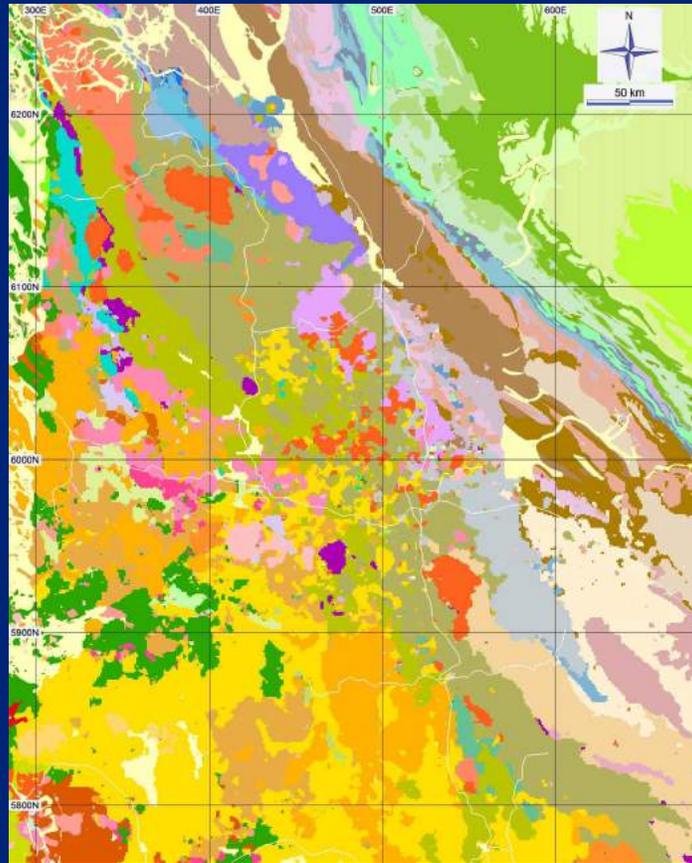




Geoscience BC Report 2009-003



**Final Report for Geoscience BC Project 2008-003**

**Using Geochemistry and Neural Networks to  
Map Geology under Glacial Cover**

*By Colin T Barnett and Peter M Williams, BW Mining*

# Using Geochemistry and Neural Networks to map Geology under Glacial Cover Geoscience BC Project 2008-003

Colin T Barnett\*      Peter M Williams†

January 2009

## Abstract

Geoscience BC's QUEST project was designed to stimulate mineral exploration in the Quesnellia Terrane of British Columbia. During 2007, about 2100 new lake and stream sediment samples were collected, and almost 5000 older drainage sediment pulps were re-assayed, to improve the geochemical data base in the project area. One of the programs initiated by QUEST in 2008 was to discover what might be learned from a systematic analysis and evaluation of the new multi-element geochemical data. This report describes the results of one such study.

Since the samples were collected from different media and analysed by different laboratories over nearly 30 years, it was first necessary to assemble the "best picks" from each sub-population, and then to relevel the various surveys, for each element, to provide uniform blends. Estimates were also made of missing data over some small areas. As a result, syntheses of the various surveys are now available on uniform grids, over a common area, for as many as 42 of the elements. These can form the basis for systematic analysis.

As a start, we have applied various clustering methods to the 42 element data. The results show marked correlations with geology. This leads to the idea of using a neural network to model the geochemistry in areas where the geology is known, and then to apply this model to infer the bedrock geology in the non-outcropping areas. The resulting inferred geology, wherever geochemistry is known, is then almost identical to mapped geology in areas of outcrop, and blends well with mapped geology along the margins, where there is no geochemistry. These results show that geochemistry combined with neural networks can provide a powerful tool for mapping bedrock geology concealed by a veneer of glacial overburden in the QUEST project area.

---

\*BW Mining, Boulder, Colorado, USA

†BW Mining, Brighton, Sussex, UK

## Contents

<b>Abstract</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Selection of elements</b>	<b>5</b>
<b>3 Levelling and blending</b>	<b>6</b>
<b>4 Missing data</b>	<b>9</b>
<b>5 Cluster analysis</b>	<b>15</b>
<b>6 Inferred geology</b>	<b>20</b>
<b>7 Conclusion</b>	<b>22</b>
<b>Acknowledgements</b>	<b>25</b>
<b>References</b>	<b>25</b>
<b>List of Figures</b>	<b>26</b>
<b>List of Tables</b>	<b>26</b>

## 1 Introduction

Geoscience BC's QUEST project was designed to stimulate mineral exploration in the Quesnellia Terrane running through Prince George, British Columbia (see Figure 1). This is a highly prospective belt of rocks that to date has had little exploration because of the till and lacustrine cover left behind by retreating glaciers (see Figure 2 and References [4, 5]).

During 2007, Geoscience BC compiled publicly available topographic, geological, geophysical, geochemical, and mineral occurrence data for a 150,000 square kilometer area centered on Prince George. In addition, new airborne gravity and electromagnetic surveys were flown to provide a geophysical framework for further exploration in this area [2]. Meanwhile, Geoscience BC also collected about 2100 new lake and stream sediment samples and re-assayed almost 5000 older drainage sediment pulps to improve the geochemical data base in the QUEST project area [3].

During 2008, while further surveys were being flown or collected to the west of QUEST, an effort was mounted to interpret and try to extract maximum information from the 2007 QUEST surveys. One of these programs, which is described in this report, was designed to analyse the new multi-element geochemical data and to see what might be learned from a systematic evaluation of these data.

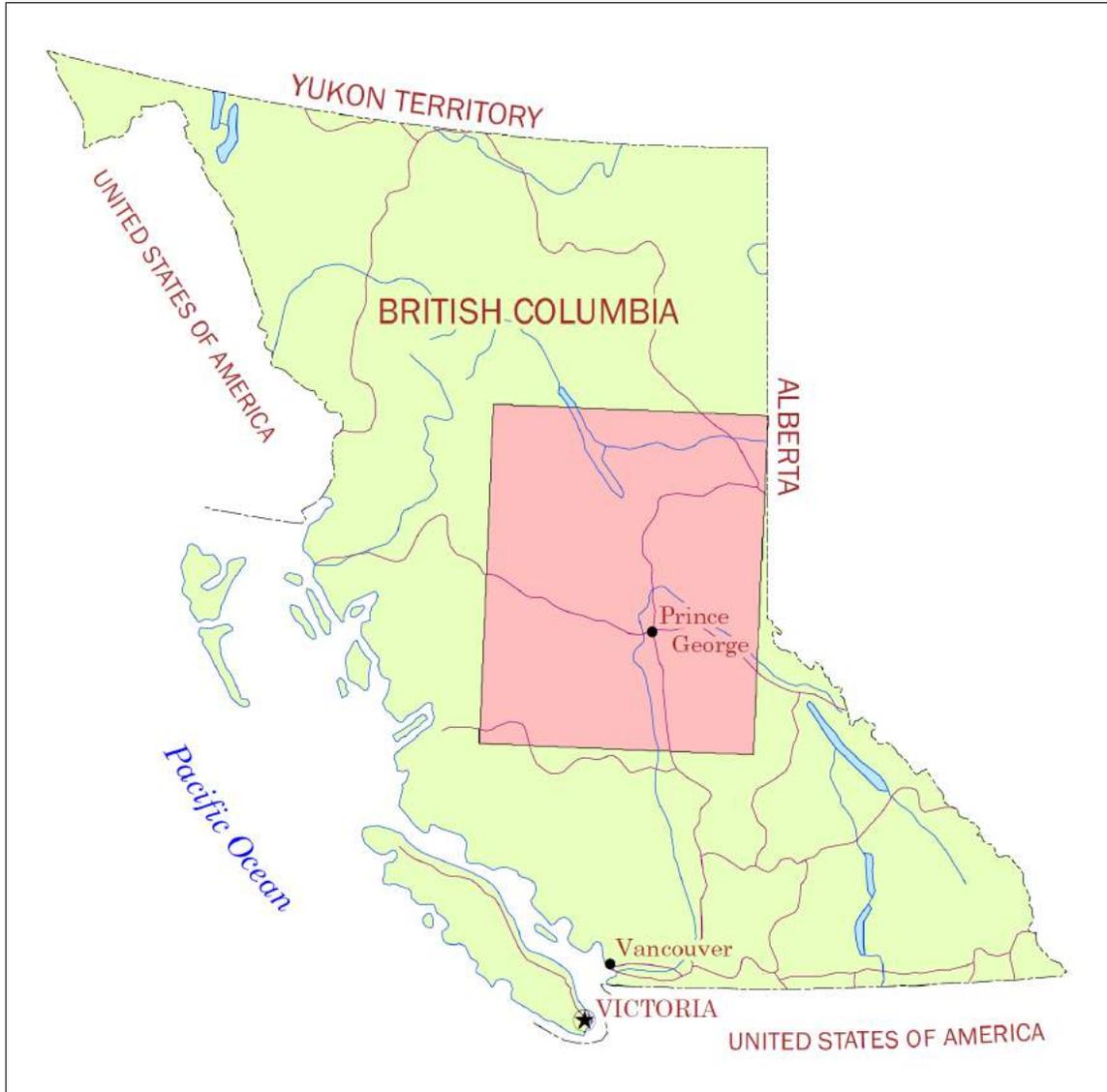


Figure 1: Location of the QUEST Project area in Central British Columbia.

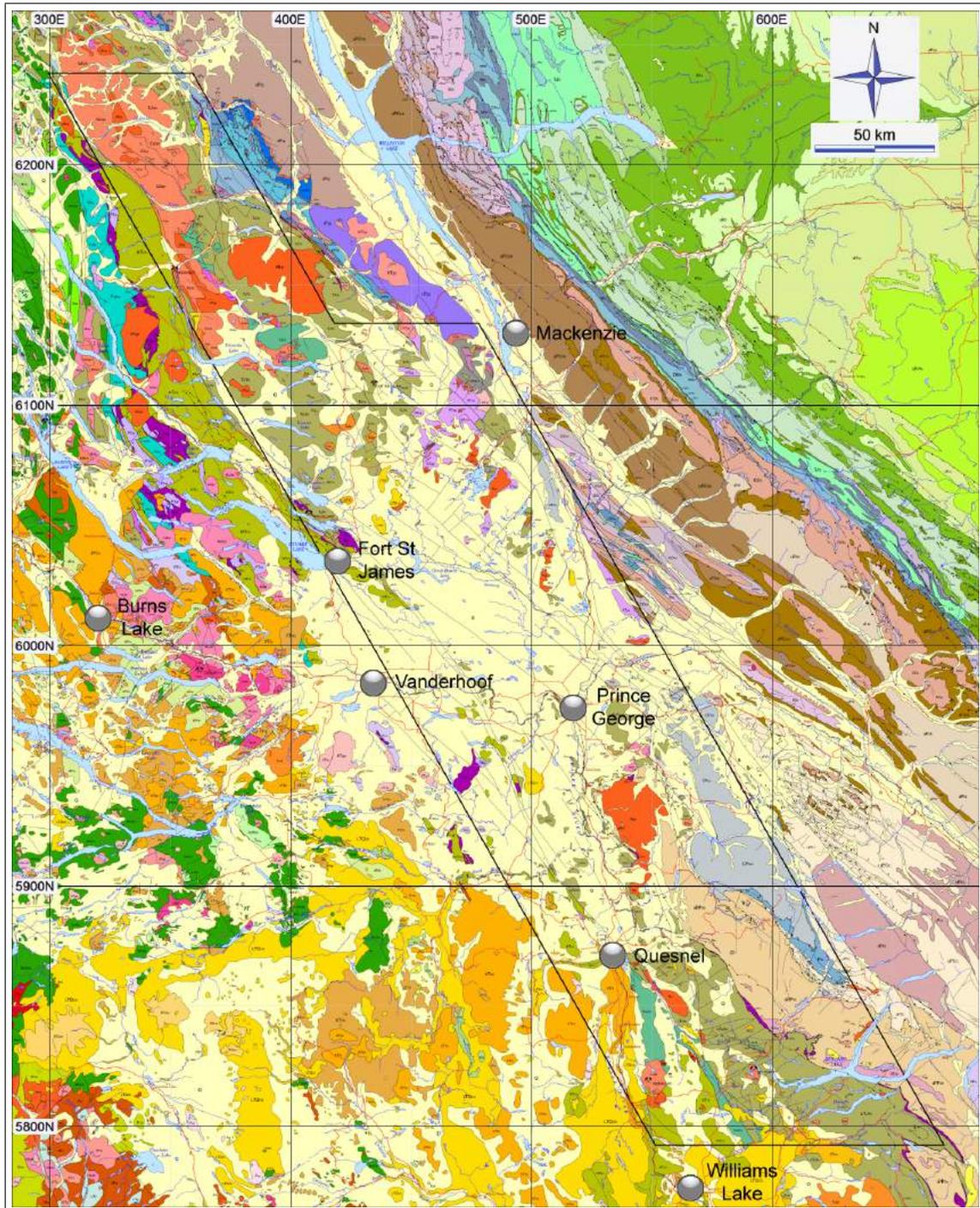


Figure 2: Surficial geological map of the QUEST project area, showing a few of the towns. The pale yellow areas represent the Quaternary overburden. The outline of the airborne gravity and electromagnetic surveys is marked in black. The map projection is UTM Zone 10 in NAD83.

Since the various geochemical surveys were collected from different sample media and analysed by different laboratory techniques over a period of nearly 30 years, the first step was to assemble the “best picks” of measured values from each sub-population for each element. Next, it was necessary to relevel the various surveys to give a satisfactory blend. Of the 42 elements treated, 17 have the same coverage, while the remaining 25 have reduced coverage to varying extents. For analysis it is more convenient if all the elements have the same coverage. The next step was to estimate values for these 25 elements over the small areas of missing data.

The resulting full suite of 42 element assays can be analysed in various ways. Clustering is a classic approach. Two such clusterings have been derived. Each resembles the mapped surficial geology. This leads to the idea of building a model of the geochemistry in the outcropping areas, where the geology is known, and applying it to infer bedrock geology in the till-covered areas. For this purpose a neural network has been used to map 42 element assays as inputs, to a probability distribution over geological formations as output. The network was trained in the region where both geology and geochemistry are known. The resulting inferred geology, wherever geochemistry is known, is almost identical to mapped geology in areas of outcrop, and blends well with mapped geology along the margins, where there is no geochemistry. These results show that geochemistry largely characterises the geological formations in exposed areas, and provides corroboration for the neural network inferences in the covered areas.

## 2 Selection of elements

The three principal analytical methods used to determine trace elements in the QUEST stream and lake sediments are atomic absorption spectrometry (AAS), aqua regia inductively coupled plasma emission–mass spectroscopy (ICP-MS), and instrumental neutron activation analysis (INAA).

AAS was the method routinely used for federal and provincial government funded drainage sediment surveys conducted before 1999, but this has now been superseded by the more sensitive ICP-MS and INAA techniques. INAA estimates the “total” element concentration, but is inadequate for measuring elements such as lead and copper, which are best determined by aqua regia ICP-MS. While aqua regia digestion is effective for dissolving gold, carbonates and sulphides in a sample, it can only partially break down alumino-silicate, oxide and other refractory minerals such as barite.

Since the various geochemical surveys in the QUEST area were collected from different sample media (lake and stream sediments) and analysed by different laboratory techniques over a period of nearly 30 years, the first step in the systematic evaluation of the geochemistry was to assemble the “best picks” of measured values from each subpopulation for each element.

Tables 1 and 2 show the preferred analytic method and detection limits for each of the 42 elements that were incorporated in the present study. Six further elements (Eu, Ta, Te, W, Yb and Zr) were found to have poor coverage and therefore could not be used in this

Element	Symbol	Detection	Unit
Silver	Ag	0.002	ppm
Aluminum	Al	0.01	%
Bismuth	Bi	0.02	ppm
Calcium	Ca	0.01	%
Cadmium	Cd	0.01	ppm
Copper	Cu	0.01	ppm
Gallium	Ga	0.2	ppm
Mercury	Hg	5	ppb
Potassium	K	0.01	%
Magnesium	Mg	0.01	%
Manganese	Mn	1	ppm
Molybdenum	Mo	0.01	ppm
Nickel	Ni	0.1	ppm
Phosphorus	P	0.001	%
Lead	Pb	0.01	ppm
Sulphur	S	0.02	%
Selenium	Se	0.1	ppm
Strontium	Sr	0.5	ppm
Titanium	Ti	0.001	%
Thallium	Tl	0.02	ppm
Vanadium	V	2	ppm
Zinc	Zn	0.1	ppm

Table 1: The 22 elements for which ICP-MS was the preferred analytic method.

study.

For gold, INAA was the preferred analytical technique, because ICP-MS uses a very low sample weight (< 1 gram). Unfortunately, the stream sediment results for the McLeod Lake quad sheet, 93J, did not include INAA information, therefore the relatively “weaker” ICP-MS determinations were substituted from this quad sheet into the final selected data set.

### 3 Levelling and blending

Figure 3 shows an image of the raw gold assays, after gridding of individual sheets, but before any levelling is applied. This illustrates the problem that we are dealing here of multiple mismatched populations. The McLeod Lake quad sheet, for example, which is roughly centred on 500E/6050N, is clearly dropped down in relation to its neighbours. There are similar mismatches with all the other elements.

There is a fundamental problem in levelling geochemical data as compared with, say, airborne geophysical data. Because of the wide turning circle of an aircraft, airborne surveys

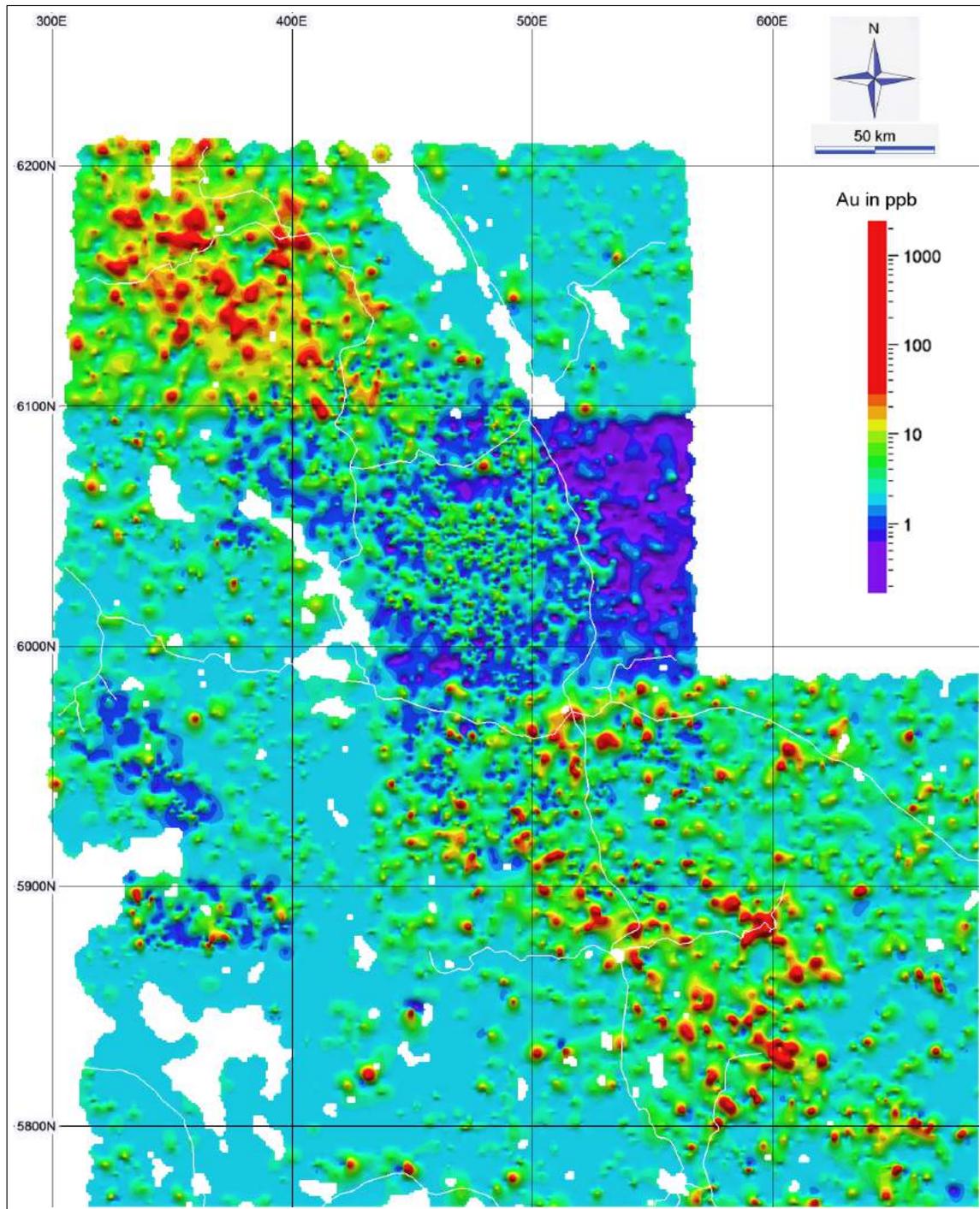


Figure 3: Raw gold assays from stream and lake sediments in the QUEST Project area.

Element	Symbol	Detection	Unit
Arsenic	As	0.5	ppm
Gold	Au	2	ppb
Barium	Ba	50	ppm
Bromine	Br	0.5	ppm
Cerium	Ce	5	ppm
Cobalt	Co	5	ppm
Chromium	Cr	20	ppm
Cesium	Cs	0.5	ppm
Iron	Fe	0.2	%
Hafnium	Hf	1	ppm
Lanthanum	La	2	ppm
Lutetium	Lu	0.2	ppm
Sodium	Na	0.02	%
Rubidium	Rb	5	ppm
Antimony	Sb	0.1	ppm
Scandium	Sc	0.2	ppm
Samarium	Sm	0.1	ppm
Terbium	Tb	0.5	ppm
Thorium	Th	0.2	ppm
Uranium	U	0.2	ppm

Table 2: The 20 elements for which INAA was the preferred analytic method.

commonly overlap, making the blending of adjoining surveys relatively easy to accomplish. The opposite is generally true of geochemical surveys, which seldom overlap. To overcome this problem, the following general procedure was applied. In order to level a pair of neighboring grids, or assemblages of grids, narrow strips were selected on either side of the join, including any overlapping region, if such existed. This provides two neighboring sets of assay samples, whose statistics can be compared. Because of the proximity of the samples, even if they are not overlapping, it can reasonably be assumed that they come from approximately the same population. This would not necessarily be true if the statistics of complete neighboring sheets were compared.

To compare the statistics of such neighboring border regions, the nine decile points were calculated for each of the two regions. An example, in the case of gold, is given in Figure 4. This shows a scatter plot of matching decile points when the procedure was applied to bordering strips of assemblages of lake and stream samples, after first applying the same procedure to border strips from the various component lake and stream regions. The straight line (which was fitted by minimising perpendicular distance, to maintain symmetry between populations) indicates the adjustments needed to one or the other or both, in level and gain, to produce a satisfactory blend. After this adjustment, a line fitted to the adjusted populations should have unit slope and pass through the origin.

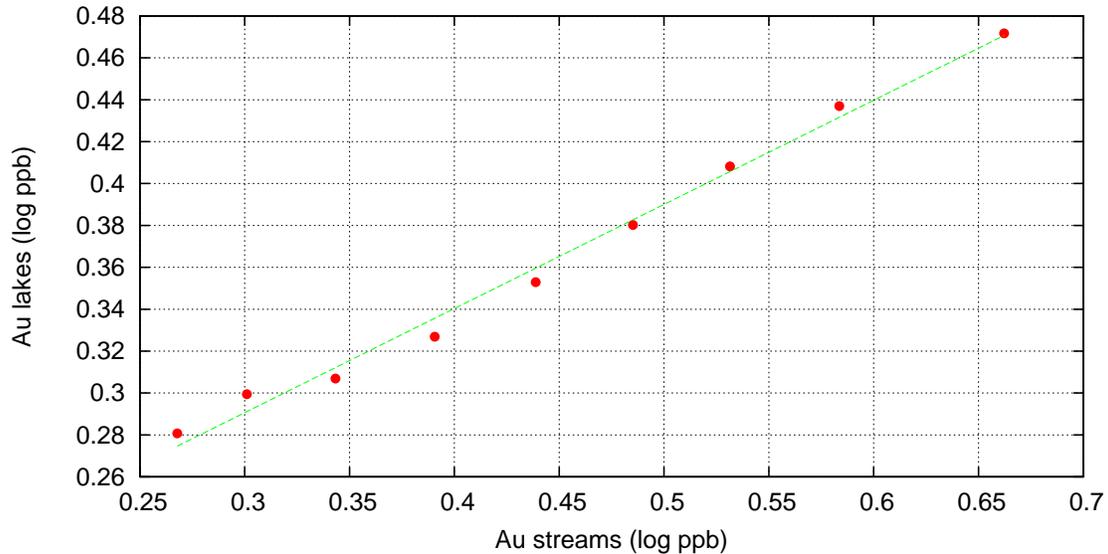


Figure 4: Scatter plot of matched decile points from neighboring strips of the streams and lakes populations. The straight line  $y = mx + c$  has parameters  $m = 0.4971$  and  $c = 0.1415$ .

Note that this correction corresponds to a linear transformation  $z' = az + b$  in log space. If  $a = 1$  this is a simple shift, corresponding to a constant positive scalar multiplication of the original assays. Otherwise, such a transformation still corresponds to a rescaling of the original assays, but one whose amplitude increases ( $a > 1$ ) or decreases ( $a < 1$ ) with the assay level. Note that, according to this procedure, the exact distributions of outliers, in the top 10% or bottom 10%, have no effect when calculating the parameters  $a$  and  $b$ . However, once these relevelling parameters have been determined by the positions of the decile points, we apply the same relevelling procedure to all samples, including outliers.

Figures 5, 6 and 7 show the result of applying the process to gold, copper and molybdenum. After relevelling, the various component grids were stitched together using a cosine taper. The overall level has been taken from the most recent stream sediment survey, collected in 2007 from quad sheet 930.

## 4 Missing data

Of the 42 elements treated in this study, 17 have the same maximum coverage, comprising the area shown equally in Figure 5 for gold and Figure 6 for copper. These 17 elements are

As, Au, Ba, Co, Cr, Cu, Fe, Hg, La, Mn, Ni, Pb, Sb, Sc, Th, U, Zn.

Each of the remaining 25 elements, namely Ag, Al, Bi, Br, Ca, Cd, Ce, Cs, Ga, Hf, K, Lu, Mg, Mo, Na, P, Rb, S, Se, Sm, Sr, Tb, Ti, Tl, V, has reduced coverage to a varying extent. An example is shown in Figure 7 for molybdenum. For the purposes of analysis, however,

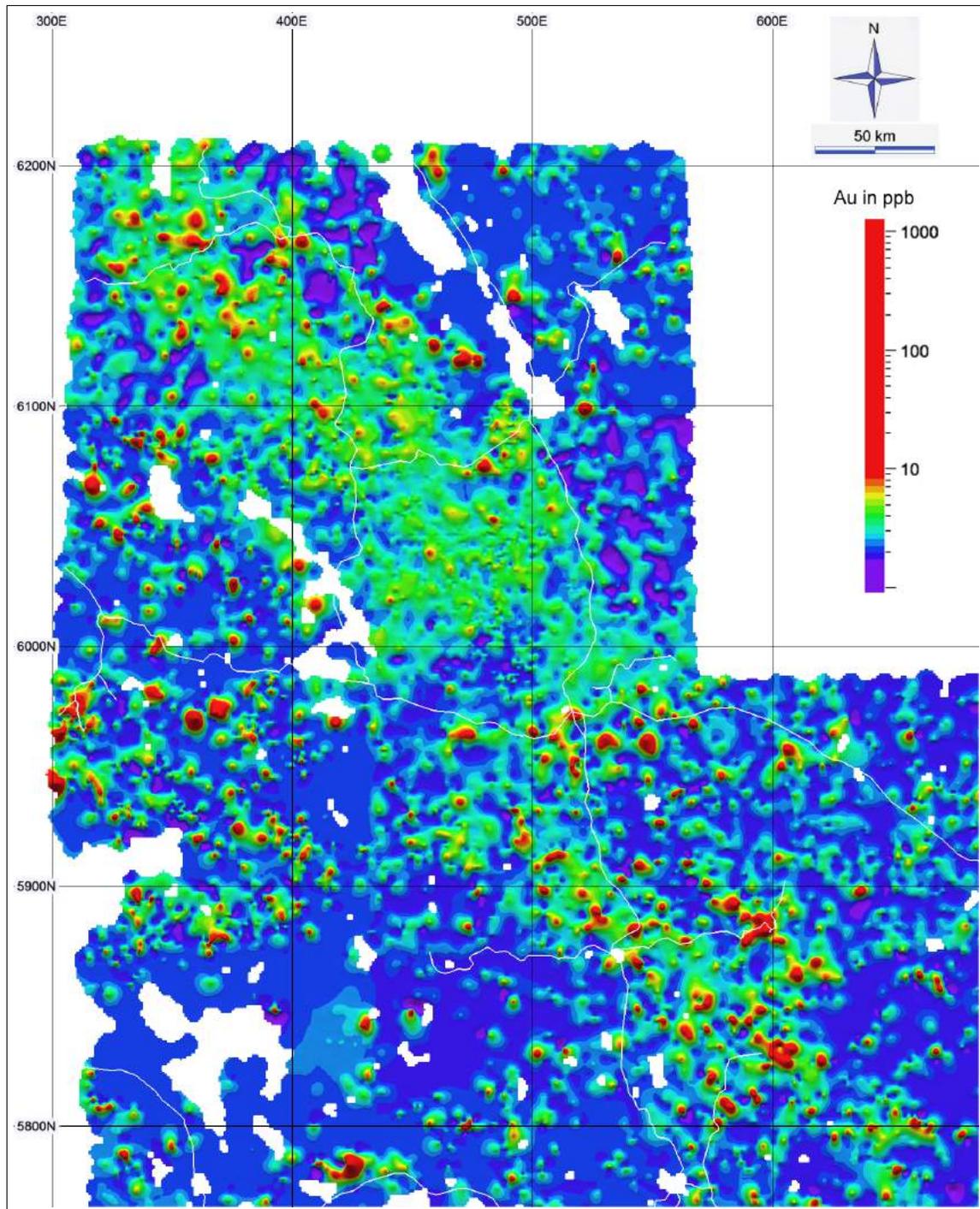


Figure 5: Levelled gold assays from stream and lake sediments.

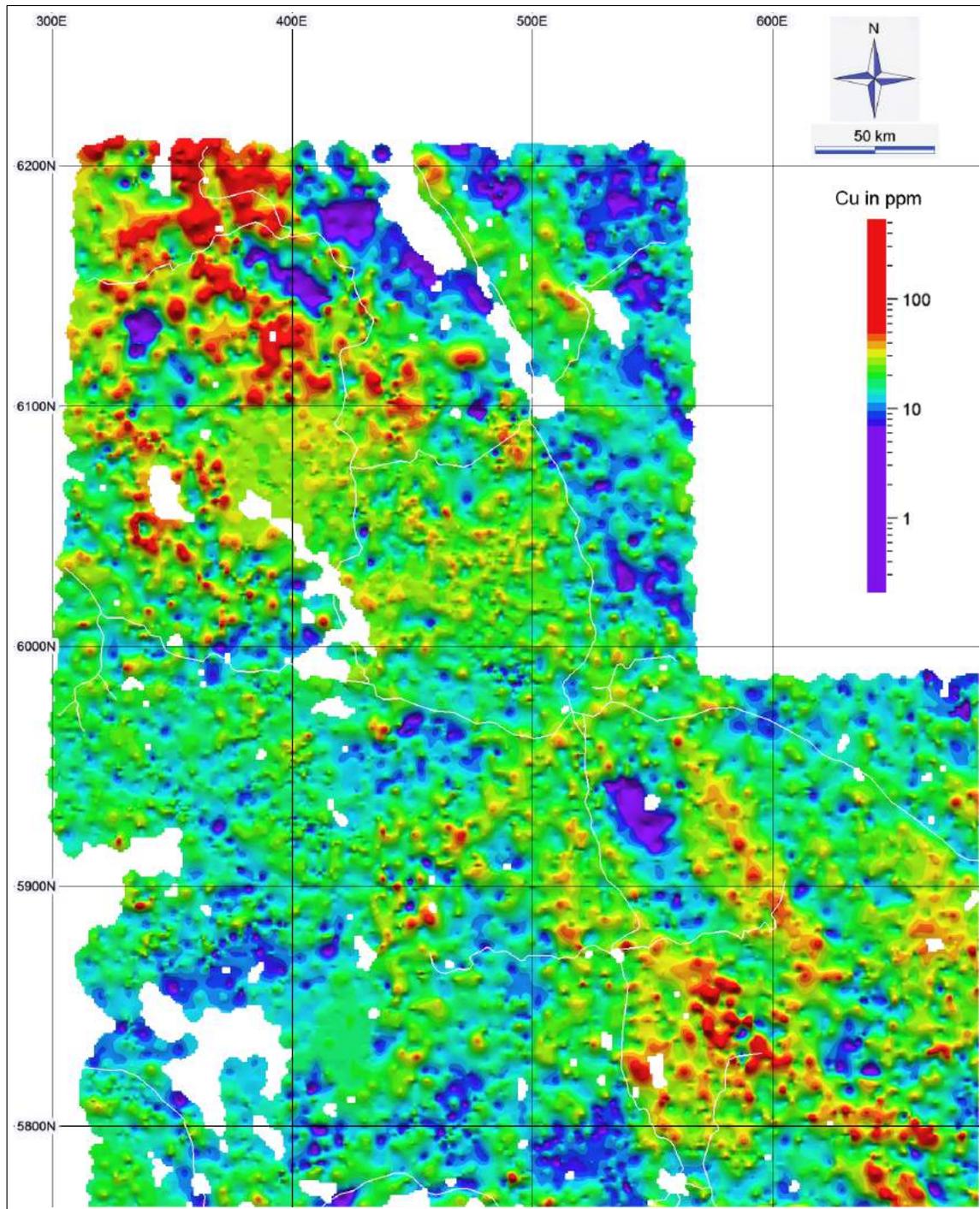


Figure 6: Levelled copper assays from stream and lake sediments.

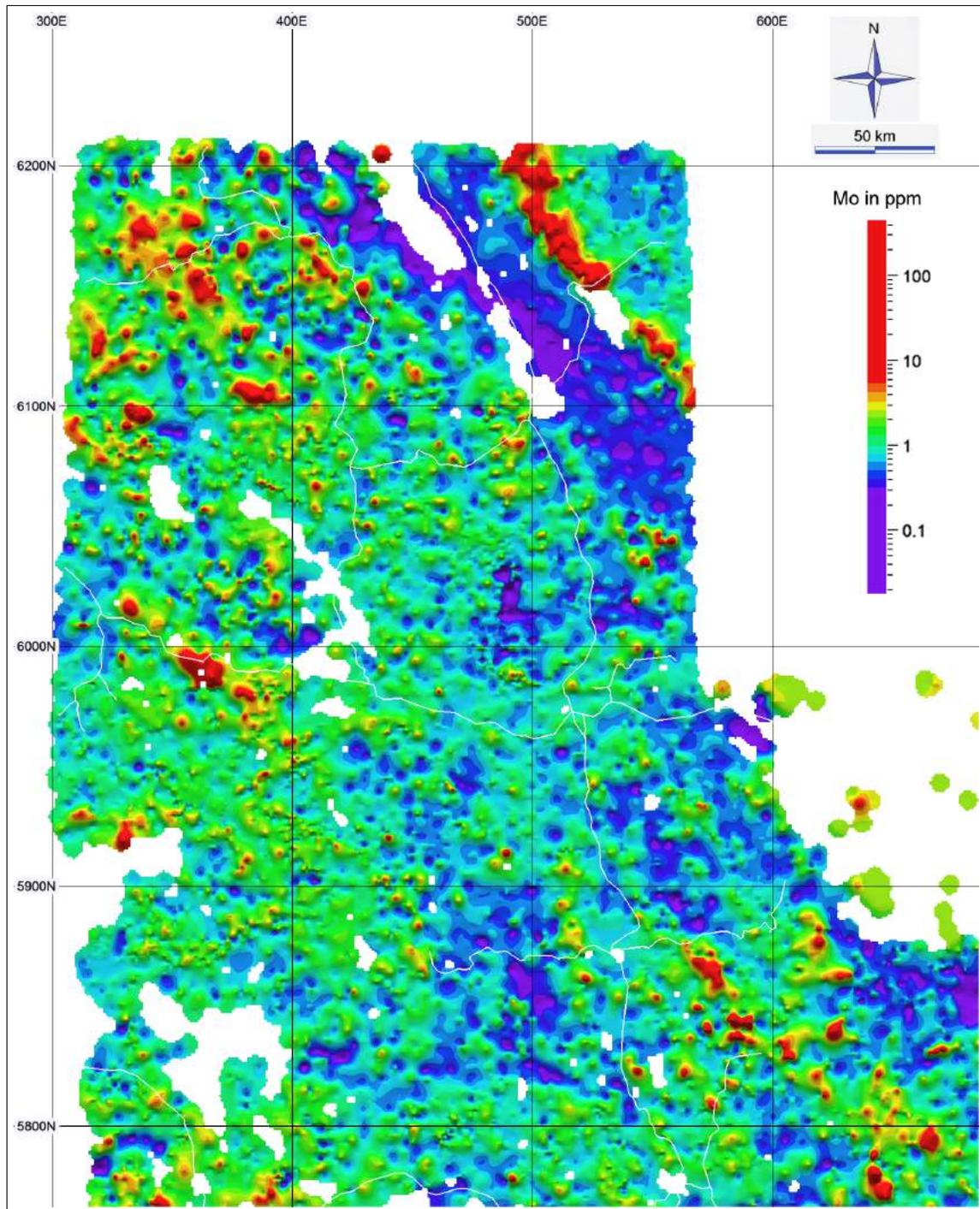


Figure 7: Levelled molybdenum assays from stream and lake sediments.

it would be convenient if all elements had the same coverage. Rather than exclude certain elements, or restrict attention to the smaller area where all 42 are present, it is better to estimate missing values if this can be achieved in a satisfactory way.

This may seem a difficult task. It is well known, however, that there are strong multi-variate statistical relations between the various elements, so that much of the information in the distribution of the full set of 42 elements may already be carried by the distribution of one or more subsets, for instance by the subset consisting of the 17 elements displayed above. The nature of these relations can be investigated by noting that there is a large area where all 42 elements are present. That area provides a basis for modelling the statistical dependency of the remaining 25 elements on the given 17.

There are a number of approaches to a missing data problem of this type. The one we chose to adopt is based on neural networks. For each of the 25 elements with restricted coverage, a neural network was trained to predict the assay value of that element, on the basis of the assay values of just the 17 elements displayed above. The result for molybdenum is shown in Figure 8.

It is important to observe that all values in Figure 8 are estimated, in the sense that they are neural network predicted values, based on just the 17 elements, including where measured values are known. Comparison of the two grids in the area of common coverage, shows just how well it is possible to approximate the measured data of Figure 7 with the predicted data of Figure 8. This correspondence provides confidence in the estimated values where no measured data are available. For analytic purposes, however, we have always used measured values when available.

## Error bars

The modelling techniques used in this approach to the missing data problem are described in [8, 9]. Thus the assay at a given location, conditional on the local measured assays of the given 17 elements, is modelled by a log normal distribution, whose local mean and standard deviation are determined by the neural network as functions of the 17 input values at the given location. Figure 8 shows the mean of the predicted distribution at each location. The standard deviation, which is also modelled, provides an error bar varying over the grid. Table 3 shows the average width, over the grid, of the error bar for each of the estimated elements. These are arranged in order of increasing size of the error bar. The ordering gives an indication of the relative ease or difficulty of predicting the given element on the basis of these 17 elements, with gallium as the easiest and titanium as the hardest. Molybdenum, for example, is seen to be one of the more difficult elements on this basis. It should be remembered, however, that these error bars are also dependent on data quality, which is variable between the elements.

The last two columns of Table 3 translate additive or subtractive logarithmic errors into upper or lower percentage errors in actual assay values. For example, the logarithmic error of 0.0333 for gallium, when raised to the power 10, corresponds to a factor of 1.080. This either divides or multiplies the assay values themselves to produce an error of between 7.4% below or 8.0% above.

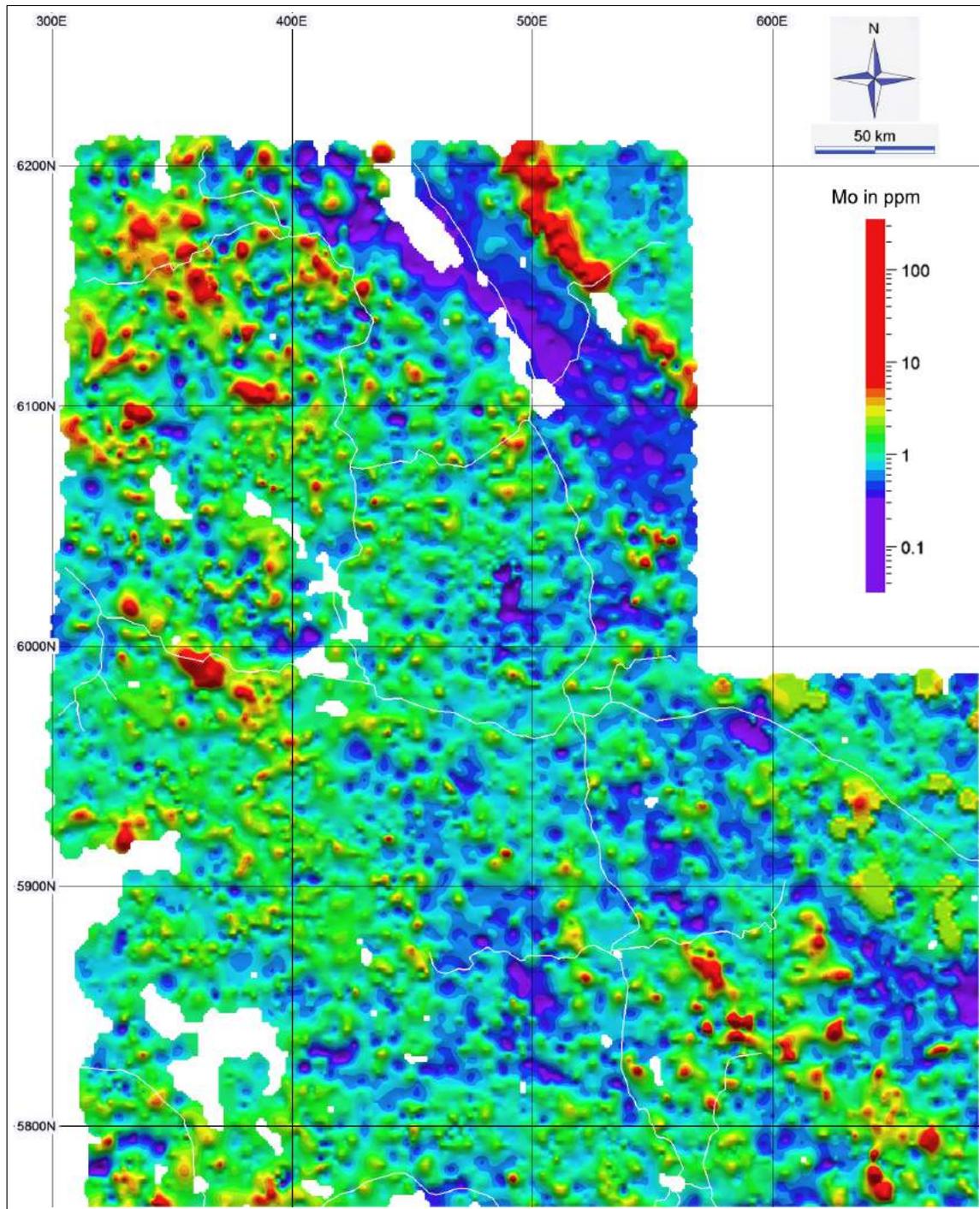


Figure 8: Measured and estimated molybdenum assays.

Element	Symbol	$\log_{10}$ error	Lower error	Upper error
Gallium	Ga	0.0333	7.4%	8.0 %
Aluminum	Al	0.0434	9.5%	10.5%
Cerium	Ce	0.0443	9.7%	10.7%
Samarium	Sm	0.0535	11.6%	13.1%
Rubidium	Rb	0.0626	13.4%	15.5%
Sodium	Na	0.0645	13.8%	16.0%
Vanadium	V	0.0700	14.9%	17.5%
Thallium	Tl	0.0726	15.4%	18.2%
Cesium	Cs	0.0740	15.7%	18.6%
Phosphorus	P	0.0746	15.8%	18.7%
Bismuth	Bi	0.0759	16.0%	19.1%
Magnesium	Mg	0.0832	17.4%	21.1%
Lutetium	Lu	0.0878	18.3%	22.4%
Hafnium	Hf	0.0922	19.1%	23.7%
Cadmium	Cd	0.0944	19.5%	24.3%
Silver	Ag	0.1027	21.1%	26.7%
Terbium	Tb	0.1074	21.9%	28.0%
Potassium	K	0.1077	22.0%	28.1%
Strontium	Sr	0.1218	24.5%	32.4%
Sulphur	S	0.1320	26.2%	35.5%
Calcium	Ca	0.1376	27.2%	37.3%
Selenium	Se	0.1394	27.5%	37.8%
Molybdenum	Mo	0.1492	29.1%	41.0%
Bromine	Br	0.1604	30.9%	44.7%
Titanium	Ti	0.1948	36.1%	56.6%

Table 3: Table showing the average sizes of error bars when estimating the 25 partially missing elements.

## 5 Cluster analysis

Once the 42 elements had been levelled and any missing samples had been estimated as described in the last two sections, it was decided to try and condense all this information into a smaller number of interpretation maps. The classical way of doing this is by cluster analysis, which attempts to partition the data into subsets which have common characteristics.

The full suite of assays, at a given point, can be represented by a vector in a 42-dimensional space. The distribution of observed data points in this high-dimensional space will be non-uniform. Data can be expected to cluster into various classes, each representing a common relationship between the abundances of the various elements. Each cluster determines a corresponding region on the map, namely the set of locations whose assay vectors belong to this cluster. Locations in the same region should be interpretable as sharing

similar geochemical properties.

Cluster analysis can be approached in a number of ways. The best known method is the  $k$ -means algorithm. For a given integer  $k$ , this aims to determine  $k$  vectors in assay space, considered as cluster centers, so that the average distance from each data vector to its nearest cluster center is minimized. Results depend significantly, however, on the method used for measuring distance in assay space. We chose Aitchison's compositional metric [1]. This also takes account of the deficit of the sample not included in the measured assays. The squared distance between two samples is then the sum of squares of the logarithms of each component, including the deficit, after subtracting their means. Figure 9 shows the result for  $k = 30$ . As might be expected, there are clear correlations with the surficial geology shown in Figure 10.

The  $k$ -means algorithm assumes that clusters are linearly separable. This can be overly restrictive. All that is needed is that members of the same cluster should be similar to each other—possibly through some connected chain of similarities—and that they should be dissimilar from members of other clusters. This is the idea behind spectral clustering, which is based on analysis of the spectrum of the graph Laplacian: specifically, of the weighted graph having assay vectors as nodes, and similarities between them as weights [7]. In this study we used symmetric normalized spectral clustering [6] with similarity derived from distance in the sense of the Aitchison metric [1]. Results for 30 clusters are shown in Figure 11.

Figure 11 again shows correlations with the mapped geology of Figure 10. To draw attention to such correlations, and to aid comparison with the  $k$ -means clustering of Figure 9, we have used the same color palette for clusters as for geological formations. The choice of specific colors for clusters was made using a simple but efficient algorithm we have developed for maximizing the overlaps between clusters and mapped formations, over all possible color permutations.

Choosing the appropriate number of clusters is not easy. We have examined results for varying numbers between 20 and 50. Broadly speaking, 20 clusters pool regions that are meaningfully separable by a larger number of clusters, while 50 clusters tend to make too many noisy distinctions. For the present clustering algorithms, the most useful number appears to be around 30. Several general criteria have been proposed in the literature for determining the optimal number of clusters. Typically, however, such criteria lead to different results, so that the problem of directly choosing the number of clusters is replaced by the problem of choosing which criterion to use. In our view, the optimal number of clusters depends on the fineness of the distinctions it is practically useful to make and, in the present case, that is probably best decided, after inspecting various results, by specific geochemical and geological expertise, rather than by appealing to a general rule.

Now that 42 elements have been levelled by this study, and extended where necessary to offer common coverage, alternative clustering techniques might be applied, and the issues discussed in the previous paragraph explored further. However, we shall not pursue that here. While there is usually likely to be some uncertainty in interpreting the meaning of clusters in data, in view of the unsupervised nature of the algorithms, clusters in the

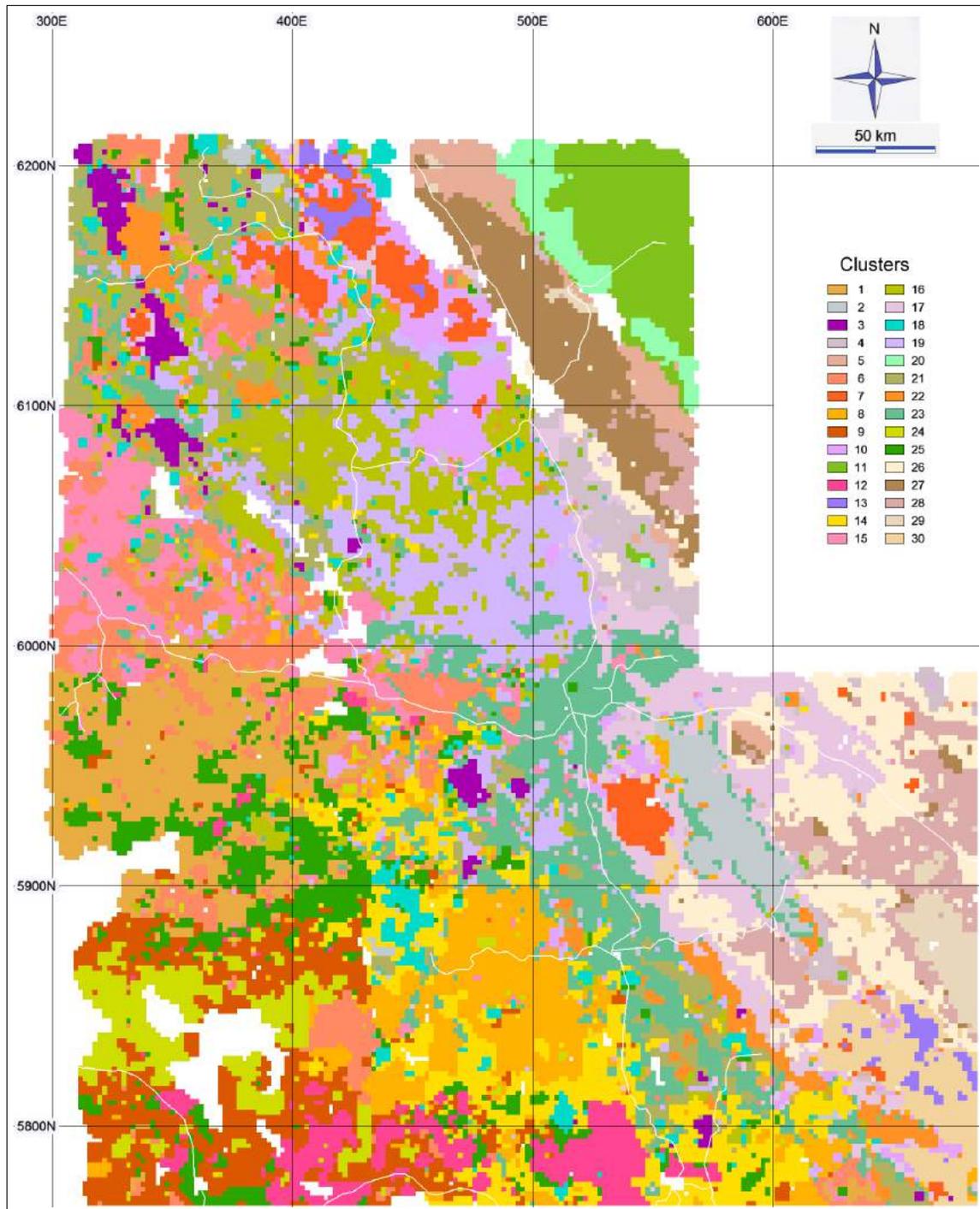


Figure 9: K-means clustering of 42 element geochemistry using 30 clusters.

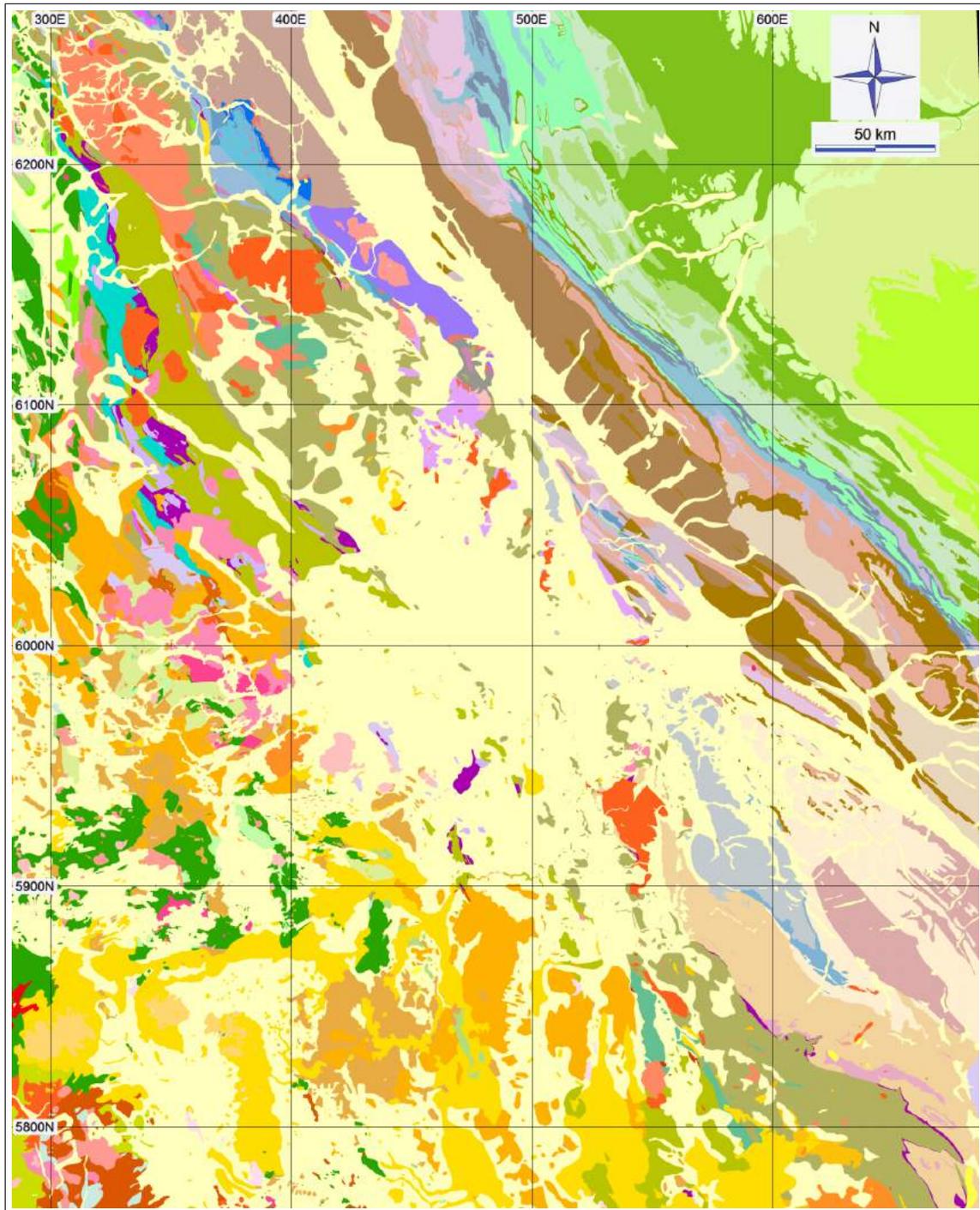


Figure 10: Mapped surficial geology with contacts and topography removed.

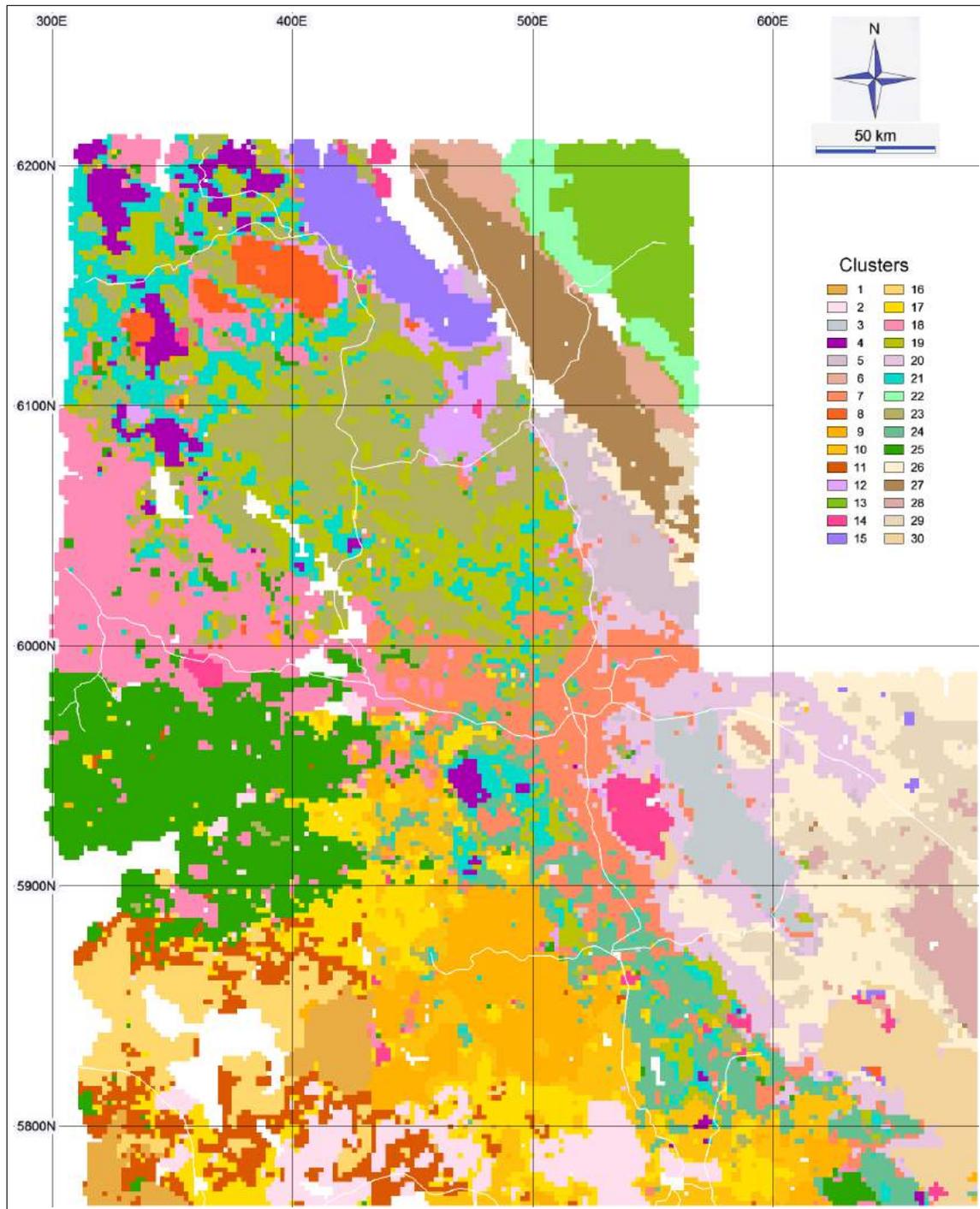


Figure 11: Spectral clustering of 42 element geochemistry using 30 clusters.

present case seem, broadly speaking, to follow the geological formations, at least at this level of detail. To that extent, much of the interest in clustering will concern the information provided about bedrock geology in areas covered by glacial drift. Fortunately, however, there are ways of exploring that issue directly, using supervised methods, to which we now turn.

## 6 Inferred geology

As we have seen, the two different cluster analysis techniques described in the previous section produced quite similar results, which resemble the mapped surficial geology. This is not at all surprising, as it is the underlying formations which gives rise to such primary patterns in the geochemistry. These results quickly lead us to the idea that we might be able to use a neural network approach to build a model of the geochemistry in the outcropping areas where the geology is known, and then to apply this model to infer the bedrock geology in the non-outcropping or till-covered areas.

Examination of the mapped geology shown in Figure 10 shows that roughly half of the region is covered by glacial till. On the other hand, we have geochemistry for 42 elements over much of this area, the exact extent being shown in Figure 9, for example. The network is therefore trained in the region where both geology and geochemistry are known, and then exploited in the covered region where only geochemistry is known. Functionally, it takes 42 element assays as inputs, and generates a corresponding probability distribution over geological formations as output. The reason for modelling probability distributions over formations is that some assay vectors may not be uniquely characteristic of a particular formation. Assay vectors corresponding to different formations will form clusters; but these may overlap to some extent, with the result that vectors in an overlap region cannot be assigned categorically to a unique formation. It should be emphasized, however, that the neural network approach does not assume that clusters have any particular shape. The region of assay space corresponding to a given formation may be dispersed and of high complexity.

There are 72 distinct formations mapped within the region for which geochemistry is available. Some of these, however, have very low abundance. It would be impractical to expect the network to predict a formation, on the basis of its geochemical signature, when the number of its instances is very small. We have therefore restricted attention to the 48 most abundant formations. These jointly account for over 98% of the area where geochemistry is available.

The results of the process are shown in Figure 12 which, at each location, displays the most probable formation. The bedrock geology can only be inferred, by means of the neural network, in regions where geochemistry is known, which explains the restricted extent of Figure 12 compared to the full extent of the mapped geology in Figure 10.

It is important to point out that Figure 12 was generated wholly by the neural network using geochemistry alone as input. A comparison of Figures 10 and 12 shows that mapped geology and neural network inferred geology are almost identical in areas of outcrop. Indeed,

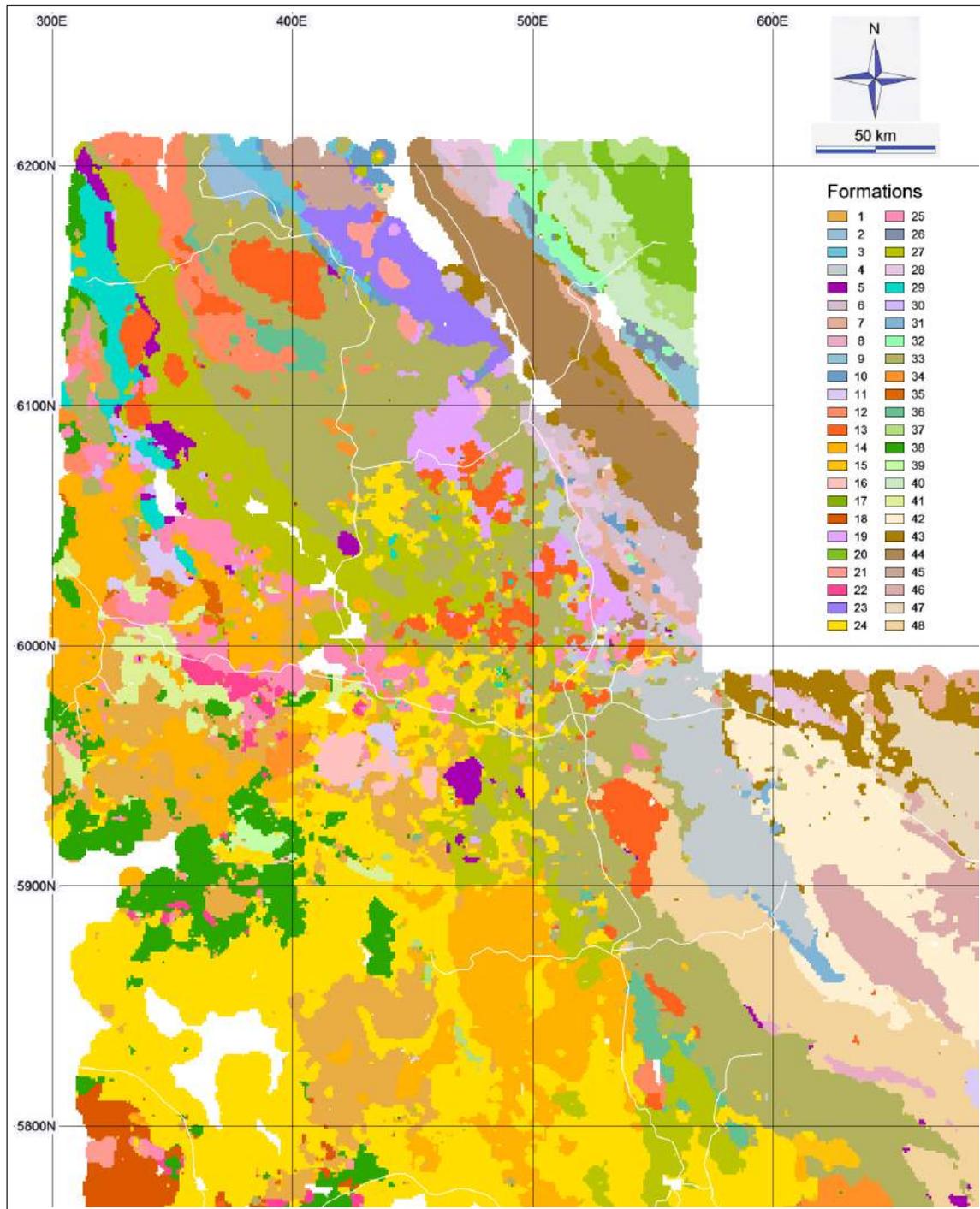


Figure 12: Inferred bedrock geology of the QUEST Project area.

where they differ, especially in poorly exposed areas, there may be grounds for preferring the neural network inferences. In any case, this near identity shows that the geochemical signature largely characterises the geological formation in exposed areas. This result provides corroboration for the neural network inferences in the covered areas. Comparatively speaking, however, in the course of inferring geology from the visible geochemical signature, greater difficulty is to be expected when the latent cause is at depth.

This observation is expressed quantitatively in our study by the probability grid shown in Figure 13, which shows the probability of the most probable formation, in other words the formation indicated as the inferred geology in Figure 12. Such probabilities are seen to be close to the maximum at or near to areas of outcrop and, generally speaking, to reduce to a minimum as depth to bedrock increases. Nonetheless, bearing in mind that this is the probability of the most probable of 48 formations, and since  $1/48 = 0.021$ , anything above deep blue is 10 times more probable than assigning a formation at random; and nothing is less than 5 times more probable, since nothing is purple. For much of the covered area, the odds are up to 100 times better than chance.

Our final image is shown in Figure 14, in which the inferred bedrock geology has been superimposed on the mapped surficial geology. It can be seen that the inferred geology blends in satisfactorily well with the mapped geology along the west, north, and northeast margins of this image where there is no geochemistry.

## 7 Conclusion

The 42 levelled grids and images of the individual elements that came out of this study are useful in their own right for further interpretation and follow-up exploration. The copper-gold deposits at Mount Milligan and Mount Polley and the molybdenum deposit at Endako, for example, show up quite clearly on the Cu and Mo images, respectively. It would be a normal starting point to follow up lookalikes in these or any of the other single-element images.

The two cluster analysis images are interesting in that they provide a straightforward means of condensing the information contained in multi-element data. They show there are clear correlations between the 42 elements and the underlying geology and were instrumental in leading to the idea of using neural networks to infer geology.

However, the most useful product that has come out of this study is the inferred bedrock geology map. With so many layers of information to draw on (i.e. 42 elements of geochemistry), the resulting map is very robust and dependable. This is shown in the coherent nature of the map, the confirmation from the known outcrop areas, and the way in which inferred formations tie together in the covered areas. Formations coming up from the south naturally taper off as they meet formations coming down from the north. This is not a crude buckshot scatter pattern, which would tend to make one suspicious, but instead appears to be a sensible and geologically meaningful result.

Perhaps of greatest interest are the small number of inferred intrusions occurring in the covered areas just north of Prince George. There is little sign of these in surface outcrop.

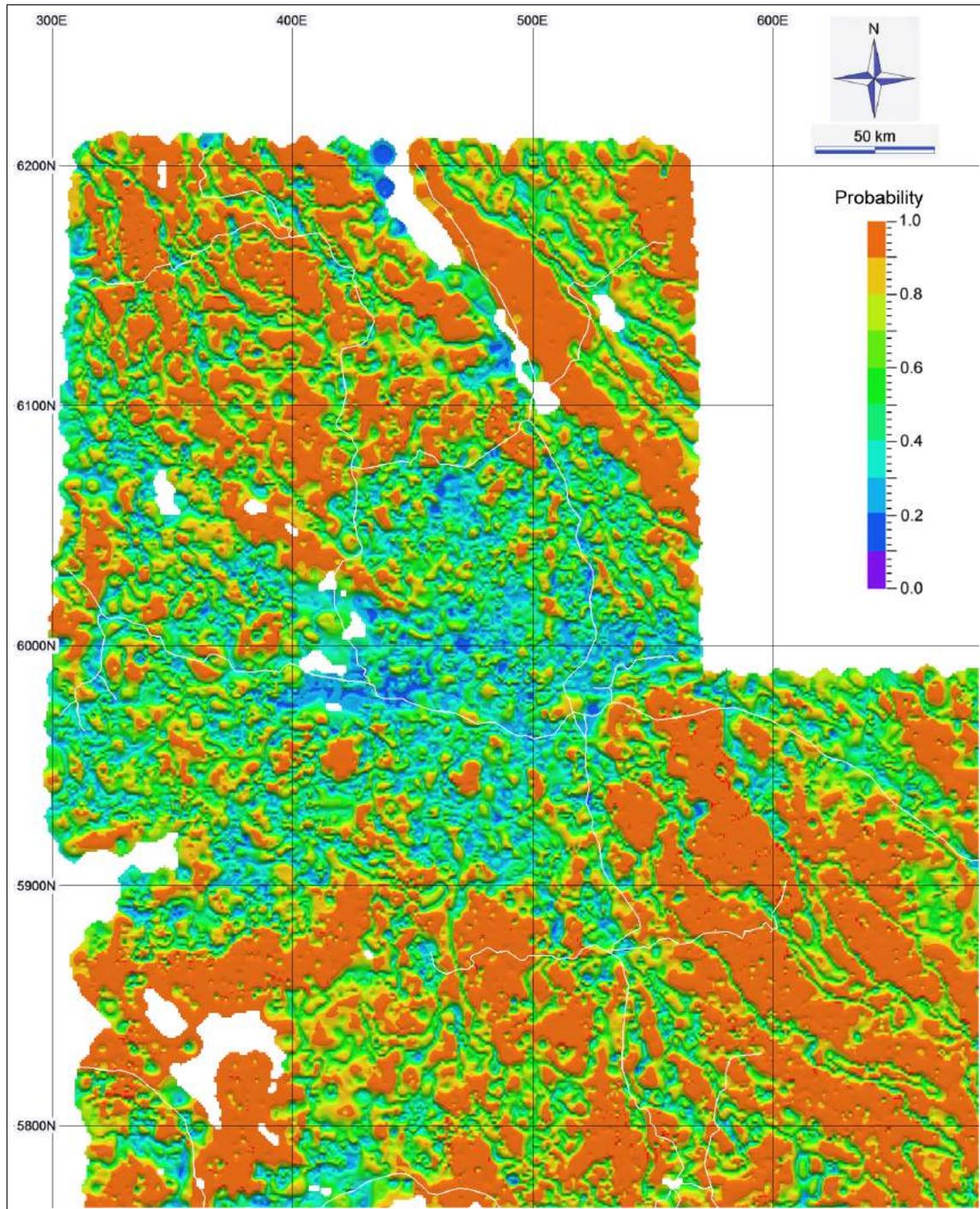


Figure 13: Probability of the inferred bedrock geology.

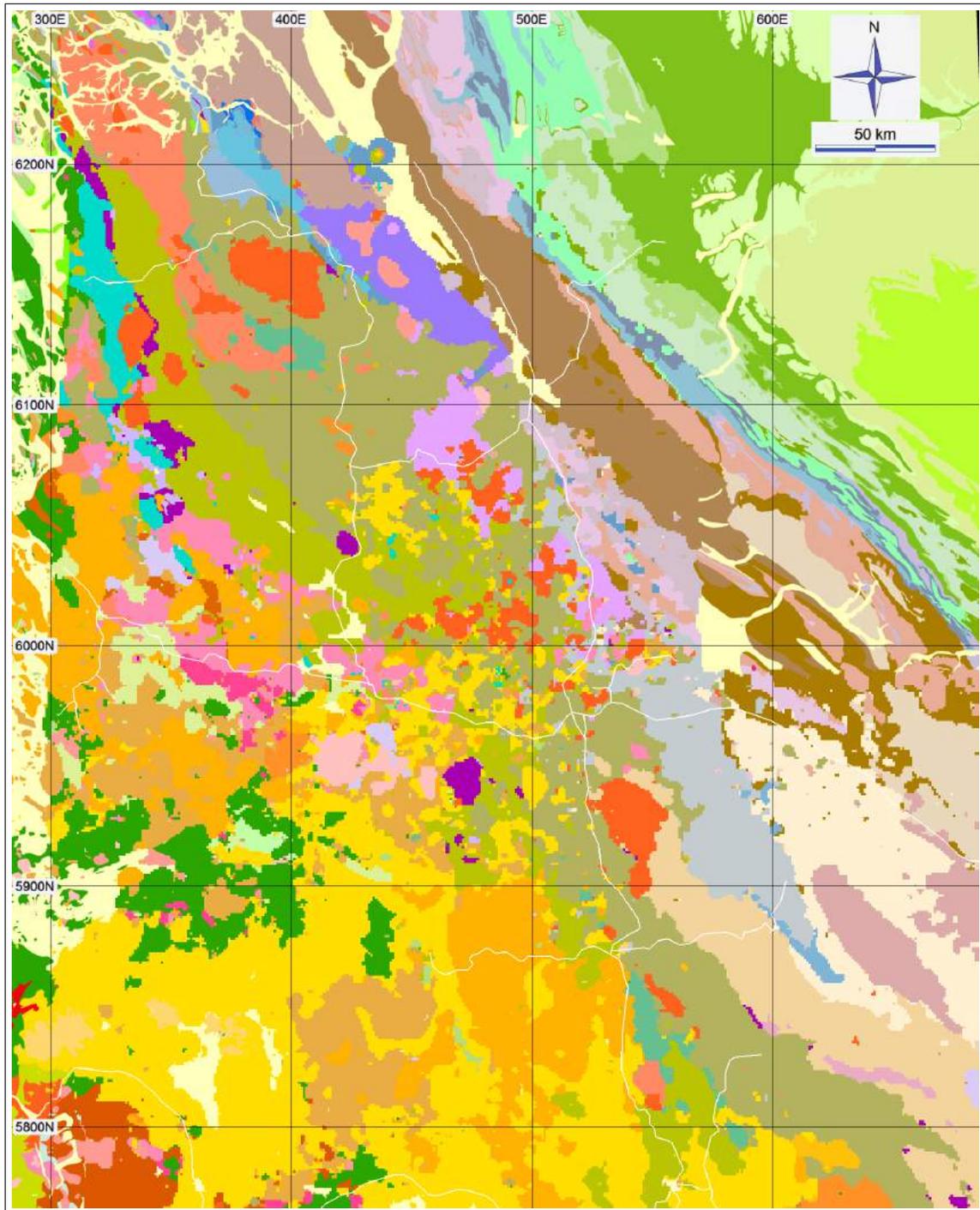


Figure 14: Inferred bedrock geology superimposed on mapped surficial geology.

Yet the 42-element geochemistry is clearly picking up the signatures of these intrusions through the glacial overburden as discrete, sharply bounded geological units.

The next logical step would be to integrate these geochemical results with the newly acquired geophysical data collected as part of the QUEST Project. The gravity, magnetic and electromagnetic data offer a completely independent view of the bedrock geology. In places they should support and corroborate the geochemical results. In other places, of course, one would expect the geophysics to pick out different patterns, based on density, susceptibility and conductivity variations that are not reflected in the geochemistry. But the results, nonetheless, should be extremely interesting and will hopefully lead to a new discovery.

## Acknowledgements

The authors would like to acknowledge the expertise and assistance provided by Wayne Jackaman of Noble Exploration Services Ltd, firstly in assembling the geochemical samples and secondly in advising on the selection of elements used in this study. They would also like to thank Stephen Williams, private GIS contractor, for compiling the surface geological map of the QUEST Project area that was used in the neural network approach described in this report. Both are members of the Geoscience BC Project Team.

## References

- [1] Aitchison, J., On criteria for measures of compositional differences, *Mathematical Geology*, **24**, 365-380, 1992.
- [2] Barnett, C. T. and Kowalczyk, P. L., Airborne electromagnetics and airborne gravity in the QUEST project area, Williams Lake to Mackenzie, British Columbia, *Geoscience BC Summary of Activities 2007*, 1-6, 2007.
- [3] Jackaman, W. and Balfour, J. S., QUEST project geochemistry: field surveys and data re-analysis (parts of NTS 93A, B, G, H, J, K, N and O), *Geoscience BC Summary of Activities 2007*, 7-10, 2007.
- [4] Logan, J., Ullrich, T., Friedman, R., Mihalynuk, M. and Sciarizza, P., Quesnel Arc—porphyry setting, potential and discoveries, *Mineral Exploration Roundup07, Vancouver, British Columbia*, abstract volume, page 34, 2007.
- [5] Massey, N. W. D., MacIntyre, D. G., Desjardins, P. J. and Cooney, R. T., Geology of British Columbia, 1:1,000,000 scale, *British Columbia Geological Survey*, Geoscience Map 2005-3, 2005.
- [6] Ng, A., Jordan, M., Weiss, Y., On spectral clustering: analysis and an algorithm, in: Dietterich, T., Becker, S., Ghahramani, Z., eds., pp. 849-856, *Advances in Neural Information Processing Systems 14*, The MIT Press, 2002.

[7] von Luxburg, U., A tutorial on spectral clustering, *Statistics and Computing*, **17**, 395–416, 2007.

[8] Williams, P. M., Bayesian regularization and pruning using a Laplace prior, *Neural Computation*, **7**, 117–143, 1995.

[9] Williams, P. M., Using neural networks to model conditional multivariate densities, *Neural Computation*, **8**, 843–854, 1996.

## List of Figures

1	Location of the QUEST Project area in Central British Columbia. . . . .	3
2	Surficial geological map of the QUEST project area, showing a few of the towns. The pale yellow areas represent the Quaternary overburden. The outline of the airborne gravity and electromagnetic surveys is marked in black. The map projection is UTM Zone 10 in NAD83. . . . .	4
3	Raw gold assays from stream and lake sediments in the QUEST Project area.	7
4	Scatter plot of matched decile points from neighboring strips of the streams and lakes populations. The straight line $y = mx + c$ has parameters $m = 0.4971$ and $c = 0.1415$ . . . . .	9
5	Levelled gold assays from stream and lake sediments. . . . .	10
6	Levelled copper assays from stream and lake sediments. . . . .	11
7	Levelled molybdenum assays from stream and lake sediments. . . . .	12
8	Measured and estimated molybdenum assays. . . . .	14
9	K-means clustering of 42 element geochemistry using 30 clusters. . . . .	17
10	Mapped surficial geology with contacts and topography removed. . . . .	18
11	Spectral clustering of 42 element geochemistry using 30 clusters. . . . .	19
12	Inferred bedrock geology of the QUEST Project area. . . . .	21
13	Probability of the inferred bedrock geology. . . . .	23
14	Inferred bedrock geology superimposed on mapped surficial geology. . . . .	24

## List of Tables

1	The 22 elements for which ICP-MS was the preferred analytic method. . . . .	6
2	The 20 elements for which INAA was the preferred analytic method. . . . .	8
3	Table showing the average sizes of error bars when estimating the 25 partially missing elements. . . . .	15