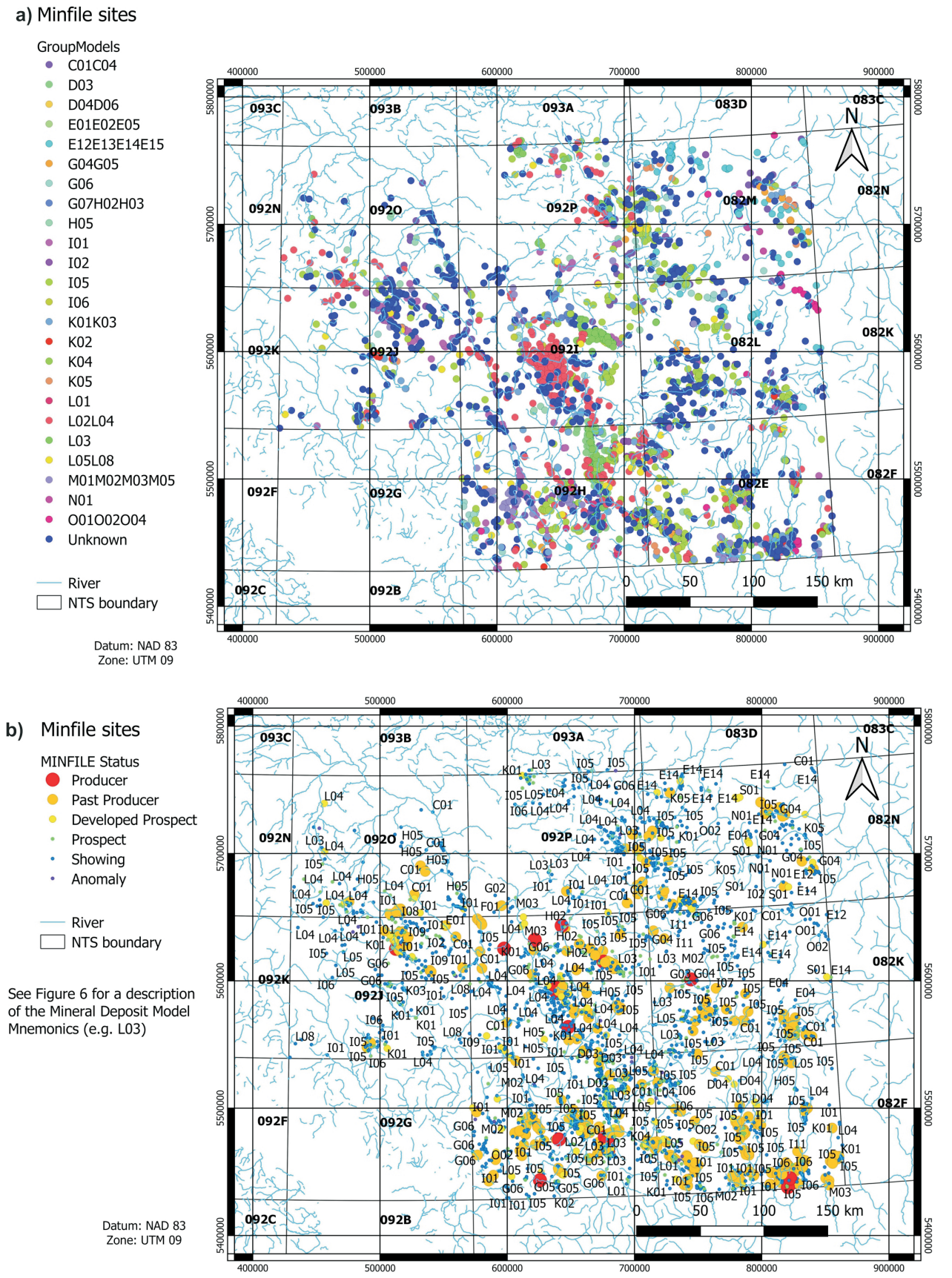


Figure 5. Geographic distribution of **a)** MINFILE sites classified by GroupModel; and **b)** MINFILE sites classified by Status and labelled by BCGS Mineral Deposit Model.

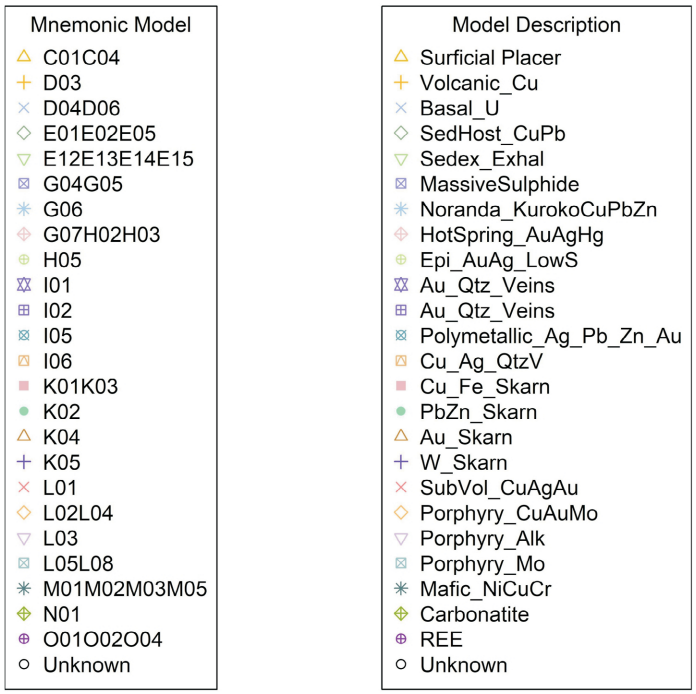


Figure 6. GroupModel classes expressed as BCGS Mineral Deposit Model mnemonics (left), with the respective BCGS Mineral Deposit Models (right).

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
E01E02E05	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	100
Unknown - train	8071
Total	**8545**

The co-ordinates resulting from a PCA 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.

Process Validation – Modelled Investigation of Geochemistry

Using the principal components derived from the clr-transformed geochemical data for stream-sediment sites that were tagged with the GroupModel class (Table 4), a GroupModel was predicted for each site that was classified as Unknown using the method of Random Forests (RF; Breiman, 2001). The Unknown class of data constituted the test set of data, except for 100 analyses that were used as part of the training dataset, as previously explained.

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 RF 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 classify correctly (categorical variables) all of the data. A more detailed description of how the Random Forest classification method was used with soil-geochemical data is provided in Harris et al. (2015).

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 log-ratio 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 or invalid. 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 kriged with the assumption of independence between the estimated classes.

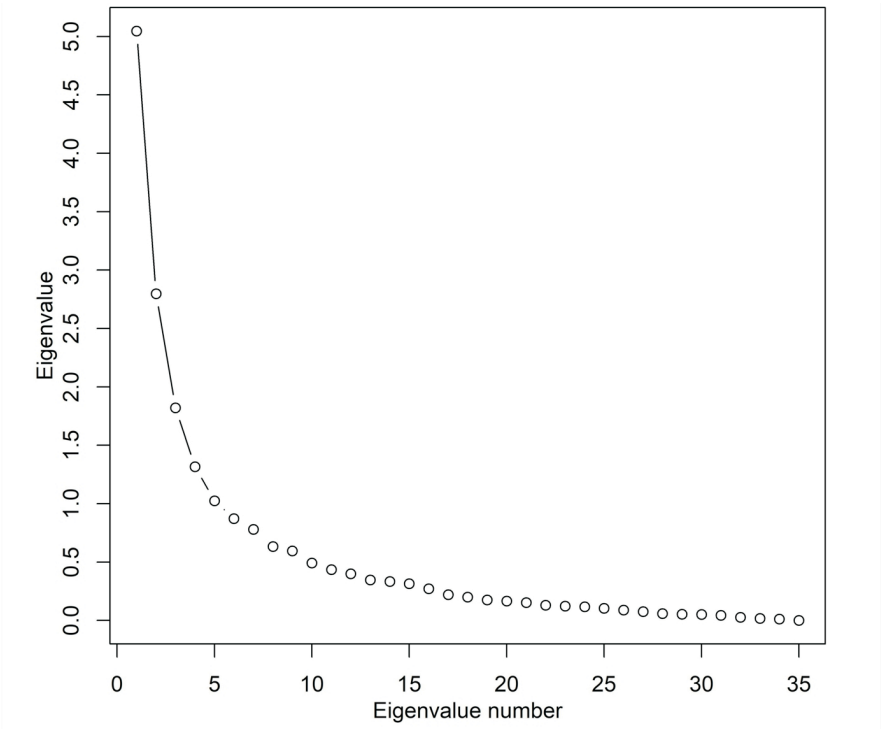
Note that kriged images based on point data have been used for validation purposes to test the sensitivity of various model input parameters and that thematically coded catchment maps will be generated with predictive results for a number of mineral-deposit types using the preferred modelling inputs,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

Results

PCA Process Discovery

A screeplot of the principal components derived from the clr-transformed data is shown in 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 undersampled geochemical or random processes. Typically, in regional geochemical surveys, principal components associated with mineral deposits are undersampled 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 paper, so only the biplots of selected principal components (PC) are shown in order to illustrate the associations between the stream-sediment sites and the elements. Table 5 shows the



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 a principal-component analysis (PCA) applied to the clr-transformed data from the QUEST-South stream-sediment geochemistry results.

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	0.6989	9.2693	0.1610	**86.5505**	0.7772	0.0600	1.7943	0.0188	0.0755	0.0027	0.4066	0.0406	0.0209	0.0055	0.0057
Ag	6.6789	**41.4826**	1.1457	0.2361	0.0384	8.7220	0.1433	2.5385	0.0037	2.1554	6.7573	**13.5810**	2.5598	0.6451	0.2579
Al	0.0073	**19.9318**	0.7834	1.3136	0.2964	3.1666	**17.9160**	0.1333	0.2477	0.0844	**13.2428**	0.6462	0.5518	0.3003	2.6763
As	**52.7714**	9.1233	**14.7048**	0.5437	4.0632	5.2089	0.4267	0.4638	1.1717	4.1879	0.4718	0.0004	0.4169	0.2066	3.5051
Ba	0.9251	5.7988	8.7258	3.6209	4.5142	0.2945	**12.0236**	3.8422	0.7422	2.6072	1.4064	0.7191	**15.0097**	0.0225	1.2733
Bi	**19.2949**	**25.3788**	8.1629	0.0126	2.7926	2.7636	3.2131	0.0355	5.5709	0.1094	6.8646	2.6774	1.4050	5.6101	0.3978
Ca	8.6571	8.0231	**39.4552**	0.3516	0.1194	4.3661	**20.8447**	4.7291	0.0026	0.0279	0.5600	0.0100	1.0093	1.8760	0.1239
Cd	**11.9913**	**37.7998**	1.1252	2.2407	2.0894	**10.6976**	0.8580	**11.7985**	1.2367	2.4223	1.9128	0.1630	1.4068	3.8286	0.8354
Co	**21.8757**	**30.7461**	**12.3658**	0.0009	2.1343	3.8330	0.3574	6.4427	0.2327	2.9409	0.0607	1.0738	0.1751	1.3216	0.8699
Cr	3.9533	**32.2648**	**14.4624**	0.7497	3.5326	1.6345	2.4365	3.5754	**19.9487**	0.0749	0.0344	2.0987	0.0104	0.4468	0.0152
Cu	**28.3649**	0.4362	1.1688	0.3700	6.9193	0.4550	0.0235	0.0691	6.3149	0.0000	5.1011	**18.4051**	0.5214	**14.4846**	2.3262
Fe	1.2387	**26.3040**	**12.2156**	0.4853	6.5318	0.3833	0.2770	7.6755	5.2405	**15.3267**	0.0553	0.1985	1.0379	0.0031	0.0386
Ga	9.5574	**26.7739**	3.5649	1.4437	4.3542	0.0613	**12.6585**	0.0971	0.9433	0.3471	6.0885	0.0241	0.2456	1.5731	0.4626
Hg	7.6828	2.7969	5.5976	2.5895	1.9206	2.7037	**11.9990**	1.2697	8.9787	2.3661	7.1758	**14.7192**	2.8291	4.6620	**14.5259**
K	**28.2095**	**10.9533**	0.3420	0.0167	7.1644	1.2161	1.6222	3.9810	**10.4682**	**20.5831**	1.8238	0.8035	1.5667	0.1013	0.0281
La	**65.9818**	0.3615	0.5050	0.0033	**10.8995**	0.9962	2.9773	1.0602	0.6095	0.4798	0.0142	0.5331	0.3707	2.1423	0.3924
Mg	**17.8416**	**38.9230**	7.0415	0.4319	6.6360	1.4551	1.5168	0.2363	0.8663	0.1638	0.0002	0.4392	0.0376	0.4997	1.5558
Mn	8.2355	0.0945	0.3812	4.9335	8.0692	0.8216	4.1058	1.4504	0.5560	**15.5973**	1.9756	**13.0761**	4.8440	0.4742	**10.5730**
Mo	0.0439	**24.8141**	0.5907	5.8816	0.9833	0.0286	5.7257	0.3953	0.5563	9.6349	**22.2267**	0.0043	5.3417	**13.4736**	0.8498
Na	0.8734	**36.4653**	**12.1305**	1.7574	0.4932	4.3404	2.1695	5.0551	0.8199	0.0628	0.9405	0.1854	**21.6942**	0.3390	0.1006
Ni	**17.1347**	**17.7046**	8.2647	0.4761	9.2980	**10.7745**	6.0195	2.1936	**19.1471**	1.2073	0.1331	0.4499	0.1205	0.2835	0.4157
P	**24.3384**	**10.6489**	2.8629	0.1690	1.9692	1.9944	0.7730	1.3455	2.4417	1.0275	4.5954	0.7438	6.0284	1.4953	0.0372
Pb	5.8067	**27.1246**	0.1257	0.2433	2.4066	7.3107	5.3710	0.2928	1.9793	0.3127	**11.3225**	5.3499	1.1474	1.8773	0.5512
S	6.8137	5.7766	**37.1747**	0.1118	**14.6116**	0.8177	0.0456	**26.4688**	3.8587	0.1187	0.0845	0.0185	0.0707	2.9206	0.0482
Sb	**61.6038**	7.4505	2.9143	0.8911	7.5966	4.5937	0.7771	0.2465	0.4769	1.5758	0.0019	0.1829	0.0369	0.3356	4.0344
Sc	8.4356	**29.2350**	6.3569	0.1566	1.6333	1.0029	4.7795	0.0556	0.0895	0.5047	5.5858	0.3722	0.0039	0.0486	0.4468
Se	**13.2096**	**10.4685**	**20.7614**	3.9385	3.3864	0.4195	0.0148	0.1247	5.5979	0.4276	5.8006	1.3523	0.0379	0.4375	0.9701
Sr	1.4017	9.9012	**46.9785**	0.1692	5.2254	3.6645	**10.3769**	6.2445	0.2397	0.1744	0.2129	0.1055	0.0038	1.5617	2.1919
Th	**68.9774**	0.1632	5.5913	0.3602	3.1495	0.0351	**11.7522**	1.0889	1.4091	0.0655	0.7424	0.0139	0.0239	0.0000	2.9220
Ti	3.8444	**50.5982**	2.2106	0.0038	0.0011	0.4688	8.9841	0.7261	0.0157	0.2574	0.0000	0.6786	1.4654	4.7206	0.7844
Tl	**33.8946**	1.3546	0.0593	1.6900	3.9299	9.2104	7.3387	3.2035	0.6633	**15.7427**	6.1739	0.4171	0.8110	0.5611	0.0018
U	**63.6091**	9.6125	2.9410	1.6616	2.8168	0.6449	1.2336	1.5284	6.6831	1.0701	0.0787	2.4354	1.1198	0.6309	1.8716
V	0.9803	**33.6136**	6.0450	0.2566	7.1988	6.5679	6.8820	0.2465	0.6066	**12.2189**	0.0013	8.1469	0.0896	0.8467	0.6873
W	**25.9920**	8.8420	**11.5808**	0.0601	**20.9871**	**24.9910**	0.0394	2.5198	0.9173	1.3099	0.2506	0.3186	1.1944	0.4243	0.0936
Zn	7.8490	1.3451	1.4823	4.4344	0.2586	**17.4519**	0.0003	4.8889	3.7955	3.3146	0.0004	3.7654	2.2082	2.0354	0.0004

relative contributions of the PCA results. The contribution of variability for each element is shown across the first 15 eigenvectors.

Figure 8a shows a biplot of PC1–PC2. The stream-sediment site scores are coded with their MINFILE GroupModel designation as shown in Figure 6. The components PC1–PC2 account for 40.7% of the variability of the data. The loadings of the elements indicate that chalcophile elements (Sb-As-Cd-Se-S-Hg) are associated with the positive PC1–negative PC2 quadrant. Symbols representing GroupModel I05 (polymetallic veins) occur within this quadrant. A group of siderophile elements (Fe-Cr-Ni-V-Cr-Co) occur within the positive PC1–PC2 quadrant. Stream-sediment sites tagged with mafic base-metal sulphide deposits (M01M02M03M05) occur in this quadrant. The loadings of W-U-Bi-Pb-La-Th-K occur across the PC2 axis and the

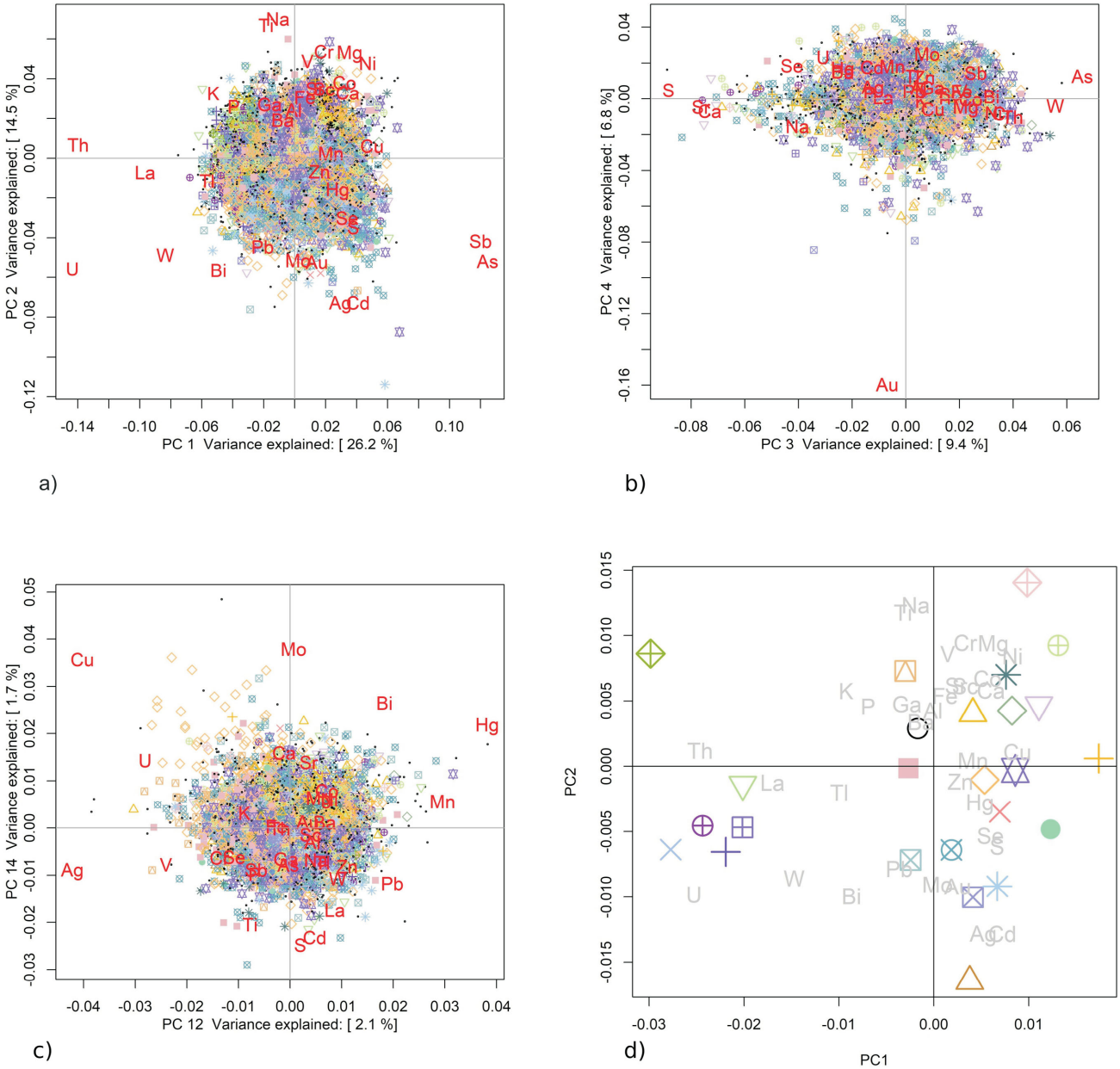


Figure 8. Biplots of **a)** PC1–PC2, showing the relative enrichment of Au along the negative PC2 axis; symbols are coloured/coded according to the legend in Figure 6; components are derived from log-centred transforms to the stream-sediment geochemistry based on a covariance structure; the first PC accounts for 26.2% of the overall variance and the second PC accounts for 14.5% of the overall variance; **b)** PC3-PC4 showing the relative enrichment of Au along the negative PC4 axis; see Figure 6 for the legend of colours and symbols; **c)** PC12-PC14 showing the relative enrichment of Cu along the positive PC14 axis and the negative PC12 axis; sites identified with relative Cu enrichment are associated with L02L04 (porphyry Cu) MINFILE designations; symbols are coloured/coded according to the legend in Figure 6; **d)** PC1-PC2 showing the mean PC1-PC2 values for each of the GroupModel classes; symbols are coloured/coded according to the legend in Figure 6.

negative PC1 axis. The relative associations of these elements likely reflect mineralized environments associated with felsic intrusive rocks. It is worth noting that the L02L04 GroupModel plots throughout the PC1–PC2 biplot. An association of L02L04 deposits is typically not associated with mafic elements and this suggests that not all GroupModel assignments, projected onto a PC1–PC2 biplot, provide insight into the multi-element associations for the different GroupModels.

Lithophile elements dominate the negative PC1–positive PC2 quadrant, including the elements Ba-Al-Ga-P-K-Ti-Na. This region of the biplot shows associations of Cu-Fe skarns (K01K03), Cu-Ag quartz veins (I06) and carbonatite (N01). The GroupModels are clearly distinct from each other, which justifies the attempt to classify the MINFILE-tagged stream-sediment sites with a GroupModel designation.

From Table 5, it is evident that most of the variability of Au is accounted for in PC4 and, for Cu, most of the variability is accounted for in PC1, PC12 and PC14. Figure 8b shows the biplot for PC3–PC4. The relative Au-enrichment trend is shown along the negative PC4 axis. Using the legend for the various MINFILE GroupModels in Figure 6, it can be seen that the relative enrichment in Au is associated with G04G05 (massive sulphides), L05L08 (Mo-rich porphyry), L02L04 (Cu-rich porphyry), L03 (alkalic porphyry), K04 (Au skarn) and I02 (Au veins). Figure 8c shows a biplot of PC12–PC14 that highlights the relative enrichment of Cu, primarily associated with L02L04 (Cu-rich porphyry).

It is difficult to see the compositional differences between the different GroupModel designations in the principal-component biplot. Figure 8d shows the mean values of PC1 and PC2 for each of the GroupModels. The relative enrichment of elements and the corresponding association with the GroupModels is evident in the biplot. Relative enrichment in siderophile elements (Fe-V-Cr-Co-Ni) occurs in the positive PC1–positive PC2 quadrant. Mineral deposits that are associated with this group of elements include mafic Ni-Cu-Cr (M01M02M03M05), sediment-hosted Cu deposits (E01E02E05), surficial placer deposits (C01C04), alkalic porphyry deposits (L03), epithermal Au deposits (H05) and volcanic redbed associated Cu deposits (D03). Elements that are dominantly chalcophile in character, including Cd-Se-Hg-Ag-Au-Mo, occur in the positive PC1–negative PC2 quadrant and are associated with porphyry Cu (L02L04), Vein-hosted Au deposits (I01), subvolcanic Cu-Ag-Au deposits (L01), massive sulphide deposits including Noranda-type deposits (G04G05, G06), polymetallic vein deposits (I05) and Au skarns (K04) occur in this quadrant. Elements that occur in the negative PC1–negative PC2 quadrant are partly chalcophile and lithophile in nature, including Pb-Bi-W-U. Mineral deposits associated with this group include Mo porphyry deposits (L05L08), W skarn deposits (K05), basal U deposits (D04D06), rare-earth element deposits (REE; O01O02O04), quartz vein Au deposits (I02) and sedimentary exhalative deposits (E12E13E14E15).

Kriged images, along with individual point scores for PC4 and PC12, are shown in Figure 9a and b, respectively. Regions of relative Au enrichment (Figure 9a) and relative Cu enrichment (Figure 9b) are clearly shown on these maps. In Figure 9a, PC4 indicates relative Au enrichment associated with negative (blue) values. In Figure 9b, PC12 indicates relative Cu enrichment associated with negative (red) values.

Process Validation – Random Forest GroupModel Prediction

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. One advantage of the random forests process is that a prior selection of variables is not required. The procedure starts with all of the variables (PC1–PC35) and then reduces the number of variables to those that provide the best nodes in the trees that are generated.

Figure 10 shows the significance of the variables derived from the random forests 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 PC2. The remaining variables show a monotonic decrease in significance.

Table 6 shows the accuracy of classification in terms of percentage, based on the training set only. The overall classification accuracy is 36.2% when model I05 (polymetallic veins) is excluded from the modelling runs. Several of the GroupModel classes show a classification accuracy of zero. The GroupModel classes that were associated with the most confusion and/or overlap were G04G05 (massive sulphide), K01K03 (Cu-Fe skarn) and L03 (alkalic porphyry Cu deposits). The confusion among these GroupModels is likely due to significant overlap of their geochemical signatures with those of other GroupModels

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 be created for each GroupModel class. Areas of contiguous elevated posterior probabilities for a given class define the 'geospatial coherence' of a

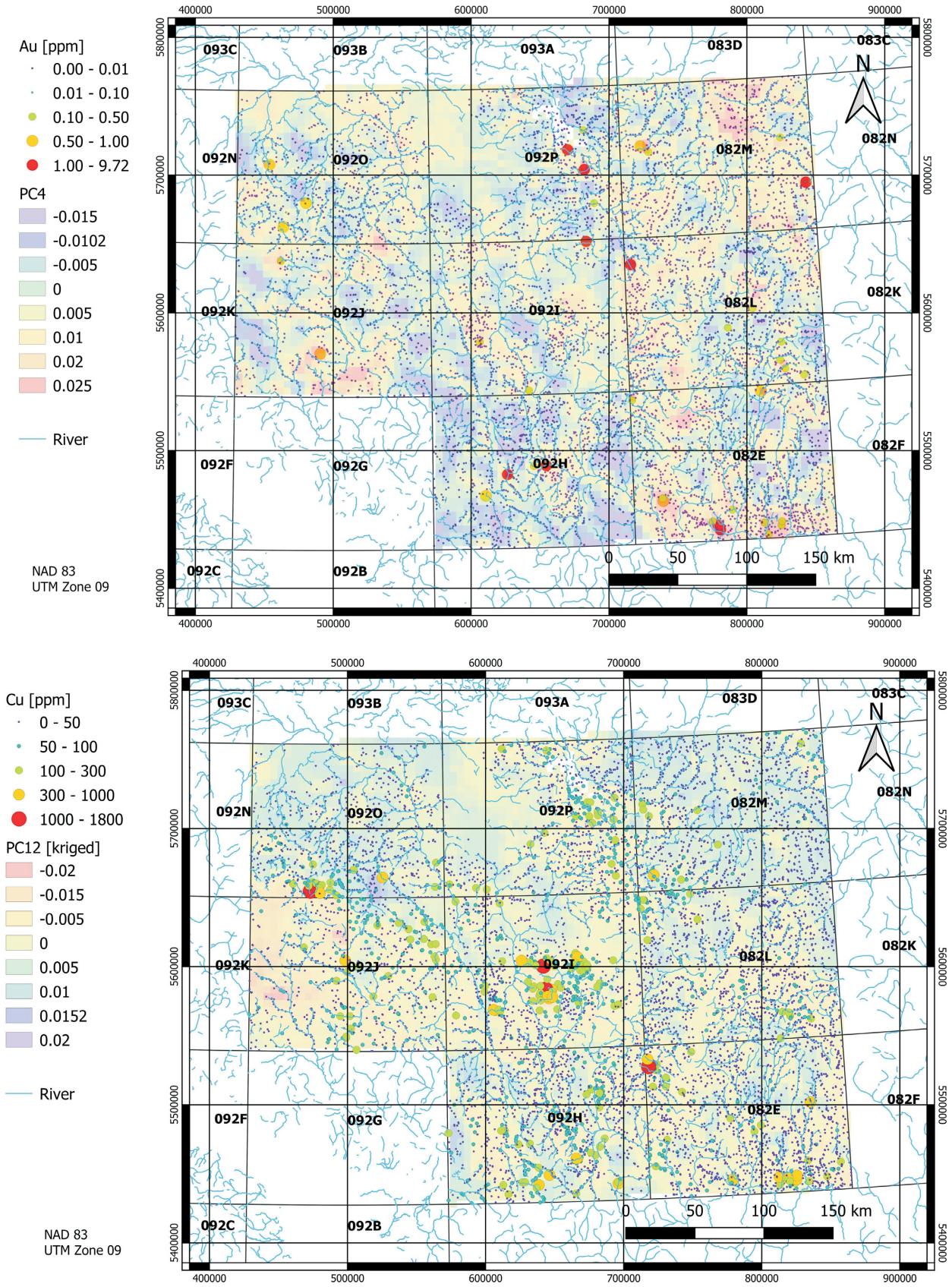


Figure 9. Geographic distribution of **a)** individual sites overlain on a kriged image of PC4, illustrating the relative enrichment of Au at selected sites across the map area; **b)** individual sites overlain on a kriged image of PC12, illustrating the relative enrichment of Cu at selected sites across the map area.

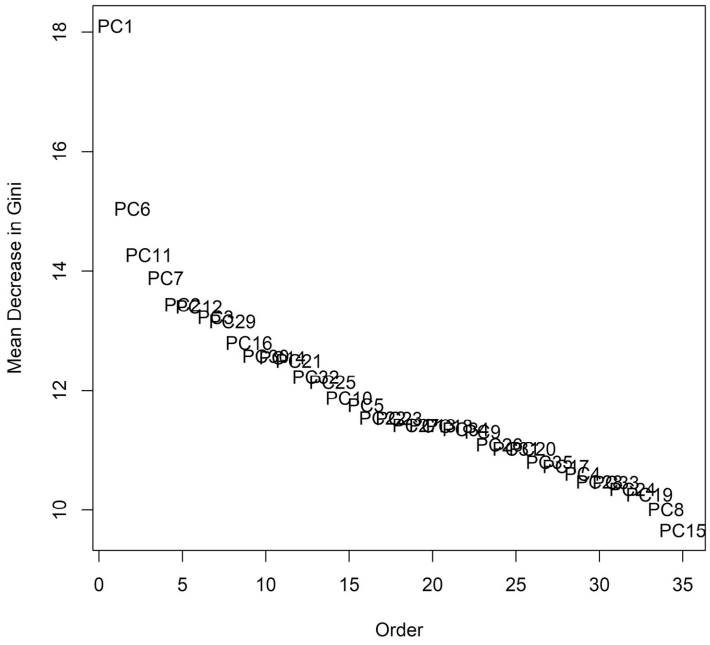


Figure 10. Plot of 'Mean Decrease in Gini' for the principal components used in the application of Random Forest prediction based on the training data [MinDep <2500 m].

GroupModel. It is expected that the maps of posterior probability can show overlap because of compositional overlap between the classes. 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.

Table 6. Accuracy matrix for the GroupModels training set, derived from the application of Random Forest classification.

GroupModel	Accuracy (%)	Class error
C01C04	61.16	0.46
D03	0	16.67
E01E02E05	0	25.00
E12E13E14E15	21.96	3.39
G04G05	8.40	7.63
G06	0.00	3.57
G07H02H03	0.00	14.29
H05	32.14	3.57
I01	26.36	1.75
I02	0	50.00
I06	0	16.67
K01K03	8.36	3.82
K02	0	33.33
K04	20.45	7.95
K05	0	14.29
L01	0	14.29
L02L04	53.69	0.75
L03	6.28	5.86
L05L08	0	7.14
M01M02M03M05	0	5.56
Unknown	65.78	0.34

A map of predictions for the GroupModel class C01C04 (surficial placer Au) deposits is shown in Figure 11. Many more sites were predicted than the actual number of MINFILE sites. The predicted sites and the interpolated image show a north-northwesterly trend and closely follow the stream/river drainage lines on the map.

Figure 12 shows a predictive map of Au skarns (K04). The interpolated map shows elevated values that coincide with the MINFILE sites and the stream-sediment sites that are classed as K04. Several areas have elevated posterior probabilities where there are no known MINFILE sites associated with K04. As explained previously, the posterior probabilities are low (<0.2), but the identified MINFILE sites and assigned random forests class coincide with the elevated kriged image of the posterior probabilities.

Figure 13 shows a predictive map of W skarns (K05). Several sites that are identified as K05 occur in the northeastern part of the area (NTS 082M). A few isolated sites occur in NTS areas 092J, 092I and 082E. The kriged image of the posterior probabilities shows an overall low prediction rate, not exceeding 0.2. Nonetheless, the elevated values in NTS area 082M coincide with the MINFILE sites. Although W was not included in the dataset, the prediction of

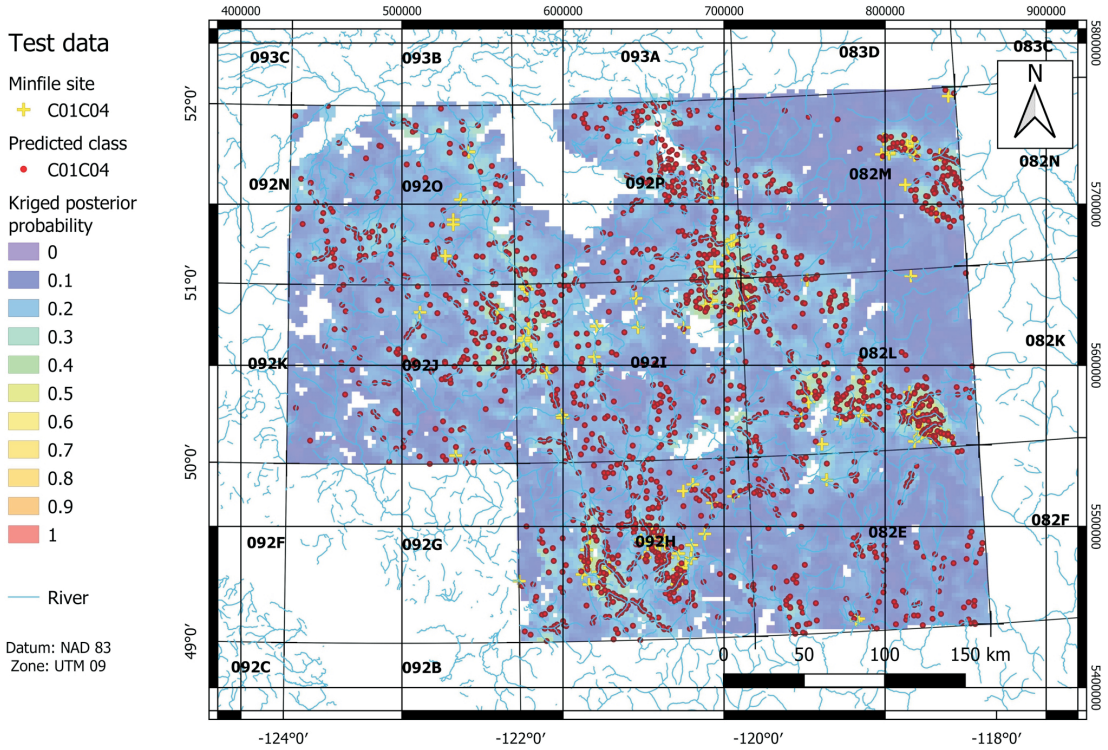


Figure 11. Geographic distribution of individual sites overlain on a kriged image of the posterior probabilities for the GroupModel C01C04 (surficial placer Au) across the map area, based on the test data and a distance threshold of 2500 m. MINFILE sites tagged as C01C04 are shown as yellow crosses. Stream-sediment sites identified as class C01C04 by random forests are shown as red dots. Areas of increased potential for C01C04 deposits are shown by colour shading.

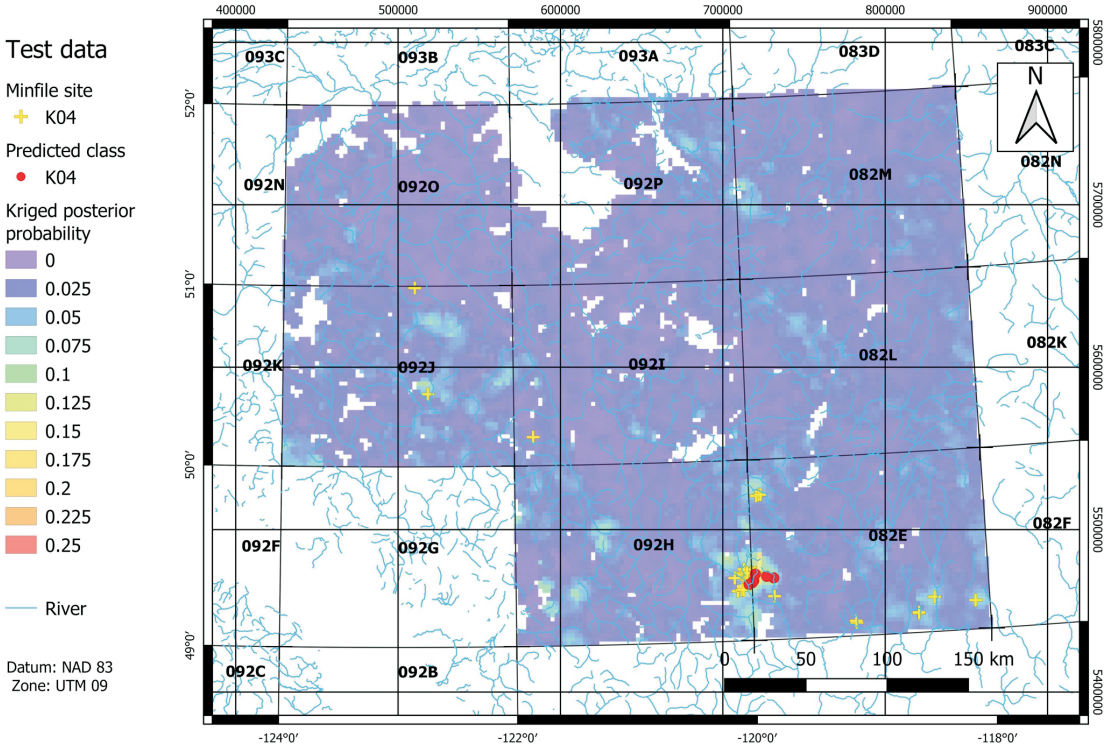


Figure 12. Geographic distribution of individual sites overlain on a kriged image of the posterior probabilities for the GroupModel K04 (Au skarn) across the map area, based on the test data and a distance threshold of 2500 m. MINFILE sites tagged as K04 are shown as yellow crosses. Stream-sediment sites identified as class K04 by random forests are shown as red dots. Areas of increased potential for K04 deposits are shown by colour shading.

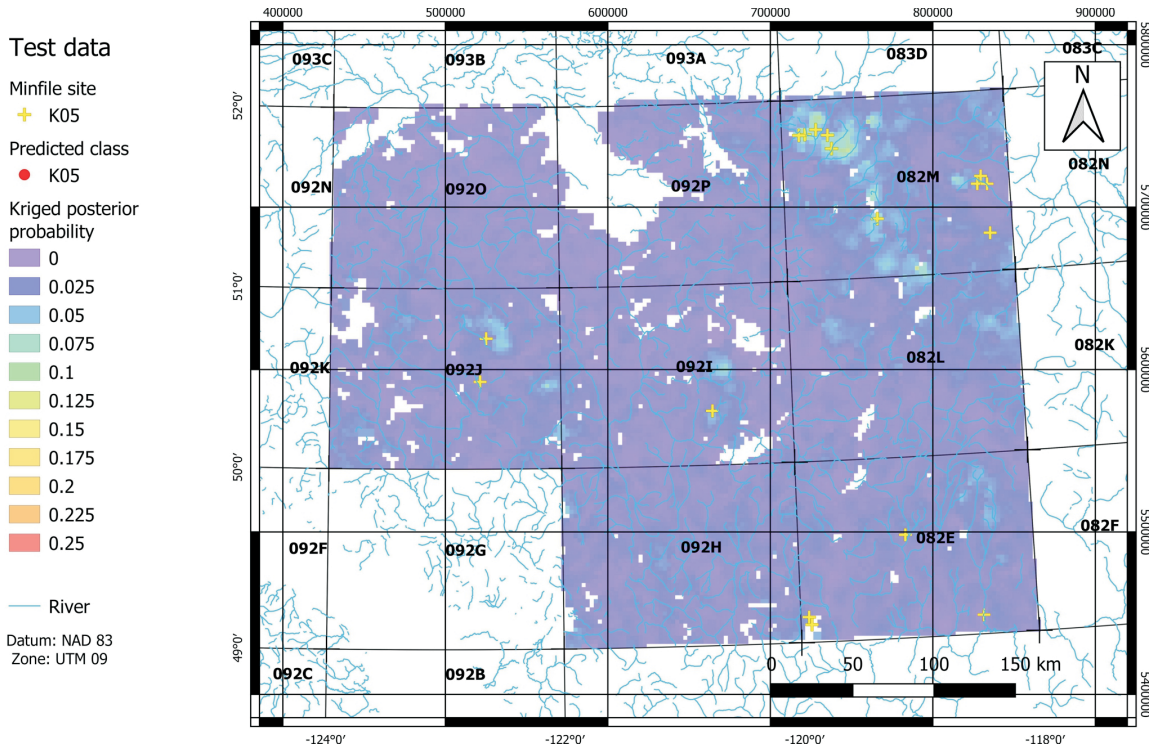


Figure 13. Geographic distribution of individual sites overlain on a kriged image of the posterior probabilities for the GroupModel K05 (W skarn) across the map area, based on the test data and a distance threshold of 2500 m. MINFILE sites tagged as K05 are shown as yellow crosses. Stream-sediment sites identified as class K05 by random forests are shown as red dots. Areas of increased potential for K05 deposits are shown by colour shading.

W skarn deposits demonstrates the unique multi-element character of these types of deposits.

Figure 14 shows a predictive map of the combined porphyry deposit models for Cu, Au and Mo (L02L04). There are clusters of L02L04 sites in the vicinity of Lornex and Highland Valley mines in 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.

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 and 5000 m yielded different and reasonable predictions. The changes in parameters will be discussed in a forthcoming report.

The predictions for the four GroupModels (C1C04, K04, K05 and L0L04) illustrate the ability to predict existing regions of known mineral-deposit potential, as well as identifying areas that have not been previously recognized as having mineral deposits.

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

1) The geochemical composition of the stream sediment associated with the mineral-deposit model 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 removed from the GroupModel classes.
2) 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 mineral deposit or distinguish between different mineral-deposit types. The method of dissolution using aqua regia is useful for sheet silicates and sulphide minerals, but aqua-regia digestion does not dissolve many silicates. Thus, some unique geochemical aspects of specific mineral-deposit types based on silicate mineral assemblages may not be recognized.
3) 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 De-

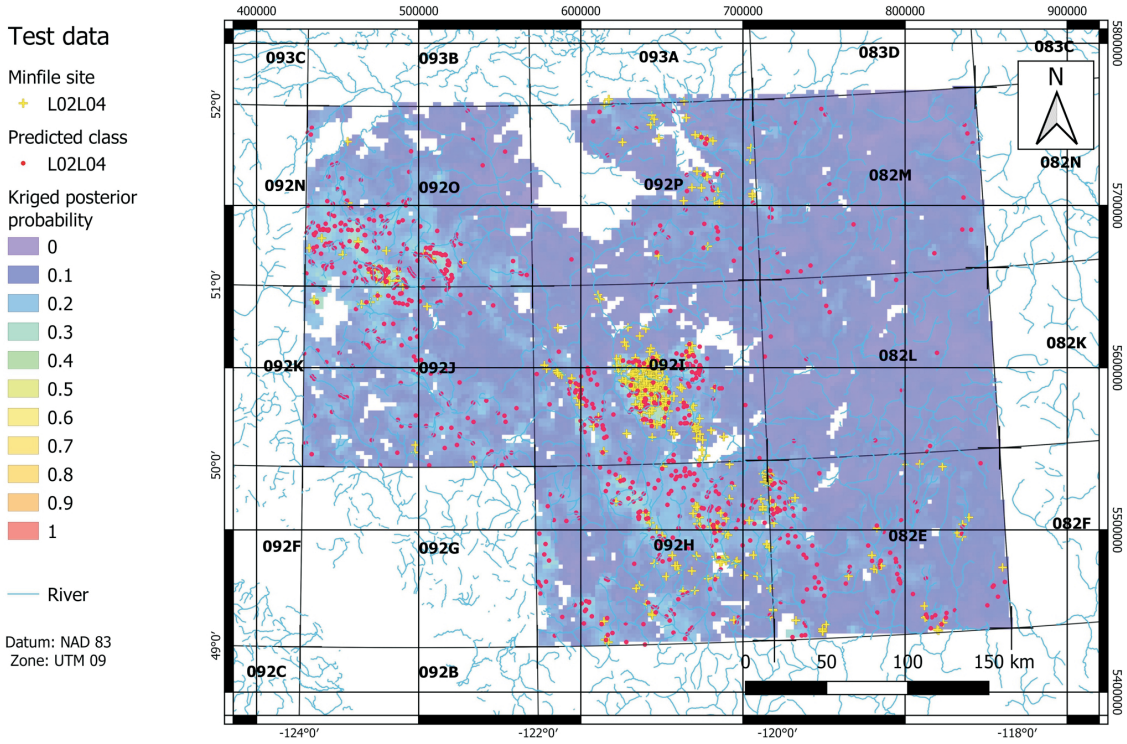


Figure 14. Geographic distribution of individual sites overlain on a kriged image of the posterior probabilities for the GroupModel L02L04 (Cu-Au-Mo porphyry) across the map area, based on the test data and a distance threshold of 2500 m. MINFILE sites tagged as L02L04 are shown as yellow crosses. Stream-sediment 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.

posit Profiles, as specified in the MINFILE field 'Deposit Type', may be incorrect or inconclusive. This can lead to misclassification errors in the subsequent application of machine-learning prediction methods.

4) 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 was 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. This requirement may be examined in subsequent work.

Given these assumptions, the results presented here indicate that various types of mineral deposit can be predicted. 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 deposit.

Further work is ongoing to provide a comprehensive picture of mineral-deposit potential, based on the BCGS Mineral Deposit Profiles.

Deliverables from this project will include files containing log-centred transformed values of the NAD 83 UTM Zone 10 co-ordinates of the stream sediments, MINFILE attributes, elements, principal-component scores, random forests votes, random forests normalized votes, random forests posterior probabilities and random forests class predictions. The files containing this information will be provided in Esri shapefile format and tab-delimited ASCII format. Kriged images will be provided in 32-bit geoTIFF format.

This paper summarizes the rationale and methodology for the prediction of mineral-deposit types based on the mineral-deposit model framework developed for BC. The use of log-ratio transforms to overcome the problem 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 forest) for predicting potential mineral-deposit sites offers a repeatable, consistent and defensible methodology, termed 'process prediction', that offers promise for the identification of prospective terrains and mineral systems. Together, they will enhance exploration strategies in the province of British Columbia.

Acknowledgments

The authors thank Geoscience BC for funding this project. Y. Cui of the BCGS is thanked for providing a peer review of this paper.

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., 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.

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., 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].

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].

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.