



CSA Global
Mining Industry Consultants

TECHNICAL REVIEW

**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
CSA Global Report № 110.2018
18 May 2018

www.csaglobal.com

Report prepared for

Client Name	Geoscience BC
Project Name/Job Code	GBCEGC01
Contact Name	Christa Pellett
Contact Title	Project Coordinator
Office Address	1101 - 750 West Pender St., Vancouver BC, V6C 2T7

Report issued by

CSA Global Office	CSA Global Canada Geosciences Ltd (Vancouver Office) Suite 610, 1155 West Pender Street Vancouver, B.C. V6E 2P4 CANADA T +1 604 681 8000
Division	Exploration

Report information

File name	R110.2018 GBCEGC01 Geoscience Technical Review - FINAL
Last edited	2018-05-18 5:07:00 PM
Report Status	Final

Author and Reviewer Signatures

Coordinating Author	Dennis Arne BSc (Hons), MSc, PhD PGeo (BC), MAIG (RPGeo)	Signature:	"Signed and Stamped"
Contributing Author	Rob Mackie BSc (Hons), MSc PGeo (BC)	Signature:	"Signed and Stamped"
Contributing Author	Chris Pennimpede BSc PGeo (BC)	Signature:	"Signed and Stamped"
Contributing Author	Eric Grunsky BSc (Hons), MSc, PhD, PGeo (BC)	Signature:	"Signed and Stamped"
Contributing Author	Matt Bodnar BSc, MSc	Signature:	"Signed"
Peer Reviewer	Adrian Martinez Vargas PhD, PGeo (BC)	Signature:	"Signed and Stamped"
CSA Global Authorisation	Daniel Wholley BSc, MAIG	Signature:	"Signed"

© Copyright 2018

Disclaimers

Purpose of this document

This Report was prepared exclusively for Geoscience BC ("the Client") by CSA Global Canada Geosciences Ltd ("CSA Global"). The quality of information, conclusions, and estimates contained in this Report are consistent with the level of the work carried out by CSA Global to date on the assignment, in accordance with the assignment specification agreed between CSA Global and the Client.

Notice to third parties

CSA Global has prepared this Report having regard to the particular needs and interests of our client, and in accordance with their instructions. This Report is not designed for any other person's particular needs or interests. Third party needs and interests may be distinctly different to the Client's needs and interests, and the Report may not be sufficient nor fit or appropriate for the purpose of the third party.

CSA Global expressly disclaims any representation or warranty to third parties regarding this Report or the conclusions or opinions set out in this Report (including without limitation any representation or warranty regarding the standard of care used in preparing this Report, or that any forward-looking statements, forecasts, opinions or projections contained in the Report will be achieved, will prove to be correct or are based on reasonable assumptions). If a third party chooses to use or rely on all or part of this Report, then any loss or damage the third party may suffer in so doing is at the third party's sole and exclusive risk.

CSA Global has created this Report using data and information provided by or on behalf of the Client and the Client's agents and contractors. Unless specifically stated otherwise, CSA Global has not independently verified that all data and information is reliable or accurate. CSA Global accepts no liability for the accuracy or completeness of that data and information, even if that data and information has been incorporated into or relied upon in creating this Report.

Results are estimates and subject to change

The interpretations and conclusions reached in this Report are based on current scientific understanding and the best evidence available to the authors at the time of writing. It is the nature of all scientific conclusions that they are founded on an assessment of probabilities and, however high these probabilities might be, they make no claim for absolute certainty.

Executive Summary

CSA Global was commissioned by Geoscience BC to compile, review and analyse new geochemical data from 14,863 archived stream sediment samples from 20 1:250,000 NTS map sheets from northwest British Columbia. The aim of the work was to undertake an integrated study of these new data using classical, as well as more advanced modern data analysis techniques to assist with targeting economic mineral deposits.

The study region is predominantly underlain by rocks of the Stikine terrane, with subordinate rocks from the Quesnel and Cache Creek terranes. Economically significant porphyry-hosted deposits, including alkalic Cu-Au, calc-alkaline porphyry Cu and low-F porphyry Mo deposits, epithermal Au-Ag, volcanic-hosted massive sulphide deposits, and polymetallic base metal and precious metal vein deposits are present in the region. Exploratory data analysis indicates that a strong mineralization signature exists in the data but that key commodity and pathfinder elements are also influenced by lithological variations, particularly in Middle Jurassic to early Cretaceous sediments of the Bowser Basin. Evidence for mineralization, lithology and metal scavenging are present in principal components generated following a centred-log ratio transformation to adjust for the effects of geochemical closure.

Several classical and new approaches to filter the data for the effects of variable lithology, metal scavenging by secondary Fe and/or Mn hydroxides, clays or organic material have been trialled. These include:

- Z-score levelling of Log_{10} geochemical data for the dominant catchment lithology
- Z-score levelling of Log_{10} geochemical data for the dominant terrane
- Multiple regression analysis of geochemical data against the areal proportions of each lithology
- Multiple regression analysis of geochemical data against principal components representing lithology or metal scavenging.

The results from these data treatments have been used to construct a series of weighted sum models (WSMs) for several mineral deposit types. These model scores have been corrected for the effects of dilution using the square root of the catchment basin area. The success of identifying known British Columbia Geological Survey (BCGS) MINFILE mineral occurrences in the study area suggests that most of the approaches listed above have been effective for porphyry Cu-Au WSMs provided they include a correction for dilution. Little improvement was observed for epithermal Au-Ag WSMs because Au does not show a lithological or metal scavenging control.

In addition to these methods, independent component analysis and two supervised learning approaches (random forests and allocation/typicality) have been assessed. Random forests and allocation/typicality computations have been undertaken using both centred-log ratio transformed data and residuals from multiple regression analysis of catchment basin lithology to assess the effect of correcting the input data for lithological effects prior to supervised learning. Random forests show the most promise in predicting the presence of further mineralization for most mineral deposit types, as well as being able to distinguish between different mineral deposit types. The use of residuals following regression against the proportion of different lithologies in the catchments basins in the computations to level for lithology does not have a positive effect on either the random forests or allocation predictions.

Quality control data from re-analyses of the samples by aqua regia digestion and mixed ICP-MS/OES undertaken between 2002 and 2017 have been reviewed. Aside from sulfur data on one map sheet, the geochemical show no significant batch-level variations requiring data levelling. Values for data falling below the lower limit of detection have been imputed and original neutron activation and fire assay data for Au have

been used in preference to new ICP-MS Au data, which are imprecise. Sample locations have been revised based on comparison with scanned images of the original topographic map sheets used during sample collection. Catchment basins for most samples are provided and the resulting polygons used to query bedrock and Quaternary geology, as well as known mineral occurrences, for each catchment basin.

Key Findings and Practical Use of the Data

Several key findings included in the report concerning the quality of the RGS data have important implications for its use by explorers and its utility in wide-ranging, sophisticated, analytical approaches.

- A) The quality of the historical sample locations in the project area was variable, with up to 50% of locations on some map sheets having to be revised based on the archived sample maps. These revised locations are a significant improvement for interpretation of the data and follow-up investigations.
- B) The quality of the re-assay data was generally good (except for Au), with little or no adjustment of the data required in compiling the dataset for the project area. This means that most of the compiled data are comparable across adjacent map sheets in the project area.
- C) An assessment of geochemical dilution as a function of basin area suggests that a significant portion of the study area was not effectively sampled by the original sampling programs and that potential for new discoveries through in-fill sampling remains for some areas. Catchments larger than 25 km² are under-sampled and warrant detailed follow-up sampling where favorable geological settings for mineral deposits exist.

The report sets out in detail tests and comparisons of several techniques that can be applied when adding value to regional stream sediment data; however, choosing the outputs which have the most practical application for explorers may be daunting. The authors have therefore identified the following products as having the most utility:

- D) Several data analysis approaches for correcting lithology and metal scavenging accompanied by a dilution correction for catchment basin area resulted in improved porphyry Cu deposit targeting models. Specifically, dilution-corrected weighted sums models using data levelled by either dominant lithology or terrane, or data obtained from multiple regression analysis against catchment lithology, identify more known porphyry Cu deposits compared to models based on raw or levelled data alone.
- E) Of the two supervised learning data analysis methods investigated, random forests show an ability to distinguish between different mineral deposit types and predict realistic areas for follow-up investigation. While these models were not independently tested in this study, they may generate plausible target catchment basins when integrated with other geological information.

The information generated during this project is contained in ArcGIS workspaces for those with geographic information systems (GIS) capability and is available on the Geoscience BC Earth Science Viewer. The main body of this report describes the methodologies used and the main assumptions associated with processing of the geochemical data to generate enhanced interpretations.

Contents

Report prepared for	1
Report issued by	1
Report information	1
Author and Reviewer Signatures	1
DISCLAIMERS	2
Purpose of this document	2
Notice to third parties	2
Results are estimates and subject to change	2
EXECUTIVE SUMMARY	3
Key Findings and Practical Use of the Data	4
1 INTRODUCTION.....	8
1.1 Context, Scope and Terms of Reference	8
1.2 Geological Setting and Mineral Deposits.....	10
1.3 Catchment Analysis Approach.....	11
2 METHODOLOGY	14
2.1 Data Compilation and Conditioning	14
2.2 Sampling and Analysis	14
2.3 Geochemical Data Quality	15
2.4 Location Validation.....	17
2.5 Derivation of Catchment Basins	18
2.6 Attribution of Lithology	20
2.7 Mineral Deposits	21
2.8 Weighted Sum Models	21
2.9 Principal Component Analysis	22
2.10 Independent Component Analysis	22
2.11 Advanced Data Analysis Methods	22
3 RESULTS	23
3.1 Raw Element Distributions	23
3.2 Conventional Levelling for the Effects of Lithology	26
3.3 Principal Component Analysis	32
3.4 Independent Component Analysis	41
3.5 Weighted Sums Models.....	46
3.6 Sampling Effectiveness	62
3.7 Advanced Data Analysis Methods	67
3.7.1 Random Forests.....	67
3.7.2 Allocation/Typicality.....	71
3.7.3 Residuals After Multiple Regression Analysis.....	74
4 SUMMARY AND DISCUSSION	82

5	RECOMMENDATIONS.....	84
6	ABBREVIATIONS AND ACRONYMS	85
7	ACKNOWLEDGEMENTS	86
8	REFERENCES.....	87
9	APPENDIX 1: ADVANCED ANALYSIS METHODS.....	89
	Random Forests Classification Algorithm	89
	Allocation /Typicality	90
10	APPENDIX 2: ASSOCIATED FILES	93
	Listing of shapefiles	93
	Listing of abbreviations.....	93

Figures

Figure 1:	Location of study area.....	9
Figure 2:	Regional geology of northwest BC	11
Figure 3:	Examples of Shewhart charts for selected standard reference materials and elements.....	16
Figure 4:	Box plot summary of S in RGS standard Red Dog plotted by map sheet	17
Figure 5:	Examples of sample location issues	18
Figure 6:	Catchment basins from northwest BC	19
Figure 7:	Distribution of raw Cu in stream sediments from northwest BC.....	24
Figure 8:	Distribution of raw Sb in stream sediments from northwest BC	25
Figure 9:	WSMs for the Golden Triangle region.....	26
Figure 10:	Tukey plots showing the distribution of lithologically-controlled major and trace elements attributed by dominant lithology in each catchment	27
Figure 11:	Tukey plots showing the distribution of commodity and pathfinder trace elements attributed by dominant lithology in each catchment	28
Figure 12:	Tukey plots showing the distribution of commodity and pathfinder trace elements attributed by terrane... <td>29</td>	29
Figure 13:	Gridded percentile images comparing raw Zn data and Zn data leveled by dominant lithology (Zn_ppm-ZLog-DomGeo), geological terrane (Zn_ppm-ZLog-DomTerrain), and all catchment lithologies (Resid_Log Zn_MR_lith)	30
Figure 14:	Gridded percentile images comparing raw Sb data and Sb data leveled by dominant lithology (Sb_ppm-ZLog-DomGeo), geological terrane (Sb_ppm-ZLog-DomTerrain), and all catchment lithologies (Resid_Log Sb_MR_lith)	31
Figure 15:	SCREE plot of eigenvalues following PCA.....	32
Figure 16:	Line plots showing the elemental loadings on the first four principal components	33
Figure 17:	Percentile gridded image of principal component 1 (PC1)	34
Figure 18:	Percentile gridded image of principal component 2 (PC2)	35
Figure 19:	Percentile gridded image of inverse principal component 3 (-PC3).....	36
Figure 20:	Percentile gridded image of inverse principal component 4 (-PC4).....	37
Figure 21:	Tukey plots showing the influence of dominant catchment lithology on the first six principal components	39
Figure 22:	Tukey plots showing the influence of terrane on the first six principal components	40
Figure 23:	Inverse IC2 defining the Bowser Basin	42
Figure 24:	IC3 associated with ultramafic rocks of the Cache Creek terrane.....	43
Figure 25:	IC6 associated with the Quesnel terrane	44
Figure 26:	Inverse IC8 associated with epithermal Au-Ag deposits	45
Figure 27:	Porphyry Cu-Au WSM using raw geochemical data corrected for dilution.....	48
Figure 28:	Epithermal Au-Ag WSM using raw geochemical data corrected for dilution.....	49

Figure 29:	Porphyry Cu-Au WSM using geochemical data levelled by dominant catchment basin lithology	50
Figure 30:	Epithermal Au-Ag WSM using geochemical data levelled by dominant catchment basin lithology	51
Figure 31:	Porphyry Cu-Au WSM using geochemical data levelled by dominant terrane	52
Figure 32:	Epithermal Au-Ag WSM using geochemical data levelled by dominant terrane	53
Figure 33:	Porphyry Cu-Au WSM using residuals following multiple regression analysis against the proportions of each lithology in catchment basins	54
Figure 34:	Epithermal Au-Ag WSM using residuals following multiple regression analysis against the proportions of each lithology in catchment basins	55
Figure 35:	Porphyry Cu-Au WSM using residuals following multiple regression analysis against principal components	56
Figure 36:	Epithermal Au-Ag WSM using residuals following multiple regression analysis against principal components	57
Figure 37:	Polymetallic Ag-Pb-Zn veins WSM using residuals following multiple regression analysis against principal components	58
Figure 38:	Magmatic Ni-Cu and ultramafic rocks WSM using residuals following multiple regression analysis against principal components	59
Figure 39:	Summary of true positives for porphyry Cu-Au and epithermal Au-Ag deposits	61
Figure 40:	Plot of raw Cu, Cu data levelled using two different methods, and a WSM for porphyry Cu-Au deposits against catchment area	63
Figure 41:	Comparison of WSM for porphyry Cu-Au deposits using data levelled for dominant lithology with dilution-correction (left) and uncorrected (right)	64
Figure 42:	Comparison of WSM for epithermal Au-Ag deposits using data levelled for dominant lithology with dilution-correction (left) and uncorrected (right)	65
Figure 43:	Map illustrating the areal extent of catchment basins >25 km ² that may not have been effectively sampled	66
Figure 44:	RF predictions for alkalic porphyry Cu-Au deposits	68
Figure 45:	RF predictions for epithermal Au-Ag deposits	69
Figure 46:	Comparison of RF predictions for epithermal Ag-Au and three types of porphyry style deposits	70
Figure 47:	Indices of typicality for porphyry Cu-Mo-Au deposits	72
Figure 48:	Indices of typicality for epithermal Au-Ag deposits	73
Figure 49:	Inverse PC1 residuals following multiple regression analysis against lithology	75
Figure 50:	Inverse PC2 residuals following multiple regression analysis against lithology	76
Figure 51:	RF predictions for alkalic porphyry Cu-Au deposits using residuals following multiple regression against catchment lithology	78
Figure 52:	RF predictions for epithermal Au-Ag deposits using residuals following multiple regression against catchment lithology	79
Figure 53:	Typicality indices for porphyry Cu-Au deposits using residuals following multiple regression against catchment lithology	80
Figure 54:	Typicality indices for epithermal Au-Ag deposits using residuals following multiple regression against catchment lithology	81

Tables

Table 1:	Summary of NTS map areas in the study area and data sources	10
Table 2:	Lithological units used in this study with number and proportion of catchment basins in which they occur as the dominant lithology	20
Table 3:	Simplified mineral deposit classification	21
Table 4:	Processing specifications for WSMs using raw and levelled data	46
Table 5:	Processing specifications for principal component WSMs	47
Table 6:	Proportions of true positives in delineating mineralization	60
Table 7:	Summary of R-scores for PCA of residual following multiple regression analysis	74

1 Introduction

1.1 Context, Scope and Terms of Reference

Regional stream-sediment geochemistry is important to mineral exploration companies, governments and First Nations communities because it helps inform mineral exploration potential in suitable geographical terranes. From 2002 to 2017 archived stream-sediment sample material were reanalyzed by inductively coupled plasma–mass spectrometry (ICP-MS) and inductively coupled plasma–atomic emission spectrometry (ICP-AES), increasing the number of elements reported and lowering detection limits compared to historical analytical methods. These improvements allow for a more rigorous assessment of the variation in sediment composition and the filtering of effects related to lithological controls and secondary scavenging, which can mask signals related to metallic metal deposits. Some of these new data covering southern and central British Columbia (BC) have been evaluated using catchment basin analysis and a variety of multivariate statistical approaches (e.g. Arne and Bluemel, 2011; Arne and Brown, 2015); however, data from much of northern BC have not yet been assessed.

This project is a regional assessment of stream-sediment geochemical data from various Geoscience BC re-analysis programs encompassing parts of 201:250 000 scale National Topographic System (NTS) map areas as shown in Figure 1 and listed in Table 1. The limits of the project were based on the following criteria:

- Areas where samples have been re-analysed by ICP-MS
- Areas with an absence of regionally extensive Quaternary cover
- Areas with a lack of previous geochemical interpretations of stream-sediment data
- Areas with a high degree of interest from mineral exploration companies.

The areas covered by the Spatsizi Plateau Wilderness Provincial Park and the Tatlatui Provincial Park have also been excluded from this study. The study area was further designed to include regions mainly underlain by the Stikine, Quesnel and Cache Creek terranes, but also includes the overlying Bowser Basin. This geological filter was applied because the fringing terranes to the east and west will have distinctly different lithological background values for many elements, which could limit the ability to “isolate” geochemical signals associated with mineralization within the targeted terranes. The adjacent terranes are also not significant hosts for base-metal or precious-metal mineralization and thus are of less immediate exploration interest.

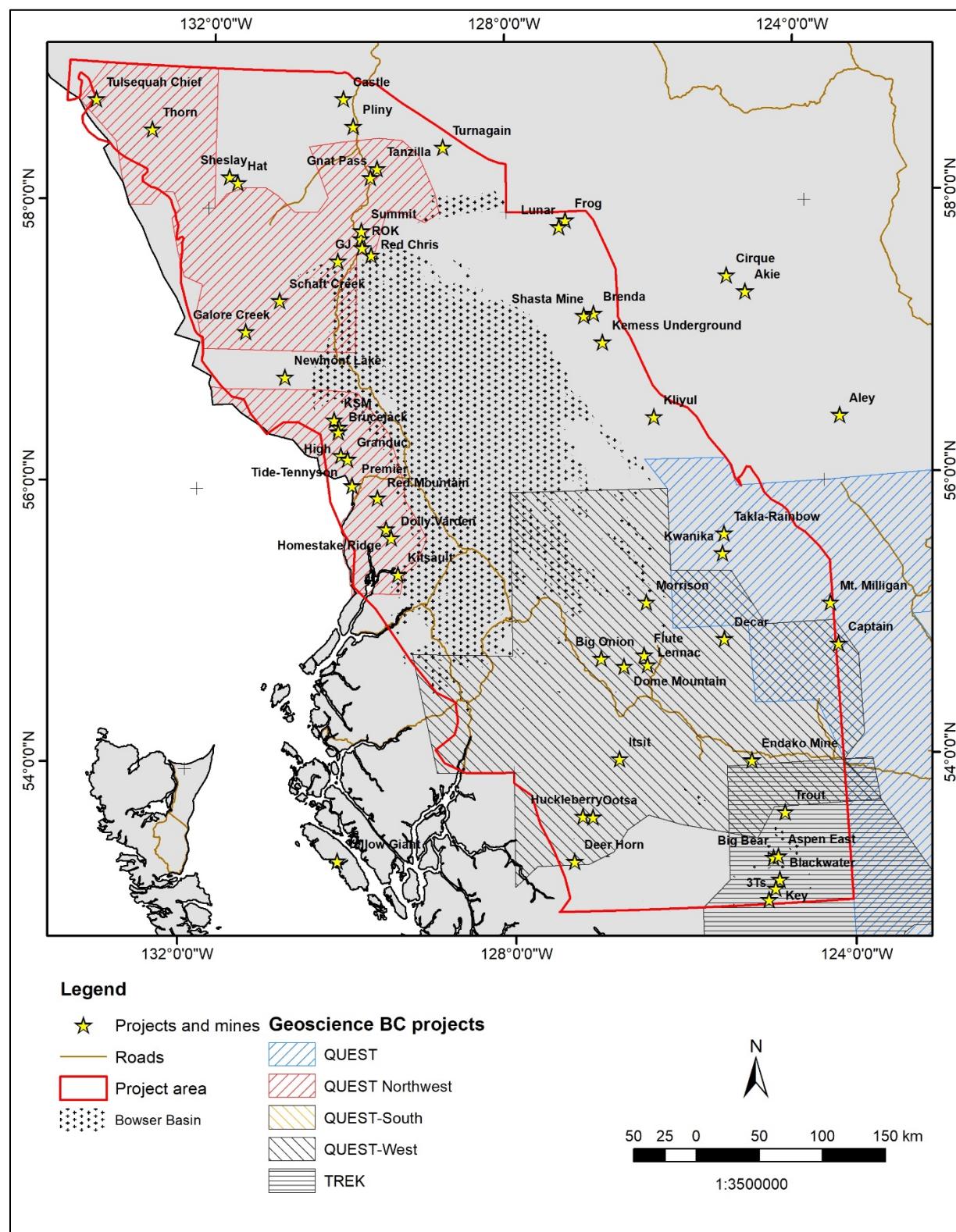


Figure 1: Location of study area

Note: Shown are the areas covered by previous Geoscience BC initiatives, the locations of major mineral deposits, and the area of bedrock covered by Bowser Basin sedimentary rocks that post-date most periods of mineralization.

Table 1: Summary of NTS map areas in the study area and data sources

NTS Map Sheet Open File Report	GBC 2008-03	GBC 2008-11	GBC 2012-05	GBC 2011-02	GBC 2009-05	GBC 2009-11	GSC OF176	GBC 2017-4
093E					x			x
093F					x	x		
093K						x	x	
093L					x			
093M					x			
093N	x							
094C					x			
094D					x			
094E					x			
103I		x						
103J		x						
103O							x	
103P							x	
104A				x				
104B				x				
104F/G				x				
104H				x				
104I				x				
104J				x				
104K			x					

1.2 Geological Setting and Mineral Deposits

The study area covers much of the north-western Stikine terrane of central and northern BC (Figure 2), including important regions of mineral exploration and development. The Stikine terrane is part of the Intermontane Belt, which stretches through BC, Yukon and into Alaska (Colpron *et al.*, 2006). It comprises a north to northwest trending allochthonous belt of dominantly Devonian to Jurassic sedimentary and volcanic rocks intruded by coeval Late Devonian, Triassic and Jurassic plutonic rocks (Gunning *et al.*, 2006). Paleozoic volcanic and sedimentary rocks of the Stikine terrane include calcalkaline to tholeiitic sequences largely dominated by basalt, marine limestone and lime mudstone. Volcanism and continued deposition of marine sediments continued through the Mesozoic until the mid-Cretaceous, following a break during the late Permian. The Stikine terrane was accreted to the adjacent Cache Creek and Quesnel terranes by the Middle Jurassic, followed by the deposition of Middle Jurassic to Early Cretaceous fluvial to marine clastic sediments that lie unconformably over the Stikine terrane in the Bowser and Sustut basins (Ricketts *et al.*, 1992). Marine sedimentation and subaerial volcanism resumed in the Late Cretaceous or Paleogene and Neogene.

Parts of the study area have previously been included in Geoscience BC projects that have focused on well-mineralized areas of the Stikine and Quesnel terranes (Figure 1). The informally named “Golden Triangle”, stretching from south of the Red Mountain Au-Ag deposit to the Red Chris Cu-Au mine, was included in the QUEST-Northwest project. This area contains volcanic-hosted massive sulphide (VHMS) deposits at Granduc and Eskay Creek, as well as porphyry Cu-Au and epithermal Au-Ag deposits such as Brucejack and KSM. The QUEST-Northwest project to the north of the Golden Triangle also encompassed the Thorn Au-Ag and Tulsequah Chief VHMS deposits. An area centred on the town of Smithers formed the basis of the QUEST-

West project that includes porphyry Cu-Au and Mo deposits, including Endako. Porphyry Cu-Au deposits in the western portion of the original QUEST project area are also included within the study area. The northern portion of the TREK project area, which contains epithermal Au-Ag deposits such as Blackwater, occurs in the southeast corner of the study area. The majority of known mineral deposits in the study area are of magmatic hydrothermal origin, associated mainly with calcalkaline to alkaline Triassic and Jurassic intrusive rocks, and to a lesser extent with Late Cretaceous–Paleogene calcalkaline intrusive rocks in the central and southern Stikine terrane (Nelson *et al.*, 2013). The main mineral deposit types that are the focus of the current investigation, therefore, are porphyry Cu-Au, epithermal Au-Ag and, less commonly, VHMS Zn-Cu-Ag-Au. Skarn and numerous polymetallic vein deposits in the study area will also have overlapping geochemical signatures with the main deposit types.

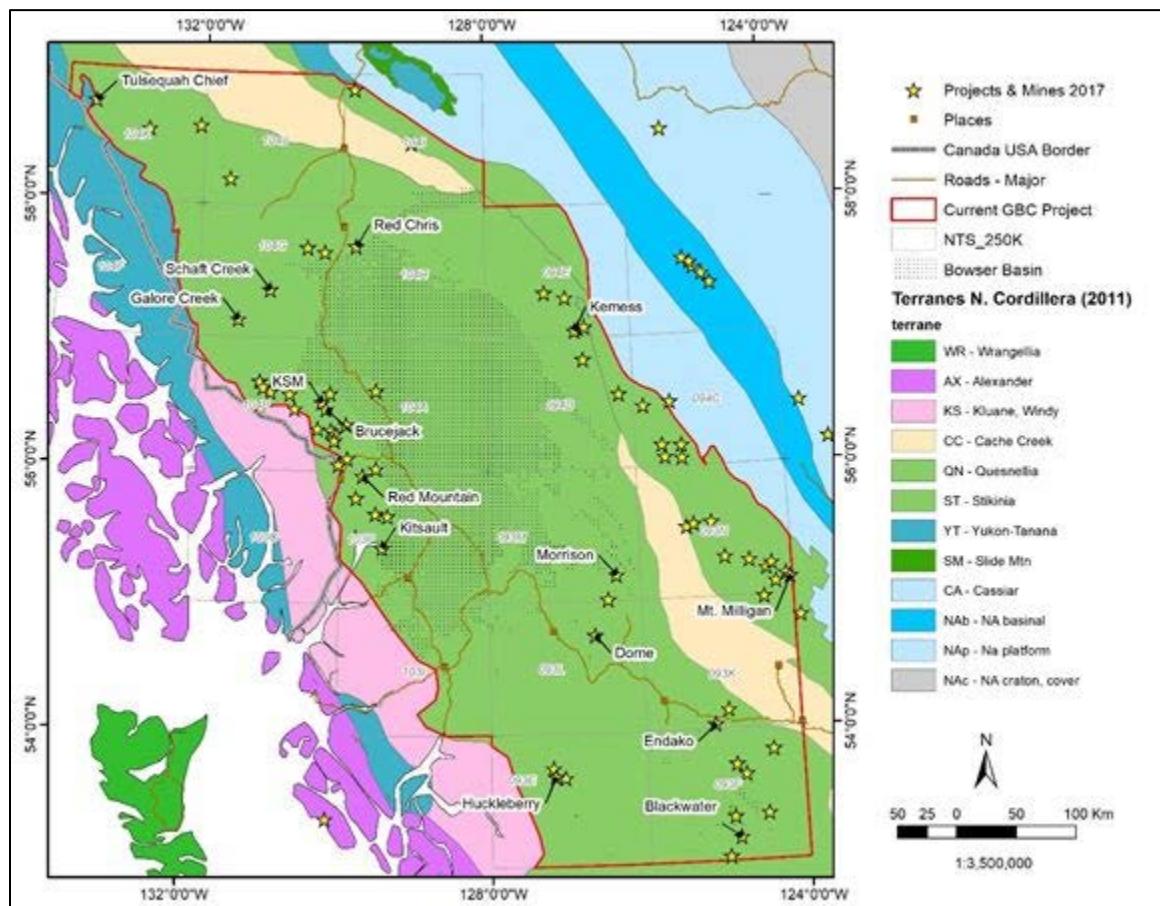


Figure 2: Regional geology of northwest BC

Note: The project area is largely confined to the Stikine, Quesnel and Cache Creek terranes.

1.3 Catchment Analysis Approach

The approach to catchment analysis used in this study is aligned with the concept of productivity described by Hawkes (1976) and further expanded on by Pan and Harris (1990) and Moon (1999). Bonham-Carter and Goodfellow (1986) demonstrated that catchment lithology was the main control on observed variation in stream sediment data from the Nahanni region of the Yukon Territory. Other effects, such as catchment area, possible adsorption of some elements onto secondary Fe or Mn hydroxides, or onto organic material, and water pH were considered minor by comparison. A similar conclusion was reached by Carranza and Hale

(1997) in a study of the main controls on stream sediment geochemistry in the Philippines. Bonham-Carter *et al.* (1987) applied a similar approach to the analysis of stream sediment data from the Cobequid Highlands of Nova Scotia and further concluded that use of the dominant lithology in the catchment basins was not as effective as the areal extent of all lithological units in the catchment using multiple regression analysis, an approach further developed by Carranza (2009). An intermediate approach that is computationally efficient is to use the presence of a single lithological unit or units to assess catchment basins in a pass/fail approach, and this may be as effective as using the entire catchment geology (Bonham-Carter *et al.*, 1987; Arne and Brown, 2015).

Arne and Bluemel (2011) levelled re-analysed stream sediment data from the Geoscience BC QUEST South project area using the dominant bedrock lithology identified in the catchments, as well as regression analysis to correct for the effects of possible metal scavenging onto secondary Fe hydroxides. Heberlein (2013) levelled newly acquired ICP-MS data for two map sheets in the Yukon using dominant bedrock lithology after demonstrating bedrock control on some elements, such as Cu. One of the fundamental assumptions of approaches using mapped or interpreted bedrock geology is that similar erosion rates affect all lithological units within a catchment area, although this is unlikely to be the case in areas of variable relief (Granger and Schaller, 2014), and that the bedrock geology is well known.

The simplified approach used by Arne and Bluemel (2011) and Heberlein (2013) may not always be appropriate in large catchment basins where multiple lithological units are to be anticipated, as argued by Bonham-Carter *et al.* (1987), nor does it account for variable erosion rates within the catchment. A spatially insignificant rock unit may contribute disproportionately to the geochemistry of a stream sediment sample from the catchment if it is relatively enriched in commodity or pathfinder elements. A more accurate approach would be to estimate a weighted background value for each catchment and element of interest using background values for individual lithological units and apply a weighting to these values based on the proportion of each unit exposed within the catchment. Such weightings assume a constant supply of sediment from each lithology and may require adjustment to account for local variations in relief and erosion weights. Topography and variable weathering effects for different lithological units are no doubt important factors in controlling the geochemical input from each lithology in a catchment basin (e.g. Mackie *et al.*, 2017), but are difficult to correct data for. Differences between calculated background metal values and those observed may indicate the presence of an anomalous metal source within a catchment basin.

Following the completion of an extensive re-analysis program completed by the Yukon Geological Survey (YGS) that resulted in new ICP-MS data for 24,279 archived regional stream sediment samples for the southern two-thirds of the Territory, a series of map products were generated by CSA Global in 2015 and 2016 targeting different mineral deposit types (Mackie *et al.*, 2015). Interpretation of the new geochemical data investigated two approaches to correct for the influence of variable bedrock lithology and metal scavenging on commodity and pathfinder elements of interest for mineral exploration.

One approach used by Mackie *et al.* (2015) levelled individual elements by the dominant bedrock lithology within the catchment basins using the approach described by Arne and Bluemel (2011) and Heberlein (2013), and as requested by the YGS. The catchment basins used were generated from a digital elevation model (DEM) by the YGS using the hydrology module in ESRI ArcMap™. Levelled data were then used to construct weighted sum models (WSMs) for specific mineral deposit types (Garrett and Grunsky, 2001). This approach requires that the sample location be accurately located on the stream that was sampled, assumes constant sediment supply from all lithological units, and requires that the geology of the catchment basins is well constrained. The influence of geochemically distinct but geographically minor lithological units is under-estimated using this approach.

As noted by Bonham-Carter *et al.* (1987) and Arne and Brown (2015), levelling the geochemical data by the dominant lithology in the catchment basin does not necessarily provide the best interpretative outcome. Therefore, the second approach used by Mackie *et al.* (2015) involved exploratory data analysis of the geochemical data using principal component analysis to identify geochemical associations related to lithology, scavenging of metals by organics, clays or secondary Fe and/or Mn hydroxides, or to mineral deposits. Individual commodity and pathfinder elements were regressed against one or more principal components in which they were prominent to normalize the data for the effects of variable lithological background geochemistry and/or the effects of metal scavenging. WSMs were generated using residuals calculated for individual samples, where appropriate or raw element data were not appropriate. Even though the catchments were not used to derive geology for the samples using this second approach, the usefulness of the resulting WSMs and correction of the data for the effects of dilution depends very much on identifying the correct catchments for further investigation. This approach relies on the main principal components clearly reflecting lithological or scavenging element associations. However, this approach has the benefit of being applicable in areas where the bedrock geology is poorly known.

The present study is designed to further this work using re-analysed geochemical data for stream sediment samples from northwest BC. A variety of interpretative methods have been applied to the data so that the results from different approaches can be compared. In addition, new advanced analytical methods have been applied, including independent component analysis and two supervised machine learning computations: allocation/typicality and random forests.

2 Methodology

2.1 Data Compilation and Conditioning

The data assembly for this project built on a previous compilation of historical regional geochemical data conducted by the BC Geological Survey (BCGS; Rukhlov and Naziri, 2015). This compilation was updated to include all the ICPMS data generated through the various recent Geoscience BC projects. A summary of NTS map areas and public reports is provided in the form of a matrix in Table 1. Note that the re-analysis data were obtained from two separate laboratories. The bulk of the re-analyses were undertaken at Bureau Veritas Minerals (formerly Acme Analytical Laboratories Ltd.; Vancouver, BC), whereas re-analyses of samples from part of NTS map area 093F were undertaken by Eco-Tech Laboratories Ltd (Kamloops, BC) in 2009. In addition, the re-analyses span 2002 to 2017; a period of 15 years. The use of two laboratories and the time span involved raises the possibility that some levelling of data to account for systematic variations within the compiled dataset may be required.

An assessment of the database indicates that 36 elements have concentrations typically greater than the lower limits of detection while providing maximum spatial coverage. These include 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, Te, Th, Ti, Tl, U, V, W and Zr. Data for Au were not included in the principal components used for regression analysis given the large number of censored data, but they were included in WSMs where relevant, and for the alternative data analysis methods (allocation/typicality and random forests analysis). This process resulted in the selection of 14,863 samples, excluding quality-control samples but including those collected within the Bowser Basin. Data for values below the lower limits of detection were imputed by the method of nearest neighbour replacement estimates (Palarea-Albaladejo *et al.*, 2014). The adjusted data were then used for subsequent multivariate statistical analyses. Multivariate analyses of data for these elements, including principal and independent component analysis following a centred-log ratio (CLR) transformation to remove the effects of closure (Aitchison, 1986), are consistent with a dominant spatial control that can be related to regional lithological and mineral deposit trends for most elements. Commodity and pathfinder elements show a general spatial correlation with major mineral deposits and districts. Except for S and loss on ignition (LOI), most raw-element and principal-component gridded maps show no abrupt variations that might be related to map area and/or survey boundaries, indicating that interpretation of the vast bulk of the geochemical data can proceed without the necessity of levelling the data for analytical batch effects.

2.2 Sampling and Analysis

The original regional geochemical survey (RGS) stream sediment samples were generally collected from first- and second-order streams under Canada's National Geochemical Reconnaissance (NGR) program at an approximate density of one sample per every 13 km² in BC (e.g. Friske *et al.*, 2003). The samples were sieved to –80 mesh (<177 µm) prior to analysis, originally using either an aqua regia digestion followed by atomic absorption spectroscopy (AAS), instrumental neutron activation analysis (INAA) or, for Au only, lead-collection fire assay. Splits of the original sieved sample material were archived in Ottawa, Ontario at the Geological Survey of Canada (GSC). This material was sampled for re-analysis using a dilute aqua-regia digestion (1 HCl:1 HNO₃:1 H₂O) at Bureau Veritas or standard aqua regia digestion at Eco-Tech Laboratories followed by analysis using both ICP-MS and ICP-AES. The original quality-control samples (field duplicates, blind or pulp duplicates and reference materials) were also sampled for re-analysis.

As demonstrated by Arne and Bluemel (2011), Arne and MacFarlane (2014), and Arne and Brown (2015), the precision of the original RGS Au data is poor, even when using up to 10 g of sample material. The original data, however, are superior to Au analysis of the archived material by ICP-MS using a 0.25–0.5 g sample aliquot (Mackie *et al.*, 2017; this paper). For this reason, the original Au data have been used for the present compilation. Duplicate Au analyses have been averaged where available.

2.3 Geochemical Data Quality

Quality-control data from the relevant Geoscience BC reanalysis projects have been provided by W. Jackaman (pers. comm., 2017). Data from a total of 1,379 field duplicate pairs, 1,394 pulp (blind) duplicate pairs, 1,476 original RGS reference materials and 2,350 laboratory reference materials have been assessed for Au, As, Cu and Mo. These elements are representative of the types of mineral deposits found in the region of most interest (porphyry Cu-Mo, epithermal Au-Ag and VHMS Cu-Zn). Descriptions could not be obtained for all the reference materials submitted with the original stream-sediment batches, although data from them can still be assessed in terms of data consistency. The same reference materials were also not used throughout the original sampling programs, so it is difficult to assess continuity beyond a 10-year period and is only possible for the Red Dog RGS reference material.

As expected, the precision and accuracy of the Au data are poor (Figure 3a, b), but improve for As, Cu and Mo. Neither the Acme DS7 (Acme Analytical Laboratories' in-house reference material) or Red Dog reference materials were designed for accurate and precise Au analyses, although they do demonstrate the amount of variability that might be observed in actual stream-sediment samples sieved to <177 µm. The control limits shown for the Red Dog reference material in Figure 3 were calculated from long-term averages, excluding outliers (R. Lett, pers. comm., 2017), and there are clearly numerous outliers that can be explained by inhomogeneity of the material and a nuggety distribution of Au particles in the reference material.

The accuracy of the Cu and As data is variable in Acme DS7 and Red Dog reference materials, respectively, but this is not surprising given that the re-analysis of the stream-sediment samples was completed between 2002 and 2017 (Figure 3c, d). Clear breaks in data for the reference materials occur where there is a break in the time sequence of the analyses; however, slight differences in the aqua regia digestions used between labs, as well as variations over time at Bureau Veritas, are not considered to be significant for the interpretation of the data, although the Au ICP-MS data must be treated with caution.

In the case of S, map sheet 093N was observed to have elevated S values in stream sediment samples compared to the surrounding map sheets. An assessment of S data from the RGS Red Dog standard reference material indicates that map sheets 093N (included in this project), 093A and 093B (both not included in the project) from Geoscience BC report 2008-03, show consistently higher S values compared to surrounding map sheets (Figure 4). The offset compared to the median S value from the Red Dog SRM from all other maps sheets included in this study is 0.28%. The S data from map sheet 093N have been reduced by a factor of 2.5 to correct for this offset, which has the effect of causing many S values from the map sheet to fall below the lower limit of detection for S of 0.02%. Data for Red Dog from other map sheets also show a spread in S values consistent with elevated data in some of those map sheets, but these have not been corrected in the present study.

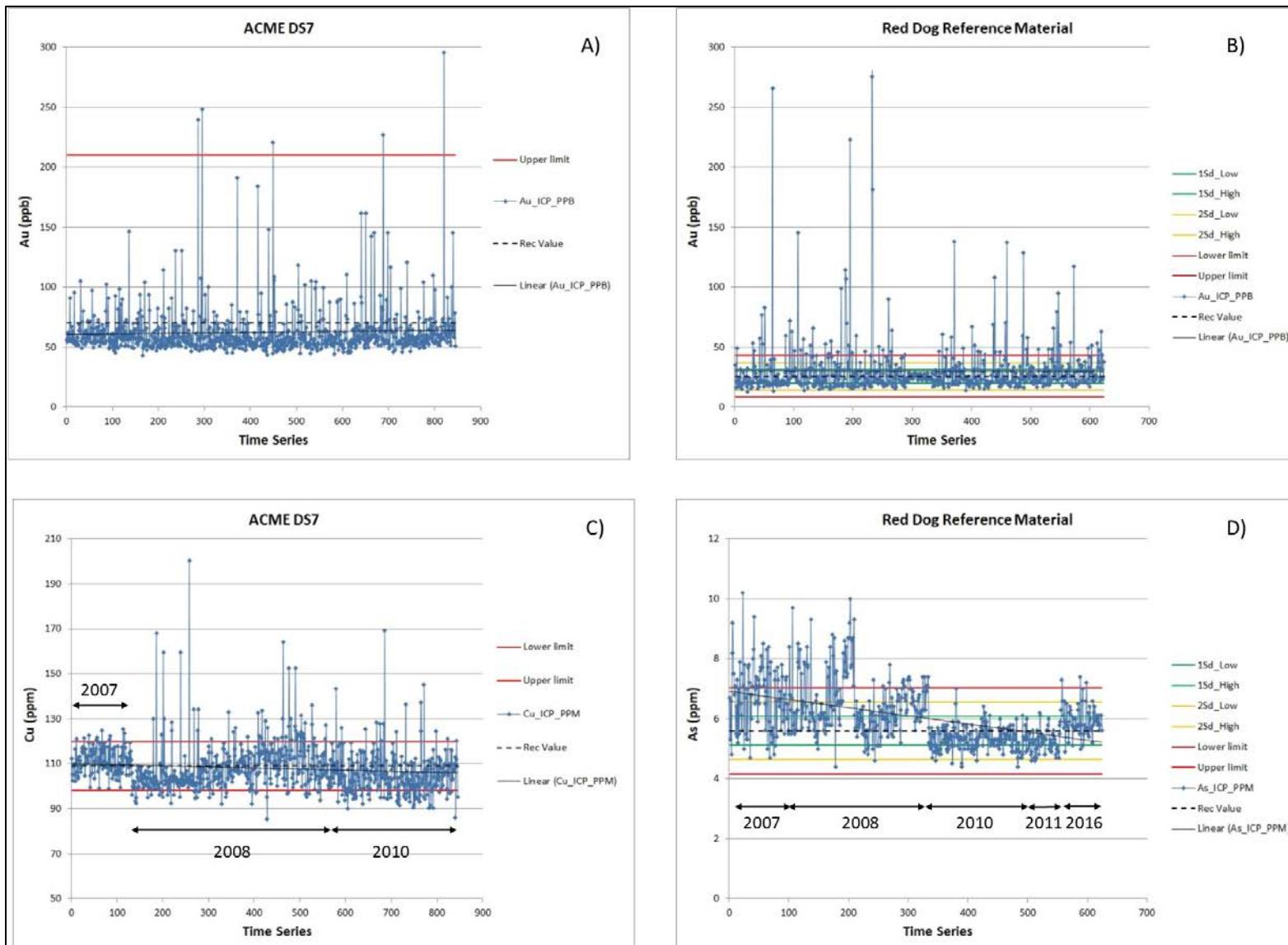


Figure 3: Examples of Shewhart charts for selected standard reference materials and elements

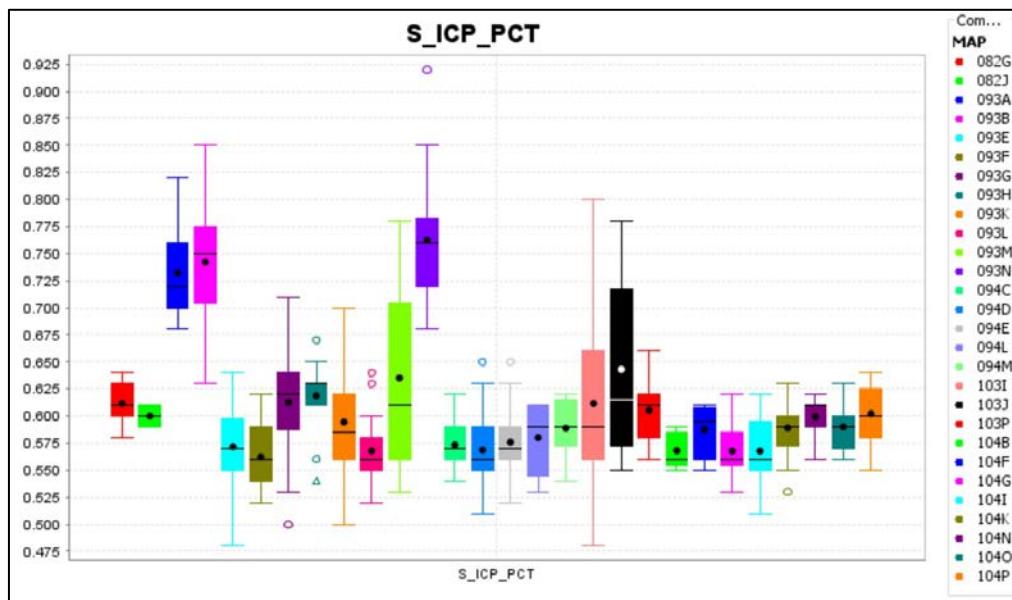


Figure 4: Box plot summary of S in RGS standard Red Dog plotted by map sheet

Note: Data from map sheets 093A, 093B and 093N are clearly elevated compared to data from the other sheets.

2.4 Location Validation

Another important data-quality consideration is the reliability of original sample location information. The generation of catchment basins for individual samples uses modern 1:20 000 terrain resource information management (TRIM) topographical and hydrological data (Cui *et al.*, 2009). Samples must therefore be accurately located within the modern topographical framework to ensure the correct catchment basin is allocated to each sample. In most cases, sample locations were manually marked onto 1:10,000 to 1:25,000 scale topographic map sheets and then the UTM grid reference locations measured off the map sheets in either NAD27 or NAD83 datums. In some cases, locations were recorded on sketch maps. In addition to the uncertainties associated with identifying sample locations prior to the availability of global positioning system (GPS) receivers, the locations of streams have sometimes either physically shifted or vary by comparison to more precise topographical data. The result is that sample locations often do not plot on the correct drainage, and validation or correction of the locations is required (Cui, 2010). This process can be time consuming as not all location errors can be rectified using automated procedures within a Geographic Information System (GIS).

An example of a typical location validation issue is presented in Figure 5. In all three instances shown on this figure, sample points (shown by grey dots) plot up to 1 km from the locations plotted by hand using red circles shown on the georeferenced original sample location map. Samples 1391 and 1390 are good examples of a common location error that would be partially fixed by using a “snap to” tool in a GIS, although the location for sample 1390 could potentially snap to the wrong stream given its distance from the correct tributary. Sample 1388 is an example of a considerable plotting error that required a significant move to its correct sample point location based on the original map. Validated, or corrected, sample locations (shown by pink dots) are the preferred sample locations based on the original sample location map.

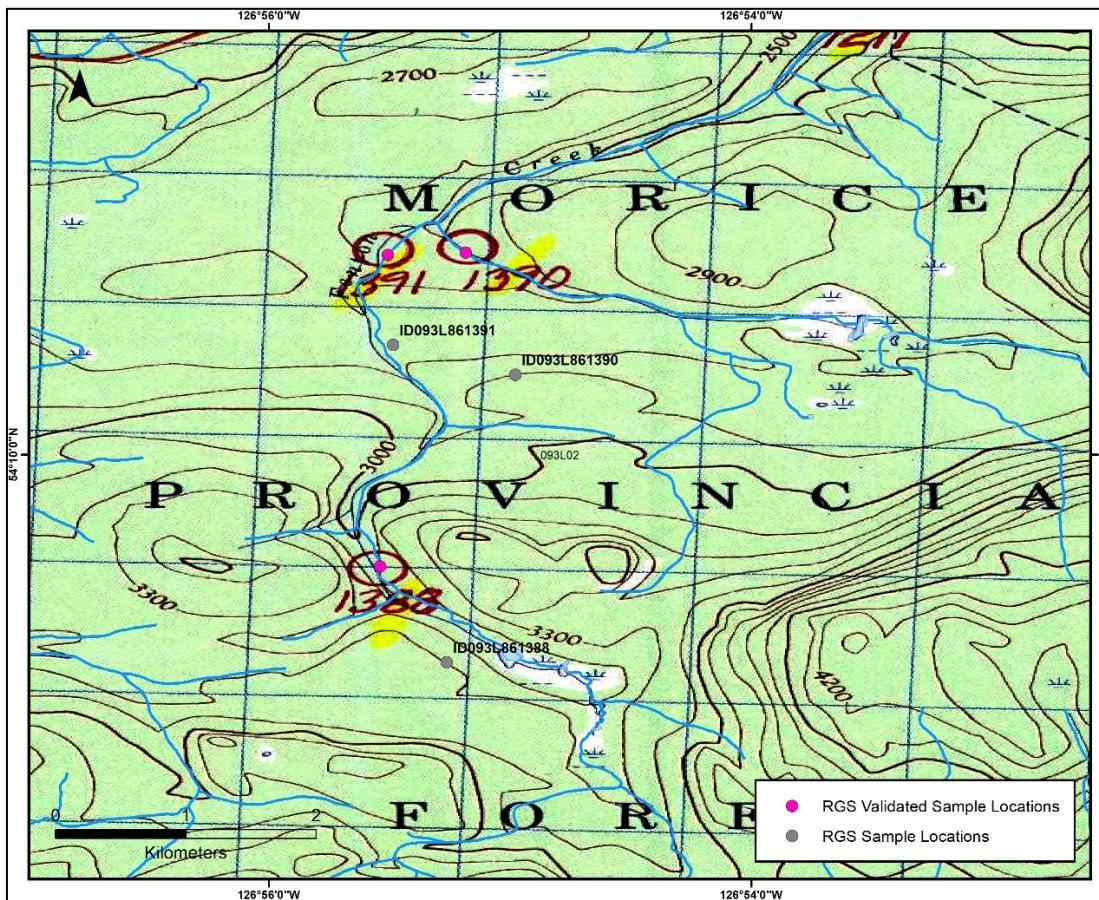


Figure 5: Examples of sample location issues

Note: Grey dots are the original sample locations recorded in the database. Pink dots are the corrected, or validated, sample locations based on the red circles on the original topographic map sheet used to record sample locations.

2.5 Derivation of Catchment Basins

Catchment basins for the validated sample locations were generated by the BCGS using the automated methodology described in Cui *et al.* (2009). This approach involved a three-stage computation:

1. Identify the root watershed for each stream sediment sample site.
2. All watersheds upstream of the root watershed are retrieved.
3. The upstream watershed boundaries are dissolved to yield a single catchment for each sample.

Note that the algorithm used extends the catchment basin downstream from the sample location to the next major stream node, although this typically represents a small area relative to the overall size of the catchment basins. Not all catchment basins in the digital files accompanying this report have associated geochemical data, either because they were not re-assayed or too many elements were missing in the re-assays.

An important distinction is that some catchments will include smaller catchment basins nested within them. It is important that individual catchment basins do not terminate at the next upstream sample, as the geochemistry of a sample point is the product of all sediment derived within the catchment that drains through the sample location. The larger catchment size will therefore impact upon the amount of dilution affecting a mineralized geochemical signal at a given sample point.

Validated catchment basins for the study area are illustrated in Figure 6. This figure demonstrates the variable density of sampling in some areas, with very tight coverage south of Ajax and areas with little to no coverage in areas of extensive alluvial deposition north of Snip and west of Kemess North. The large area to the east of Red Chris is provincial wilderness area (Spatsizi Plateau Wilderness Provincial Park and the Tatlatui Provincial Park) and so excluded from this study. Most of the map sheet in the south-east corner of the study area that includes the Blackwater Au deposit has also been excluded due to a limited element suite available for analysis.

Catchment basins were ultimately defined for 15,448 samples. They range in area from 0.1 km² to 390 km² and have a log-normal distribution. Two hundred catchments were not generated by the BCGS due to location or other issues related to the process of generating watersheds from the TRIM data. These 200 catchments were manually generated using the hydrology module in MapInfo/Discover and merged with the originally generated product. Most of these sample points are within the Bowser Basin.

Given the wide range in basin size, a direct correction using a basin area as proposed by Hawkes (1976) was considered too significant and so the square root of the catchment area was used for the correction. The dilution-corrected values are residuals calculated from linear regression of WSM scores multiplied by the square root of the catchment area plotted against catchment area.

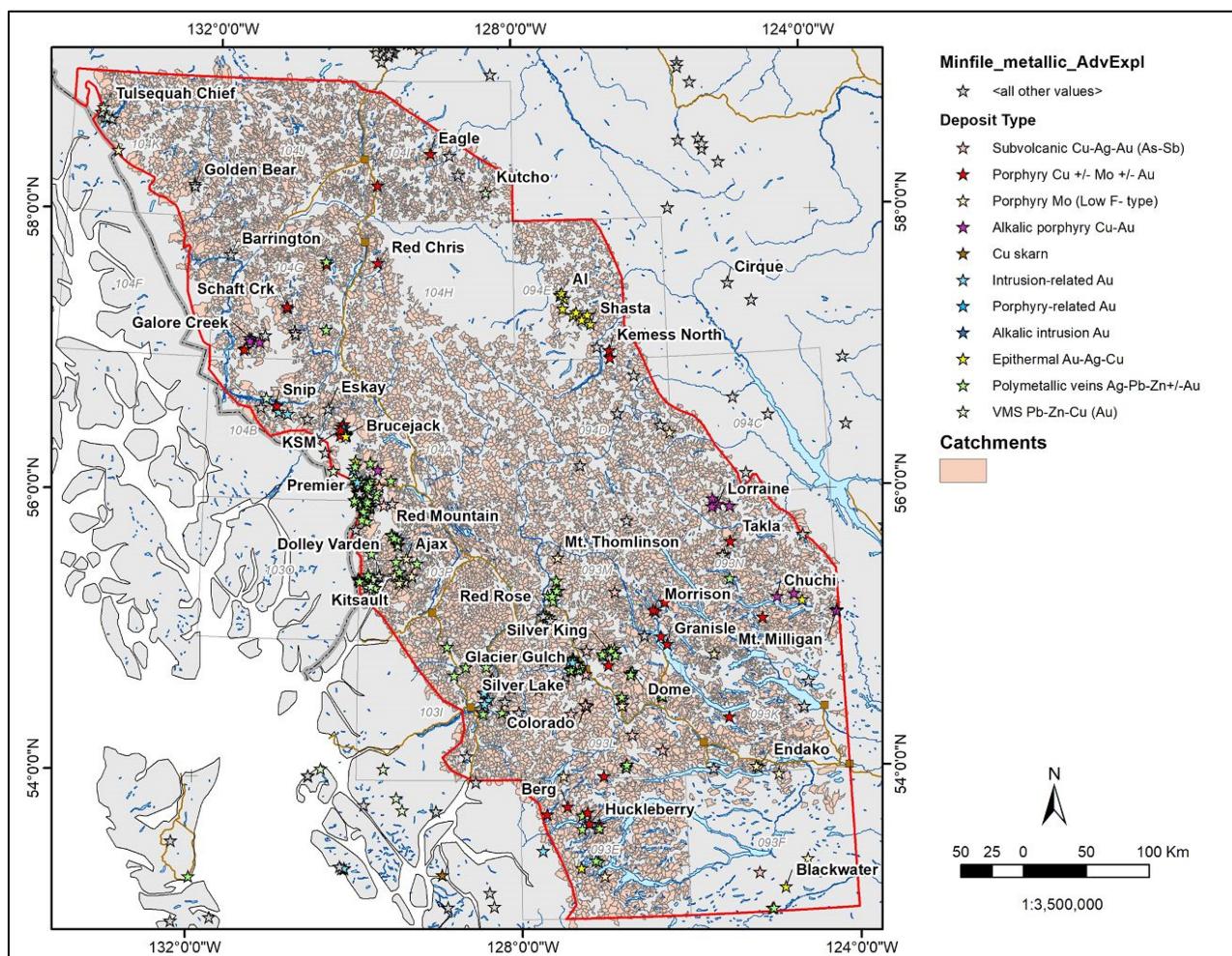


Figure 6: Catchment basins from northwest BC

2.6 Attribution of Lithology

Catchment lithology, rather than geological formations, are the most informative measure of control on stream sediment geochemistry. After catchments were produced for each stream sediment sample, the BCGS bedrock geology and Quaternary geology were combined to reflect the geology of each catchment. This merger was completed in ArcGIS using the spatial analyst and statistics tools. The bedrock geology was erased in areas of overlapping Quaternary geology as the Quaternary “cover” would be the dominant source of geochemical input into a catchment where it occurs. The Quaternary layer was merged into the bedrock layer to produce a complete layer file once the overlapping bedrock layer had been erased. The geological units were simplified to reflect basic lithology and all Quaternary units were lumped in a “Quaternary” field. A tabulate intersection tool was then applied to the catchments to produce the proportions of each lithology for each catchment basin using the unique sample ID of each sample point. The dominant lithology was taken to be that lithological unit with the highest proportion of area in the catchment (Table 2).

Table 2: Lithological units used in this study with number and proportion of catchment basins in which they occur as the dominant lithology

Rock class/lithology	Map code	Number	Proportion
Metamorphic			
Argillite	m_argillite	0	0.0%
Carbonate	m_carbonate	0	0.0%
Clastic	m_clastic	0	0.0%
Felsic	m_felsic	0	0.0%
Mafic	m_mafic	143	0.9%
Ultramafic	m_umafic	31	0.2%
Undivided	m_undivided	79	0.5%
Plutonic			
Felsic	p_felsic	1,816	11.6%
Intermediate	p_interm	793	5.1%
Mafic	p_mafic	235	1.5%
Ultramafic	p_umafic	1	0.0%
Undivided	p_undivided	382	2.4%
Plutonic/Volcanic			
Mafic	pv_mafic	125	0.8%
Sedimentary			
Argillite	s_argillite	1,040	6.6%
Carbonate	s_carbonate	144	0.9%
Chert	s_chert	149	1.0%
Clastic	s_clastic	2,123	13.6%
Undivided	s_undivided	2,705	17.3%
Volcanic			
Volcaniclastic	v_clastic	311	2.0%
Felsic	v_felsic	97	0.6%
Felsic-intermediate	v_felsic-interm	7	0.0%
Intermediate	v_interm	1,191	7.6%
Intermediate-felsic	v_interm_felsic	59	0.4%
Mafic	v_mafic	1,114	7.1%
Mafic-felsic	v_mafic_felsic	38	0.2%
Mafic-intermediate	v_mafic_interm	44	0.3%
Undivided	v_undivided	842	5.4%
Quaternary	q_Quaternary	2,178	13.9%
TOTAL		15,647	100%

This same process was also used to identify the dominant geological terrane in each catchment area by utilizing the publicly available BCGS terrane layer file and completing a tabulated intersection query.

The most common dominant rock class in the catchment basins is igneous, of which plutonic and volcanic rocks make up just over 40% of dominant lithologies in roughly equal proportions. The plutonic rocks are predominantly of felsic compositions whereas the volcanic rocks are predominantly intermediate to mafic in composition. Sedimentary rocks are the next most common dominant lithology, with clastic or undivided lithologies the most abundant. Quaternary deposits make up nearly 14% of the dominant catchment lithologies and these would consist predominantly of till or alluvium. Metamorphic rocks form the least common rock class in the catchment basins.

2.7 Mineral Deposits

Mineral deposits in the study area were taken from the most recent British Columbia Ministry of Energy, Mines & Petroleum Resources MINFILE database (<http://www.empr.gov.bc.ca/Mining/Geoscience/MINFILE/Pages/default.aspx>).

Mineral occurrences are subdivided into Past Producers, Developed Prospects, Prospects, Showings and Anomalies. The MINFILE database contains a great many mineral deposit types that are non-metallic and not strictly relevant to testing geochemical models in the current study. The simplified listing of mineral deposit types used in this investigation are listed in Table 3.

Table 3: Simplified mineral deposit classification

File name	Full name	Number
NoDep	No deposit known	15,069
Porph_	Porphyry CuMoAu	51
Porp_1	Porphyry Alkalic	26
Porp_2	Porphyry Moly	36
Int_Re	Intrusion Related Au	9
SubVol	Subvolcanic	33
Epithe	Epithermal	44
PolyMe	Polymetallic Veins	177
VMS	VMS	16
Skarn_	Skarn - Base Metals	5
UM_CrN	Mafic CrNiCuPGE	10

2.8 Weighted Sum Models

WSMs have been calculated using the methodology of Garrett and Grunsky (2001). The elements selected for inclusion and their weightings (importance rankings) are based on our experience as to what commodity and pathfinder elements are most relevant for the different mineral deposit. Allowance must also be made for de-coupling of some elements from the primary element association as a function of secondary dispersion during weathering and transport. These models are constructed on a trial and error basis using a qualitative comparison to the distribution of known mineral deposits to optimize the elements and their weightings for each mineral deposit class. Negative importance rankings can be used to remove effects associated with processes not related to mineralization, such as metal scavenging and lithology. WSMs are expert-driven models that are subjective and dependent upon the experience of the person constructing them.

2.9 Principal Component Analysis

Principal component analysis (PCA) was undertaken independently on two datasets. The first dataset consisted of the raw data for which values below the lower limit of detection had been imputed using the nearest neighbour approach, and for which some S data had been adjusted for systematic variations discussed in Section 2.3. These data underwent a CLR transformation to correct for the effects of geochemical closure prior to PCA and were used for regression analysis of individual elements as well as for advanced analytical methods described in Sections 2.10 and 2.11.

The second dataset consisted of residuals for all elements following multiple regression analysis against the proportions of different lithologies attributed to each catchment basin from the available BCGS data, as discussed in Section 2.6. Not all elements will show a strong lithological control, but all underwent regression analysis to ensure uniform treatment of the data. These residuals underwent PCA prior to inclusion in the alternative analytical methods described in Section 2.11.

2.10 Independent Component Analysis

Independent component analysis (ICA) is a computational method used to separate signals within a multivariate dataset. It assumes that each source is independent and has a non-Gaussian distribution (Hyvarinen and Oja, 2000). A CLR transform was applied to the data followed by an ICA (R package fastICA). These components may show patterns that are associated with specific lithologies and/or districts containing mineral deposits (Table 3).

2.11 Advanced Data Analysis Methods

A description of random forests and allocation/typicality taken from Harris and Grunsky (2015) and Grunsky (1991), respectively, is provided in [Appendix 1](#) at the end of this report.

3 Results

3.1 Raw Element Distributions

A number of commodity and pathfinder elements are conspicuously elevated in the vicinity of the Golden Triangle area, including Au, Ag, As, Bi, Cd, Cu, Hg, Mo, Pb, S, Sb, Se, Te, as well as locally in the area of known metallic mineral deposits. In particular, Cu and Ag are elevated in areas of the Quesnel terrane containing porphyry Cu deposits. However, some elements (i.e. As, Sb, Te, Hg and Zn) are also elevated slightly in samples taken from catchment basins within the Bowser Basin, which post-dates most mineralization, indicating that their use as pathfinder elements for metallic mineral deposits must be used with caution. Examples of gridded percentile images for Cu and Sb are illustrated in Figure 7 and Figure 8, respectively.

While the raw element gridded images are informative, they reflect nothing that would not have already been obvious, at least for some elements such as As and Cu, in the original stream sediment data. Further processing of the raw data is required to reveal subtle geochemical signatures associated with mineral deposits that may not have been obvious in the original geochemical data, or in the re-analysed data for the same samples.

Combining the raw elements Ag, As, Sb and Cu to produce a WSM for mineralization in the Golden Triangle area does produce a good fit to known mineral deposits, but also produces elevated weighted sums scores within the Bowser Basin (Figure 9). Other elements that are also elevated within the Bowser Basin catchments include Hg, Co, Ni, Mg, Zn, Te and, locally, Pb. The reason for this geochemical response in the Bowser Basin is not immediately clear but is unlikely to reflect erosion of mineralized material from the Stikine terrane into the Bowser Basin during its deposition. The main provenance of Bowser Basin sediments is believed to have been from oceanic crust in the Cache Creek terrane located to the northeast of the Bowser Basin (Cookenboo, 1993). The spatial distribution of raw data for commodity and important pathfinder elements indicates that there is a lithological control on the geochemical data that may be related to specific terranes, particularly in the Bowser Basin.

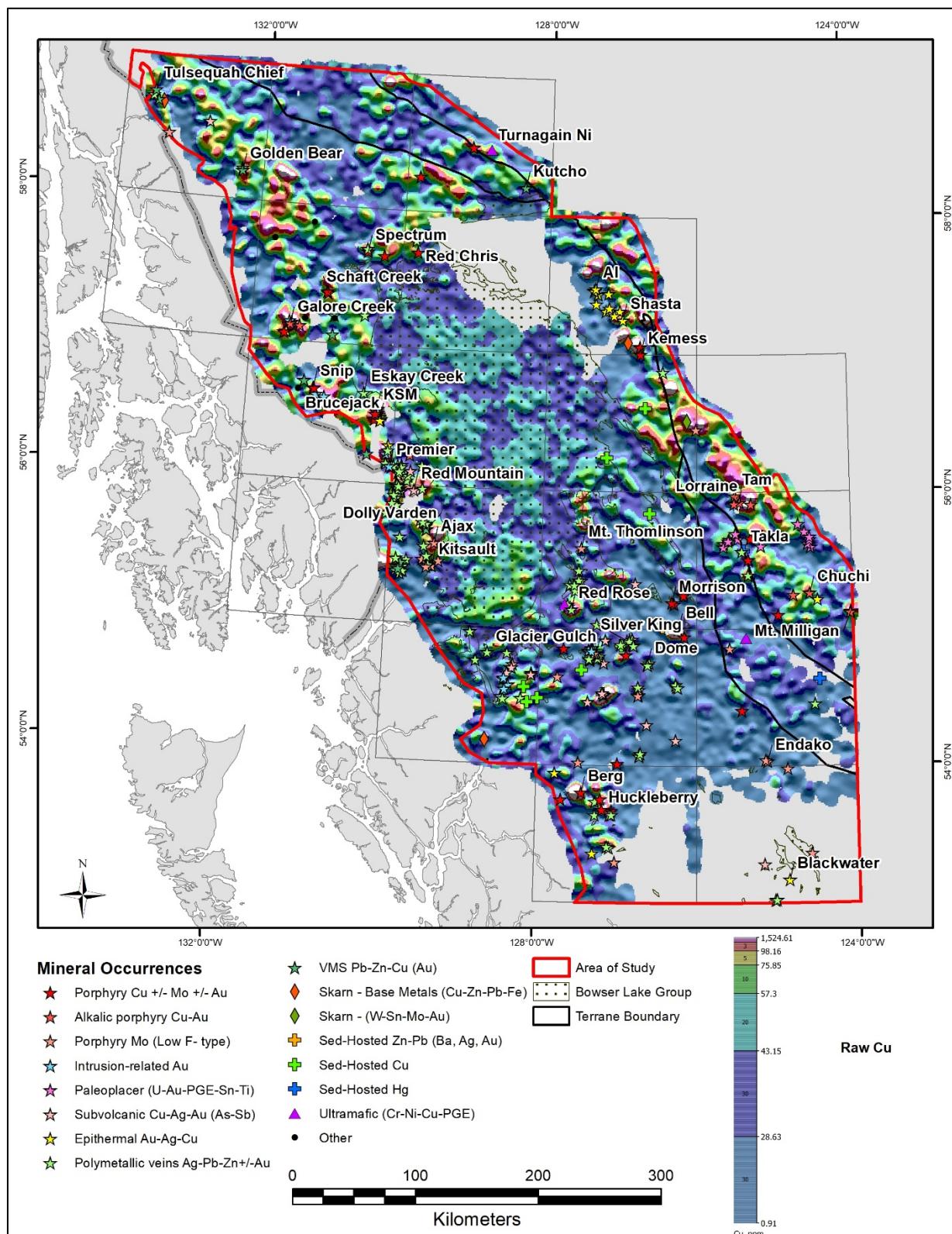


Figure 7: Distribution of raw Cu in stream sediments from northwest BC

Note: The hatched area in the figure is the area underlain by Bowser Basin sediments.

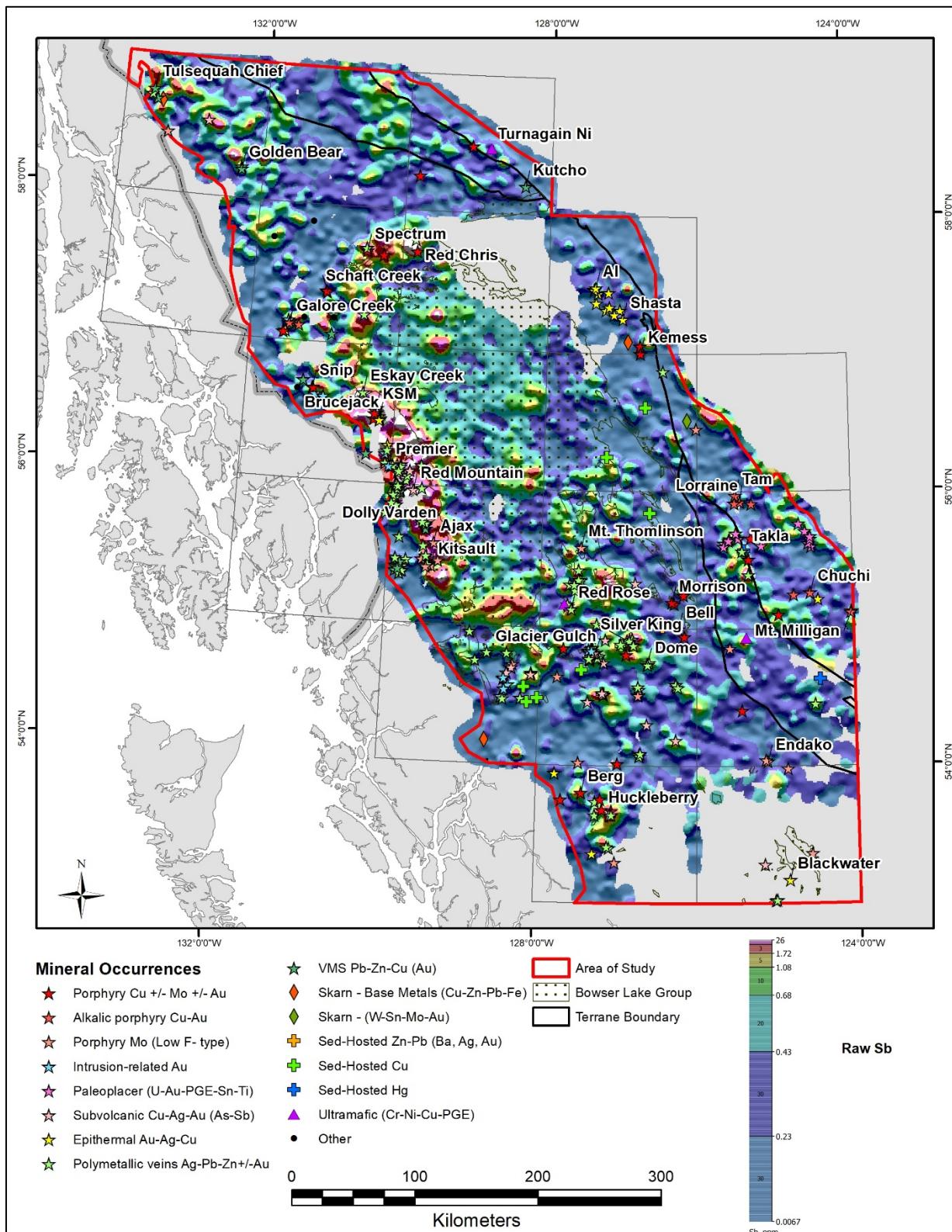


Figure 8: Distribution of raw Sb in stream sediments from northwest BC

Note: The hatched area in the figure is the area underlain by Bowser Basin sediments,

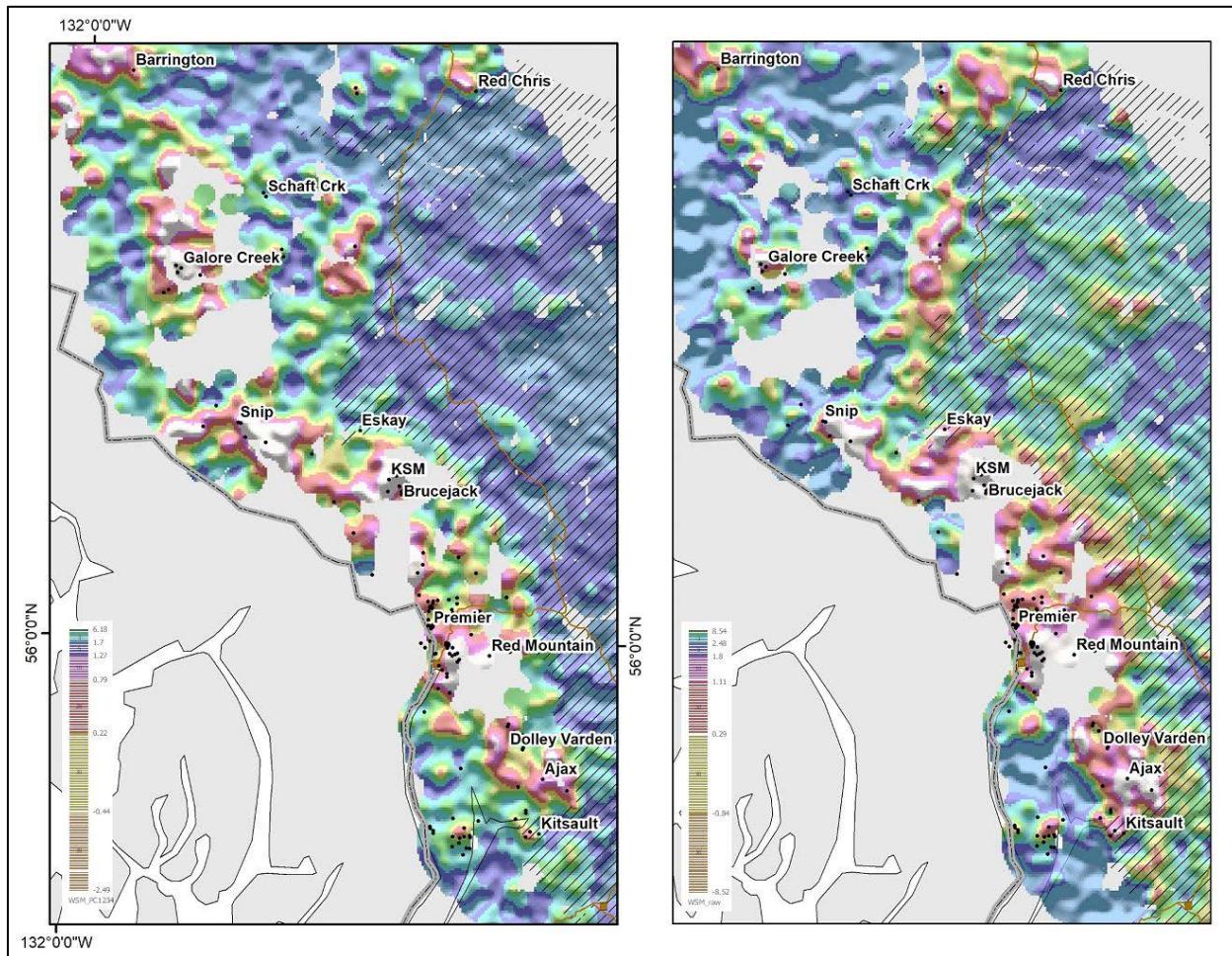


Figure 9: WSMs for the Golden Triangle region

Note: The gridded percentile image to the right is based on raw Ag, As, Sb and Cu data. The gridded percentile image to the left was constructed using principal components, as described in Section 3.5.

3.2 Conventional Levelling for the Effects of Lithology

Analysis of Tukey plots showing the distribution of major and trace elements known to be strongly controlled by lithology and attributed based on the dominant lithology in each catchment basin supports the assertion based on the spatial distribution of some elements that there are strong lithological controls on the geochemical data (Figure 10 and Figure 11). This is in keeping with previous investigations into stream sediment geochemical data (e.g. Bonham-Carter and Goodfellow, 1986; Bonham-Carter *et al.*, 1987; Carranza and Hale, 1997; Arne and Brown, 2015). This lithological control extends to many commodity and pathfinder elements, as indicated in Section 3.1, and is also evident in data attributed based on terrane (Figure 12).

Z-score levelling following a Log_{10} transformation of the trace element data serves to reduce the terrane effects on key commodity and pathfinder elements, irrespective of whether levelling is based on dominant lithology or terrane (Figure 13 and Figure 14). Positive residuals for individual elements following multiple regression analysis against the varying proportions of lithologies in the catchments also results in a similar distribution of key commodity and pathfinder elements (Figure 13 and Figure 14). However, the effectiveness of these approaches depends upon the reliability with which the catchment geology is understood.

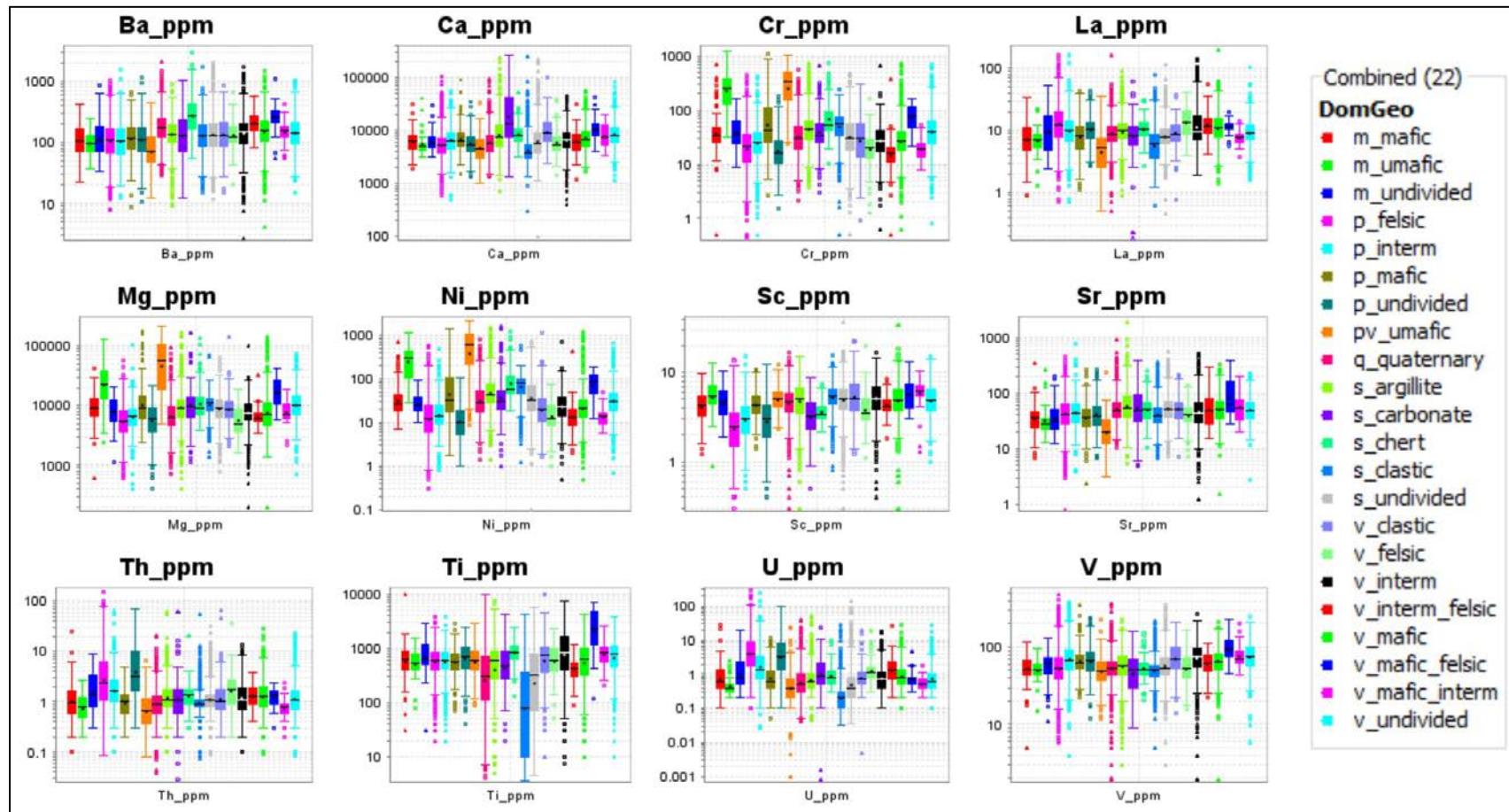


Figure 10: Tukey plots showing the distribution of lithologically-controlled major and trace elements attributed by dominant lithology in each catchment

Note, for example, that Ni, Cr and Mg are elevated in catchment basins in which the dominant lithology is ultramafic rock, and that U and Th are elevated in felsic plutonic rocks and undefined plutonic rocks. See Table 2 for a full description of lithologies.

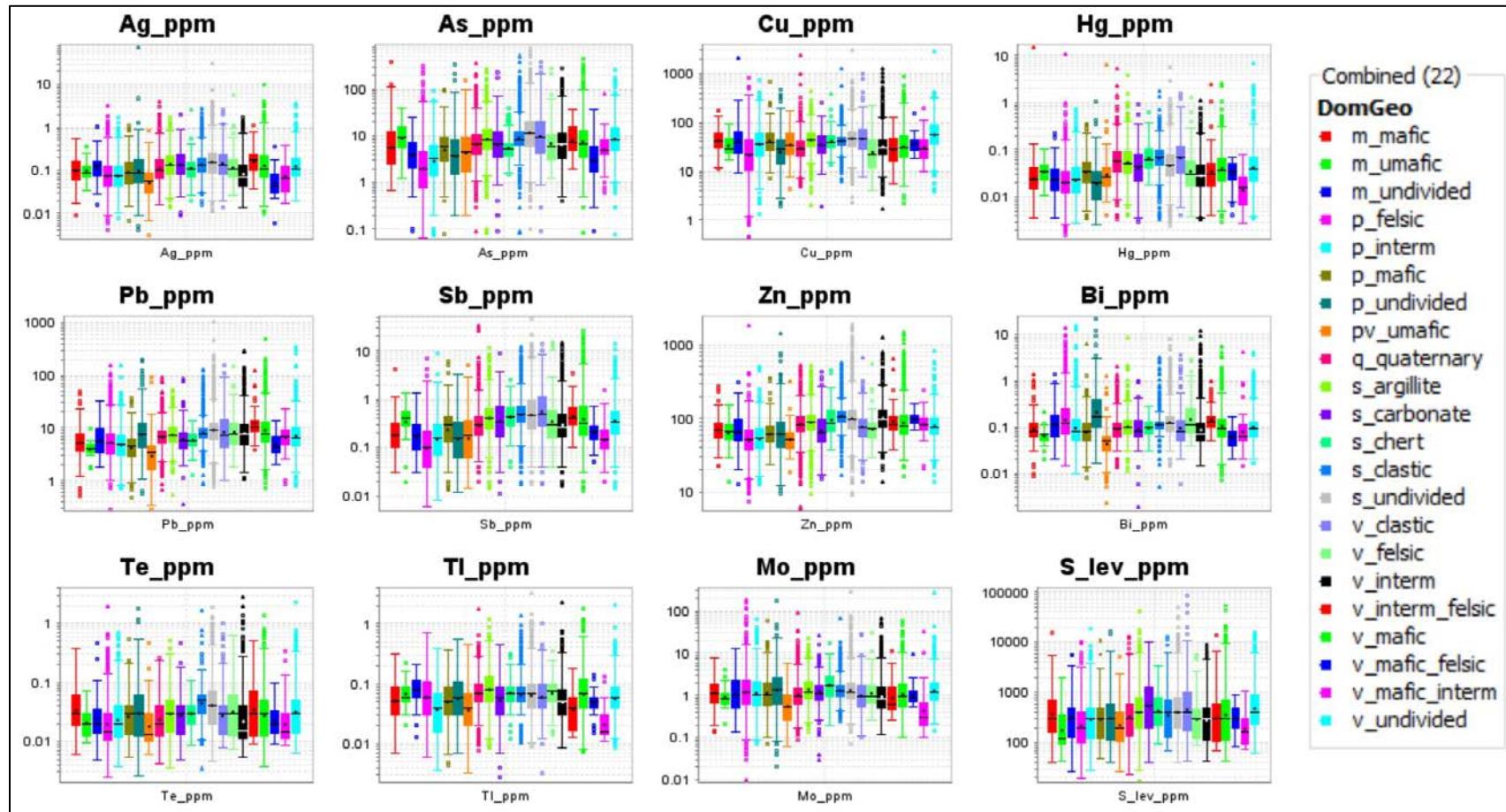


Figure 11: Tukey plots showing the distribution of commodity and pathfinder trace elements attributed by dominant lithology in each catchment

Note that Ag, Hg, Zn and Sb are elevated in sedimentary rocks relative to plutonic and volcanic rocks. See Table 2 for a full description of lithologies.

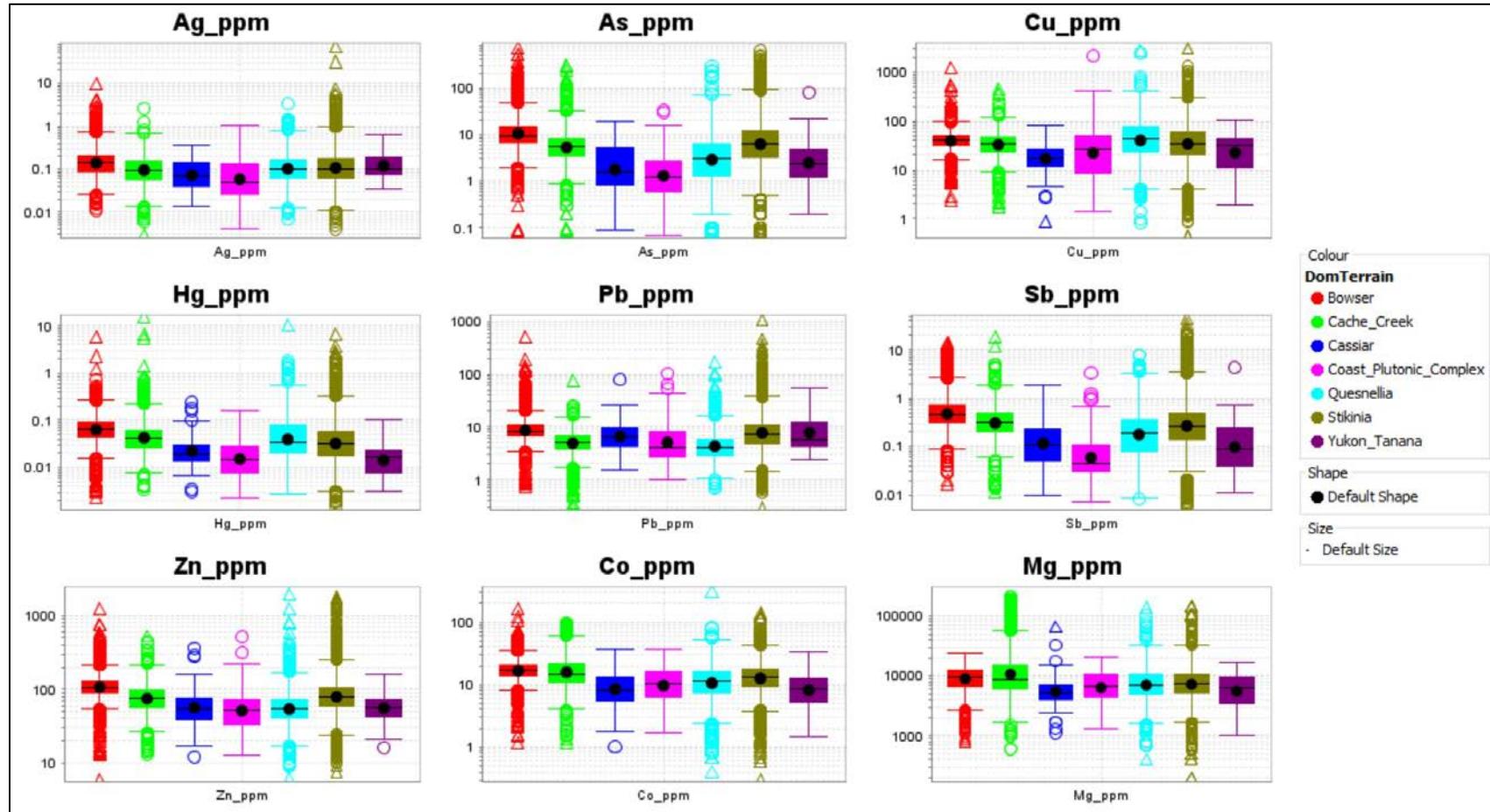


Figure 12: Tukey plots showing the distribution of commodity and pathfinder trace elements attributed by terrane

Note that As, Hg, Zn and Sb are, on average, elevated in both the Bowser Basin and samples from the Quesnel terrane

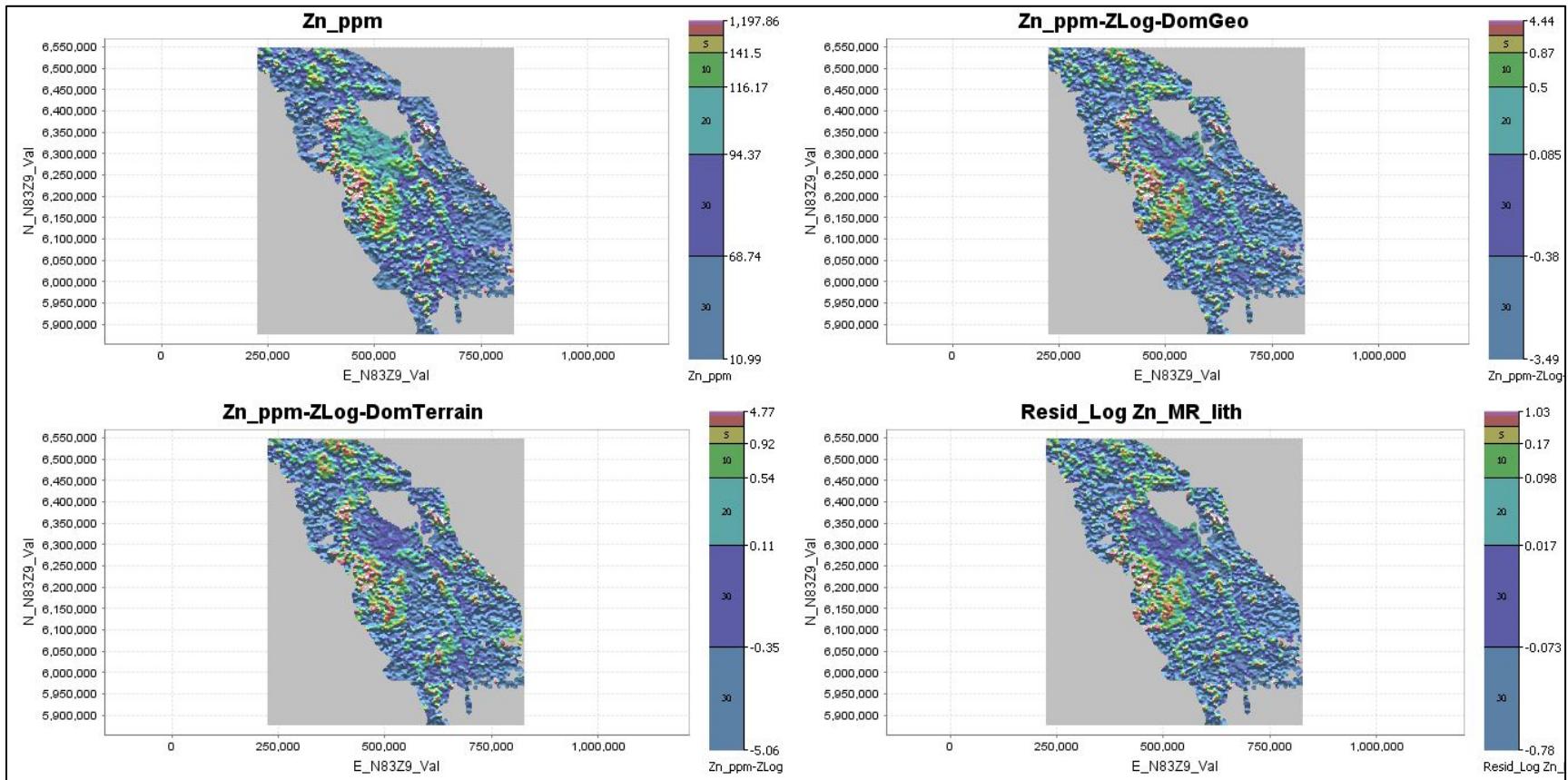


Figure 13: Gridded percentile images comparing raw Zn data and Zn data leveled by dominant lithology (Zn_ppm-ZLog-DomGeo), geological terrane (Zn_ppm-ZLog-DomTerrain), and all catchment lithologies (Resid_Log_Zn_MR_lith)

Note the overall decrease in elevated Zn from samples in the Bowser Basin

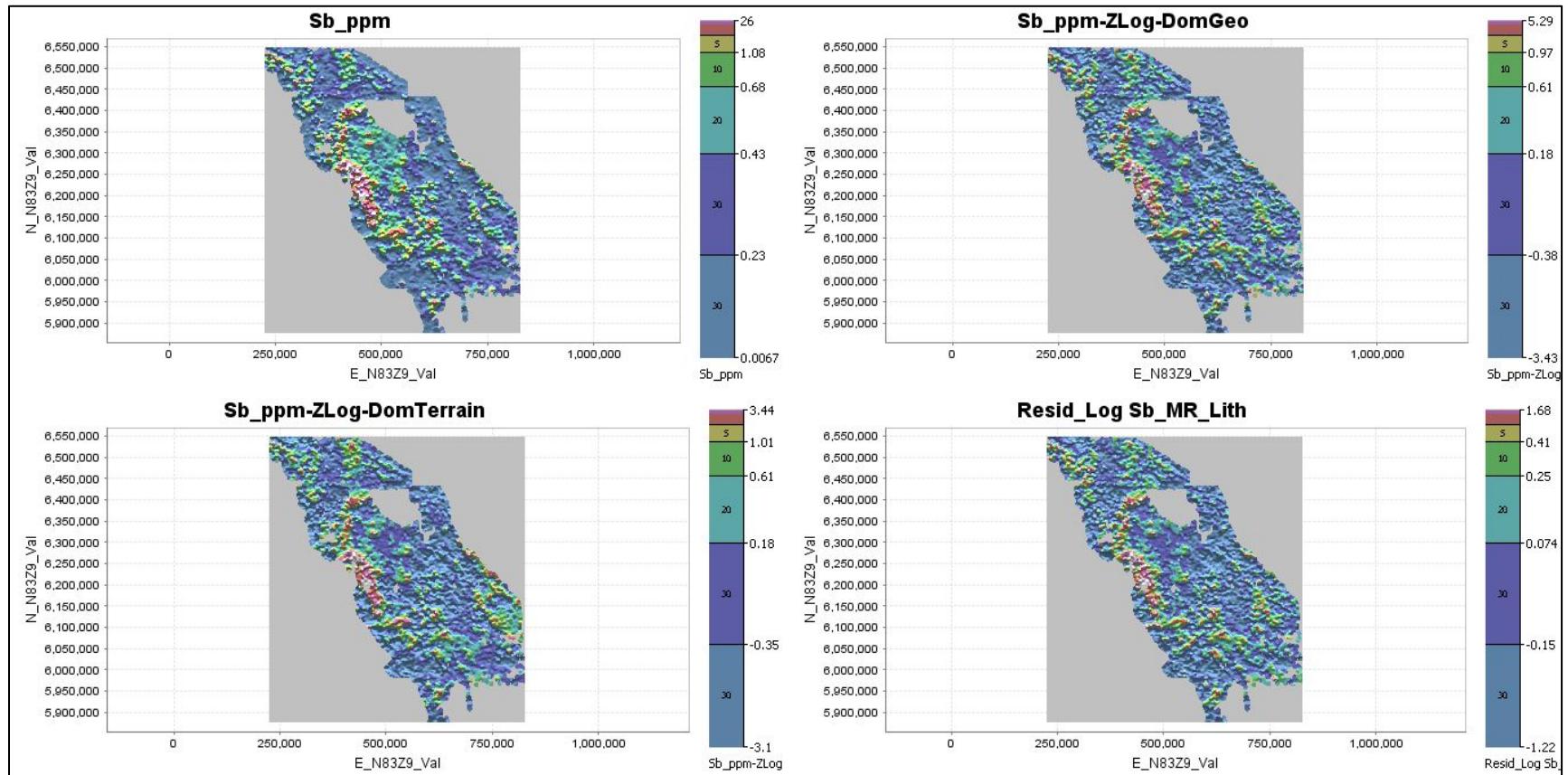


Figure 14: Gridded percentile images comparing raw Sb data and Sb data leveled by dominant lithology (*Sb_ppm-ZLog-DomGeo*), geological terrane (*Sb_ppm-ZLog-DomTerrain*), and all catchment lithologies (*Resid_Log_Sb_MR_lith*)

Note the overall decrease in elevated Sb from samples in the Bowser Basin.

3.3 Principal Component Analysis

The compiled data underwent PCA following a CLR transformation of the data. A SCREE plot of eigenvalues for each principal component (Figure 15; shown with Au data included) indicates that the first two principal components account for 37% of the variability within the data. The amount of variability explained by the data drop significantly for the next three principal components before tapering off to increasingly low eigenvalues with subsequent principal components. It is estimated that the bulk of the variability in the data (78%) is accounted for by the first 11 principal components. The dominant components (PC1 to PC11) in Figure 15 can be interpreted as representing potentially significant processes and the lesser components (PC12 to PC34) can be interpreted as “under-sampled” processes or noise (Grunsky *et al.*, 2014).

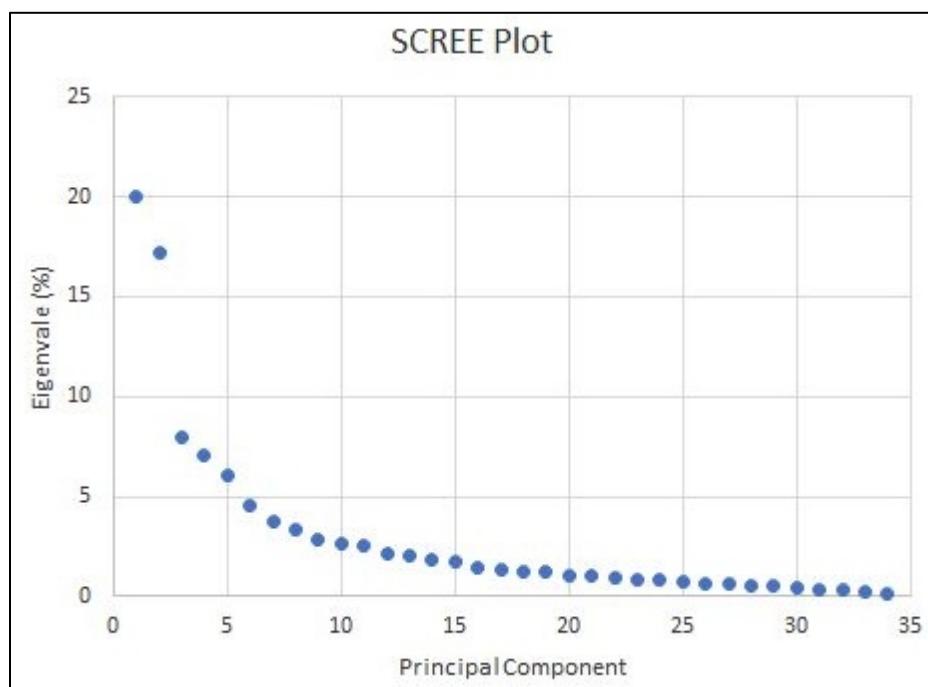


Figure 15: SCREE plot of eigenvalues following PCA

Note: A significant component of variability in the data is encapsulated in the first two principal components

Unusually for the first principal component in a regional dataset, this component is positively weighted by commodity and pathfinder elements (Figure 16), although some of these are clearly elevated in the Bowser Basin, as discussed previously in Section 3.1 (Figure 17). PC2, which is orientated orthogonally to PC1 in multi-dimensional space, shows the variability in composition between felsic and mafic rocks. Inverse PC2 clearly defines the Bowser Basin (Figure 18). PC3 has positive base metal loadings that are spatially associated with polymetallic base metal veins as well as young basaltic rocks. The negative loadings on this principal component show a positive correlation with LOI data, suggesting metal scavenging involving secondary Fe and Mn hydroxides, despite loadings by some commodity elements and S (Figure 19). Negative loadings on PC4 show the strongest spatial association with known porphyry Cu-style mineralization, whereas the positive loadings also display a positive correlation with LOI suggestive of metal scavenging onto organics.



Figure 16: Line plots showing the elemental loadings on the first four principal components

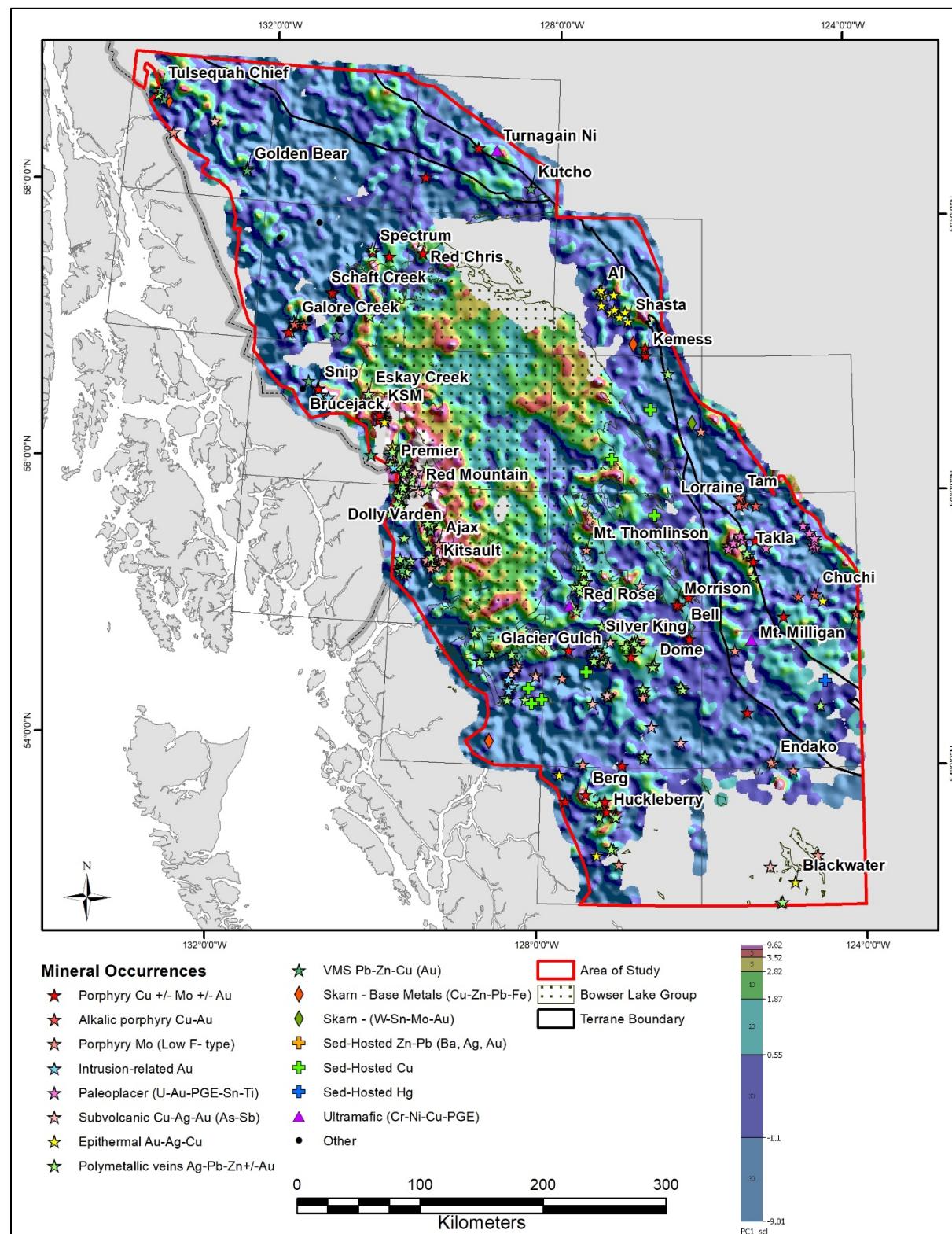


Figure 17: Percentile gridded image of principal component 1 (PC1)

Note: PC1 is loaded by commodity and trace elements elevated in the Bowser Basin.

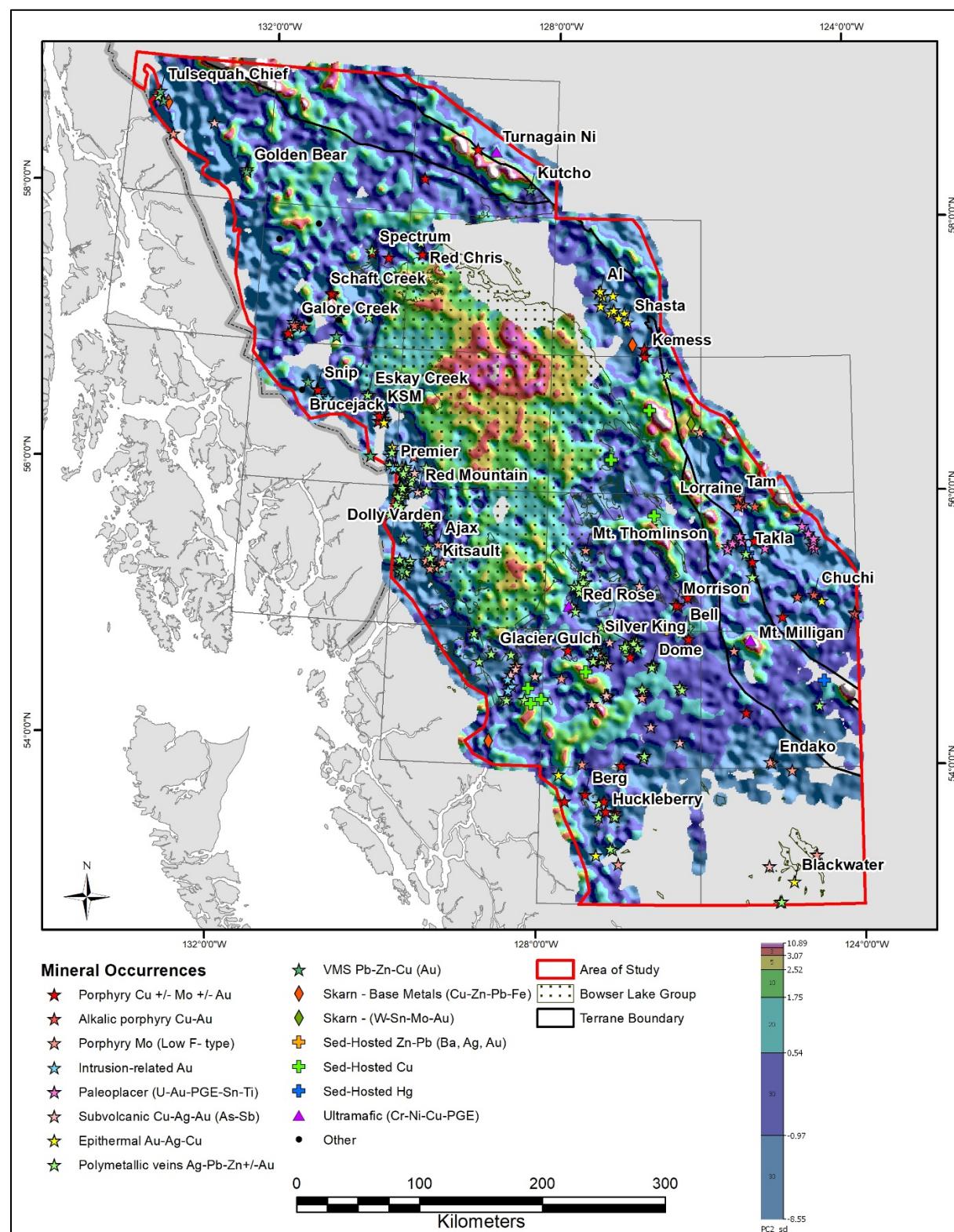


Figure 18: Percentile gridded image of principal component 2 (PC2)

Note: PC2 is loaded by mafic elements that are elevated in the Bowser Basin sediments.

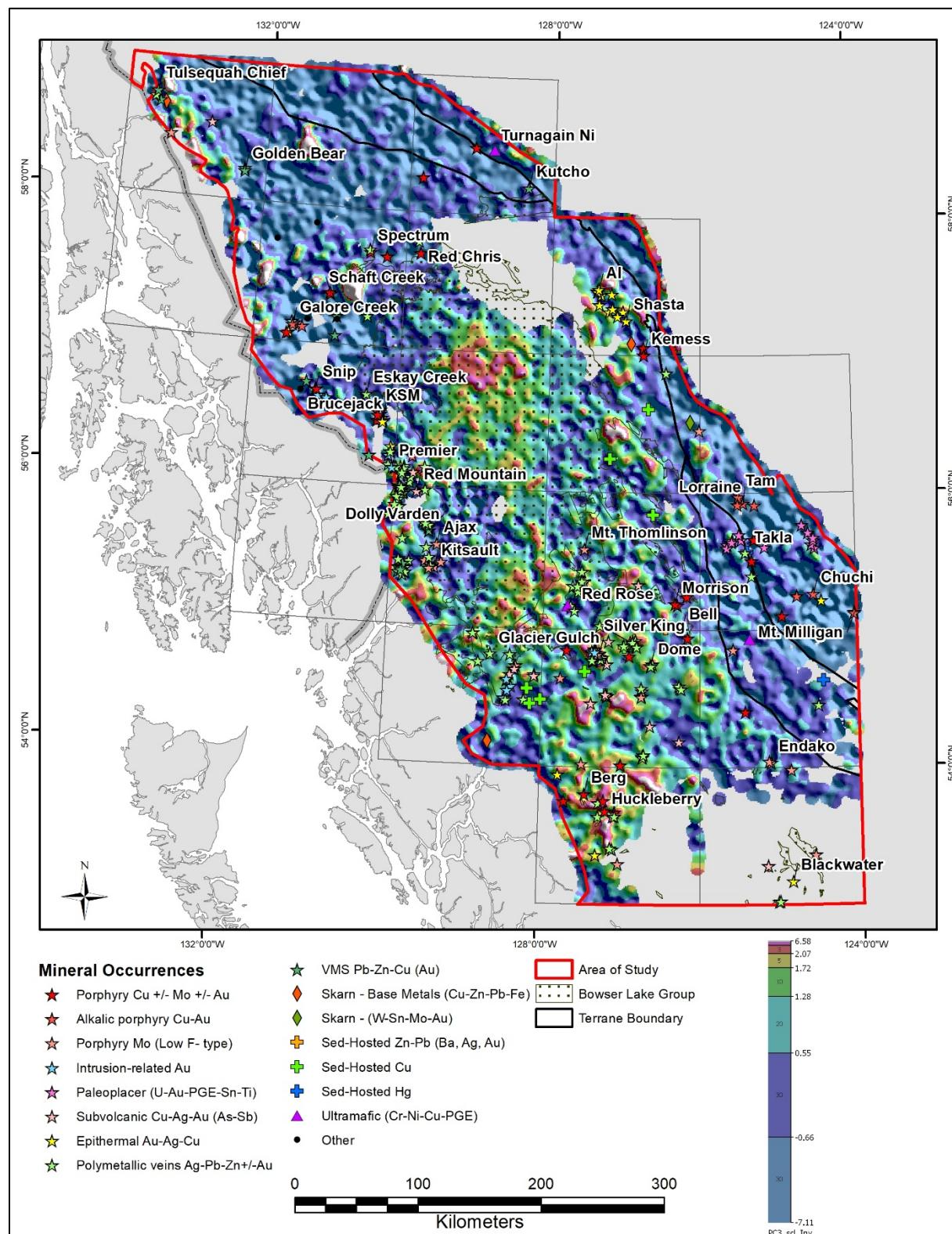


Figure 19: Percentile gridded image of inverse principal component 3 (-PC3)

Note: Inverse PC3 is loaded by elements that are highest in samples with elevated LOI values (i.e. are associated with possible scavenging of metals).

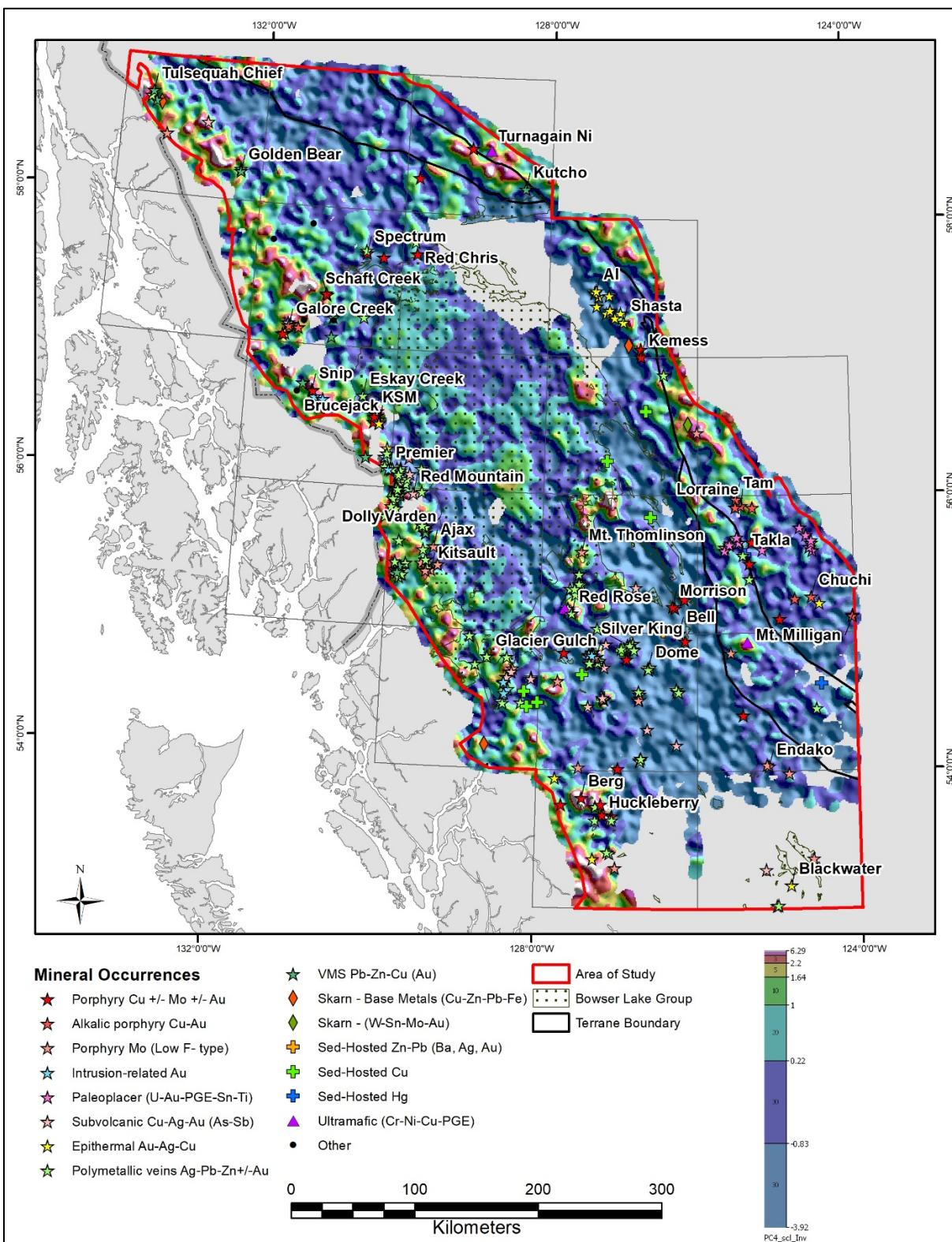


Figure 20: Percentile gridded image of inverse principal component 4 (-PC4)

Note: Inverse PC4 is loaded by elements such as Cu and Te that are highest in samples spatially associated with porphyry Cu deposits.
The mineral deposit legend is the same as that shown on Figure 8.

The highest-order principal components therefore reflect the lithological and terrane variation previously noted in the raw element data but have the advantage of grouping elements according to their correlations. This reduces the number of variables needed for assessment, an important consideration given the large multi-element datasets now routinely generated by laboratories. It also allows influences such as lithological variation and metal scavenging by secondary Fe and/or Mn hydroxides, clays or organics to be assessed simultaneously.

As illustrated in Figure 21, the first six principal components representing the bulk of variability in the data show differences that can in part be attributed to lithological variation. For example, PC1 is elevated in sedimentary rocks relative to other lithological units, whereas PC2 and PC4 are lowest in ultramafic rocks. Given that the geological terranes are defined, in part, by their lithological assemblages, it is therefore not surprising that the terranes also show an influence on the first six principal components (Figure 22). These observations raise the possibility that levelling of data for the effects of lithological variation or metal scavenging could be undertaken using principal components rather than mapped geology, as well as providing a direct indication of mineralization from multivariate signatures. Such an approach has the advantage of using the geochemical data to define mineralization, lithological variation or metal scavenging signatures within the dataset and so can be applied to areas where geological and/or topographical data are either not available or unreliable. In the next section, we compare WSMs generated using conventionally levelled geochemical data with those generated using principal components.

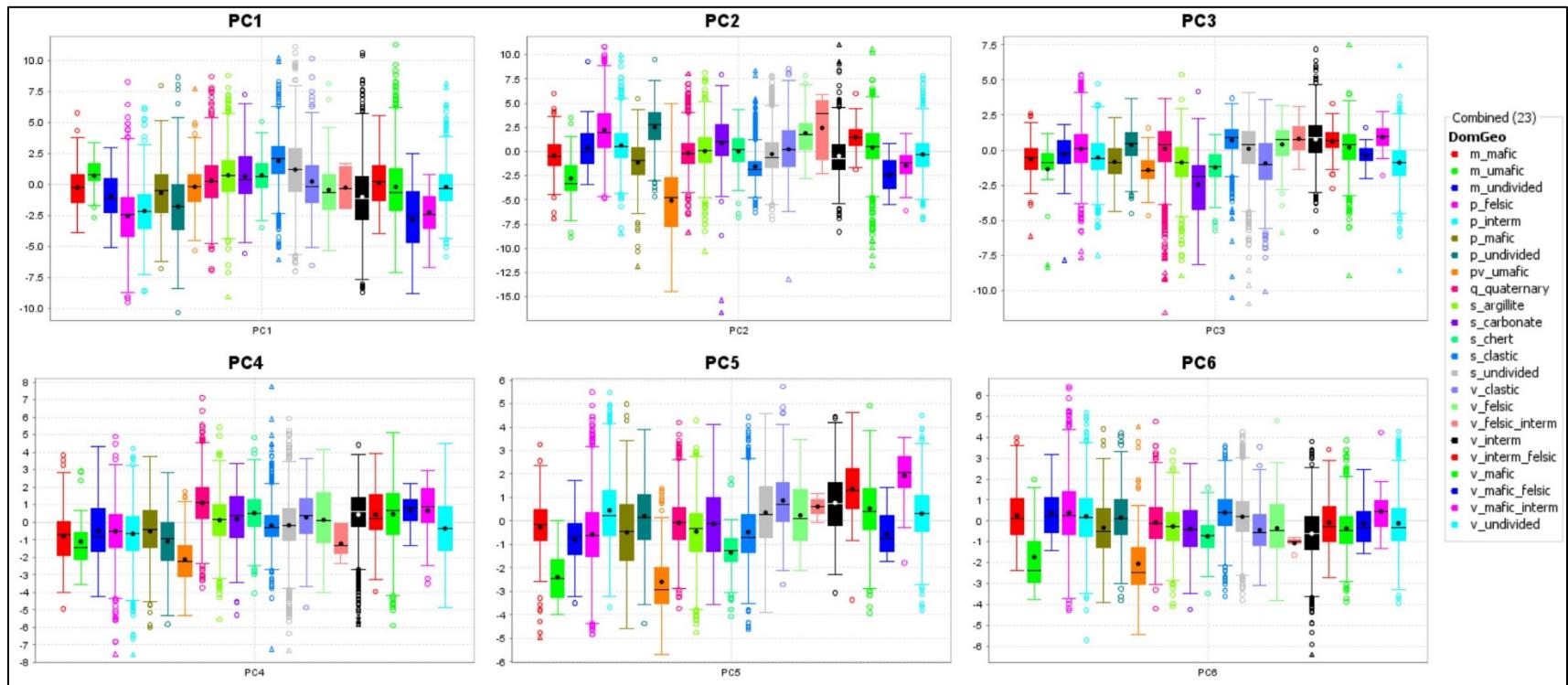


Figure 21: Tukey plots showing the influence of dominant catchment lithology on the first six principal components

Note that, on average, PC1 is elevated in sedimentary rocks, and PC2 and PC4 lowest in ultramafic rocks. See Table 2 for a full description of lithologies.

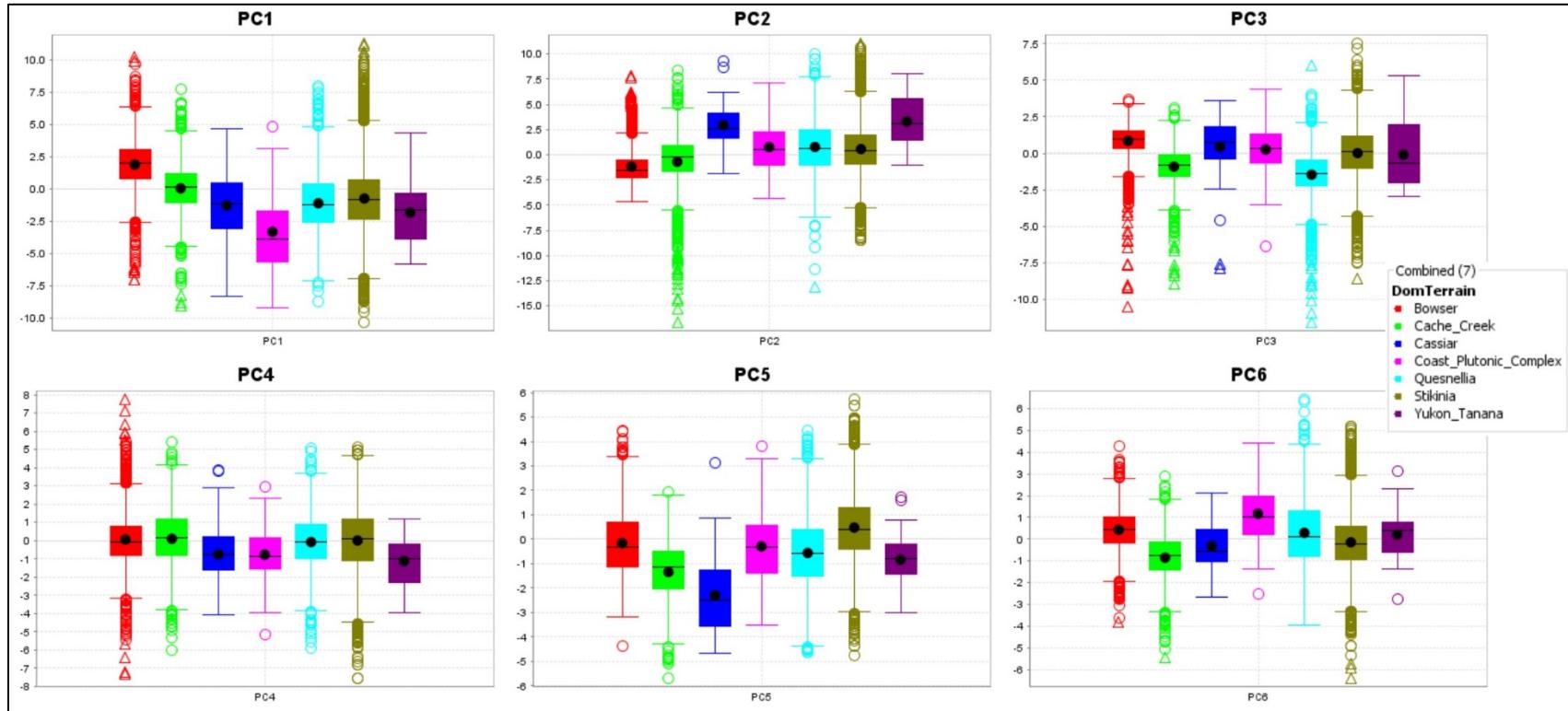


Figure 22: Tukey plots showing the influence of terrane on the first six principal components

Note that, on average, PC1 is elevated in the Bowser Basin, and PC2 has the most lower outliers in samples from Cache Creek terrane.

3.4 Independent Component Analysis

ICA was undertaken on the CLR elements to try and separate different lithological, metal scavenging and mineral deposit signatures. The ability to resolve different multivariate geochemical signals depends on each signal behaving independently and having a non-Gaussian distribution, neither of which criteria may have been honoured completely in this data set. For example, the spatial association between ultramafic rocks and magmatic Ni-Cu deposits means neither is independent. A similar argument can be made for the presence of both porphyry Cu-Au and epithermal Au-Ag mineralization associated with the same hydrothermal magmatic system.

The number of significant principal components used was 11 (Figure 15). Consequently, 11 independent components were calculated as an initial estimate of identifiable processes. Some of the independent components appear to resolve specific lithological or mineral deposit groups. The inverse (negative values) of IC2 appears to outline the extent of the Bowser Basin (Figure 23), whereas IC3 defines ultramafic rocks within the Cache Creek terrane (Figure 24). IC6 is broadly related spatially with alkalic porphyry Cu-Au systems and carbonate rocks of the Quesnel terrane (Figure 25), as well as with carbonate lithologies, and IC8 shows a good spatial association with epithermal Au-Ag mineralization (Figure 26). In fact, the upper 90th percentile of inverse IC8 captures a higher proportion of known epithermal Au-Ag deposits (54%) than any of the WSM discussed in Section 3.5.

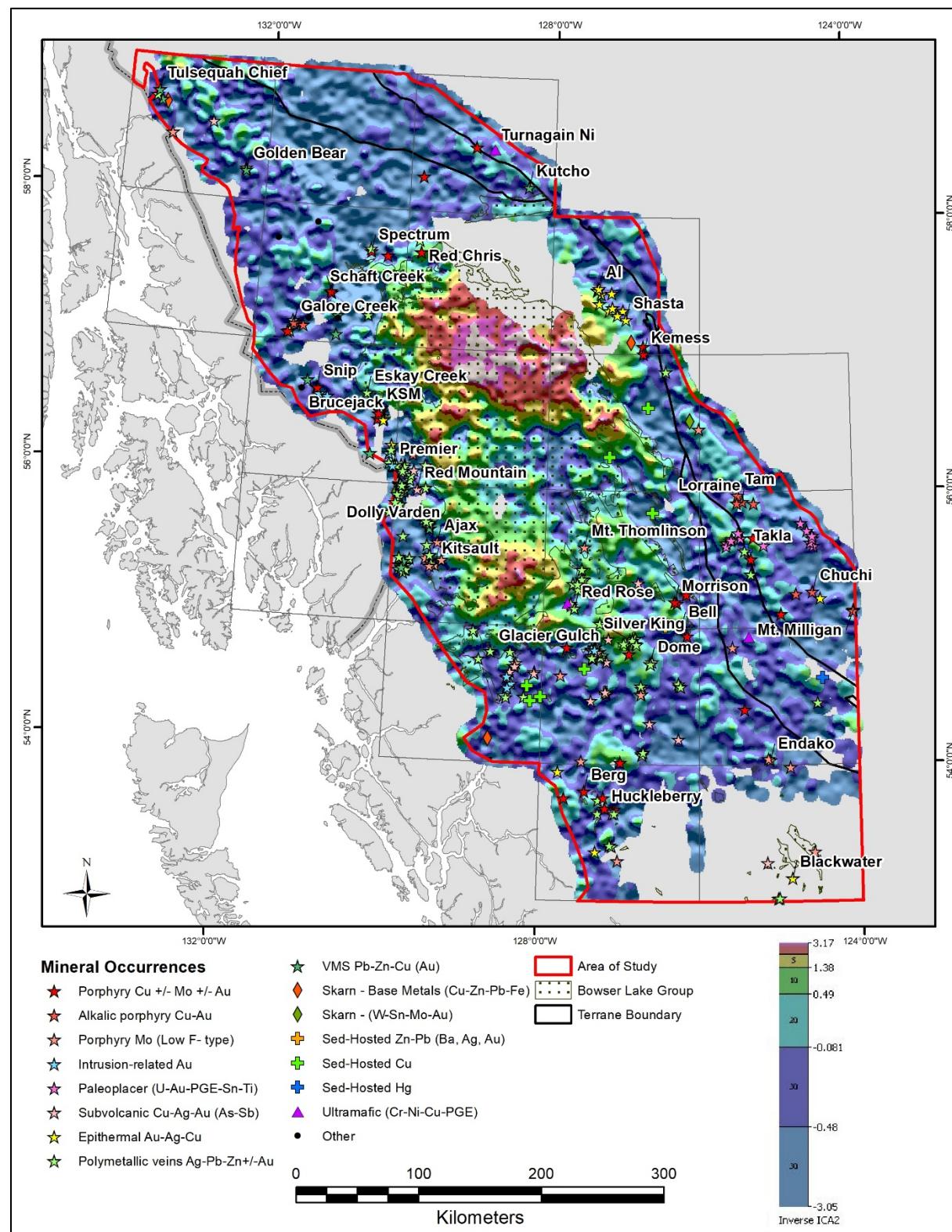


Figure 23: Inverse ICA2 defining the Bowser Basin

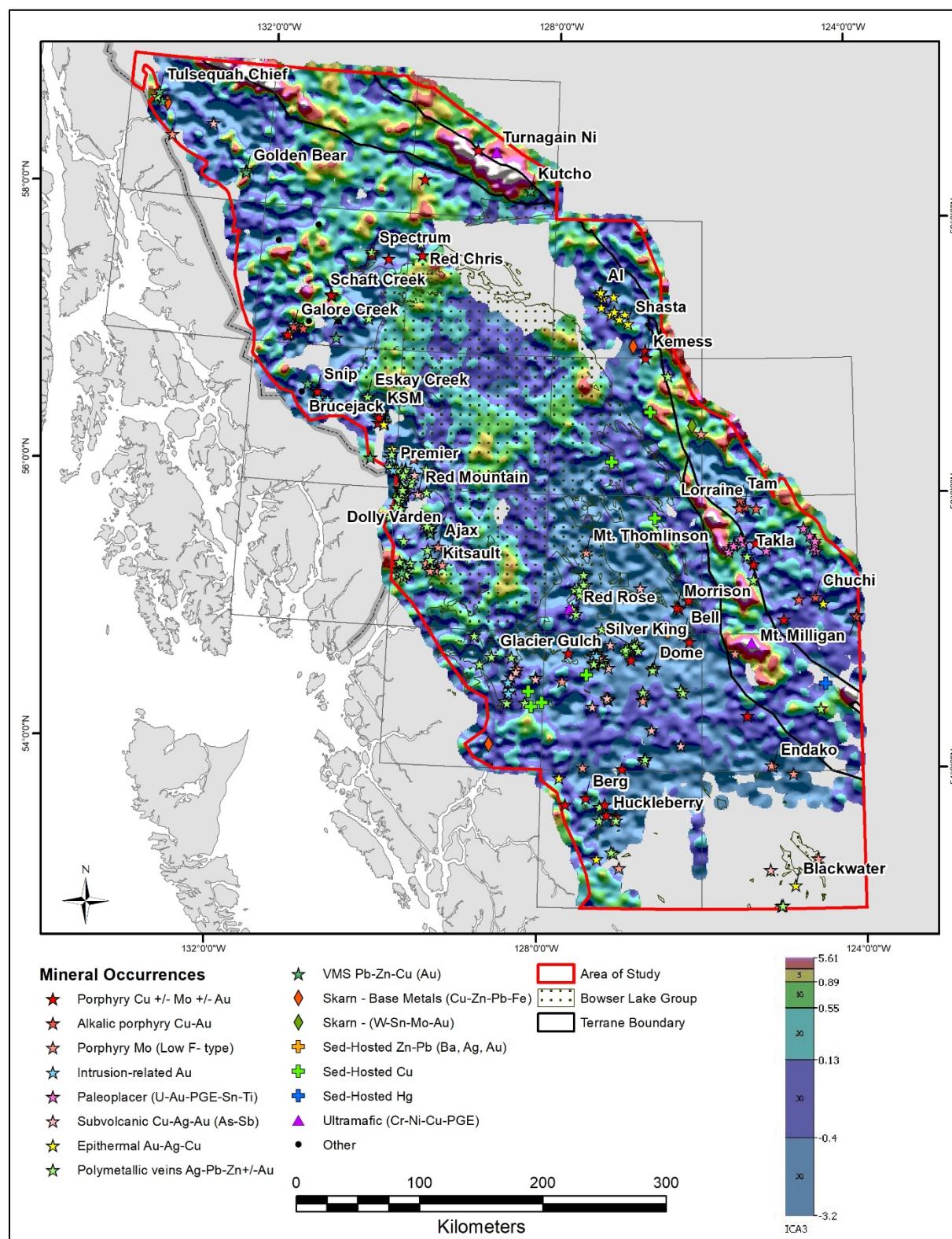


Figure 24: IC3 associated with ultramafic rocks of the Cache Creek terrane

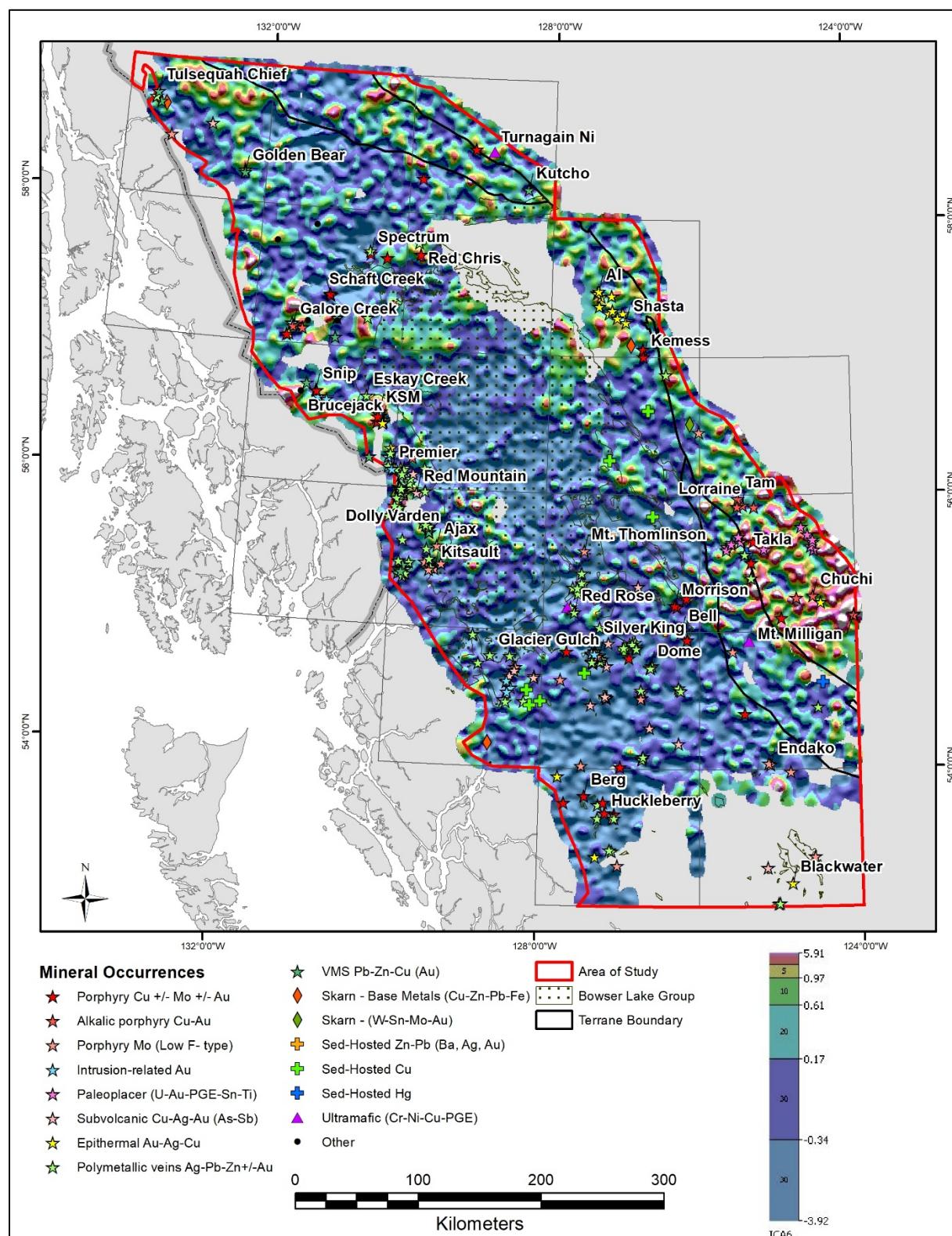


Figure 25: IC6 associated with the Quesnel terrane

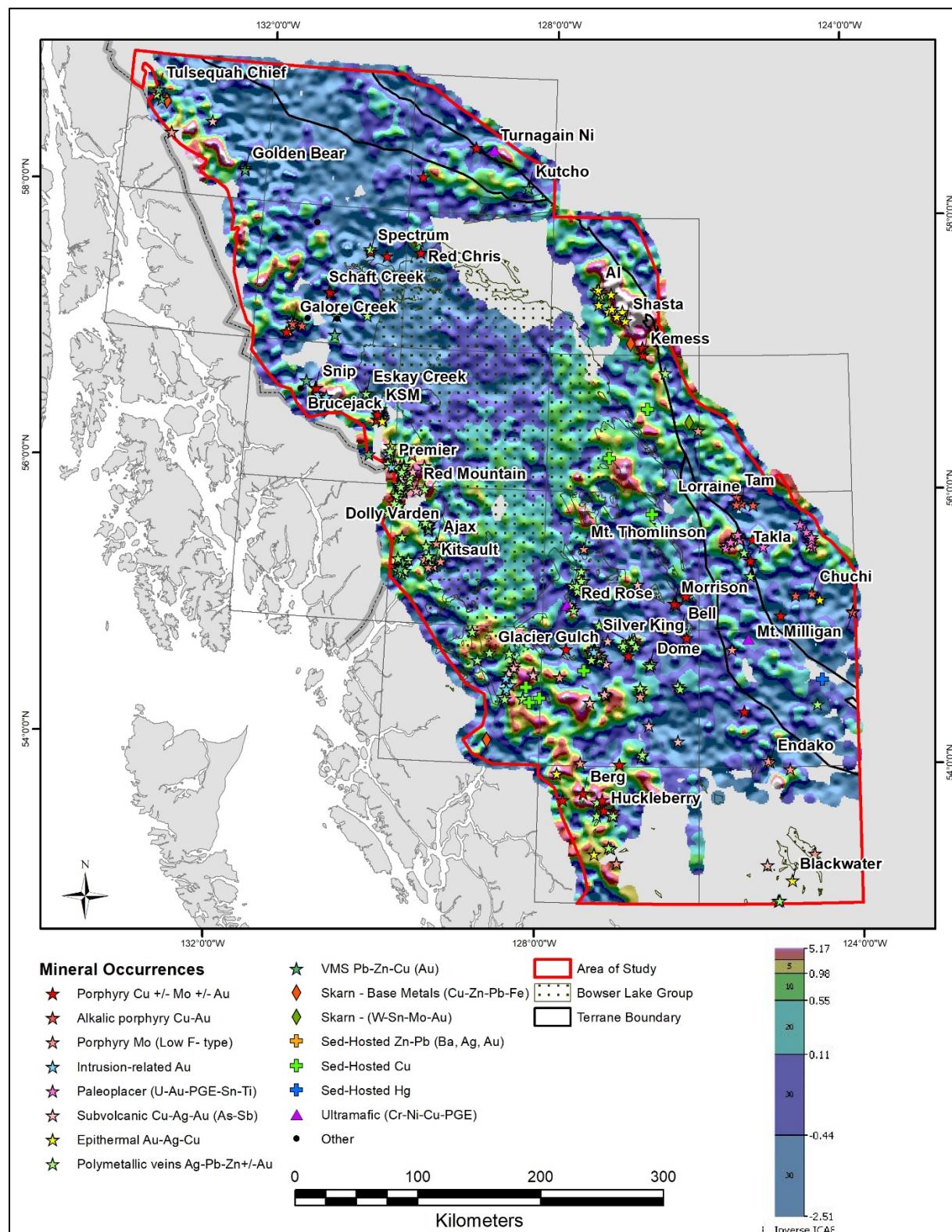


Figure 26: Inverse IC8 associated with epithermal Au-Ag deposits

3.5 Weighted Sums Models

WSMs for porphyry Cu-Au and epithermal Au-Ag deposits were generated using both raw geochemical data and data levelled in a variety of ways to minimize the effects of lithological variation and metal scavenging (Table 4). These include the following products:

- Raw data
- Geochemical data levelled by dominant catchment basin lithology
- Geochemical data levelled by dominant terrane
- Residuals following regression analysis of individual elements against the proportions of different lithologies in each catchment basin
- Either principal components or residuals following regression of individual elements against principal components interpreted to reflect the influence of lithology or metal scavenging.

Raw element WSMs for the polymetallic vein occurrences used the following importance rankings (in parentheses): Cd (1), Ag (1), Sb (1), As (1), Pb (1), Zn (1), Co (-1), Mg (-1), Ni (-1), Sc (-1). The negative rankings for elements associated with mafic and ultramafic rocks are designed to remove magmatic Ni-Cu deposits from these WSMs. Those used for the magmatic Ni-Cu deposits were Ni (1), Cr (1) and Cu (1).

The principal components generated for regression analysis did not incorporate Au data given its strongly censored nature affecting correlations with other elements, but raw Au data were incorporated into WSMs where appropriate. A dilution correction has also been applied to the WSMs by multiplying the WSM score by the square root of the catchment basin area, as described in Section 2.5. A wide selection of WSMs have been generated for only the porphyry Cu-Au and epithermal Au-Ag mineral deposit types to allow direct comparison of the different data processing methods used. Descriptions of the processing applied to the principal components for four different mineral deposit types are summarized in Table 5. Note that the weights (importance rankings) applied to each element are the same for each processing method so that a comparison of how effective each method works will be a function of the processing method used to prepare the data for analysis, not the weightings applied to the model. Final WSMs are presented in Figure 27 to Figure 38.

Table 4: Processing specifications for WSMs using raw and levelled data

Deposit type	Parameters	Importance ranking			
		Raw data	Dominant lithology	Dominant terrane	Multiple regression
Porphyry Cu-Au	Log Au	1	1	1	1
	Cu	5	5	5	5
	Bi	2	2	2	2
	Mo	1	1	1	1
	Te	2	2	2	2
	As	1	1	1	1
Epithermal Ag-Au	Log Au	4	4	4	4
	Ag	2	2	2	2
	Sb	2	2	2	2
	As	2	2	2	2

Table 5: Processing specifications for principal component WSMs

Deposit type	Parameters	Ranking
Porphyry Cu-Au	Log Au	1
	Cu residuals after PC2 and PC3	5
	Bi residuals after PC3	2
	Mo residuals after PC1 and 3	1
	Te residuals after PC1 and 3	2
	As residuals after PC1	1
	PC2 (remove Co, Mg, Ni, Sc)	-3
Polymetallic Ag-Zn-Pb	PC1	2
	PC2 (remove Co, Mg, Ni, Sc)	-2
	Pb residuals after inverse PC3	2
	Zn residuals after inverse PC3	1
Epithermal Ag-Au	Log Au	4
	Ag residuals after PC1	2
	Sb residuals after PC1	2
	As residuals after PC1	2
Magmatic Ni-Cu	Ni residuals after PC2	3
	Cr residuals after PC2	2
	Cu residuals after PC6	1
	PC1	-2

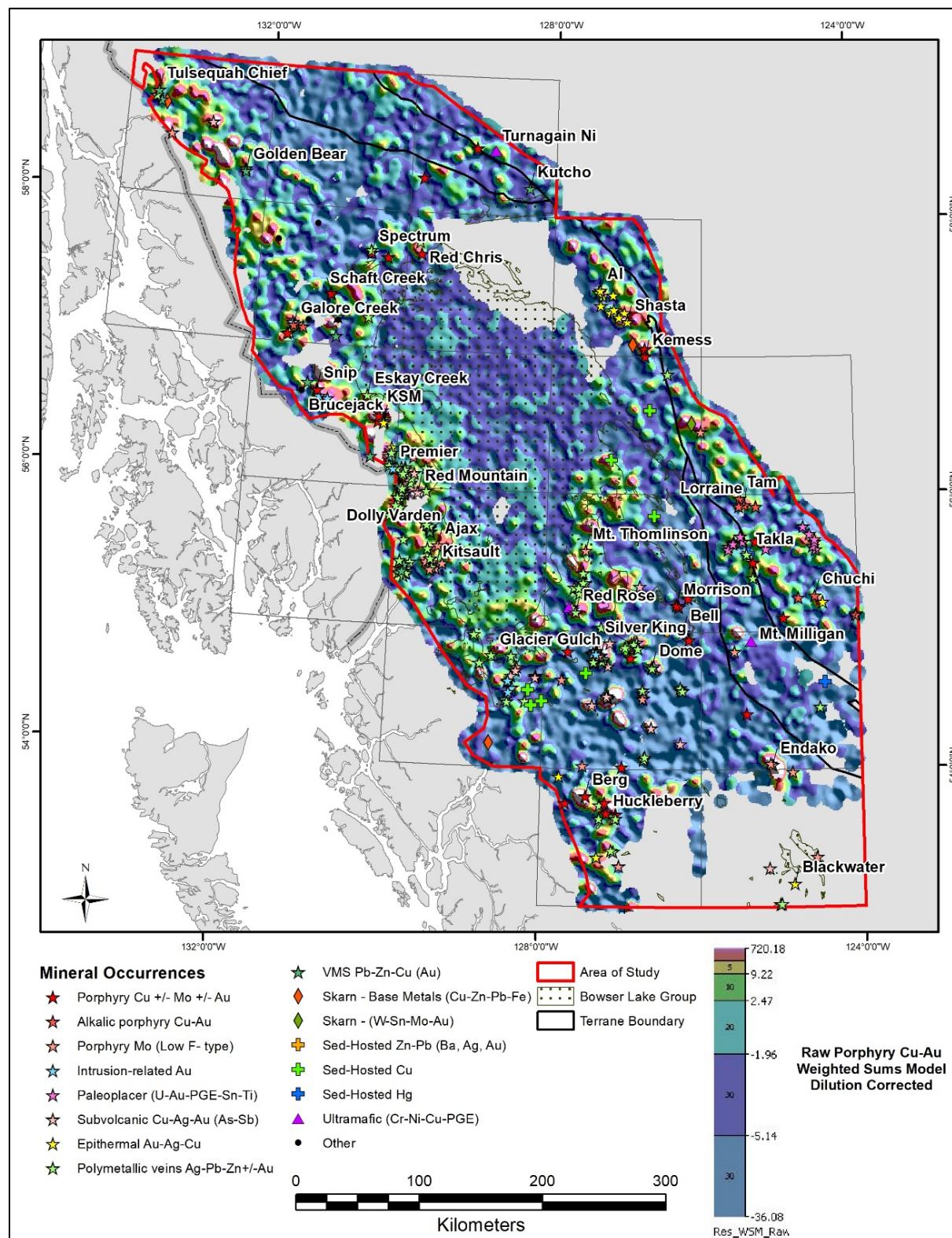


Figure 27: Porphyry Cu-Au WSM using raw geochemical data corrected for dilution

Note: See Table 4 for details.

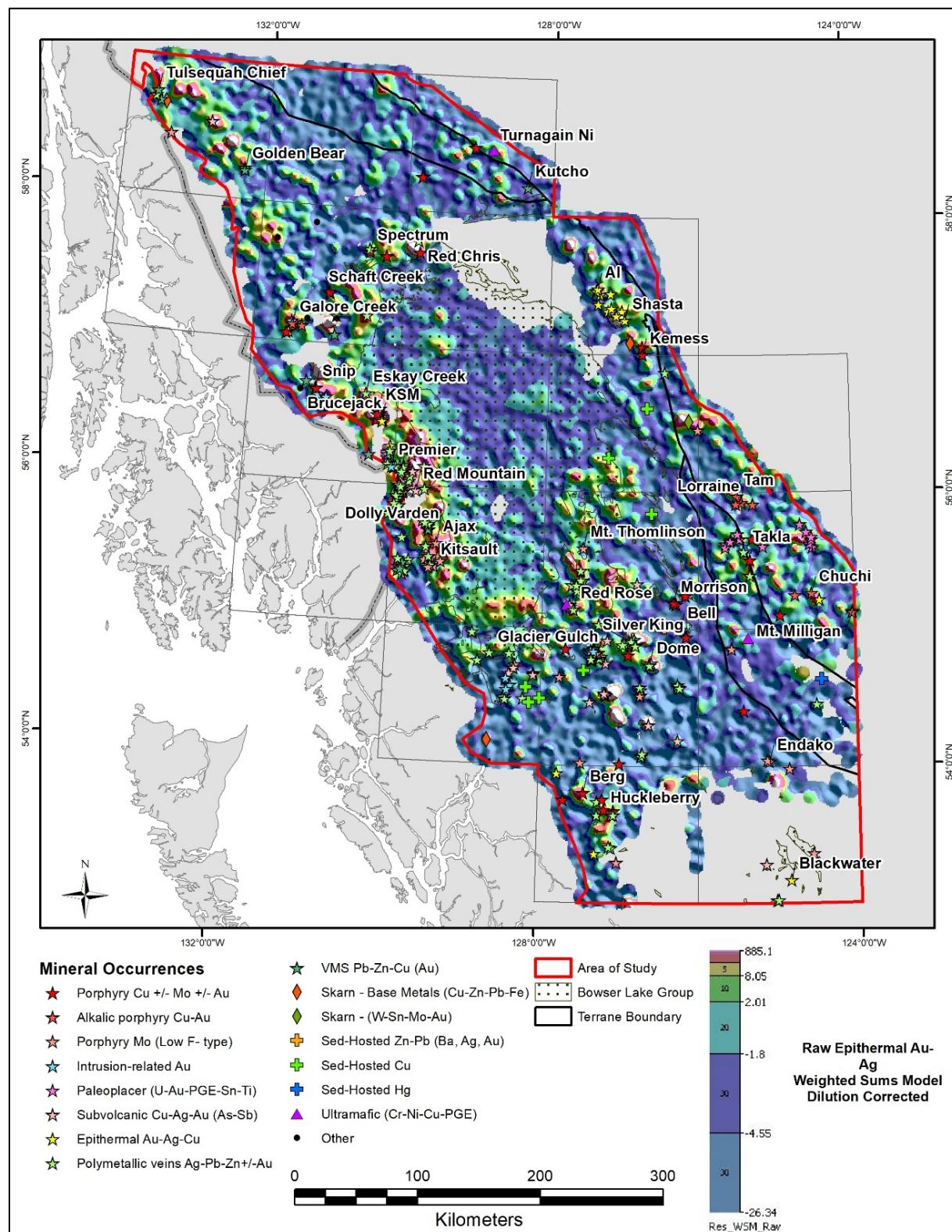


Figure 28: Epithermal Au-Ag WSM using raw geochemical data corrected for dilution

Note: See Table 4 for details.

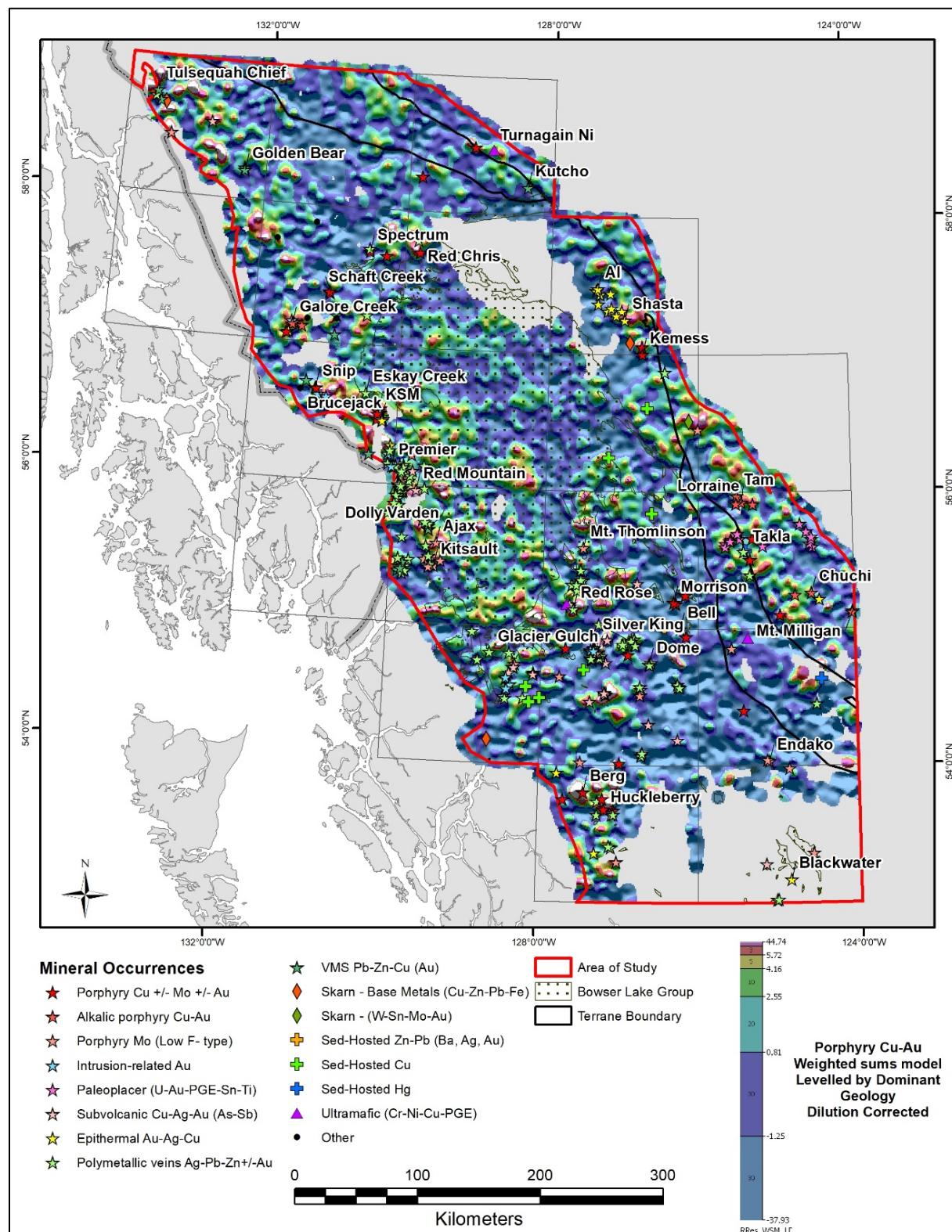


Figure 29: Porphyry Cu-Au WSM using geochemical data levelled by dominant catchment basin lithology

Note: See Table 4 for details.

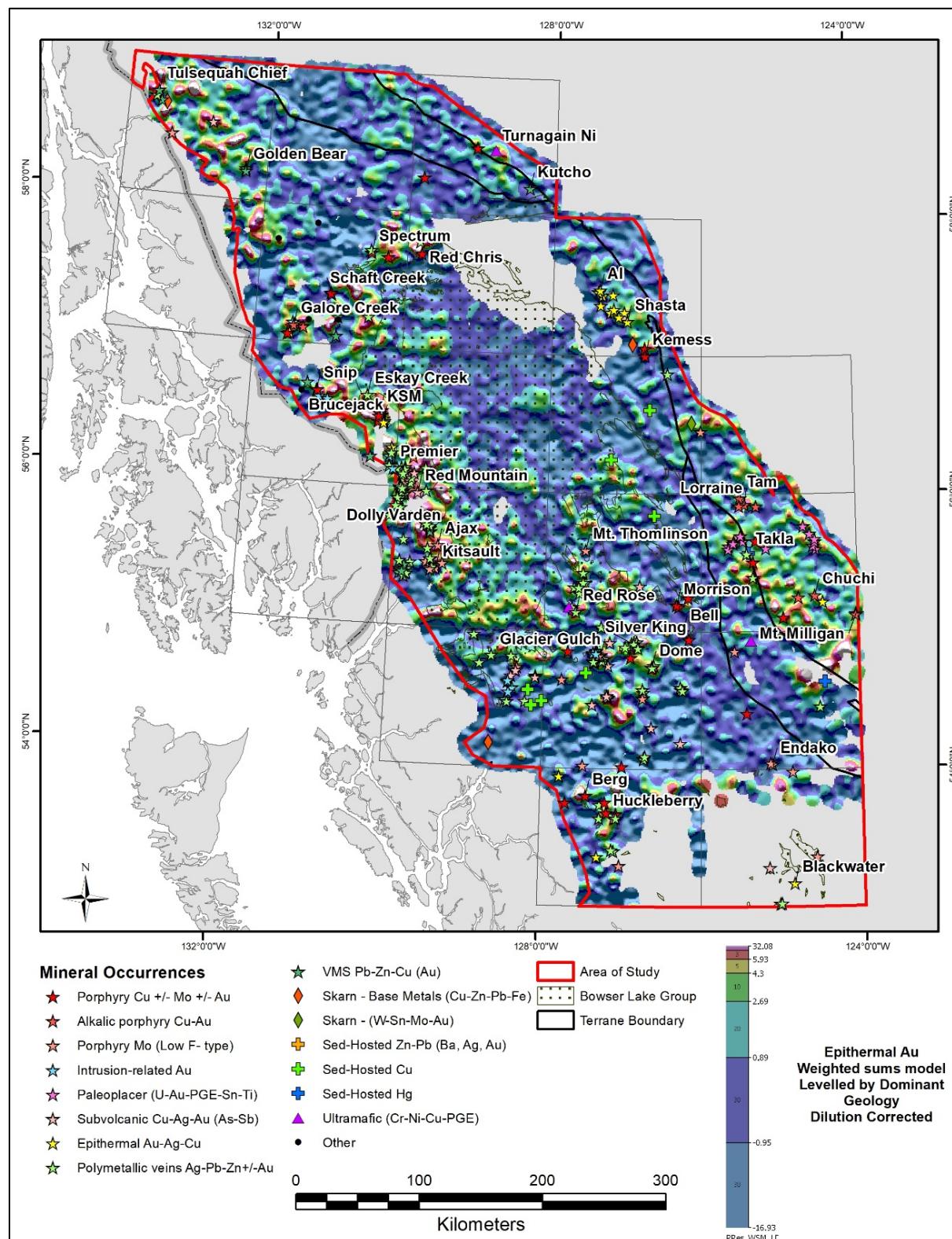


Figure 30: Epithermal Au-Ag WSM using geochemical data levelled by dominant catchment basin lithology

Note: See Table 4 for details.

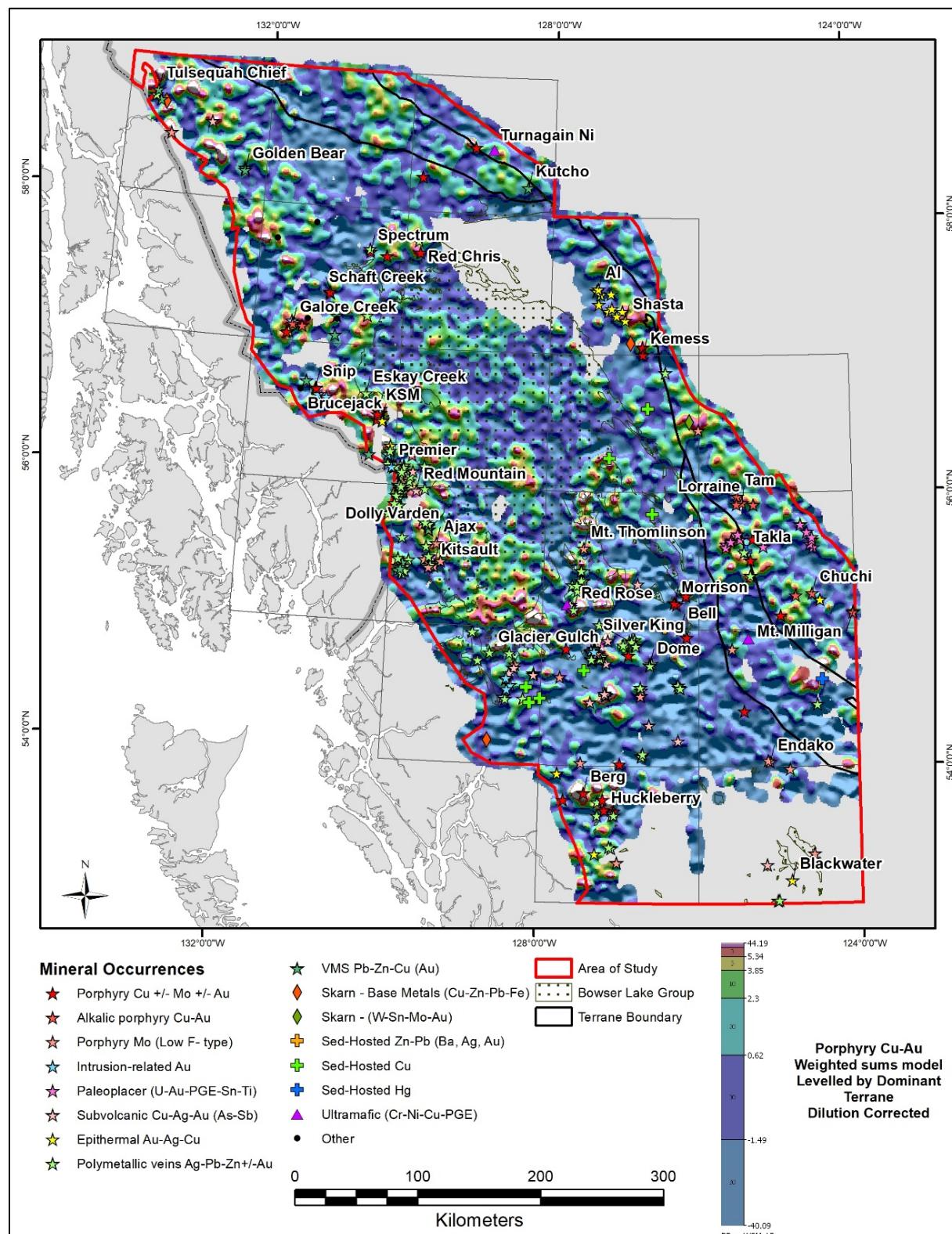


Figure 31: Porphyry Cu-Au WSM using geochemical data levelled by dominant terrane

Note: See Table 4 for details.

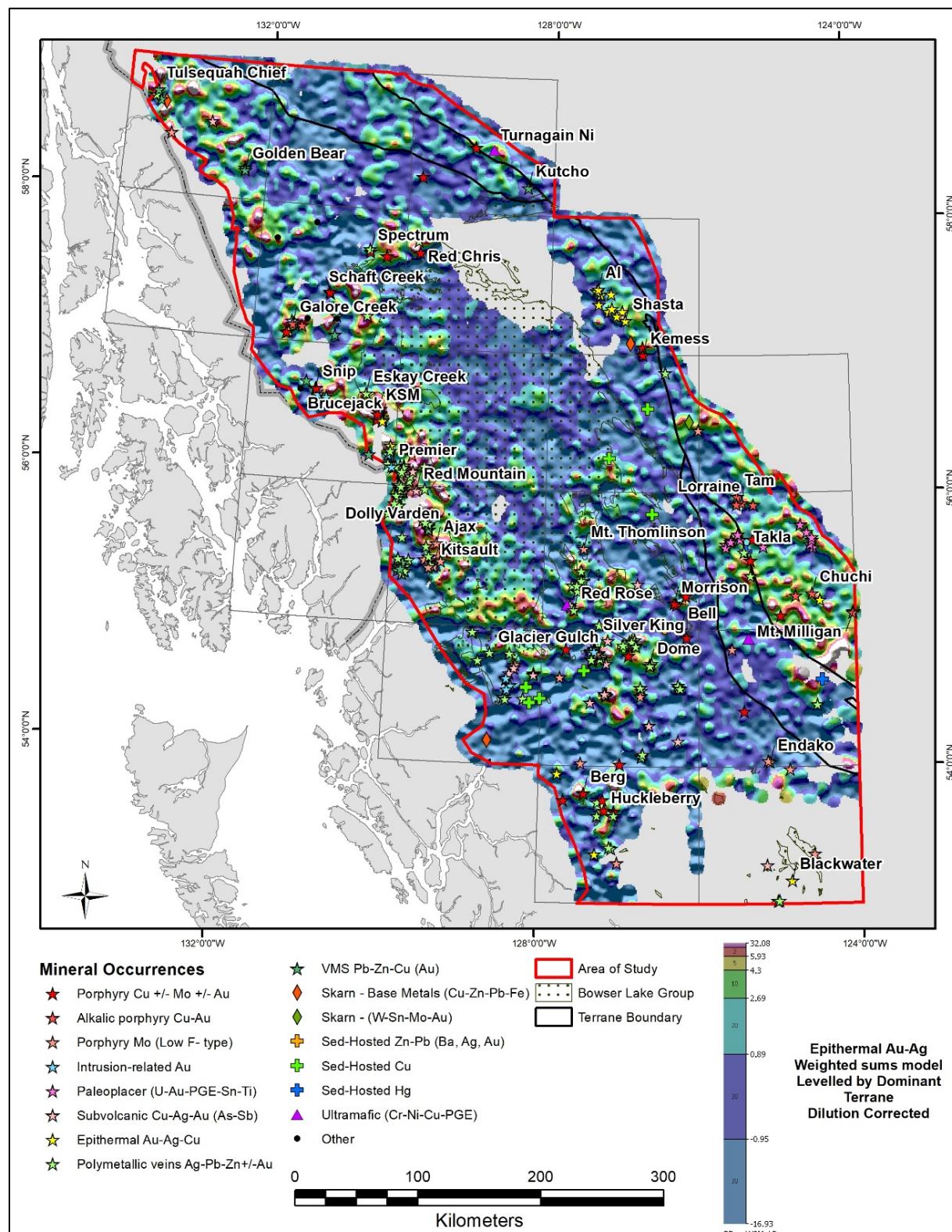


Figure 32: Epithermal Au-Ag WSM using geochemical data levelled by dominant terrane

Note: See Table 4 for details.

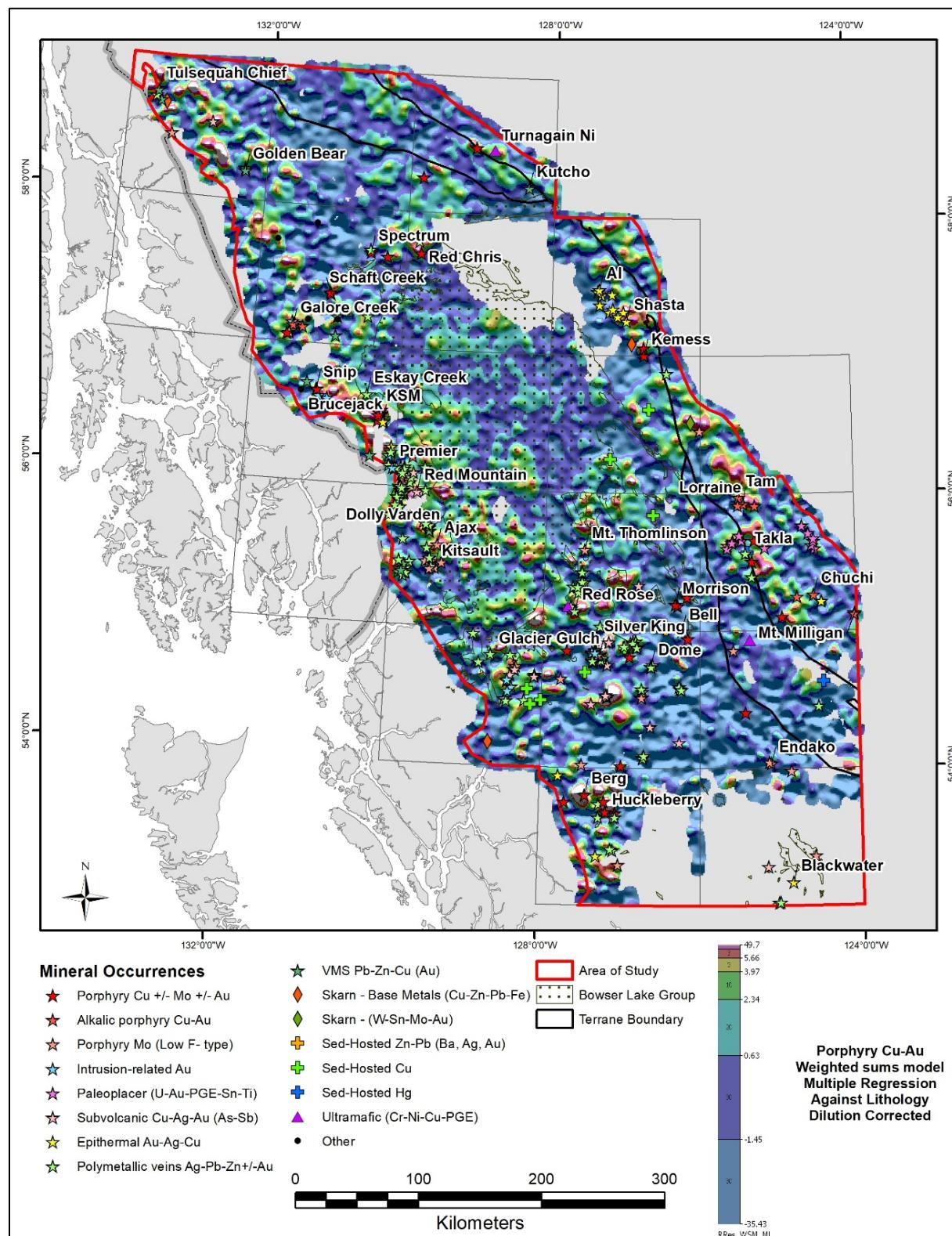


Figure 33: Porphyry Cu-Au WSM using residuals following multiple regression analysis against the proportions of each lithology in catchment basins

Note: See Table 4 for details.

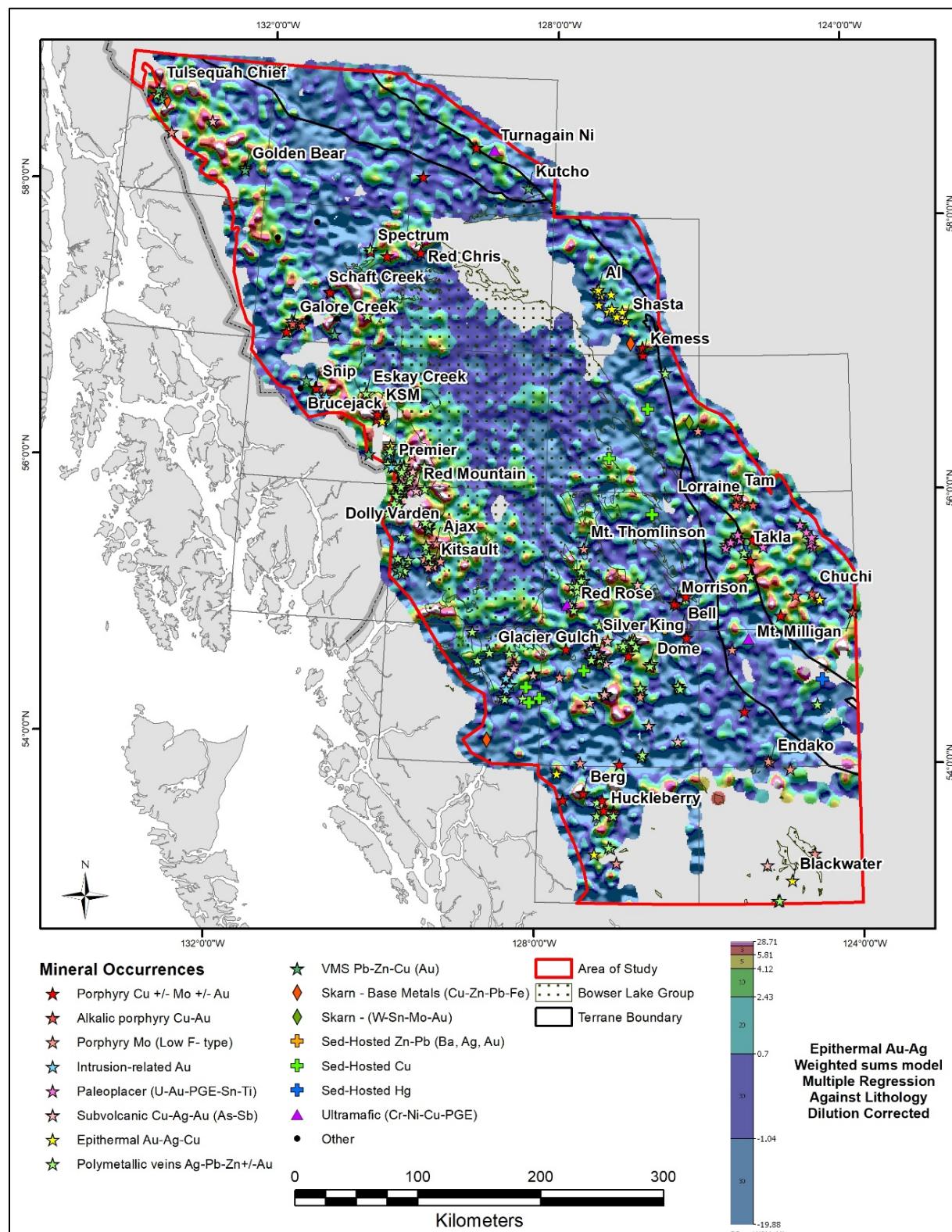


Figure 34: Epithermal Au-Ag WSM using residuals following multiple regression analysis against the proportions of each lithology in catchment basins

Note: See Table 4 for details.

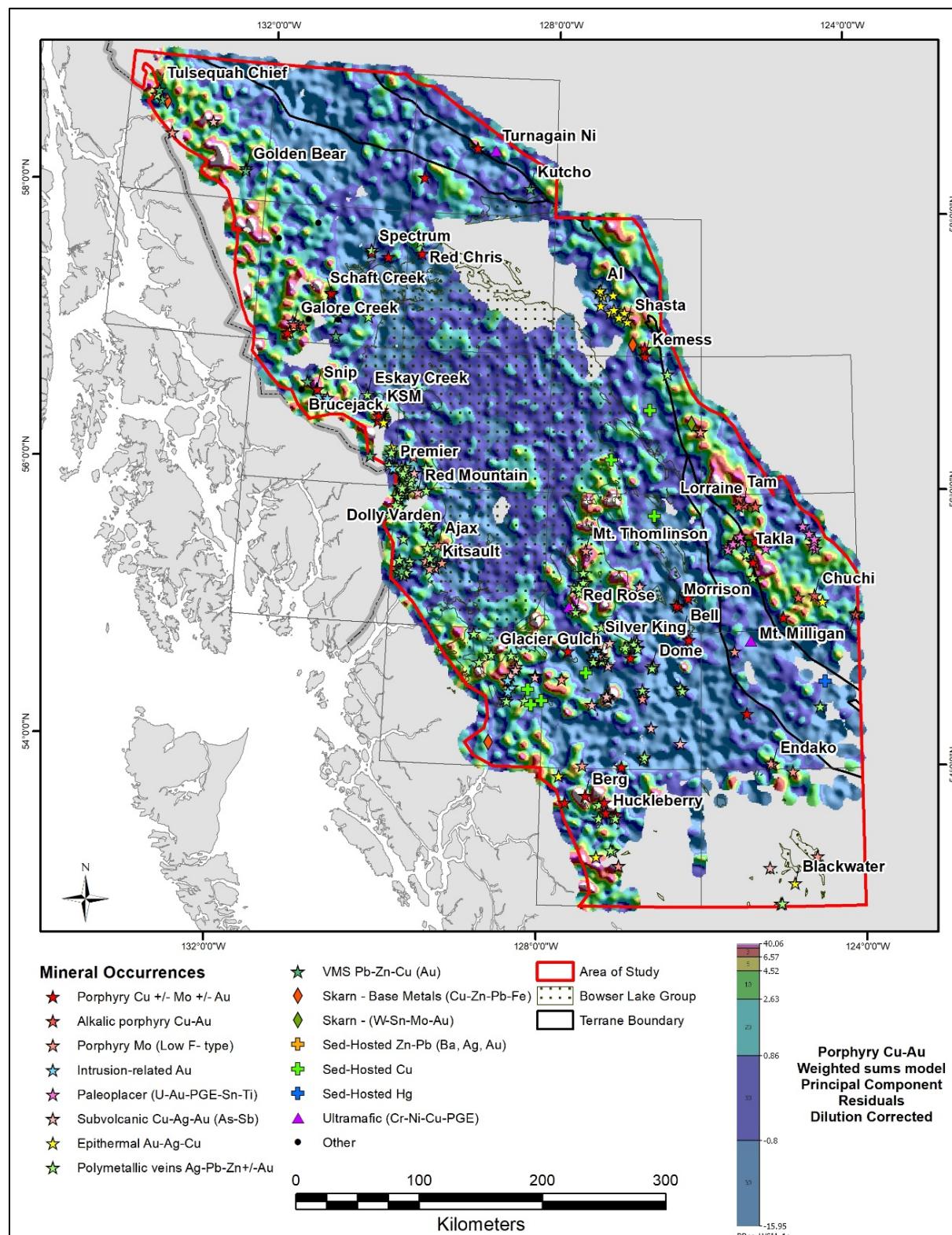


Figure 35: Porphyry Cu-Au WSM using residuals following multiple regression analysis against principal components.

Note: See Table 5 for details.

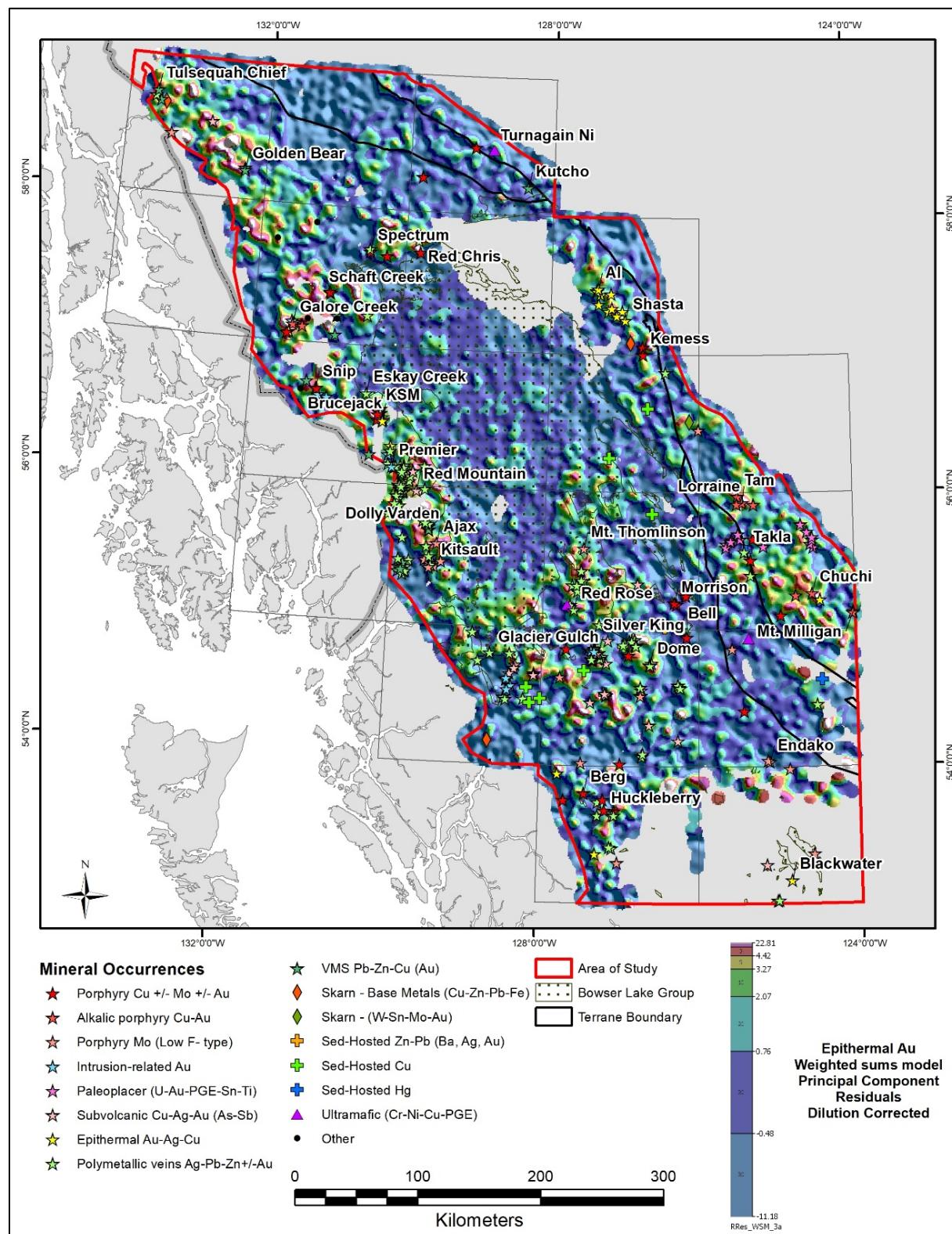


Figure 36: Epithermal Au-Ag WSM using residuals following multiple regression analysis against principal components.

Note: See Table 5 for details.

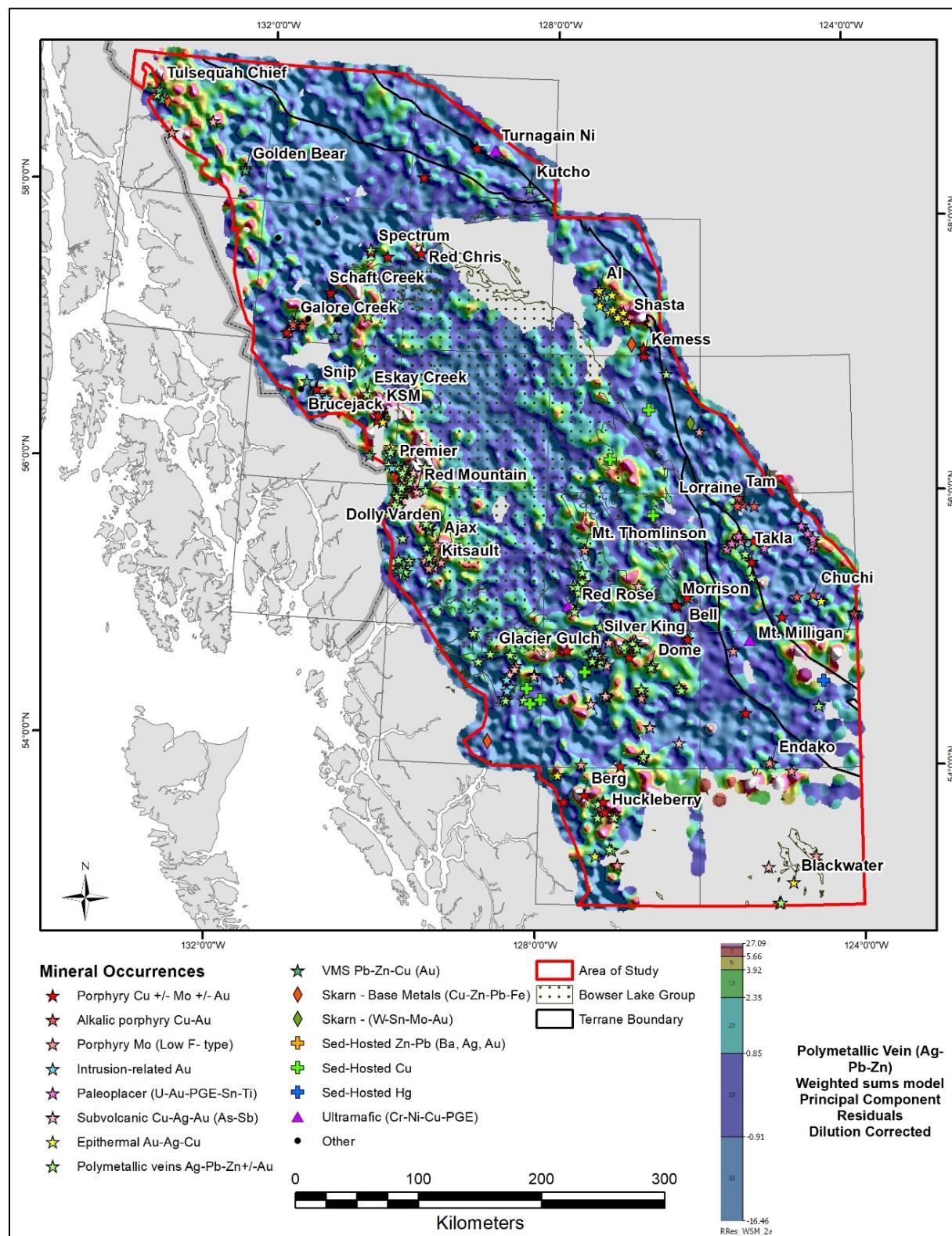


Figure 37: Polymetallic Ag-Pb-Zn veins WSM using residuals following multiple regression analysis against principal components

Note: See Table 5 for details.

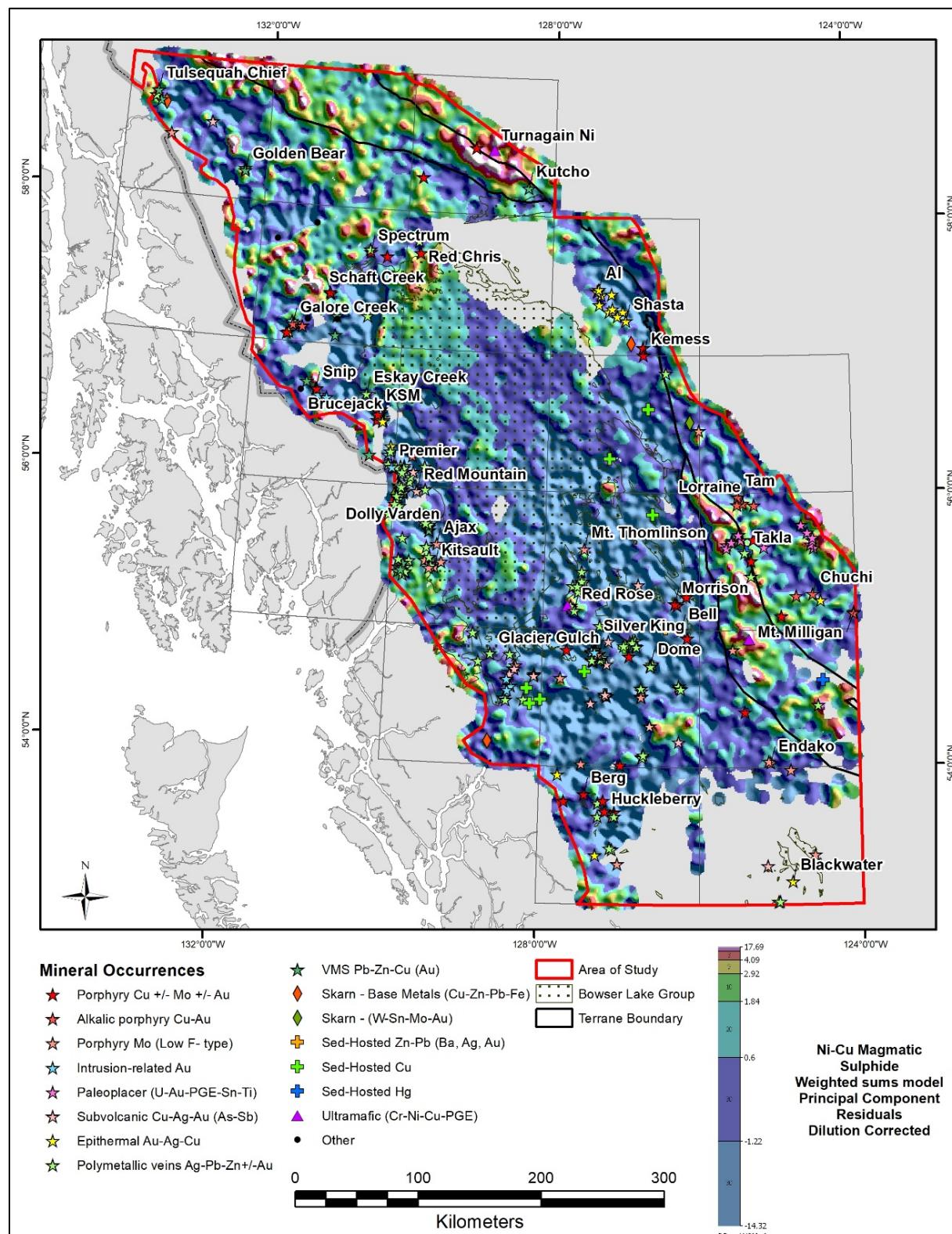


Figure 38: Magmatic Ni-Cu and ultramafic rocks WSM using residuals following multiple regression analysis against principal components.

Note: See Table 5 for details.

The efficacy of the different WSM to correctly identify mineralized catchment basins has been assessed for the porphyry Cu-Au and epithermal Au-Ag WSMs by determining the number of developed prospects, past producers and prospects captured by the upper 90th percentile of their respective scores as a proportion of the total number of MINFILE occurrences in that mineral deposit class. In addition, the proportion of each mineral deposit class compared to all metallic mineral deposits captured within the upper 90th percentile of data from each WSM score has also been determined. The former is a function of the number of true positives in the upper 90th percentile of data (i.e. the number of mineral occurrences correctly identified), whereas the latter is a function of the number of false positives (i.e. mineral occurrences not consistent with the WSMs). These proportions are listed in Table 6. WSMs that are uncorrected and corrected (DC) for the effects of dilution as a function of catchment basin area are included.

Table 6: Proportions of true positives in delineating mineralization

Deposit type	Processing method	Proportion of True Positives (%)	Proportion of Total Mineral deposits (%)
Porphyry Cu-Au	Raw data	26	28
	Levelled by dominant lithology	23	27
	Levelled by dominant terrane	24	29
	Multiple regression against geology	23	28
	Regression against principal components	23	36
	Raw data with dilution correction – DC	25	27
	Levelled by dominant lithology – DC	41	26
	Levelled by dominant terrane – DC	42	26
	Multiple regression against geology – DC	40	26
	Regression against principal components – DC	31	29
Epithermal Au-Ag	Raw data	38	7
	Levelled by dominant lithology	34	6
	Levelled by dominant terrane	35	6
	Multiple regression against geology	30	7
	Regression against principal components	25	8
	Raw data with dilution correction – DC	32	7
	Levelled by dominant lithology – DC	33	4
	Levelled by dominant terrane – DC	38	4
	Multiple regression against geology – DC	33	4
	Regression against principal components – DC	26	5

The data in Table 6 show a difference in the response to various data processing methods for porphyry Cu-Au and epithermal Au-Ag deposits. Using processed geochemical data to generate WSMs for porphyry Cu-Au deposits has no demonstrable improvement in the proportion of true positives unless accompanied by a dilution correction, although there is a reduction in the number of false positives where the data used in the WSMs have been regressed against principal components. These data processing methods have resulted in a maximum improvement of 62% and 29%, respectively, for true positives and reduction in false positives for porphyry Cu-Au deposits (Figure 39). Simple data levelling methods have the greatest increase in true positives when accompanied by a dilution correction, but at the cost of more false positives for porphyry Cu-Au deposits. The use of principal components for regression analysis without a dilution correction best reduces the number of false positives, but at the cost of fewer true positives. This reflects the effect of lithology on Cu distribution and the ability of the various data processing methods to correct for this effect.

Copper is strongly weighted in the porphyry Cu-Au WSM (Table 4 and Table 5). It is clear for the porphyry Cu-Au WSM that the dilution correction has the most beneficial effect, both in terms of capturing more true positives, as well as mineral deposits in general.

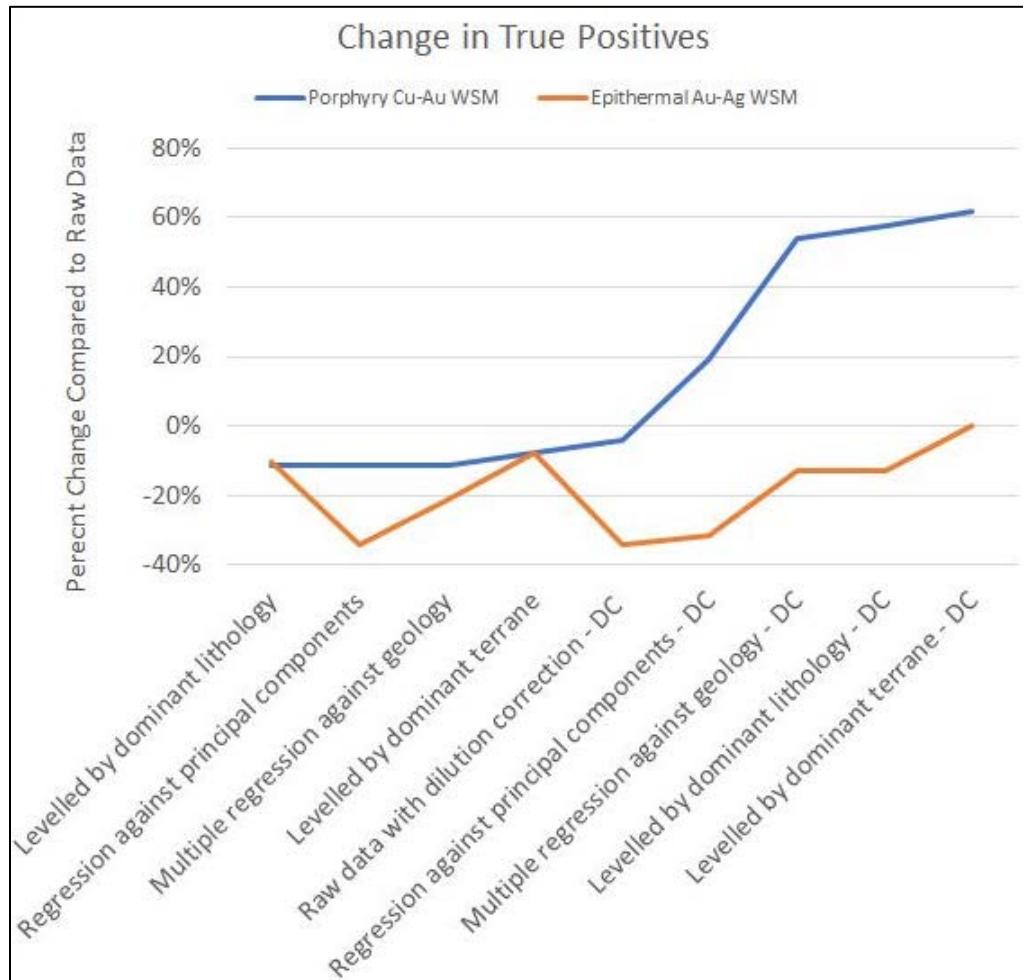


Figure 39: Summary of true positives for porphyry Cu-Au and epithermal Au-Ag deposits

Note: "DC" refers to dilution-corrected processing methods. See Table 6 for details.

By contrast, there is minimal improvement in predictability for epithermal Au-Ag deposits irrespective of the data processing technique used. Levelling the data with or without dilution correction even degrades the predictability of the WSMs. This reflects the very strong weighting of Au in the epithermal Au-Ag WSM (Table 4 and Table 5). Gold shows little lithological or metal scavenging control in this project area, and so the various data processing methods examined have minimal impact on the performance of the WSM used here. Gold is also one of the least precise elements available for incorporation into the WSM. Therefore, whether the various data processing methods discussed here will produce an improvement in WSM for mineral exploration targeting depends upon the extent to which the elements used to construct the models are influenced by lithology or metal scavenging. These processes can be identified through exploratory data analysis (EDA).

WSMs have also been created for raw magmatic Ni-Cu and polymetallic vein deposits and these scores have been compared to WSMs for the same deposit types, both with and without a dilution correction. There is a

slight increase in the percentage of true positives in the upper 90th percentile when the elements in the magmatic Ni-Cu WSM are regressed against principal components. There is also a slight decrease in the number of false positives if a dilution correction is also incorporated. The number of false positives is decreased, as is the number of true positives, where the data for the polymetallic vein deposits are regressed against principal components compared to the raw data WSM. The dilution-corrected data are virtually the same for both raw and regressed data.

3.6 Sampling Effectiveness

The final WSM products generated for this report have been corrected for the effects of dilution using catchment area. The reason for this correction is evident where raw Cu, Cu levelled by two different methods, and porphyry Cu-Au WSM scores are plotted against catchment area (Figure 40). Note that 24 catchments with estimated areas >400 km² have been excluded from the plot, as have catchments having an area <0.1 km². Catchments with areas >100 km² may be related either to locational errors that we have been unable to rectify or were ill-advised samples.

The exponential decay of metal values with increasing catchment basin area is a commonly observed phenomenon due to the increasing influence of background-derived sediment on the geochemistry of the stream sediment sample. The data points shown in Figure 40 have also been attributed with the highest ranking MINFILE occurrence in the catchment basin. The plot illustrates the subtle or non-existent response of geochemical data in large catchment basins.

Levelled WSM products for two major mineral deposit types have been selected to compare the dilution-corrected and uncorrected WSMs (Figure 41 and Figure 42). The effect of correcting for dilution is to subtly enhance catchment basins with elevated WSM scores associated with large catchment areas and de-emphasize catchment basins with elevated WSM scores associated with small catchment areas.

Examination of many plots like Figure 40 suggests that most metal values or interpretive products derived from them have largely decayed to background values at catchments areas around 25 km². Samples from catchment basins with areas greater than this (~8% of catchment basins) are dominated by background sediment, raising the possibility that they have not been effectively sampled. Scope therefore exists to re-sample these larger catchment areas to detect new geochemical anomalies. The extent of the area that may not have been effectively sampled is illustrated in Figure 43.

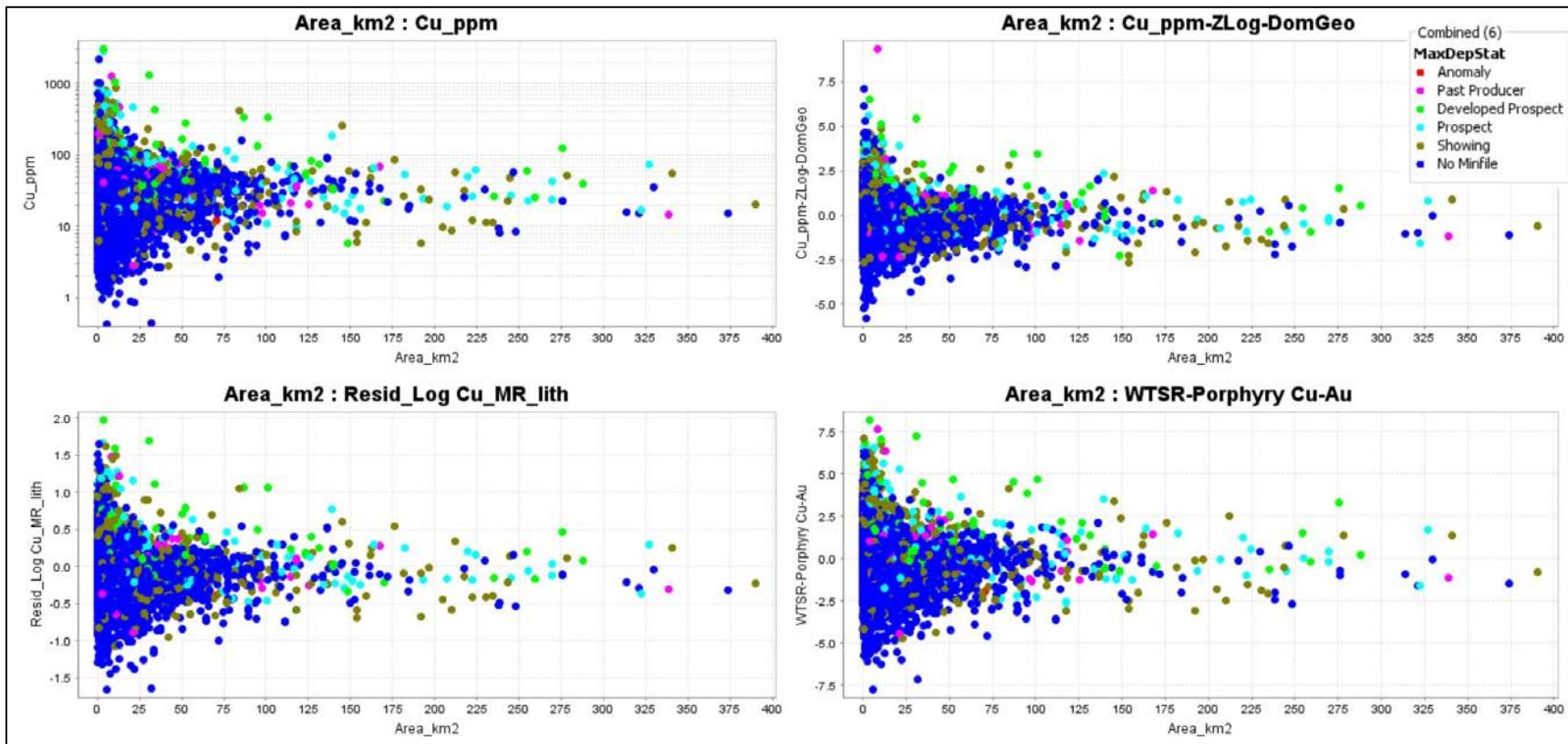


Figure 40: Plot of raw Cu, Cu data levelled using two different methods, and a WSM for porphyry Cu-Au deposits against catchment area

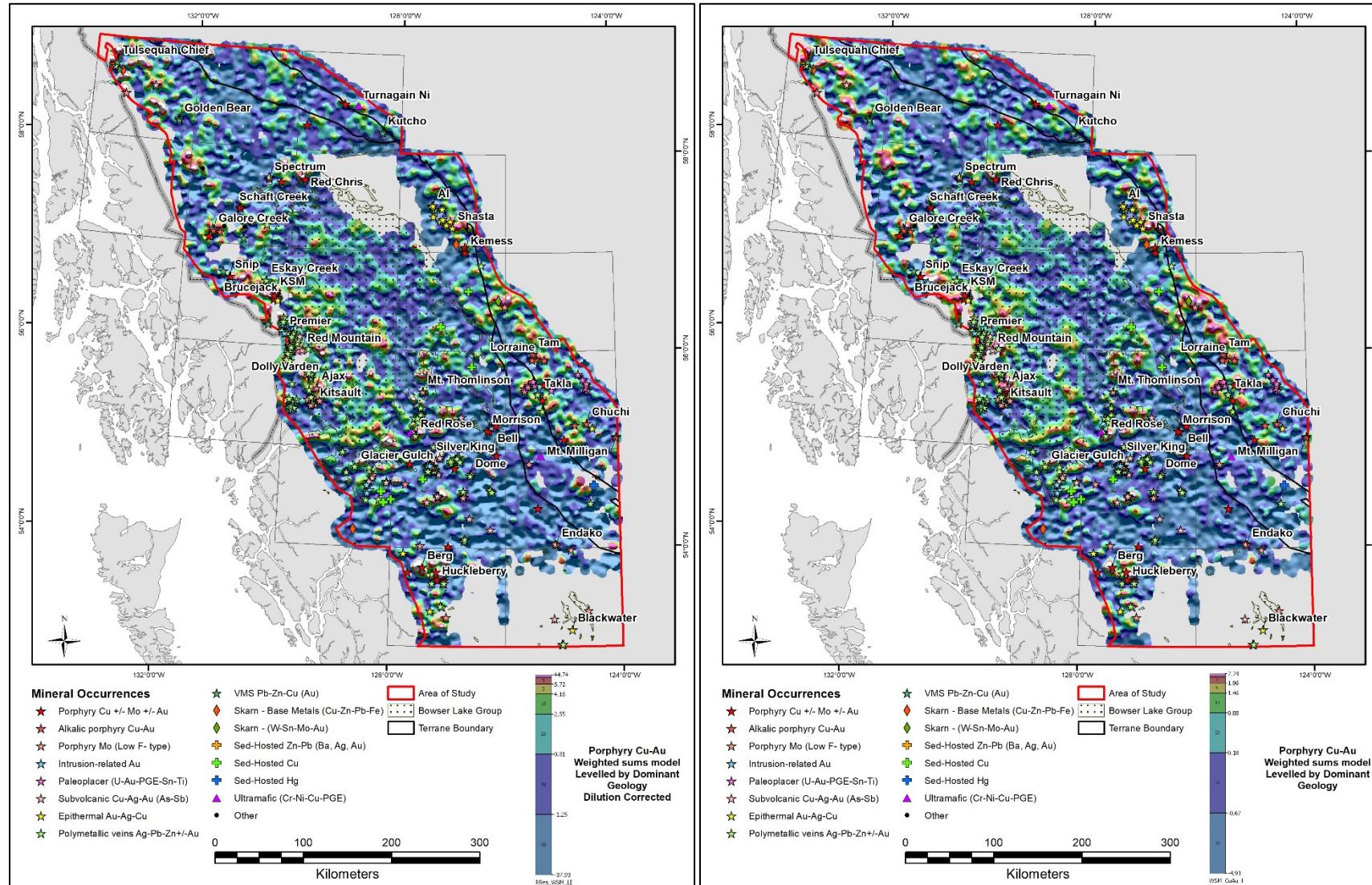


Figure 41: Comparison of WSM for porphyry Cu-Au deposits using data levelled for dominant lithology with dilution-correction (left) and uncorrected (right)

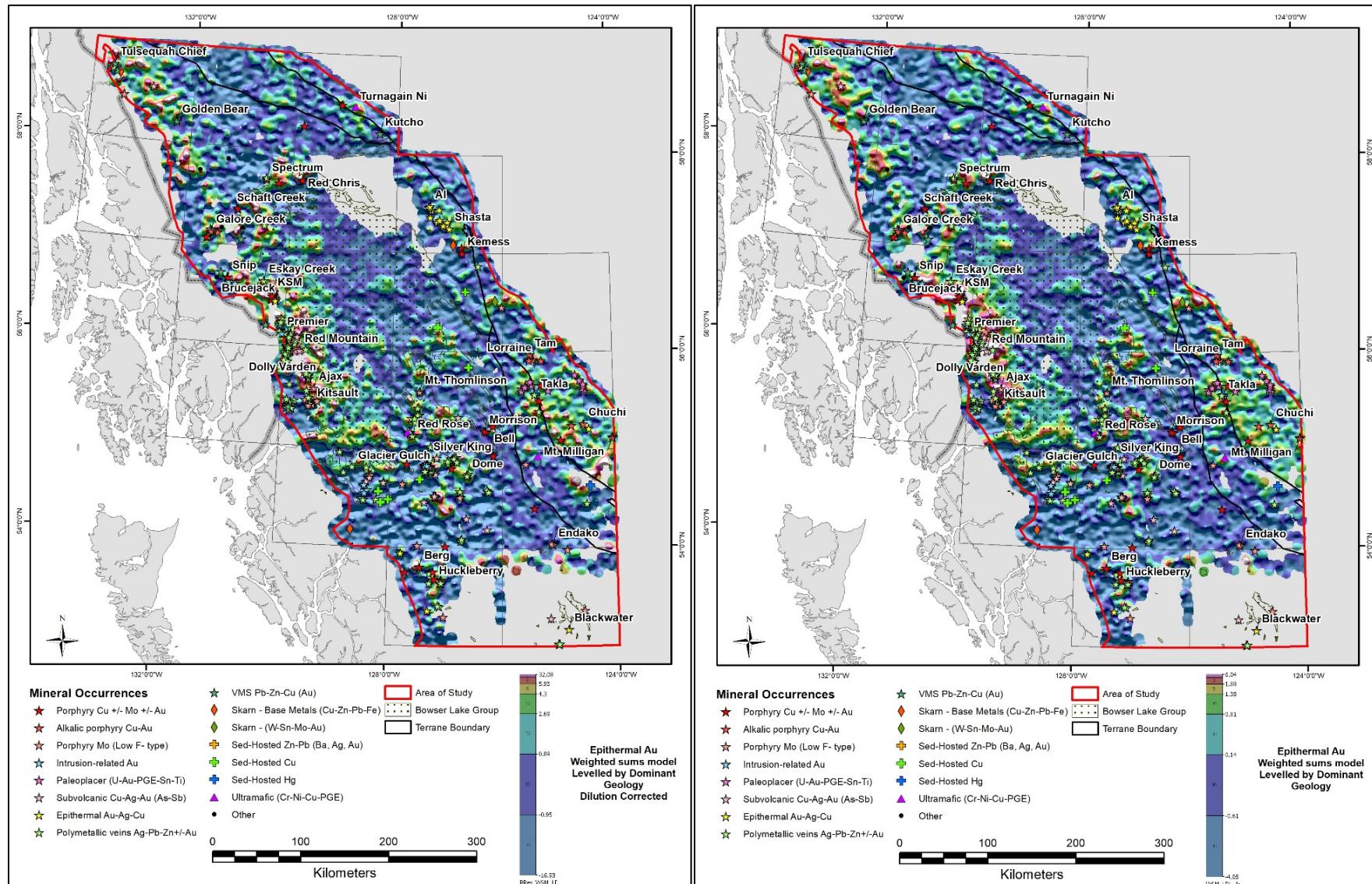


Figure 42: Comparison of WSM for epithermal Au-Ag deposits using data levelled for dominant lithology with dilution-correction (left) and uncorrected (right)

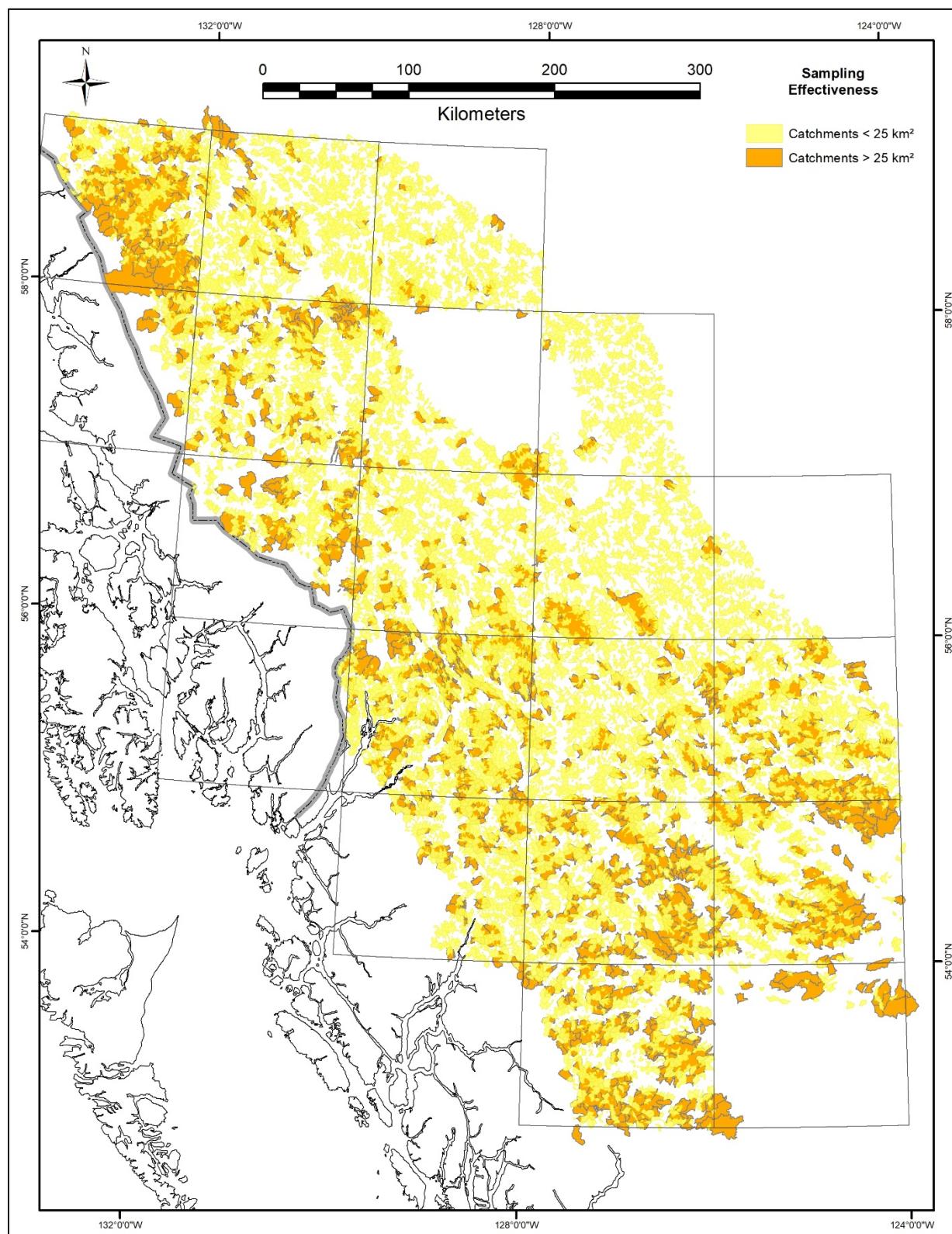


Figure 43: Map illustrating the areal extent of catchment basins >25 km² that may not have been effectively sampled

3.7 Advanced Data Analysis Methods

Unlike the WSM described in the preceding section, the advanced data analysis methods investigated in this study were supervised, as they required a formal training dataset based on known mineral occurrences. Two different types of data inputs were used in the case of both RF and typicality methods. The first, and more conventional approach, was to use CLR data for PCA as inputs. This led to the identification of an unrealistically high number of prospective catchments in the case of the polymetallic Ag-Pb-Zn vein deposits using both approaches. Therefore, the second approach used residuals following multiple regression analysis of each element against the proportions of each lithology within individual catchment basins to minimize the effects of variable lithology on the random forests and typicality classifications to see if they could be improved. An independent assessment of the efficacy of the supervised approaches is not possible because all known mineral deposits were used for the initial training set. However, visual inspection of results from both approaches with the WSM maps can be made to determine whether the unsupervised and supervised methodologies generate similar predictions for some mineral deposit types.

3.7.1 Random Forests

RF predictions have been generated for the following mineral deposit types:

- Porph_Alk = Alkalic porphyry Cu-Au
- Epi_Au = Epithermal Au-Ag-Cu
- Int_Au = Intrusion-related Au
- PMV = Polymetallic veins Ag-Pb-Zn±Au
- Porph_Cu = Porphyry Cu±Mo±Au
- Porph_Mo = Porphyry Mo (Low F-type)
- Skarn_BM = Skarn – base metals (Cu-Zn-Pb-Fe)
- SubVolc = Subvolcanic Cu-Ag-Au (As-Sb)
- UM_PGE = Ultramafic (Cr-Ni-Cu-PGE)
- VMS = VMS Pb-Zn-Cu (Au).

An initial training dataset included the mineral deposits listed above as well as 100, randomly selected samples in which no deposits were indicated in the MINFILE. The RF model generated by this first test was then used to classify all samples for which no mineralization had been identified in their catchment basins (test 2). Only the normalized votes from catchment basins classed as unmineralized are presented in the following figures.

Only the alkalic porphyry Cu-Au and epithermal Au-Ag mineral deposit predictions are presented in this section in keeping with the previous sections of the report. These are presented in Figure 44 and Figure 45, respectively. Note that the percentile gridded images present only normalized votes for samples for which there are no known mineral occurrences, in contrast to previous images which contain all data. These predictions provide very clear and focused predictions of further mineral potential for these (and other) deposit types. The RF predictions are also able to distinguish the different varieties of porphyry-style mineralization and provide predictions in which there is minimal overlap (Figure 46).

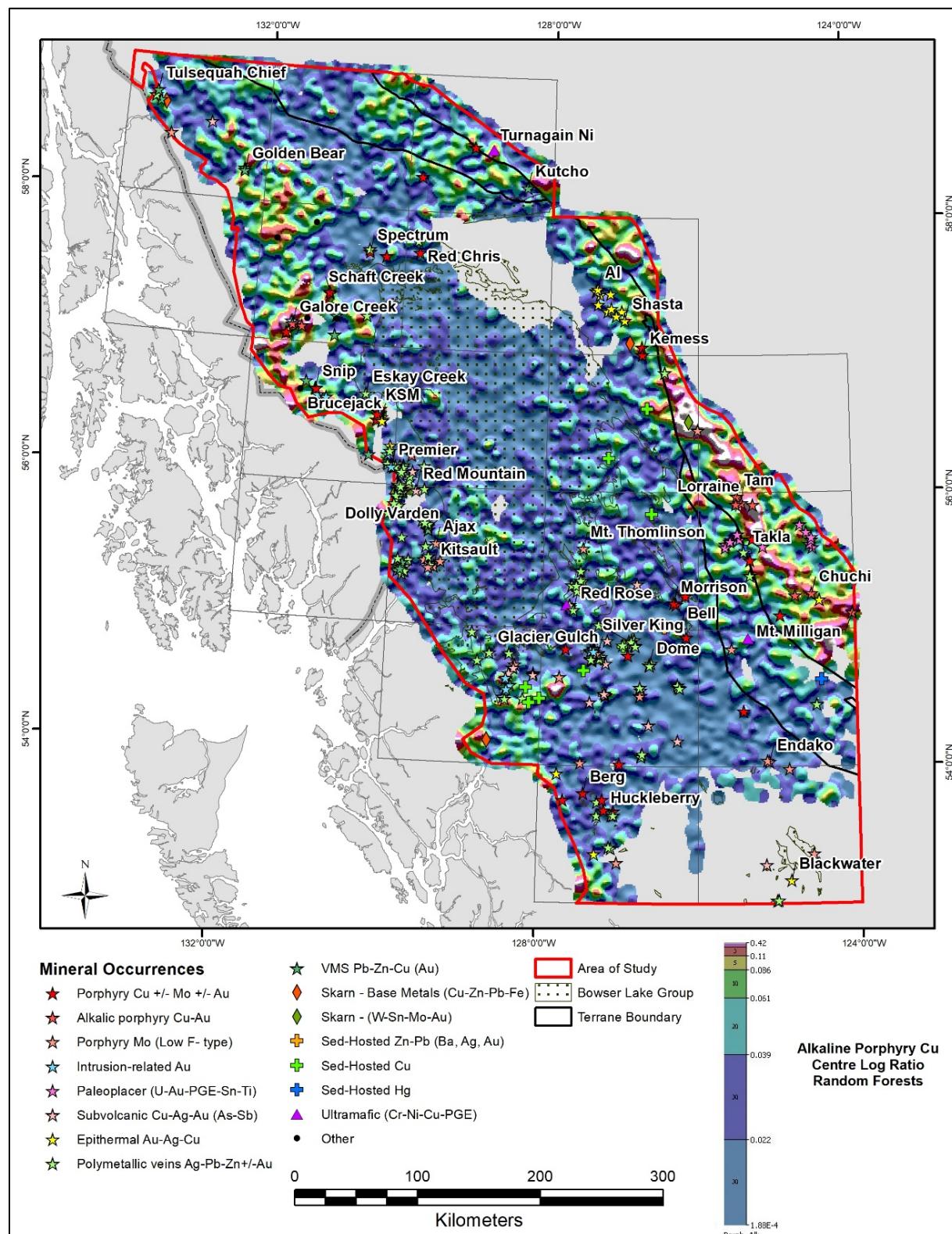


Figure 44: RF predictions for alkalic porphyry Cu-Au deposits

Note: Only data from samples in which no known mineral occurrences occur in their catchments are shown.

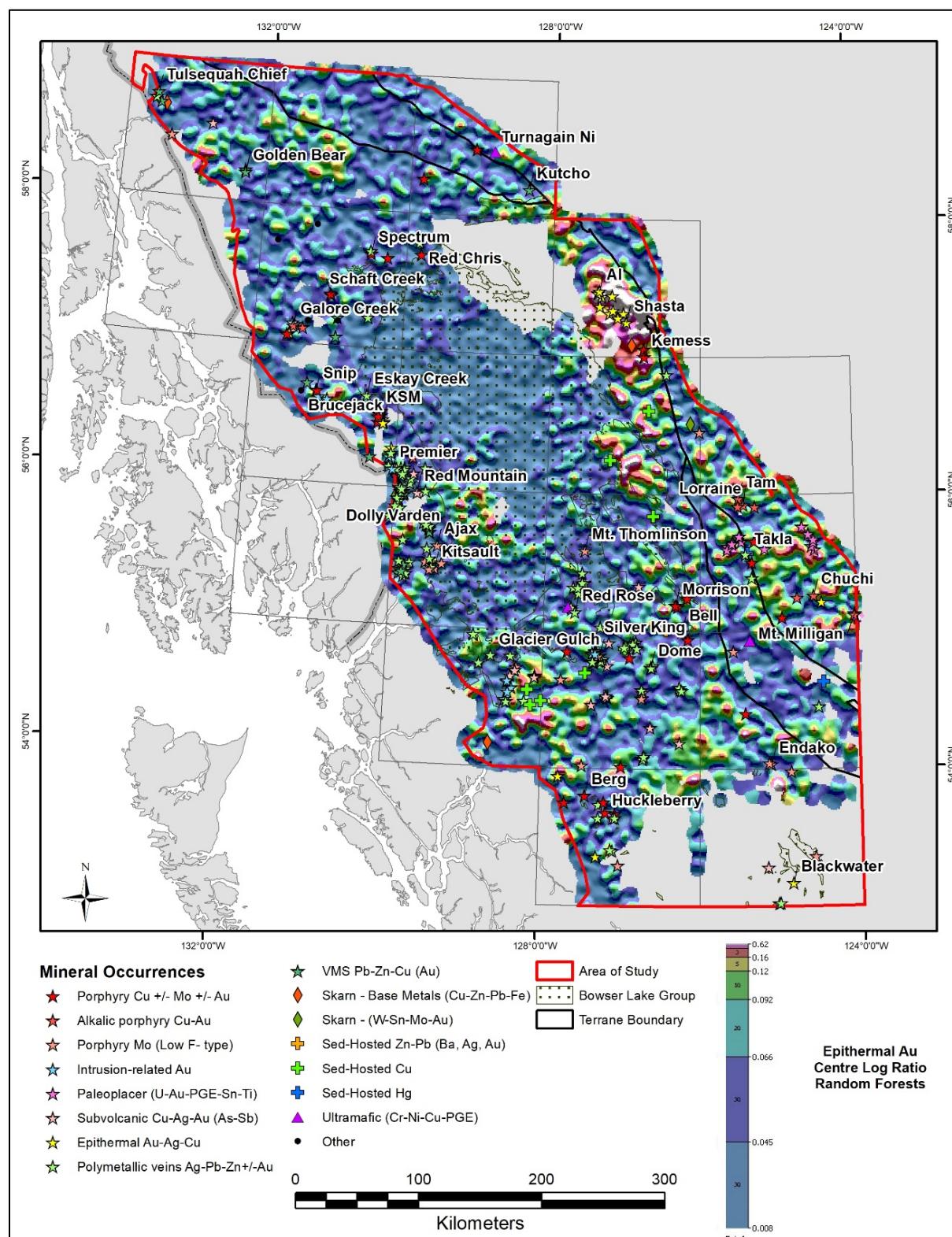


Figure 45: RF predictions for epithermal Au-Ag deposits

Note: Only data from samples in which no known mineral occurrences occur in their catchments are shown.

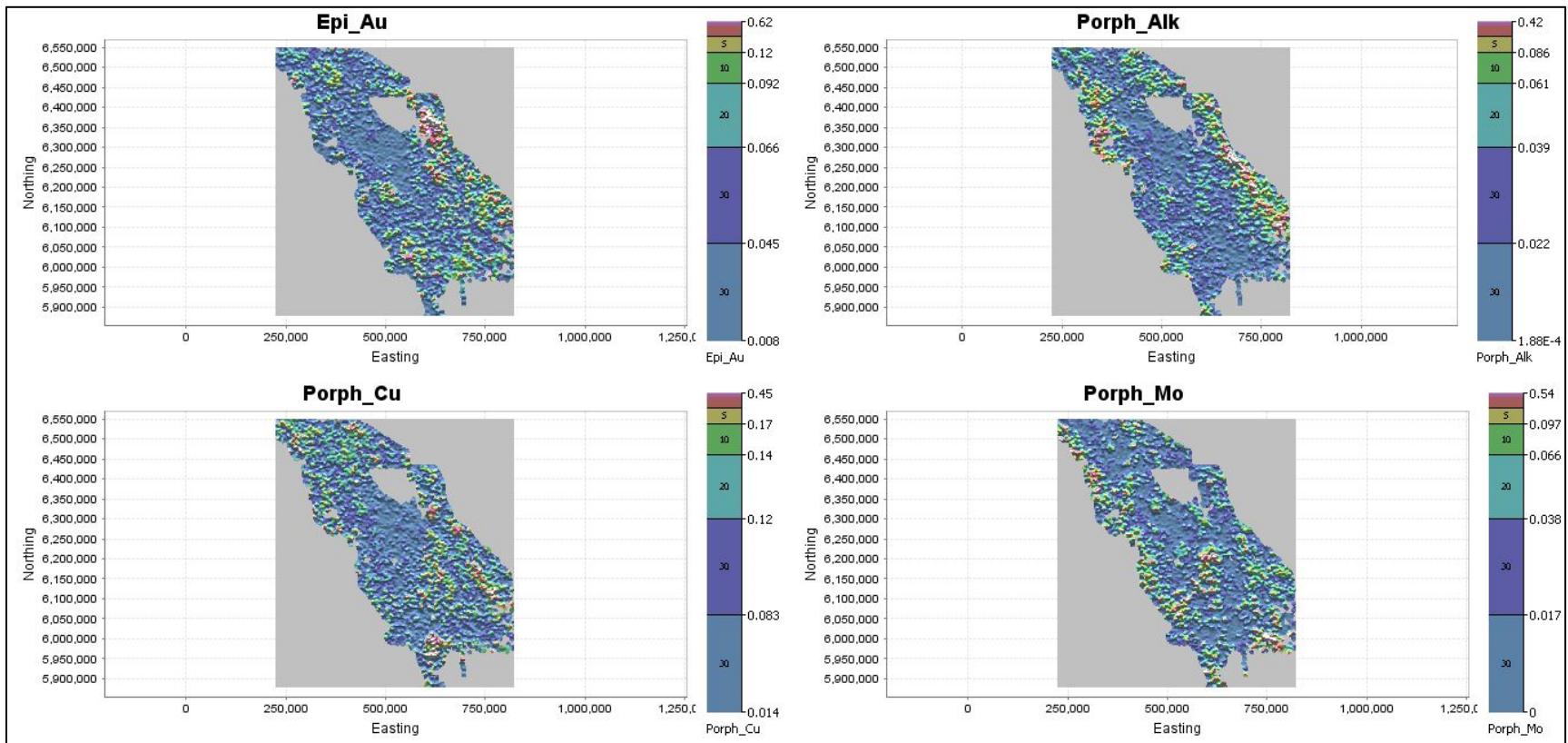


Figure 46: Comparison of RF predictions for epithermal Ag-Au and three types of porphyry style deposits

Note: Only data from samples in which no known mineral occurrences occur in their catchments are shown. Scale bars are percentile breaks for normalized votes.

3.7.2 Allocation/Typicality

Typicality indices are available for the following deposit types:

- Porph_Alk = Alkalic porphyry Cu-Au
- Epi_Au = Epithermal Au-Ag-Cu
- PMV = Polymetallic veins Ag-Pb-Zn±Au
- Porph_Cu = Porphyry Cu±Mo±Au
- Porph_Mo = Porphyry Mo (Low F- type)
- SubVolc = Subvolcanic Cu-Ag-Au (As-Sb)
- VMS = VMS Pb-Zn-Cu (Au).

An analysis of variance was carried out to determine which principal components (based on a CLR transform) are the best discriminators between the MINFILE deposit types. The best 11 principal components were used.

Using the R package “rgr”, estimates of covariance for each mineral deposit type were then calculated followed by a measure of the Mahalanobis distance from each sample site to each mineral deposit centroid. Using the degrees of freedom and the F-distribution with a confidence of .95, a measure of typicality was determined for each class. A site composition can belong to none, one or more than one classes based on the overlap of covariance of the classes and the similarity of the composition to these class covariances.

The typicality indices appear to show less selectivity for some deposit types, such as porphyry Cu-Mo-Au (Figure 47), epithermal Au-Ag (Figure 48), and polymetallic vein deposits compared to the RF predictions for the same deposit types. Polymetallic veins, for example, have widely distributed elevated typicality indices that extend into areas underlain by Bowser Basin sediments and are not geologically reasonable. This may reflect the mis-allocation of elevated metals associated with the Bowser Basin sediments given the broad range of metals associated with this broad class of deposit type. In contrast, porphyry Mo, alkalic porphyry Cu-Au, sub-volcanic precious metal and VMS base metal deposits show more restricted typicality indices compared to random forests predictions.

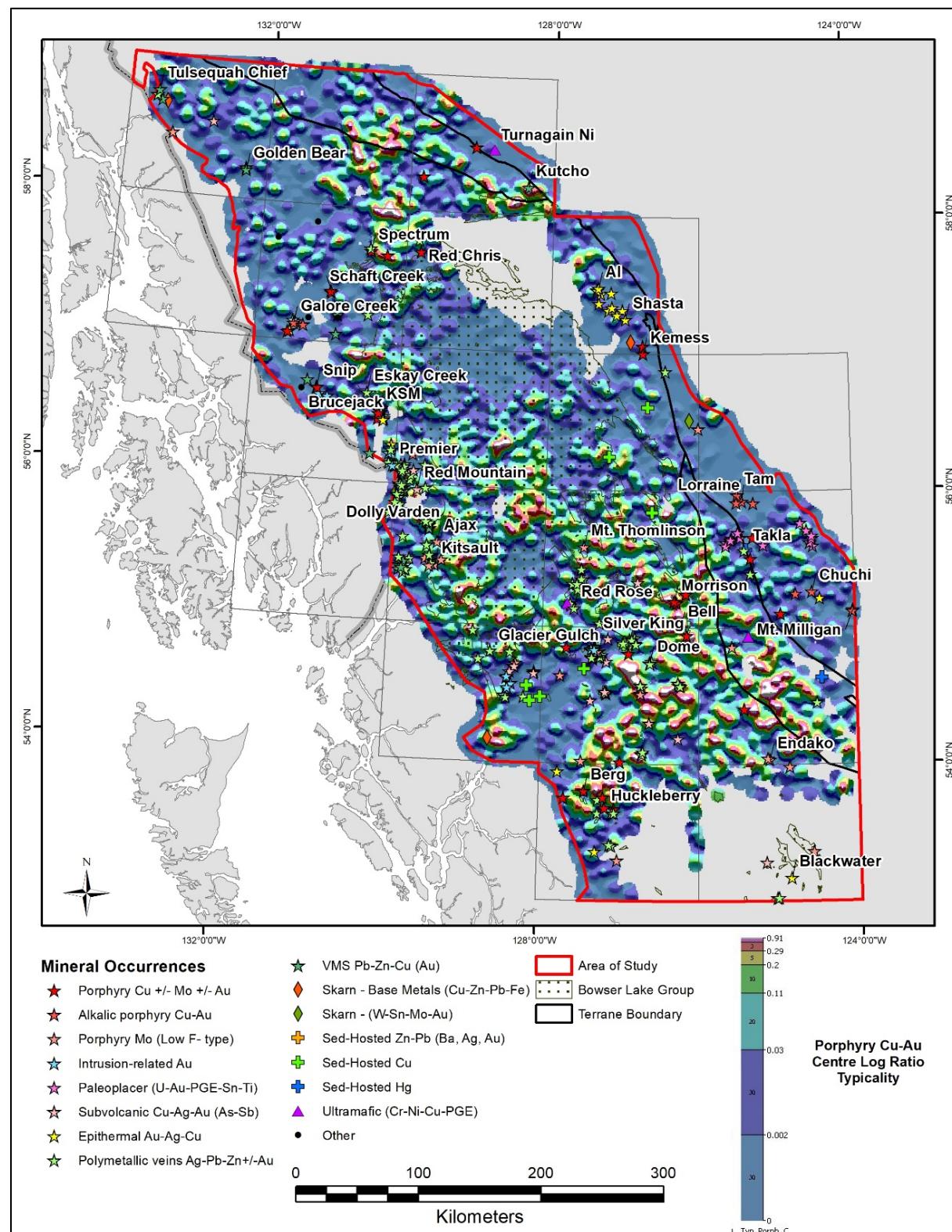


Figure 47: Indices of typicality for porphyry Cu-Mo-Au deposits

Note that it is inherent in the data that there are numerous values with "0" indices for some deposit types.

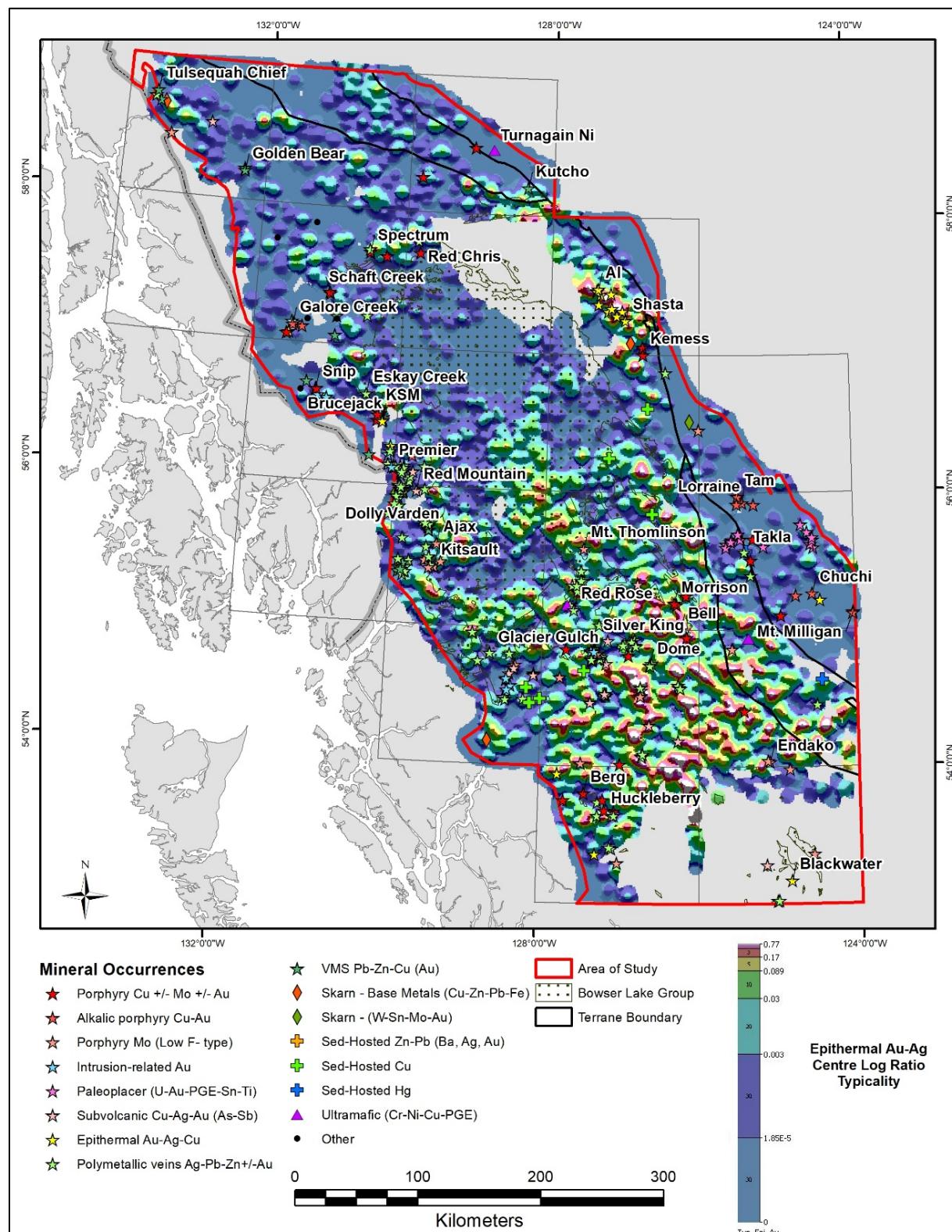


Figure 48: Indices of typicality for epithermal Au-Ag deposits

Note that it is inherent in the data that there are numerous values with "0" indices for some deposit types.

3.7.3 Residuals After Multiple Regression Analysis

Residuals from multiple regression of Log₁₀-transformed geochemical data against the proportion of different lithological units within each catchment basin were analysed to test the effects of using data levelled for the effects of lithology on advanced analytical methods. The residuals from multiple regression analysis were selected for this purpose given that the various approaches to data levelling produced similar outcomes and the multiple regression method is one of the most comprehensive classical approaches (e.g. Bonham-Carter *et al.*, 1987; Carranza, 2009). The residuals underwent PCA prior to their use in typicality and RF calculations.

The resulting principal components were also assessed to determine if they provided a better response to known mineralization compared to PCA of CLR data discussed in Section 3.3. Examination of a SCREE plot (not shown) indicates that the first eight principal components account for the bulk of the variation (70%) in the data. The R-scores are more difficult to interpret given some correction of lithological effects prior to PCA, but some lithological influence remains in the data, possibly because the bedrock geology is not accurate at the catchment basin scale, or perhaps due to the multiple regression method not adequately accounting for the lithological variation. Nonetheless, the geochemical response to both mineralization and metal scavenging appears to exert a stronger influence on the resulting principal components.

A summary of R-scores for individual principal components is summarized in Table 7. Some of the elemental associations are difficult to interpret as they appear to represent a mix of lithological, metal scavenging and mineralization effects. However, negative loadings on the first two principal components display strong mineralization signatures that were only evident in PC1 when using the CLR data without levelling for the effects of lithology. Percentile gridded images for inverse PC1 and inverse PC2 are presented in Figure 49 and Figure 50.

Table 7: Summary of R-scores for PCA of residual following multiple regression analysis

Principal component	Element association	Interpretation
+PC1	Th, Ti, U, Na, La	Felsic lithology
-PC1	Te, As, Mo, Se, Ag, Cu, Co, Cd, Zn	Mineralization +/- lithology
+PC2	Mg, V, Cr, Ga, Al, Co	Mafic lithology (Bowser Basin)
-PC2	Sb, Cd, Tl, Mo, Ag, Pb, Bi	Mineralization
+PC3	U, La, Na, Ti, Th, K, Sr	Felsic lithology
-PC3	Ni, Cr, Sb, Co, Mg, Hg	Ultramafic lithology
+PC4	Ca, Sr, S, Se, Hg, P, Ba	Metal scavenging/carbonate lithology
-PC4	Th, Bi, K, Pb, Fe, Co	Mixed lithology, metal scavenging, mineralization
+PC5	Tl, Ba, Ni, Hg	Uncertain
-PC5	Au, Fe, Te, V, Cu, As	Mineralization
+PC6	Cr, Au, Ni, K, Se, S, Cu	Lithology/mineralization
-PC6	Mn, Sc, Fe, Zn, Ba, Pb	Metal scavenging
+PC7	La, Sb, P, Cr, Ti, Hg, Ni	Uncertain
-PC7	Te, K, Bi, Sr, Al, Ba, Na	Mineralization/lithology
+PC8	Al, Ga, Sc, U, Ti, Mo, Cd, Zn, Ag	Lithology/mineralization
-PC8	Ba, Th, K, Sb, Sr, As, Ca	Mineralization/lithology

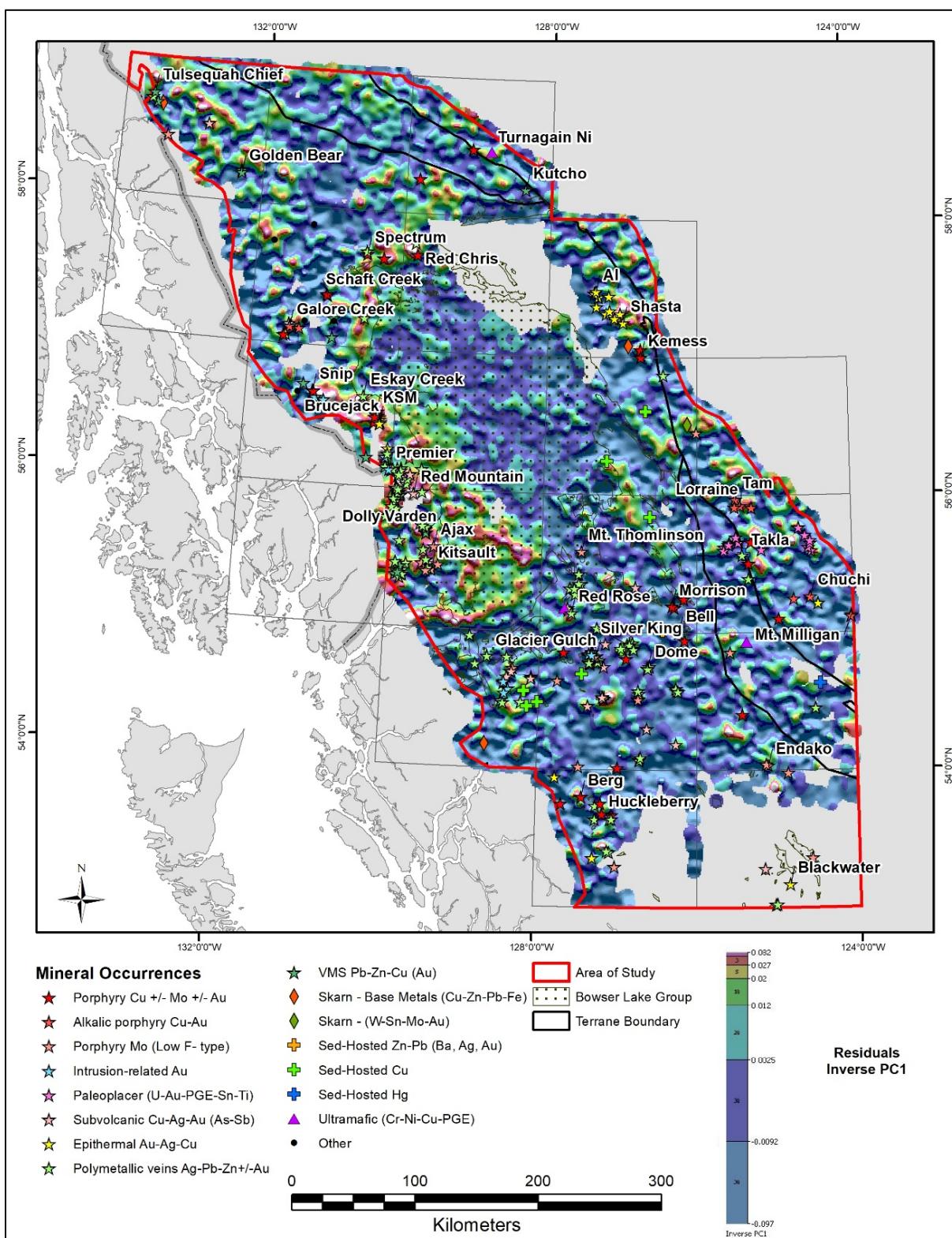


Figure 49: Inverse PC1 residuals following multiple regression analysis against lithology

Note: Elevated values reflect the presence of porphyry Cu-Au mineralization, as well as having an elevated response in the southwest Bowser Basin.

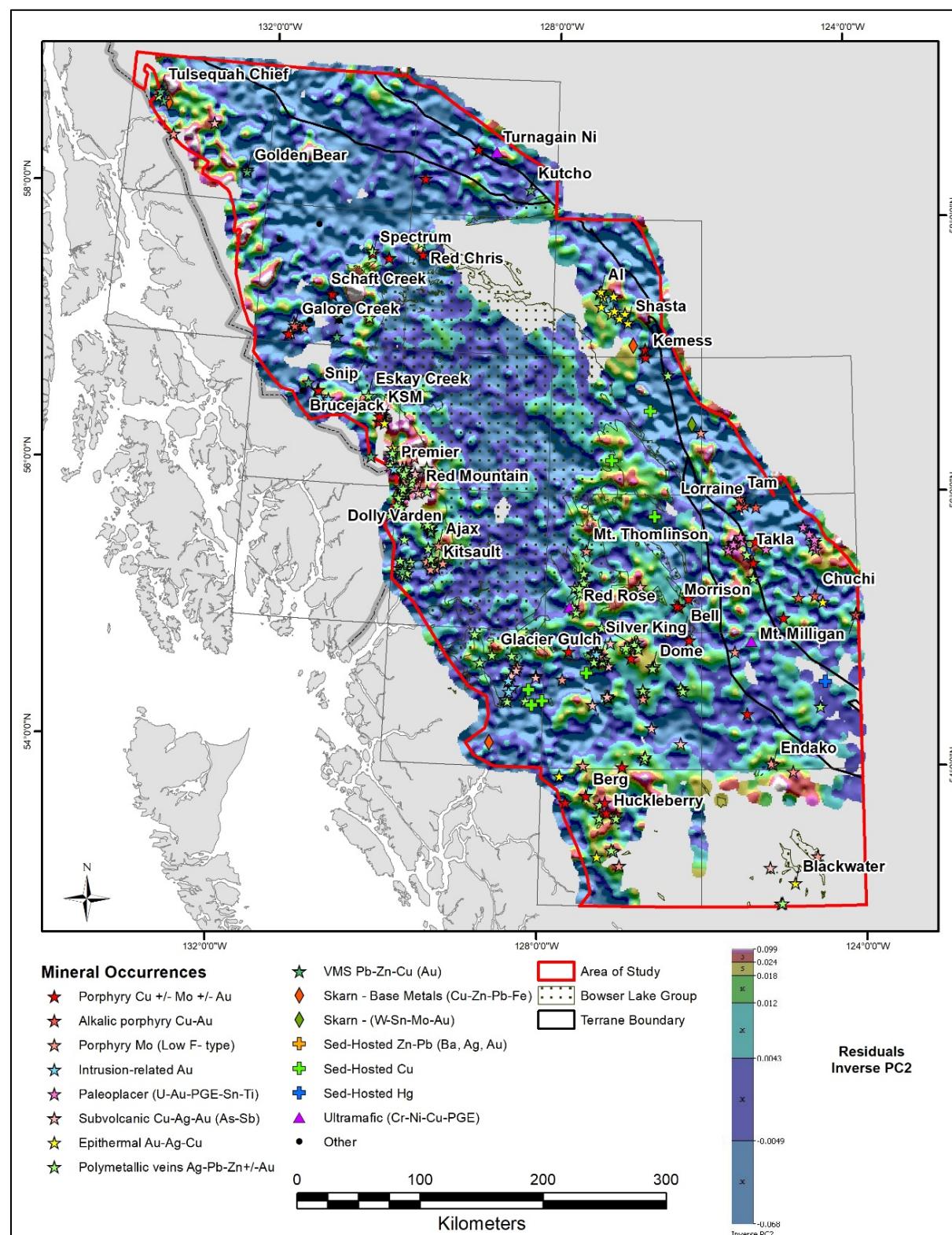


Figure 50: Inverse PC2 residuals following multiple regression analysis against lithology

Note: Elevated values reflect the presence of porphyry Cu-Mo-Au mineralization without the influence of lithology.

The residual principal components have been used for the advanced analytical methods to see if prior levelling for lithological variation might reduce mis-allocation of data to mineral deposit predictions. In the case of random forests predictions for alkalic porphyry Cu-Au (Figure 51) and epithermal Au-Ag deposits (Figure 52), the use of residuals following multiple regression analysis against catchment lithology generates similar predictions to those using CLR transformed data, suggesting that the random forests calculations can identify multivariate signatures associated with mineralization and largely filter out the effects of lithology. In detail, use of the residuals has resulted in predictions for alkalic porphyry Cu-Au deposits within the Bowser Basin that are not geologically reasonable, suggesting that artefacts have been introduced into the process and the predictions have been somewhat degraded through their use.

A similar conclusion is reached where residuals were used for allocation. The typicality indices are similar for both the porphyry Cu-Au (Figure 53) and epithermal Au-Ag (Figure 54) deposits, but there are a number of high typicality indices in the Bowser Basin that are not geologically reasonable predictions. It appears likely therefore that the use of residuals following multiple regression analysis against catchment lithology has degraded the outcome compared to the use of CLR data, possibly due to the presence of metal scavenging that may have been emphasized in the residuals once the effects of lithology were minimized.

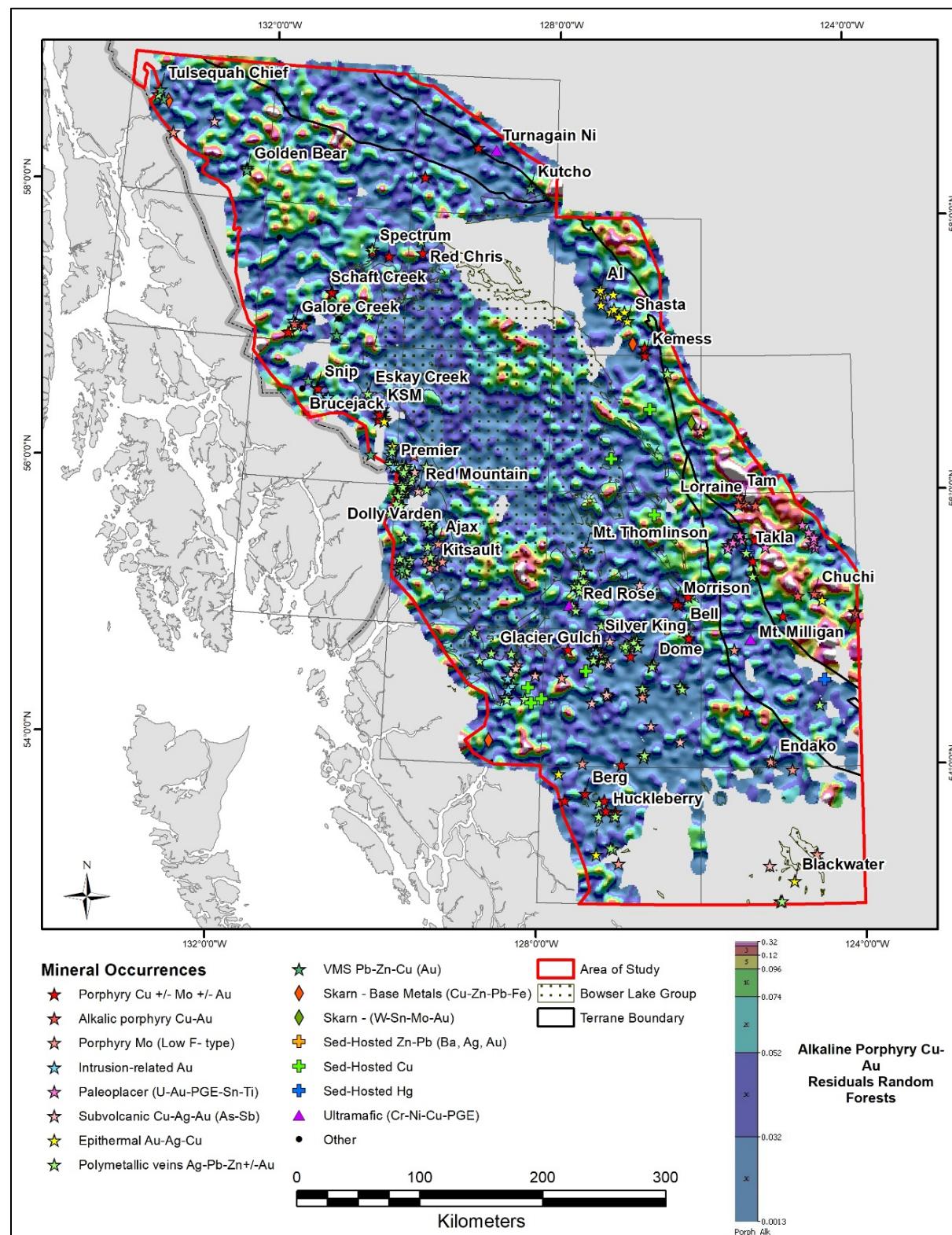


Figure 51: RF predictions for alkalic porphyry Cu-Au deposits using residuals following multiple regression against catchment lithology

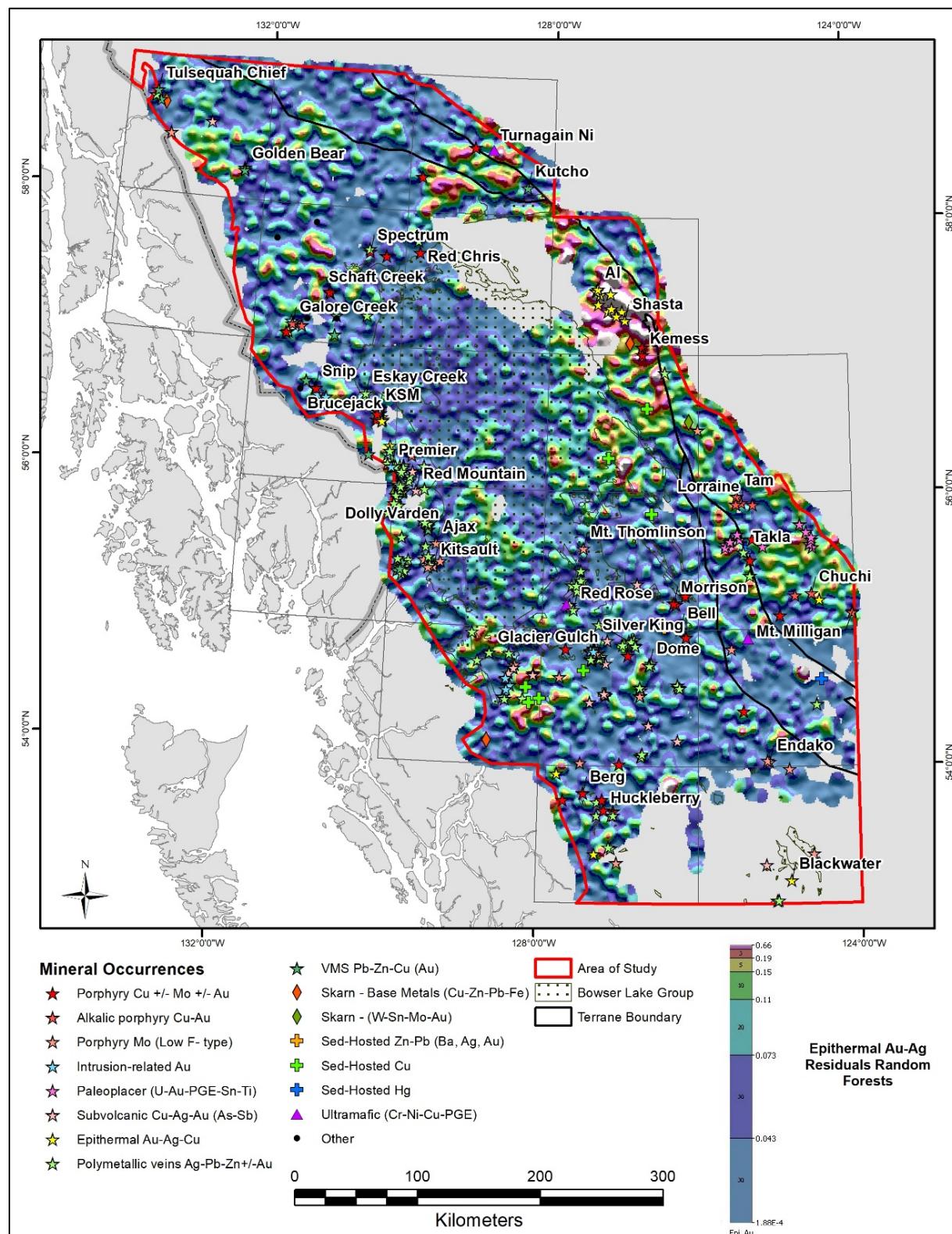


Figure 52: RF predictions for epithermal Au-Ag deposits using residuals following multiple regression against catchment lithology

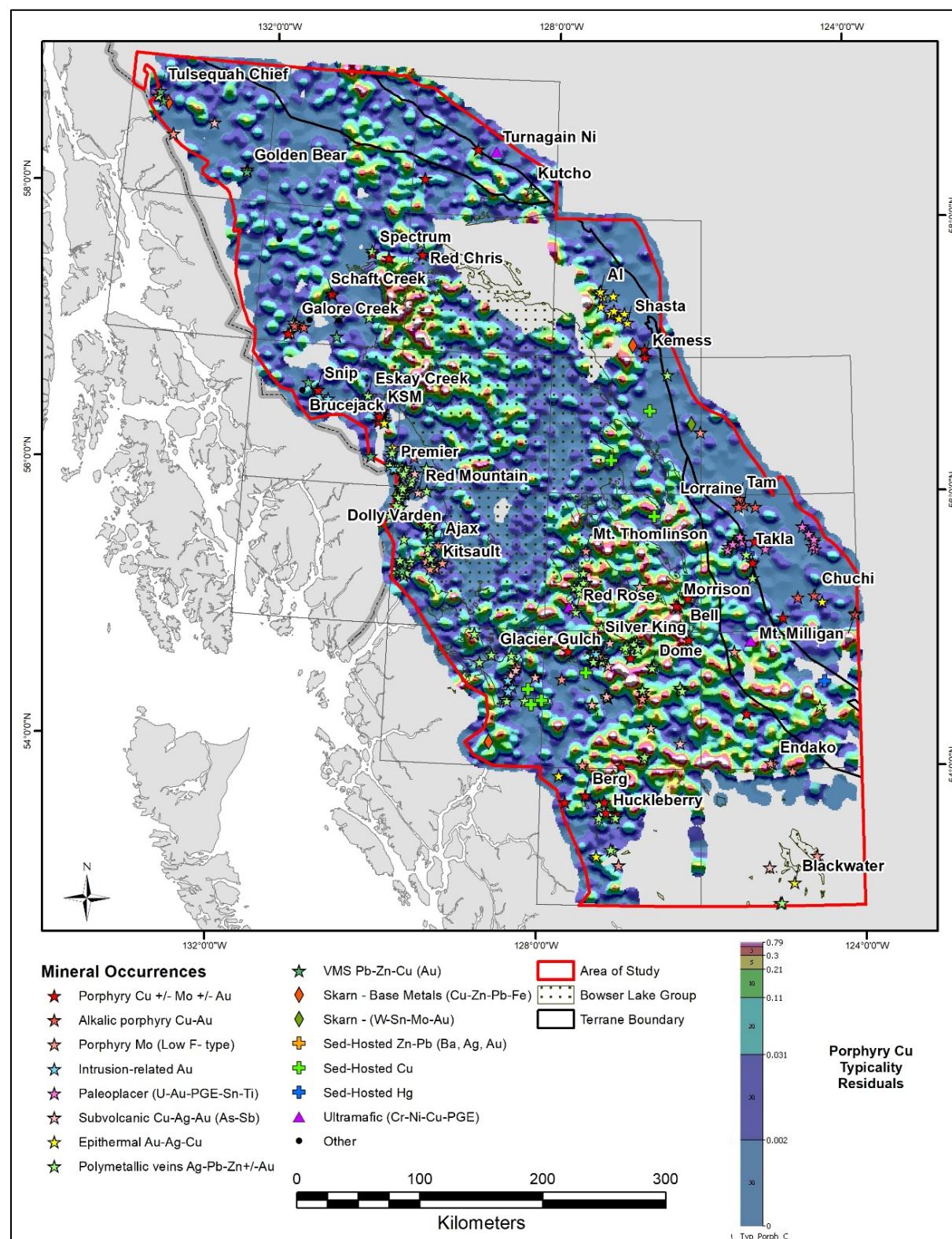


Figure 53: Typicality indices for porphyry Cu-Au deposits using residuals following multiple regression against catchment lithology

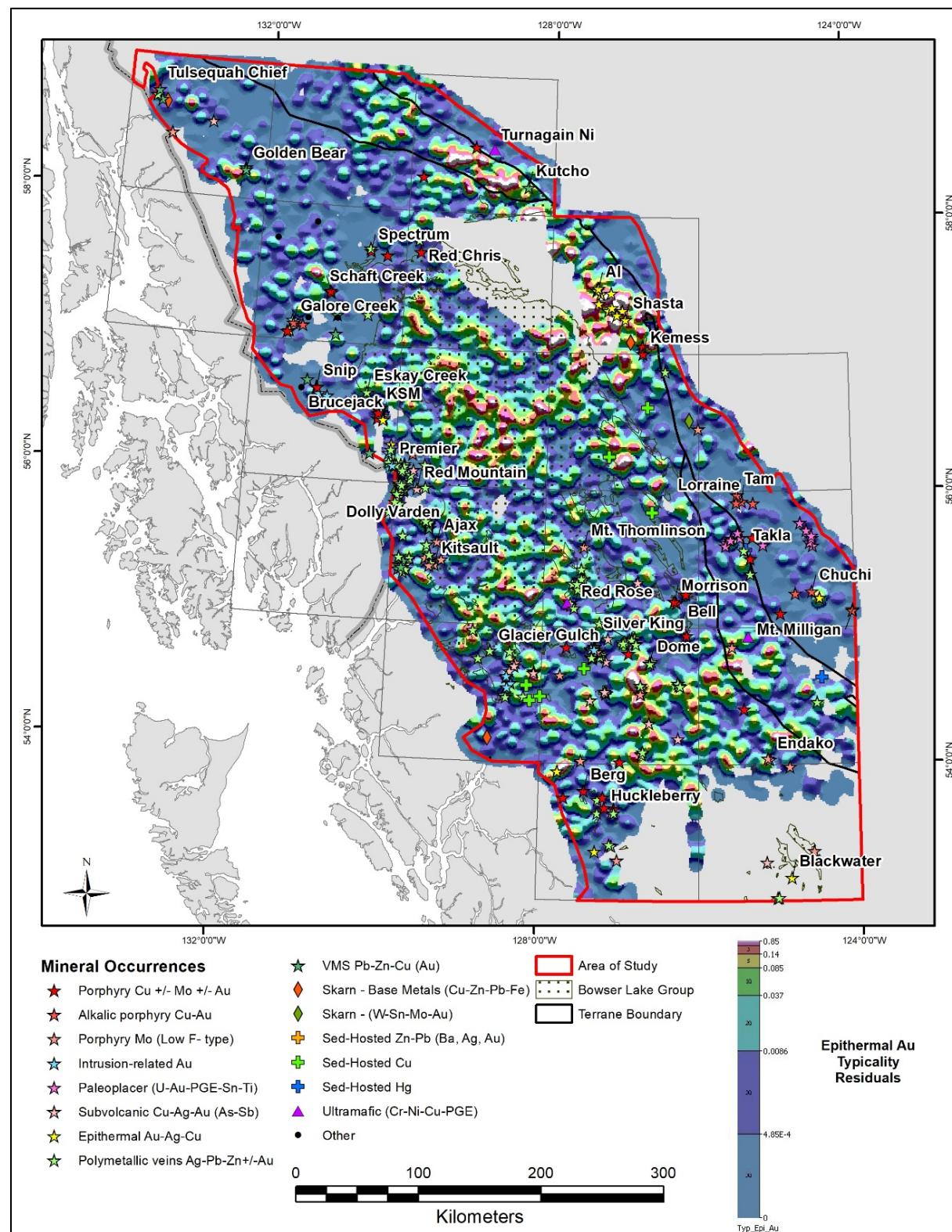


Figure 54: Typicality indices for epithermal Au-Ag deposits using residuals following multiple regression against catchment lithology

4 Summary and Discussion

There are several approaches to the processing of stream sediment data to correct for the effects of variable bedrock lithology that might interfere with the geochemical response of mineralization within catchment basins. Many of these classical approaches have previously been described in the literature (see Section 1.3), but all assume that the geology of the catchment basins from which stream sediment was derived is accurately known and that each lithological unit contributes sediment proportional to its area within the catchment. Additional corrections for the effects of metal scavenging and sediment dilution can also be made to geochemical data levelled for variable lithology. Several classical corrections to the data have been made for porphyry Cu-Au WSM in this study but only result in an improvement over the use of raw geochemical data where accompanied by a dilution correction. This outcome reflects the presence of a strong primary mineralization signature within the data and minimal effects from lithology on the elements used to construct the WSM.

An alternative to these classical approaches is to undertake PCA of the stream sediment data following a CLR transformation of the data to account for the effects of geochemical closure (Aitchison, 1986; Buccianti and Grunsky, 2014). This approach allows the influence of various natural processes on the geochemical data to be divorced from its spatial location, as well as the uncertainties associated with interpreted geology and assumptions regarding weathering and erosion. Principal components can sometimes be difficult to interpret, but typically lithological effects dominate in regional datasets (e.g. Grunsky *et al.*, 2009; Grunsky, 2010), with metal scavenging and mineralization providing secondary influences (Bonham-Carter and Goodfellow, 1986). The northwest BC stream sediment dataset presented in this report is unusual in that a strong mineralization signature is present in the dominant principal components. This allows the principal components to be used as direct indicators of mineralization in WSMs. Further refinement of the WSMs has been obtained by subtracting the influence of lithology evident in some principal components. Individual elements have also been regressed against principal components that are interpreted to represent either lithological or metal-scavenging controls prior to being used in WSMs. A statistical assessment of the proportion of true positives predicted by classical levelling approaches and the use of principal components indicates that principal components provided a superior result for the porphyry Cu-Au WSMs in respect to slightly reducing the proportion of false positives. A better response might be obtained for the classical data levelling methods where the weightings for levelled data are optimized for each method.

A statistical comparison of performance of WSMs for porphyry Cu-Au and epithermal Au-Ag deposits illustrates the importance of mineral deposit type in influencing how effective the various data processing methods will be. Whether the various methods discussed here will produce an improvement in WSMs for mineral exploration targeting depends upon the extent to which the elements used to construct the models are influenced by lithology or metal scavenging. As Cu is strongly weighted in the porphyry Cu-Au WSMs and shows some degree of lithological control, processing of the data to remove the effect of lithology has a positive outcome in the predictability of the model when accompanied by a dilution correction. By contrast, WSMs in which Au is strongly weighted (e.g. epithermal Au-Ag) are not significantly improved by data levelling methods to account for lithological or metal scavenging effects. Correction for the effects of dilution based on catchment area had the greatest overall impact on the porphyry Cu-Au WSMs by increasing the number of observed true positives based on known mineral occurrences. Clearly, a standard approach to data processing will have different outcomes for different mineral deposit types having a range of element associations and the methodology used must be evaluated on its merits for each situation. EDA is required to gain an understanding of processes that have influenced the composition of stream sediment samples within a particular project area.

WSMs are expert-driven as the parameters used and the weightings applied are selected by the user based on experience and, in this instance, by qualitative comparison to known mineral occurrences. The approach therefore entails a degree of subjectivity. Alternative treatments of the data using alternative analytical methods, including IAC, RF and allocation/typicality computations, were also investigated. ICA seeks to isolate individual geochemical signals within the multivariate “noise” and was able to isolate distinct lithological units as well as a hydrothermal magmatic mineralization signal that predicted known epithermal Au-Ag mineralization better than the classical data treatments that were employed. Allocation and RF are supervised multivariate techniques that require the presence of known deposits (or classes) to define a multivariate signature for mineralization, but the subjectivity of the observer is removed, and the outcomes are entirely data-driven. One limitation of the supervised approaches is that the known occurrences must be consistently and correctly attributed to their respective classes. For this reason, only Past Producers, Advanced Prospects and Prospects were used for supervised learning as mineral occurrences in more advanced stages of exploitation are likely to be well understood in terms of mineral deposit classification. There must also be sufficient members of each class to define the covariance within the data for that class. The number of deposit types included in the allocation/typicality computations is limited for this reason.

One conclusion of this investigation is that allocation/typicality and RF computations produce significantly different results for most mineral deposit types. Predictions from the two methods could not be compared quantitatively in the same way that different processing methods for data levelling were because in both instances all known mineral occurrences were used for training, but the RF outputs are more geologically reasonable. Mineral deposit types having relatively common base metal associations such as the polymetallic veins are over-represented by typicality indices but are less prominent in RF predictions. In the case of the typicality indices, some of the elevated values occur within the Bowser Basin and so may reflect the mis-allocation of lithologically-controlled data to mineral deposit signatures. An attempt to filter this effect by using geochemical data already levelled for the effects of lithology proved counterproductive – more geologically reasonable outcomes were produced using the CLR transform data. It appears that the random forests computations distinguish the geochemical signature for mineral deposits from those produced by lithological variations, even where these variations involve commodity and pathfinder elements characteristic of some mineral deposits.

Deliverables from this project include a series of shape files containing a compilation of re-assay and historical geochemical data for individual stream sediment samples linked to their catchment basin polygons. Included in various shapefiles are computational outputs, including centred-log ratio values, principal components, WSMs, independent components, levelled scores, residuals from regression analysis, random forests normalized votes and typicality indices. Percentile classes at 50%, 70%, 90%, 95% and 98% have also been supplied for the most relevant outputs. The catchments have also been attributed with known mineral occurrences, with the highest-ranking occurrence class indicated. There is much more information available than has been presented in this report, and sufficient data are provided for the user to generate their own models, gridded images and maps for areas of specific interest. Further refinements in the interpretation of the data are possible and a more rigorous statistical assessment of outcomes required. A description of fields included in the various data files is provided in [Appendix 2](#).

5 Recommendations

An important contribution of this project has been the validation of sample locations so that catchment basins can be defined and attributed with bedrock rock and Quaternary geology. However, CSA Global makes no warranty that the validated sample locations are correct. They have been located using a “best efforts” approach and users of the data should satisfy themselves that the locations of individual samples of interest to them are accurate.

A qualitative assessment of dilution suggests that catchment basins $>25 \text{ km}^2$ may not have been effectively sampled. This includes approximately 8% of the samples but, as the catchment basins involved are large, they represent much more than 8% of the project area. When combined with areas that were not initially sampled, this represents a significant area in a highly prospective region that may have been under-explored. There is scope for further sampling in some areas. However, this general statement must be tempered by examination of the density and effectiveness of sampling in specific areas of interest.

Many of the anomalous geochemical scores generated by classical methods are related to known mineral occurrences. The implication is that similar responses not related to known occurrences may be due to undiscovered mineralization. Other explanations for anomalous geochemical responses exist, including the possibility of incorrectly located samples, contamination, or the presence of volumetrically small but geochemically significant lithological units, such as metalliferous black shales, that geochemically mimic the signal associated with known mineral deposits. The authors make no representation that geochemical anomalies reflect undiscovered mineral deposits.

Of the different WSMs produced during this investigation, only the porphyry Cu-Au models that have been corrected for dilution show a positive response in predicting the presence of known mineral deposits. None of the WSMs generated for epithermal Au-Ag deposits resulted in improved outcomes and the WSMs generated using raw data for this deposit type appears to be as, if not more, effective than the models generated using data levelled for the effects of lithology and/or metal scavenging. Although not tested as extensively as the porphyry Cu-Au and epithermal Au-Ag deposits, the WSMs for magmatic Ni-Cu and polymetallic vein deposits generated using data regressed against principal components and corrected for dilution are likely to be the best predictors of prospective catchment basins for these deposit types. Both deposit types are both strongly influenced by bedrock geology and are therefore likely to be improved by correction for these effects. However, sufficient data are provided to allow the user to generate their own WSMs with different weightings that may be more effective within a localized area.

As one of the stated aims of this project was to investigate different classical data analysis methods and alternative data analysis techniques, it has resulted in a plethora of interpretive products that may appear confusing and at times conflicting. What we are looking for in the data are consistent patterns of anomalism that can be related either to lithology, possible metal scavenging, or known mineral deposits. The consistent presence of anomalous samples irrespective of the interpretative method used should be taken as a good indication that the anomaly is real and worthy of further investigation. Elevated values that occur in only one interpretive product should be treated with caution as they likely represent analytical artefacts.

For all the reasons listed previously, the authors recommend that users of the data conduct their own follow-up sampling and analysis to confirm the geochemical anomalies identified by this work.

6 Abbreviations and Acronyms

AAS	atomic absorption spectroscopy
BC	British Columbia
BCGS	British Columbia Geological Survey
CLR	centred-log ratio
DEM	digital elevation model
EDA	exploratory data analysis
GIS	Geographic Information System
GPS	global positioning system
GSC	Geological Survey of Canada
ICA	independent component analysis
ICP-MS	inductively coupled plasma–mass spectrometry
ICP-AES	inductively coupled plasma–atomic emission spectrometry
INAA	instrumental neutron activation analysis
km	kilometre
km ²	square kilometres
LOI	loss on ignition
NGR	National Geochemical Reconnaissance
NTS	National Topographic System
OOB	out-of-bag
PCA	principal component analysis
RF	Random Forests
RGS	regional geochemical survey
VHMS	volcanic-hosted massive sulphide
WSM	weighted sum model
YGS	Yukon Geological Survey
TRIM	terrain resource information management

7 Acknowledgements

The authors wish to thank Geoscience BC for their funded support of this project. The cost of re-sampling in this region is high and their efforts to re-analyse archived material and to provide enhanced interpretations of the new data in a form that is readily accessible are appreciated. Yao Cui of the BCGS kindly provided the catchment basins based on our validated sample locations. Wayne Jackaman supplied the quality control geochemical data collected during the re-analysis programs. Adrian Martinez Vargas of CSA Global undertook the peer review of the report prior to its release. We hope that this work ultimately contributes to new mineral discoveries in the region.

8 References

- AITCHISON, J., 1986. The Statistical Analysis of Compositional Data. Methuen, New York.
- ARNE, D.C. & BROWN, O., 2015. Catchment analysis applied to the interpretation of new stream sediment data from northern Vancouver Island, Canada (NTS 102I and 92L). Geoscience BC Report 2015-4, 41 p
- ARNE, D. & MACFARLANE, B., 2014. Reproducibility of gold analyses in stream sediment samples from the White Gold District and Dawson Range, Yukon Territory, Canada. *Explore*, **164**, 1-10.
- ARNE, D.C. & BLUEMEL, E.B., 2011. Catchment analysis and interpretation of stream sediment data from QUEST South, British Columbia, Geoscience BC, Report 2011-5, 25 p
- BONHAM-CARTER, G.F & 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*, **25**, 139-155.
- BONHAM-CARTER, G.F., ROGERS, P.J. & ELLWOOD, D.J., 1987. Catchment basin analysis applied to surficial geochemical data, Cobequid Highlands, Nova Scotia. *Journal of Geochemical Exploration*, **29**, 259-278.
- BREIMAN, L., 2001. Random Forests. *Machine Learning*, **45**, 5 -32.
- BUCCANTI, A. & GRUNSKY, E., 2014. Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes? *Journal of Geochemical Exploration*, **141**, 1-5.
- CARRANZA, E.J.M., 2009. Catchment analysis of stream sediment anomalies. In: Hale, M., (editor), *Handbook of Exploration and Environmental Geochemistry*, **11**, 115-144.
- 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*, **60**, 157-171.
- COOKENBOO, H.O., 1993. Lithofacies, provenance, and diagenesis of Jura-Cretaceous strata of the northern Bowen Basin, British Columbia. Unpublished PhD thesis, University of British Columbia, 234 p
- CUI, Y., 2010. Regional geochemical survey: Validation and refitting of stream sample locations. In: Geological Fieldwork 2010, BC Ministry of Forests, Mines and Lands, Paper 2011-1, 169-179.
- CUI, Y., ECKSTRAND, H. & 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, Paper 2009-1, 231-238.
- COLPRON, M., NELSON, J.L. AND MURPHY, D.C., 2006, A tectonostratigraphic framework for the pericratonic terranes of the northern Canadian Cordillera, In Paleozoic Evolution and Metallogeny of Pericratonic Terranes at the Ancient Pacific Margin of North America Canadian and Alaskan Cordillera, Colpron, M. and Nelson, J. (eds.), *Geological Association of Canada Special Paper* **45**, 1–23.
- FRISKE, P.W.B., JACKAMAN, W., LETT, R.E.W., MCCURDY, M.W., 2003, Regional stream sediment and water data, Fort Fraser, British Columbia. Geological Survey of Canada, Open File 1766.
- GARRETT, R.G., 1990. A robust multivariate procedure with applications to geochemical data. In Agterberg, F.P. & Bonham-Carter, G.F. (eds) Statistical Applications in the Earth Sciences. Geological Survey of Canada Paper 89-9, 309–318.
- GARRETT, R.G. & GRUNSKY, E.C., 2001. Weighted sums – knowledge based empirical indices for use in exploration geochemistry. *Geochemistry: Exploration, Environment, Analysis*, **1**, 135–141.
- GRANGER, D.E. & SCHALLER, M., 2014. Cosmogenic nucleides and erosion at the watershed scale. *Elements*, **10**, 369-373.
- GRUNSKY, E.C., 1991. Numerical methods, techniques and strategies for the evaluation and interpretation of geochemical data; Curtin University Perth, Australia, June, 1991, 89 p., Short Course. Handbook.

- GRUNSKY, E.C., 2010. The interpretation of geochemical survey data. *Geochemistry: Exploration, Environment, Analysis*, **10**, 27-74.
- GRUNSKY, E.C., DREW, L.J. & SUTPHIN, D.M., 2009, Process recognition in multi-element soil and stream-sediment geochemical data. *Applied Geochemistry*, **10**, 1602-1616.
- GRUNSKY, E.C., MUELLER, U.A., CORRIGAN, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. *Journal of Geochemical Exploration*, **141**: 15-41.
- GUNNING, M.H., HODDER, R.W. AND NELSON, J.L., 2006, Contrasting volcanic styles and their tectonic implications for the Paleozoic Stikine assemblage, western Stikine terrane, northwestern British Columbia, In Paleozoic Evolution and Metallogeny of Pericratonic Terranes at the Ancient Pacific Margin of North America Canadian and Alaskan Cordillera, Colpron, M. and Nelson, J. (eds.), *Geological Association of Canada Special Paper* **45**, 201-227.
- HARRIS, J. R., & GRUNSKY, E. C., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers & Geosciences*, **80**, 9-25.
- HAWKES, H.E., 1976. The downstream dilution of stream sediment anomalies. *Journal of Geochemical Exploration*, **6**, 345-358.
- HEBERLEIN, D.R., 2013. Enhanced interpretation of RGS data using catchment basin analysis and weighted sums modelling: examples from map sheets NTS 105M, 105O and part of 105P. Yukon Geological Survey, Open File 2013-15, 18 p
- HYVARINEN, A. & OJA, E., 2000. Independent Component Analysis: Algorithms and Applications, *Neural Networks*, **13**, 4-5.
- MACKIE, R.A., ARNE, D.C. & BROWN, O., 2015. Enhanced interpretation of regional stream sediment geochemistry from Yukon: catchment basin analysis and weighted sums modelling. Yukon Geological Survey, Open File 2015-10, 9 p
- MACKIE, R.A., ARNE, D.C. & PENNIMPEDE, C., 2017. Assessment of Yukon regional stream sediment catchment basin and geochemical data quality; Yukon Geological Survey, Open File 2017-4, 36 p
- MOON, C.J., 1999. Towards a quantitative model of downstream dilution of point source geochemical anomalies. *Journal of Geochemical Exploration*, **65**, 111-132.
- NELSON, J.L., COLPRON, M. AND ISRAEL, S., 2013, The cordillera of British Columbia, Yukon and Alaska: tectonics and metallogeny; In Tectonic, Metallogeny, and Discovery: The North American Cordillera and Similar Accretionary Settings, Colpron, M. Bissig, T., Rusk, B.G. and Thompson, J.F.H. (eds.), *Society of Economic Geology Special Publication* **17**, 53-109.
- PALAREA-ALBALADEJO, J., MARTÍN-FERNÁNDEZ, J.A., & BUCCANTI, A., 2014. Compositional methods for estimating elemental concentrations below the limit of detection in practice using R. *Journal of Geochemical Exploration*, **141**, 71-77.
- PAN, G. & HARRIS, D.P., 1990. Quantitative analysis of anomalous sources and geochemical signatures in the Walker Lake quadrangle of Nevada and California. *Journal of Geochemical Exploration*, **38**, 299-321.
- RICKETTS, B.D., EVENCHICK, C.A., ANDERSON, R.G. AND MURPHY, D.C., 1992, Bowser Basin, northern British Columbia: constraints on the timing of subsidence and Stikine terrane-North American terrane interactions. *Geology*, **20**, 1119-1122.
- RUKHLOV, A. AND NAZIRI, M., 2015, Regional geochemical survey database 2015, British Columbia Geological Survey, Ministry of Energy and Mines, Geofile 2015-3.
- SMITH, R.E., CAMPBELL, N.A., AND LITCHFIELD, R., 1984. Multivariate Statistical Techniques Applied to Pisolitic Laterite Geochemistry at Golden Grove, Western Australia, *Journal of Geochemical Exploration*, **22**, 193-216.

9 Appendix 1: Advanced Analysis Methods

Random Forests Classification Algorithm

Random forests (RF) is an ensemble, multiple decision tree classifier that offers a number of advantages for mineral prospectivity modelling:

- Data (evidence layers) can be binary, categorical or continuous (in this study, the 12 evidence (raster) layers are all continuous).
- Performs internal cross-validation through bootstrapping which provides a robust estimate of classification accuracy through out-of-bag (OOB) estimates.
- Is a non-parametric classifier and is relatively insensitive to outliers in the training data.
- Requires minimal user input (m – the number of decision trees, and n – the number of variables for each decision tree).
- Produces a classification map showing permissive areas for exploration for each mineral class but more importantly, probability maps (strength of membership to each mineral class – in this study areas with higher prospectivity for Au and other commodities) that serve as the actual prospectivity maps.
- Ranks the input variables with respect to their importance in the predictions.

RF was originally developed by L. Breiman and A. Cutler at the University of California, Berkeley (Breiman, 2001). Training data (i.e. locations of mineral occurrences) are required for this approach, similar to other data-driven approaches. Input parameters into the RF classifier are minimal and include only the number of variables (in this study – evidence layers) for each tree and the number of trees to create. The number of variables is set to the square root of the total number of variables and the number of trees must be set through experimentation. The OOB error stabilizes at a given number of trees and this number is used. Each tree employs a bagging process (i.e. bootstrap sampling) where approximately two-thirds of the training areas (pixels) are randomly selected with replacement and these are used for generating the classification (in-bag data) and the remaining one-third (OOB) is used for validation. This random sampling with replacement of the training dataset is undertaken for every tree. In-bag data are used to create multiple decision trees which are applied to produce independent classifications and the OOB data is used to validate the classification by calculating an OOB error. Compared to cross-validation, the OOB error is unbiased and a good estimate for the generalization error. At each node of the individual decision tree, the best split is chosen from a random sample of variables. Each tree is grown to the maximum extent with no pruning (deletion of branches from each tree). We used the Gini Index to determine the impurity at each node:

$$\text{Gini Index} = 1 - \sum_c (p^2 (c|t))$$

Where:

- c = number of classes (e.g. Au, non-deposits)
- t = node of a tree
- p = relative frequency of c (a given class).

The stop criteria for splitting each node is based on the minimum of samples in a node (we used 1) and the minimum impurity in a node (we used 0 allowing full growth of the decision trees, i.e. no pruning).

An ensemble of trees (predictions) is created and a voting procedure is employed to assign the majority class to each pixel in the final prediction map. RF is less sensitive to noise or over-fitting and there is no need for cross-validation as it is performed internally (e.g. OOB). However, as with any supervised classification

method, an independent check training dataset of occurrences are still required to calculate an unbiased and more robust estimate of classification accuracy. In addition to the classification map generated by RF, a probability map is also generated that shows the strength of membership for each mineral prospect class. These probability maps and not the classified map itself serve as the mineral prospectivity maps for each mineral type and will be discussed below. Additionally, the probability of RF membership to each class can be used to evaluate the uncertainty of the RF classification.

Another very useful aspect of RF is it calculates the importance (predictive power) of each variable in the classification process. This is accomplished by the following process:

- For each tree, the OOB samples are permuted in the respective variable and then put down the tree and the number of correct classifications are calculated (nP).
- The in-bag training samples (original) are put down the tree and the number of correct classifications are calculated (nC).
- Calculate $Nc-Np$.
- The average of these differences of the accuracies for all trees is the raw importance of each variable. However, to provide a more robust estimate, the raw variable importance is divided by the respective standard deviation creating a normalized variable importance value.

A high normalized value has a high importance for the entire RF and vice versa for a low number.

The main point of ensemble classifiers such as RF is that the process produces not just one prediction (decision tree) but from many predictions which are then combined. This is extremely beneficial as this process will help to reduce the variance as the results are less dependent on peculiarities of a single training dataset. Furthermore, a more robust estimate of the overall classification accuracy is achieved.

In this study to perform the RF classification the R package, “randomForest” was used. The RF parameters we use in this study include n (number of variable to create each tree) which was set to the square root of the total number of variables and m (number of decision trees) through experimentation, which was set to 500.

Allocation /Typicality

The use of supervised classification schemes provides a means of obtaining a statistical measure of the distinction of different classes of data. Through the creation of reference groups, supervised methods calculate the statistical distinctiveness of each class. A consequence of the establishment and testing of the reference groups is the ability to test unknown samples for possible membership into none, one or more than one of the groups. This type of procedure is known as allocation.

Allocation requires an estimate of covariance within each class. Of the mineral deposit types listed in Section 2.7, intrusive-related Au, magmatic Ni-Cu and base metal skarns were deemed to contain too few known occurrences and were dropped from the process.

Allocation procedures (also known as classification procedures) and measures of typicality make it possible to predict the probability of an unknown sample belonging to a set of reference groups. The method assumes the variables are the same between the reference groups and the unknown sample(s). Allocation can also be used to test the group membership of samples used to create the reference groups in a supervised classification such as linear discriminant analysis. Garrett (1990) provides a useful summary of the methods of allocation.

Allocation procedures work on the basis of measuring the distance of a sample from each reference group centroid. By using the covariance estimates of the populations of samples used for the reference groups, the

Mahalanobis distance can be computed between each unknown sample and the group centroids. The covariance matrix contains the characteristics of the dispersion of the reference population. If the Mahalanobis distance of a sample, with respect to a particular reference group centroid, is within the dispersion matrix defined by the reference group covariances, then the probability of that sample belonging to that reference group is greater than zero. The closer a sample is to a group centroid, the higher the probability of membership, however, any sample with a probability >0 can be considered to be similar to the reference group. The dispersion between groups can overlap which results in some samples having probabilities of belonging to more than one reference group. Similarly, if a sample has a Mahalanobis distance outside the dispersion matrix then it has a zero probability of belonging to any of the reference groups, which is the definition of typicality. A significant test is the determination of whether or not the groups have similar covariances. If they do not, then comparison of samples must be modified according the different covariance matrices.

Posterior Probability

Allocation of individual samples through the computation of posterior probabilities is described briefly in the following discussion.

For the estimation of posterior probability of membership in the kth population is given by:

$$pr(k; \bar{x}_m) = p_k f(x_m; P_k) / \sum_{j=1}^g p_j f(\bar{x}_m; P_j)$$

Where:

- $pr(k; \bar{x}_m)$ is the posterior probability of membership in the kth group with group vector (centroid) \bar{x}_m
- p_k is the prior probability that the unknown sample belongs to the kth group
- $f(\bar{x}_m; P_k)$ is the value of the probability density function for the group vector \bar{x}_m .

The posterior probability is the probability of belonging to the kth group divided by the sum the probabilities belonging to all g groups. An unknown sample is then assigned to the reference group with the highest posterior probability (smallest Mahalanobis distance).

Index of Typicality

Given a covariance (dispersion) matrix S , k groups, with samples composed of p variables, the D^2 generalized Mahalanobis distance is calculated such that:

$$D_k^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_{ik})' S_k^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{ik})$$

Where:

- k is the kth group
- i is the ith sample.

For the index of typicality, the unknown sample is provisionally allocated to the jth of the g groups such that:

$$D_j^2 + \ln |S_j| = \min[D_k^2 + \ln |S_k|]$$

Where:

- $k=1,\dots,j,\dots,g$ reference groups
- $|S_j|$ is the determinant of the covariance of the jth group.

If the covariances of the groups are equal, then the natural logarithmic term is dropped.

The probability of group membership of each sample is predicted for each reference group using the statistic:

$$F = \frac{(N - g - p + 1) n_k}{p (N - g) (n_k + 1)} D_k^2$$

Where:

- N= total number of samples over the groups being tested
- g= number of groups
- n_k = number of samples in group k
- p= number of variables.

This statistic is distributed as F with p (numerator) and (N-g-p+1) denominator degrees of freedom. This is a “predictive” probability rather than an “estimative” probability that would normally be computed using the χ^2 distribution.

The posterior probabilities force a fit of the sample to be allocated to at least one of the groups. The index of typicality does not force a fit. Samples that have probabilities in more than one group indicate that they overlap between the groups.

The application of posterior probabilities and indices of typicality have obvious advantages in an exploration program where the geochemical characteristics of the commodity being sought are known. Samples that classify within the background populations can be recognized immediately. Samples from areas which have high typicality indices for any of the target groups are obvious areas for follow-up exploration. Maps of target group membership probability can be created which may outline areas of high mineralization potential. Such an approach was taken by Smith *et al.* (1984). Samples that have typicality indices of zero for any of the target or background groups should be investigated for their likelihood of representing some previously unrecognized target or background geochemical model. The use of posterior probabilities can assist in “forcing” a fit to one of the reference groups to examine which group an unknown sample is the closest.

Both alternative data analysis methods were applied to the raw data as well as data levelled for catchment lithology to see if there were significant differences in the predictions for the two datasets. A similar response for both datasets would indicate that the CLR transformation had moderated background effects and attenuated the geochemical signal associated with mineralization.

10 Appendix 2: Associated Files

The following is a listing of digital files and abbreviated header descriptions distributed with this report:

Listing of shapefiles

GBC_NWBC_Compilation – original compiled geochemical data

GBC_NWBC_PCA_WSM – principal components with WSMs following regression analysis

GBC_NWBC_RAW_WSM – raw data WSMs

GBC_NWBC_Levelled_WSM – Z-score data levelled by dominant lithology of terrane and resulting WSMs

GBC_NWBC_MRL_WSM – WSM constructed following multiple regression against catchment lithology

GBC_NWBC_ICA – results of independent component analysis following CLR

GBC_NWBC_RF_PCA_CLR – Random forests normalized votes following CLR and PCA for mineral deposit types

GBC_NWBC_TYP_PCA_CLR – Typicality indices following CLR and PCA for selected mineral deposit types

Listing of abbreviations

PC	Principal component
Inv	Inverted principal component
scl	Scaled
RRes	Robust residual
Res/Resid	Residual
CLR/clr	Centred-log ratio
Val	Validated location
DC	Dilution-corrected
WSM	Weighted sums model
per	Percentile class
Text	Description of percentile class
Epithe	Epithermal Au-Ag
PolyMe	Polymetallic vein
Porph	Porphyry Cu-Au-Mo
UM_CrN	Mafic Cr-Ni-Cu-PGE
ZLog	Z-score levelling after a Log transformation
DomGeo/DL	Dominant lithology
DomTerrane/DT	Dominant terrane
LDL	Levelled by dominant lithology

LDT	Levelled by dominant terrane
MRL/MR_lith	Multiple regression against lithology
LOI_LevMean	Mean-levelled Loss on Ignition
LOI_LevMed	Mean-levelled Loss on Ignition
S_d25	Sulphur data from map sheet 093N levelled by dividing by a factor of 2.5
S_d25_scl	Sulphur data from map sheet 093N levelled by dividing by a factor of 2.5 and scaled



Australia • Canada • Indonesia • Russia
Singapore • South Africa • United Kingdom

csaglobal.com

