Research Review Day, Harbour Centre
September 17, 2009

Target generation from QUEST geochemical Data using the CSIRO SOM analysis

Peter Kowalczyk
GBC Project Team
kowalczyk@geosciencebc.com

Geoscience BC Report 2009-14
"An Investigation Using SiroSOM for the Analysis of QUEST Stream-Sediment and Lake-Sediment Geochemical Data"

Released Sept 16, 2009
http:\\www.geosciencebc.com

The problem

- The Quest geochemical data release comprises 15000+ samples, Lake and Stream sediments, collected in a varied background of different rock packages, glacial and fluvial overburden.
- 42 different elements have been analyzed and reported upon.
- How does one summarize this data, and identify more than the most obvious of anomalies, using the full information content of the data.

Geoscience BC

SOM: Self Organizing Maps

Wikipedia says...

A self-organizing map (SOM) ... is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map.

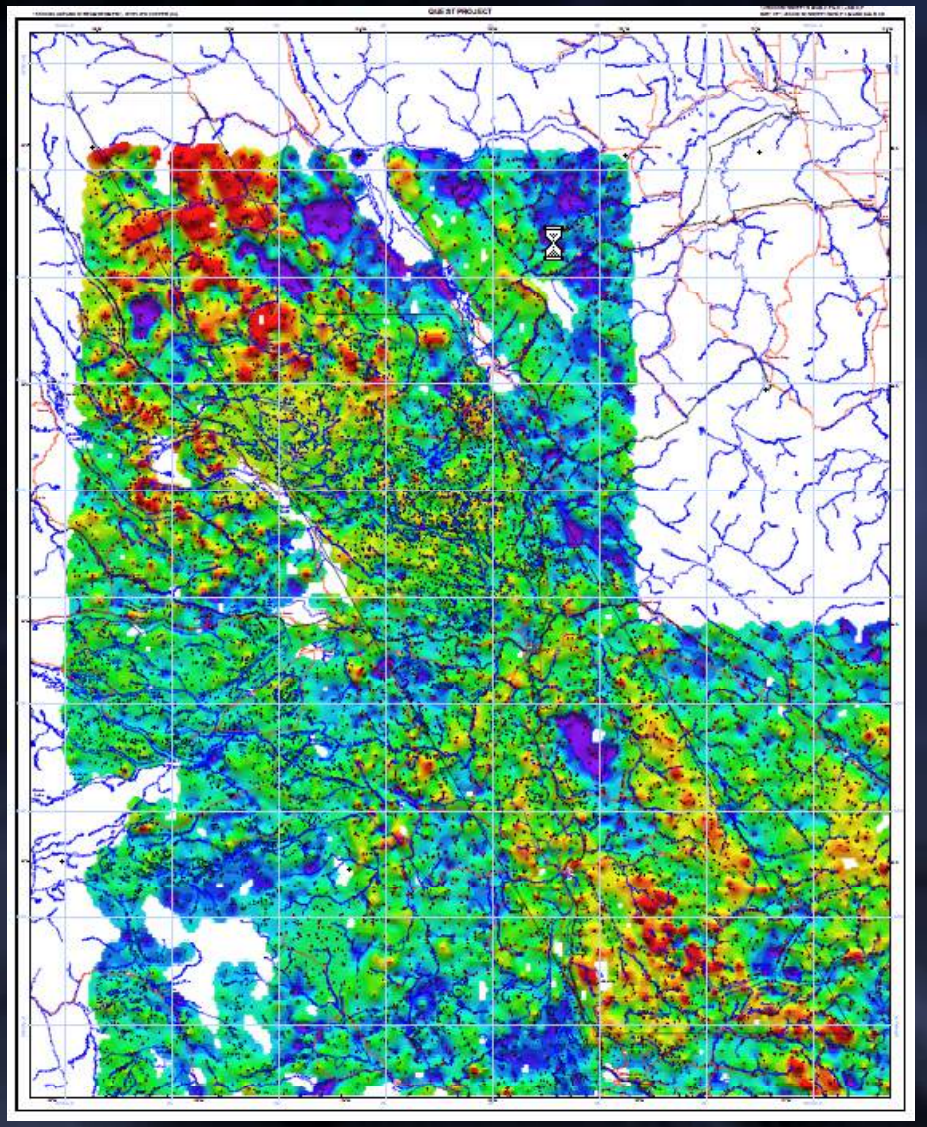
SOM forms a ... map where similar samples are mapped close together and dissimilar apart.

SOM may be considered a nonlinear generalization of principal components analysis (PCA)

It has been shown, using both artificial and real geophysical data, that SOM has many advantages over the conventional feature extraction methods such as Empirical Orthogonal Functions (EOF) or PCA

SOM analysis process

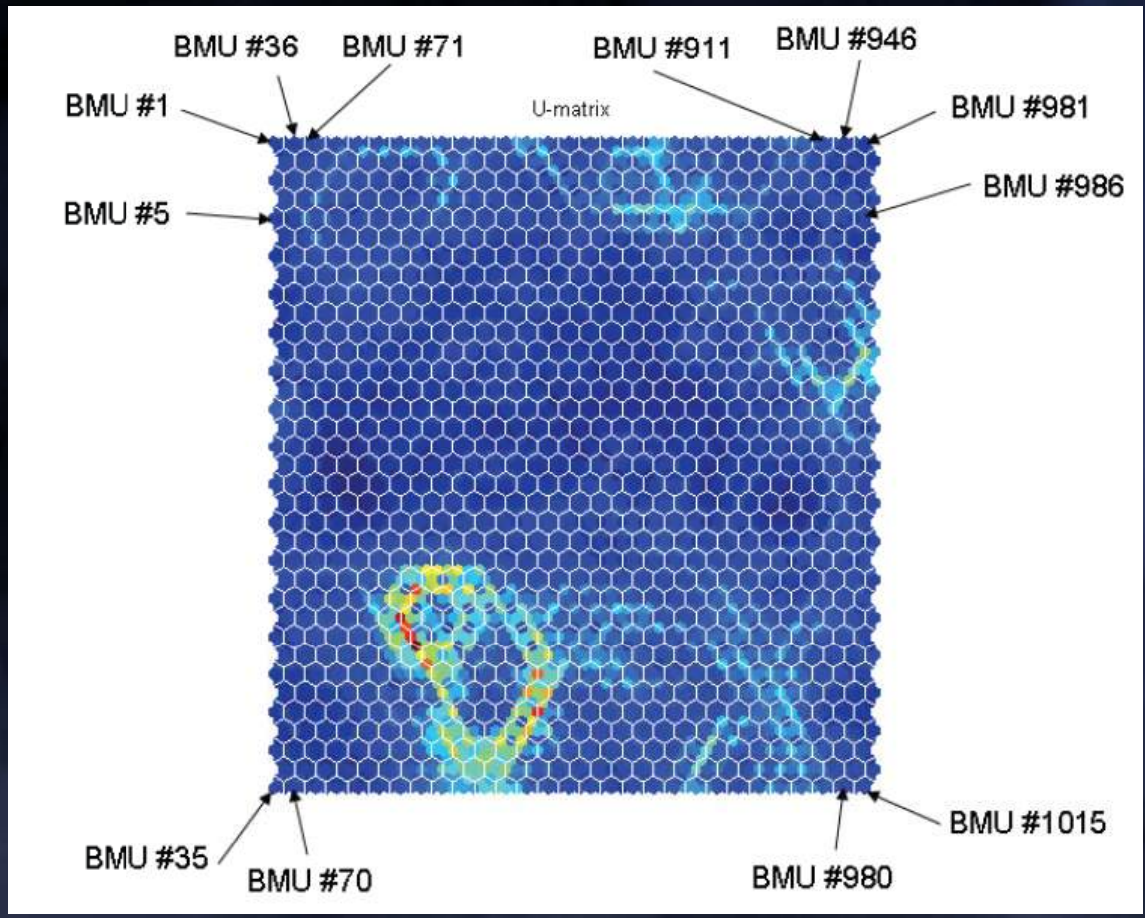
- Start with the data
  - Use best summary data: levelled grids prepared by Barnett and Williams (GBC report 2009-3)
  - Extract a value for each element for each sample site
  - This produces 15000+ sample sites with 42 elements for each site.
  - note: this is not the original raw data



Geoscience BC

SOM analysis process

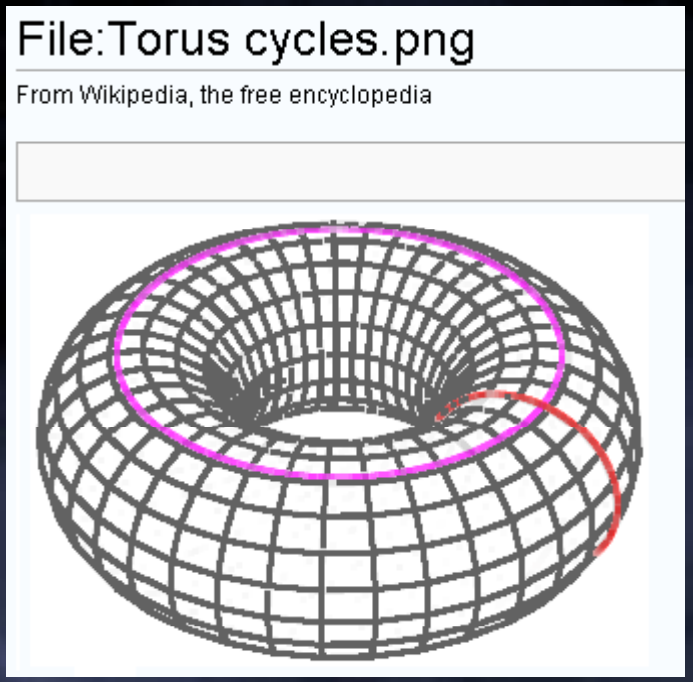
- Each data point is mapped to a Best Matching Unit (BMU) by the SOM algorithm. This is a new 2D coordinate system for the 42 element geochemical vector.
- 15000+ samples have been mapped to 1015 BMU's
- Samples mapped to a BMU are similar to each other
- BMUs close to each on the the U-matrix map are similar. The U-matrix displays high values between BMUs that are dissimilar, and small values between similar BMUs.

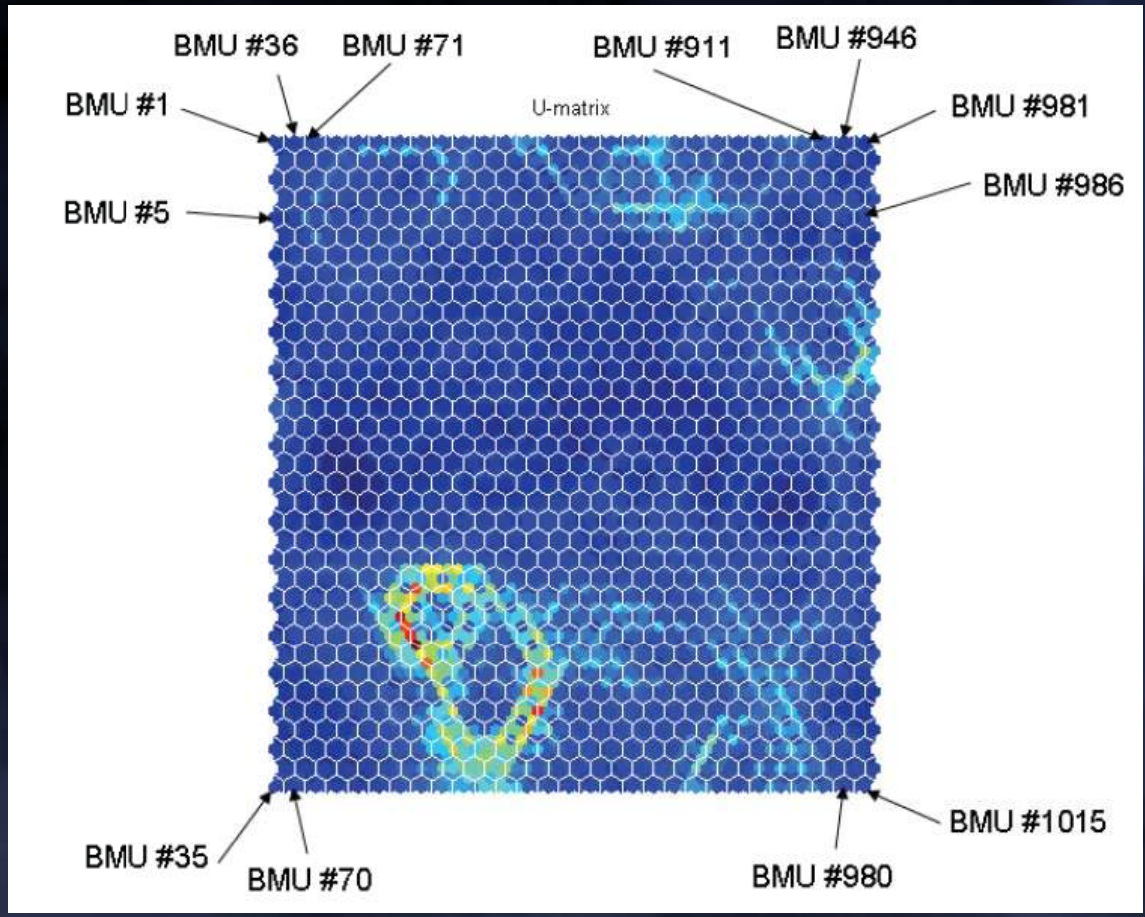


Geoscience BC

SOM analysis process

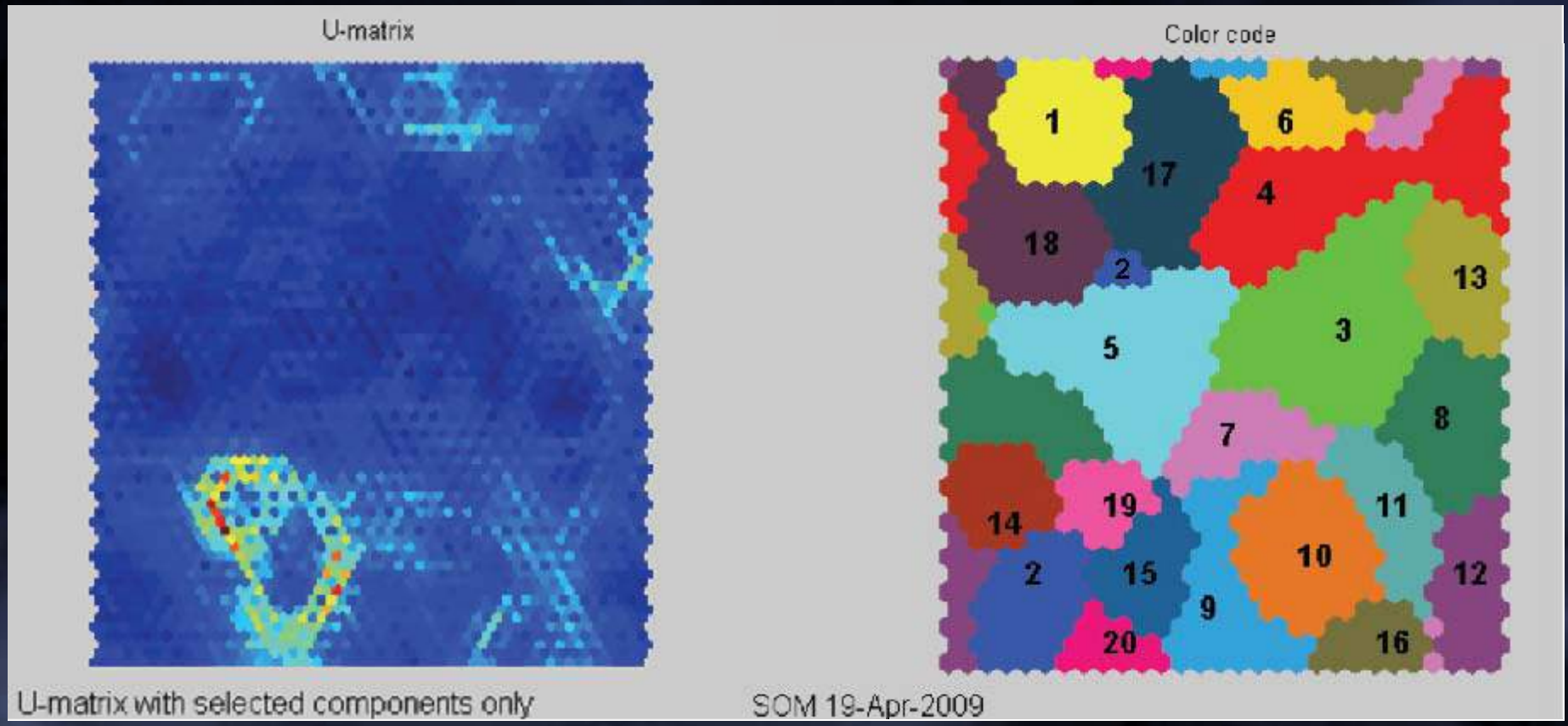
- Note that the U-matrix was mapped onto a toroidal projection surface, the top wraps to the bottom, and the right side to the left side



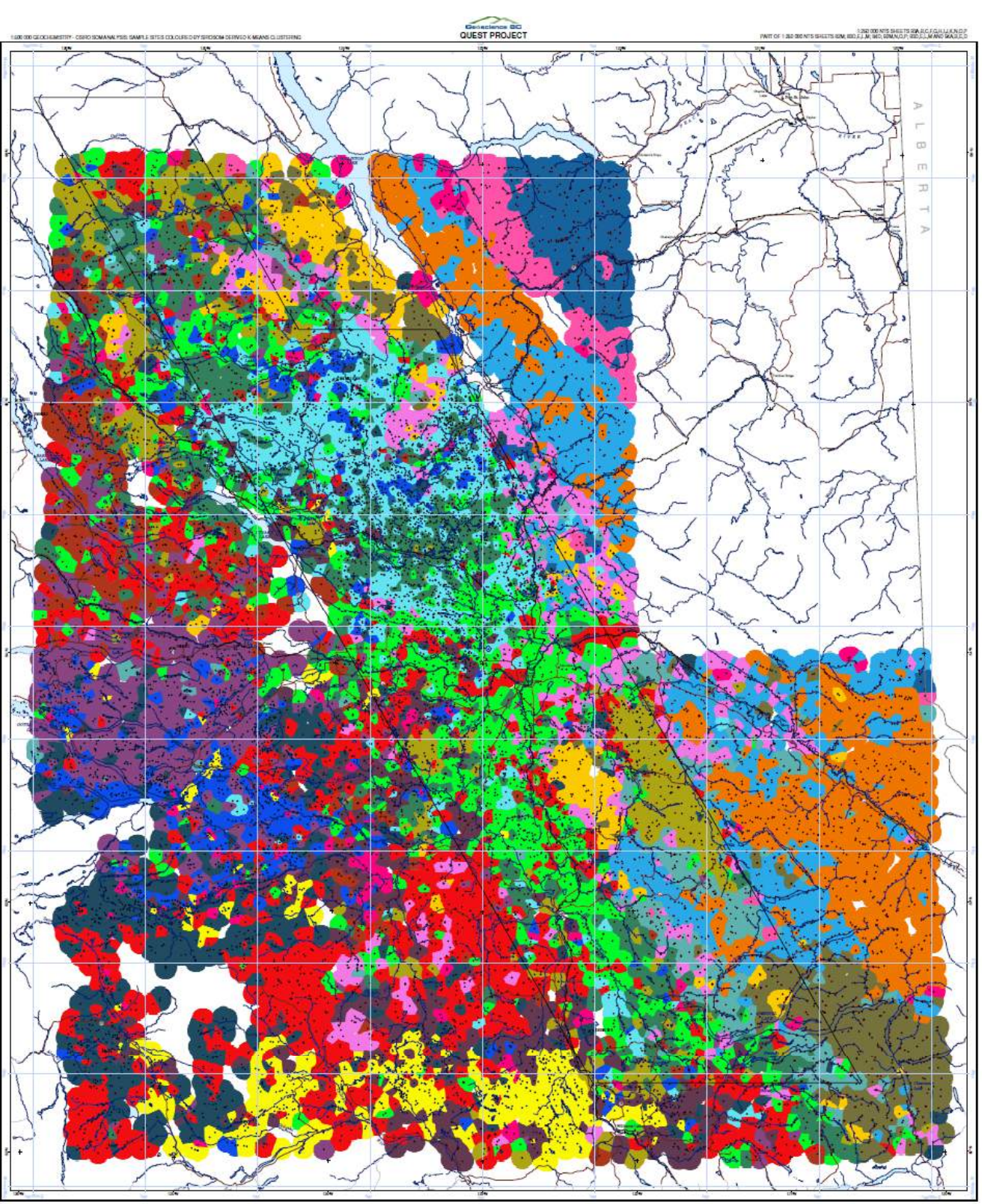


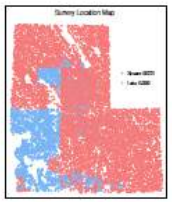
SOM analysis process

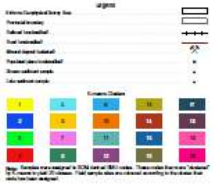
- A geochemical value for each BMU can be calculated using the average of all the samples mapped to it.
- These values are used to cluster the BMUs into classes using K-means



Geoscience BC







GEOSCIENCE BC - QUEST - GEOCHEMISTRY - CSIRO SOM ANALYSIS

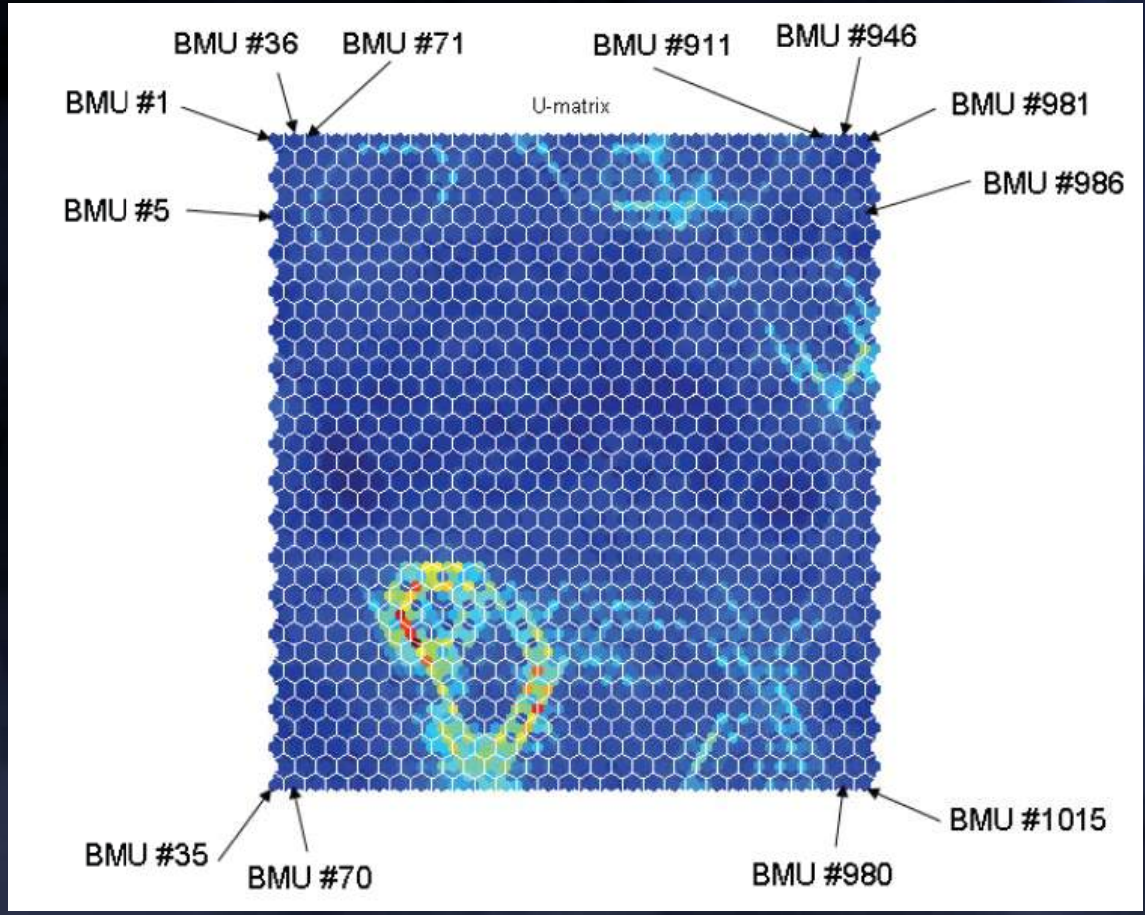
- The map of K-means clusters shows samples that are similar
- Note clustering was done on the BMU values, not the samples
- GBC Map:

2009_14_01_Kmeans.pdf

Geoscience BC

SOM analysis process

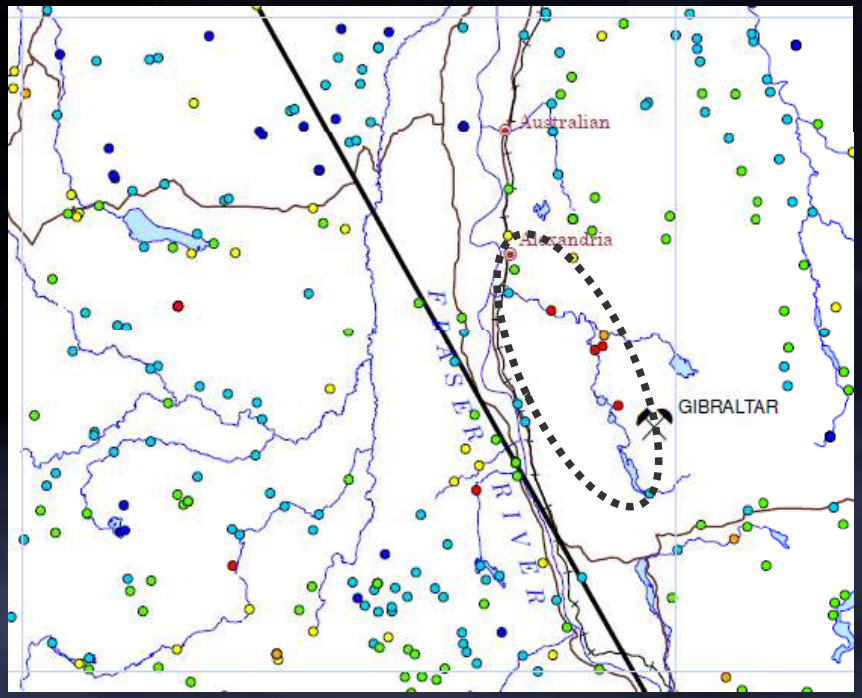
- Each data point is mapped to a Best Matching Unit (BMU) by the SOM algorithm. This is a new 2D coordinate system for the 42 element geochemical vector.
- The map is not perfect, each sample's 42 vector coordinate is a little different from that of the BMU.
- The distance of a sample from its BMU in 42 vector space is the Quantization Error (Q-Error).
- Samples with a high Q-Error are unusual, and may be mineralized.



Geoscience BC

SOM analysis process

- Quantization error is calculated for each sample when it is mapped to a Best Matching Unit (BMU) on the SOM U-matrix.
  - In this example note that the samples down-stream from Gibraltar mine have a high Q-Error. This indicates they are atypical geochemical samples – i.e. they are anomalous.

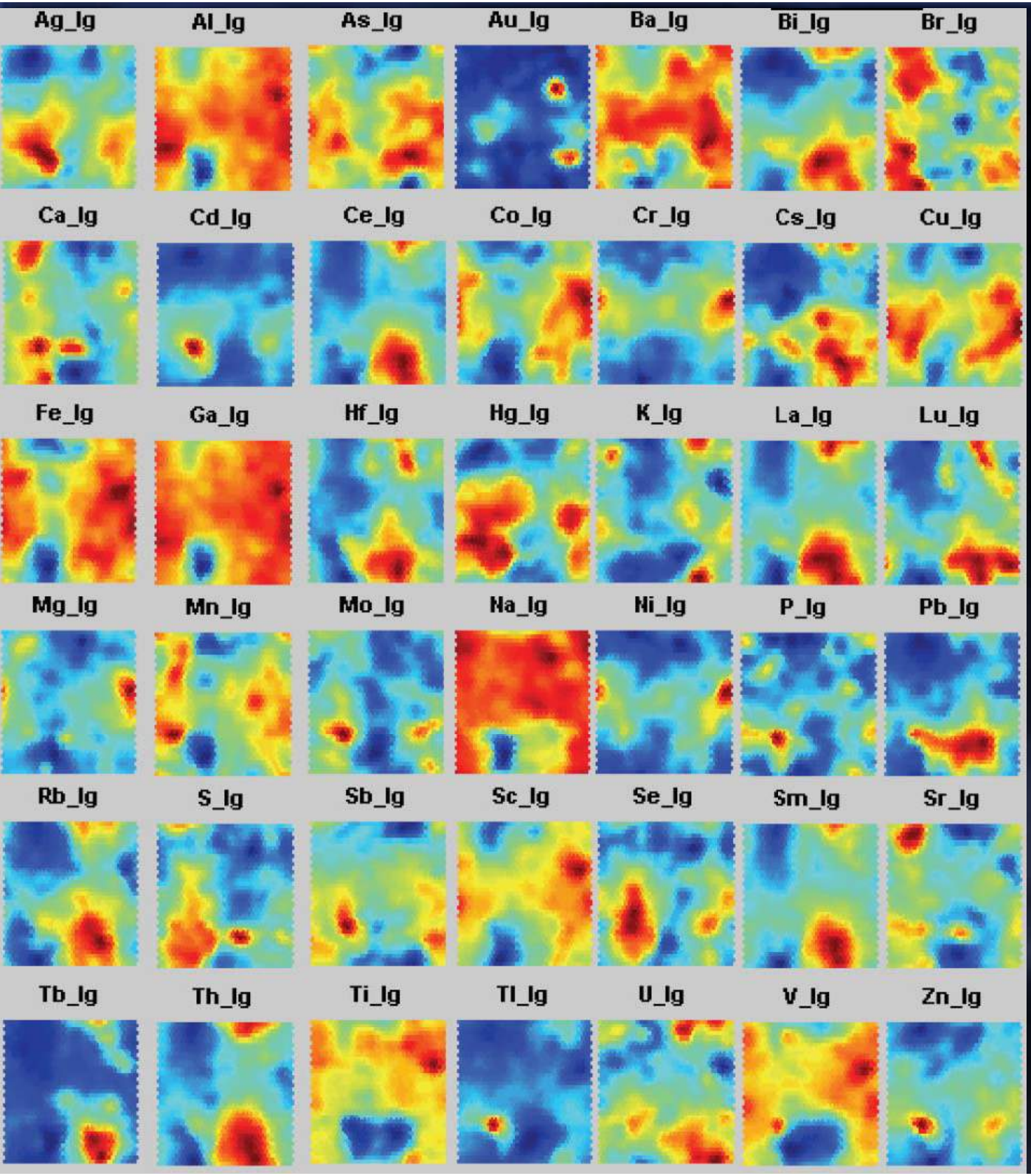


Quantization Error Value Breakdown

> 6	
5 - 6	
4 - 5	
3 - 4	
2 - 3	
0 - 2	

Extracted from GBC map 2009_14_02_QError.pdf

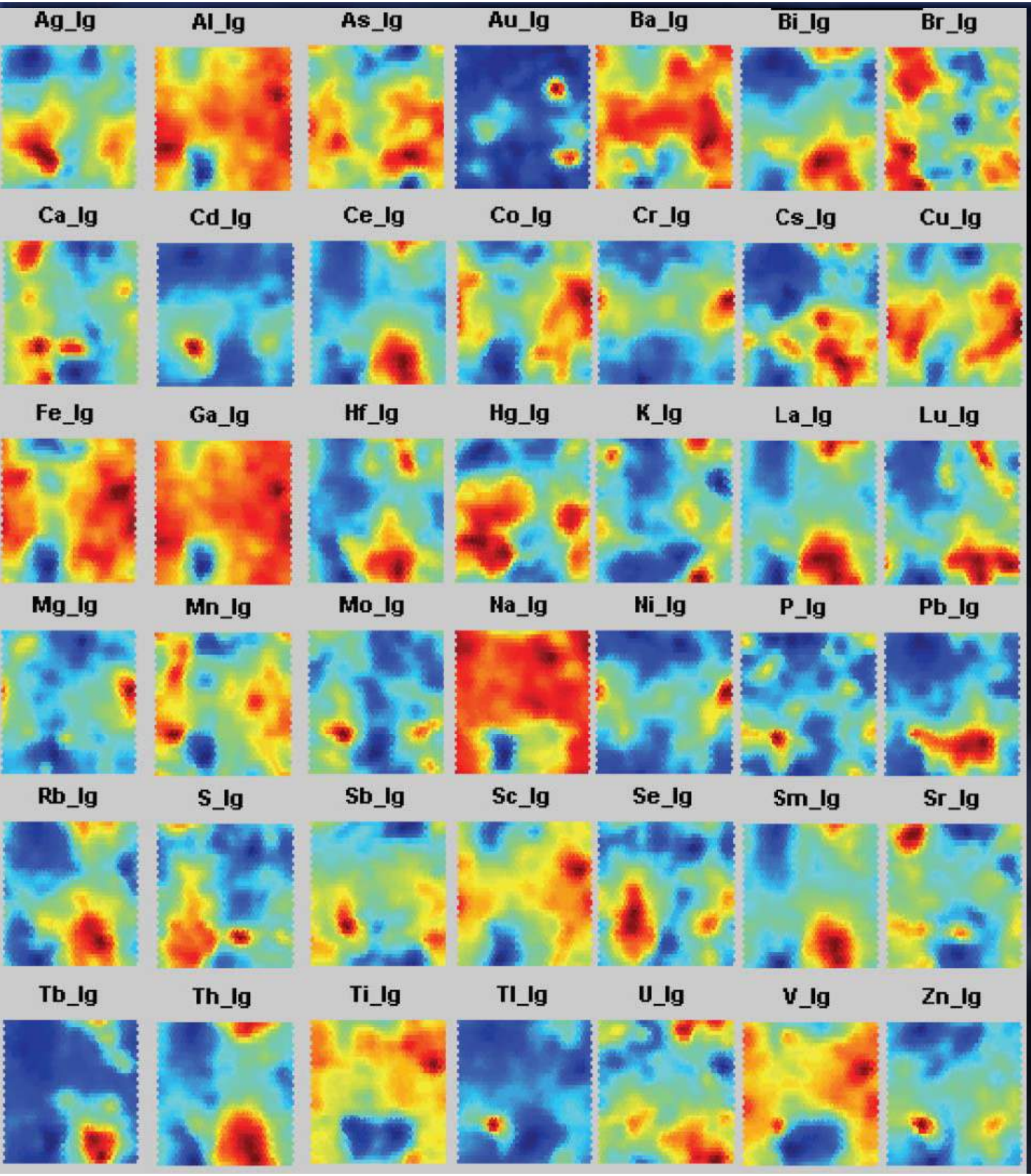
Geoscience BC



U-matrix component plots

The elemental values for each BMU can be shown on the U-matrix. This illustrates the elemental associations of samples that report to each BMU.

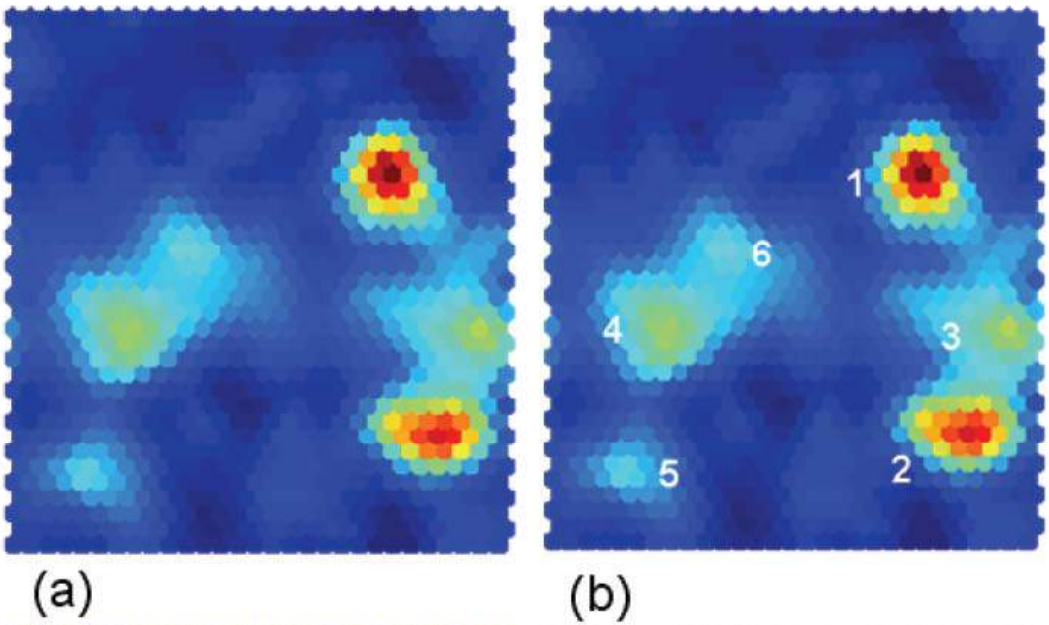
Geoscience BC



The K means clustering has identified BMUs that have similar elemental associations

These clusters represent different populations of geochemical samples, and anomalous samples can be defined for each k-means cluster

Geoscience BC



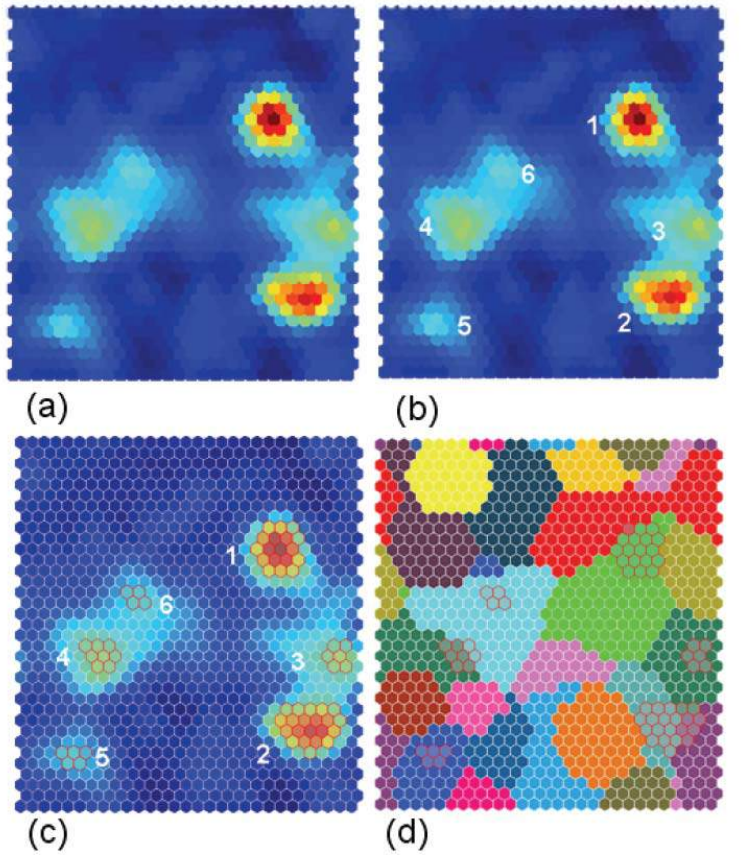
Targeting gold:

Six groups of BMU's are identified as anomalous in gold

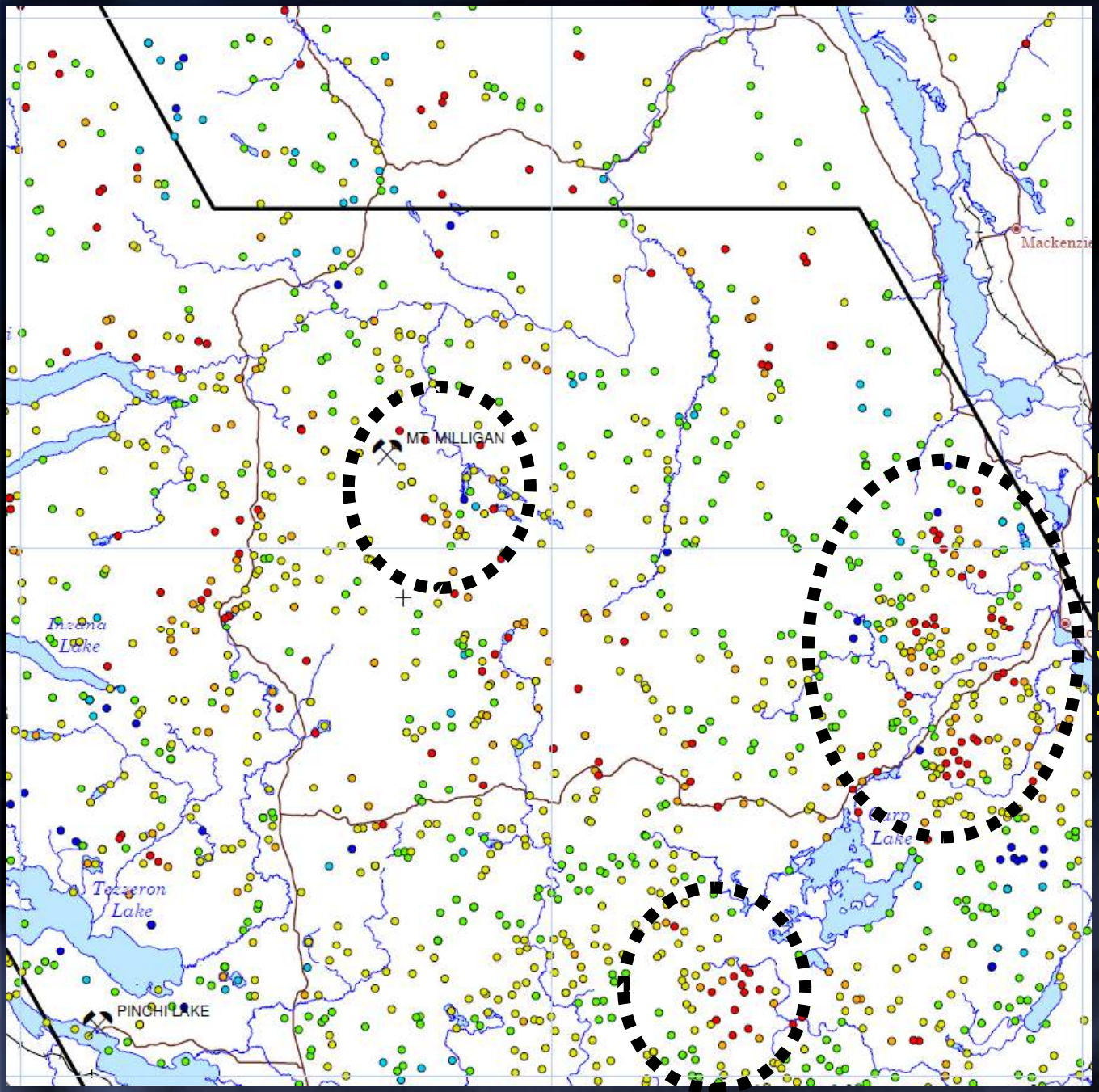
These BMU's are selected, and then plotted on a map.

Geoscience BC

Group 6 of the anomalous gold BMUs all plot around Mt. Milligan



Geoscience BC



Gold anomalism:

All samples are normalized by their standard deviation from the norm of the BMU they map to. Samples of high interest will be anomalous samples from BMUs in K-means clusters with higher than average gold, or samples with a high Q-error.

Note that BMUs with low gold values will still contain highly anomalous samples! Anomalous samples in clusters with little gold probably represent low, but locally anomalous values of gold and are unlikely to be economic gold targets.

Gold Anomalism Value Breakdown

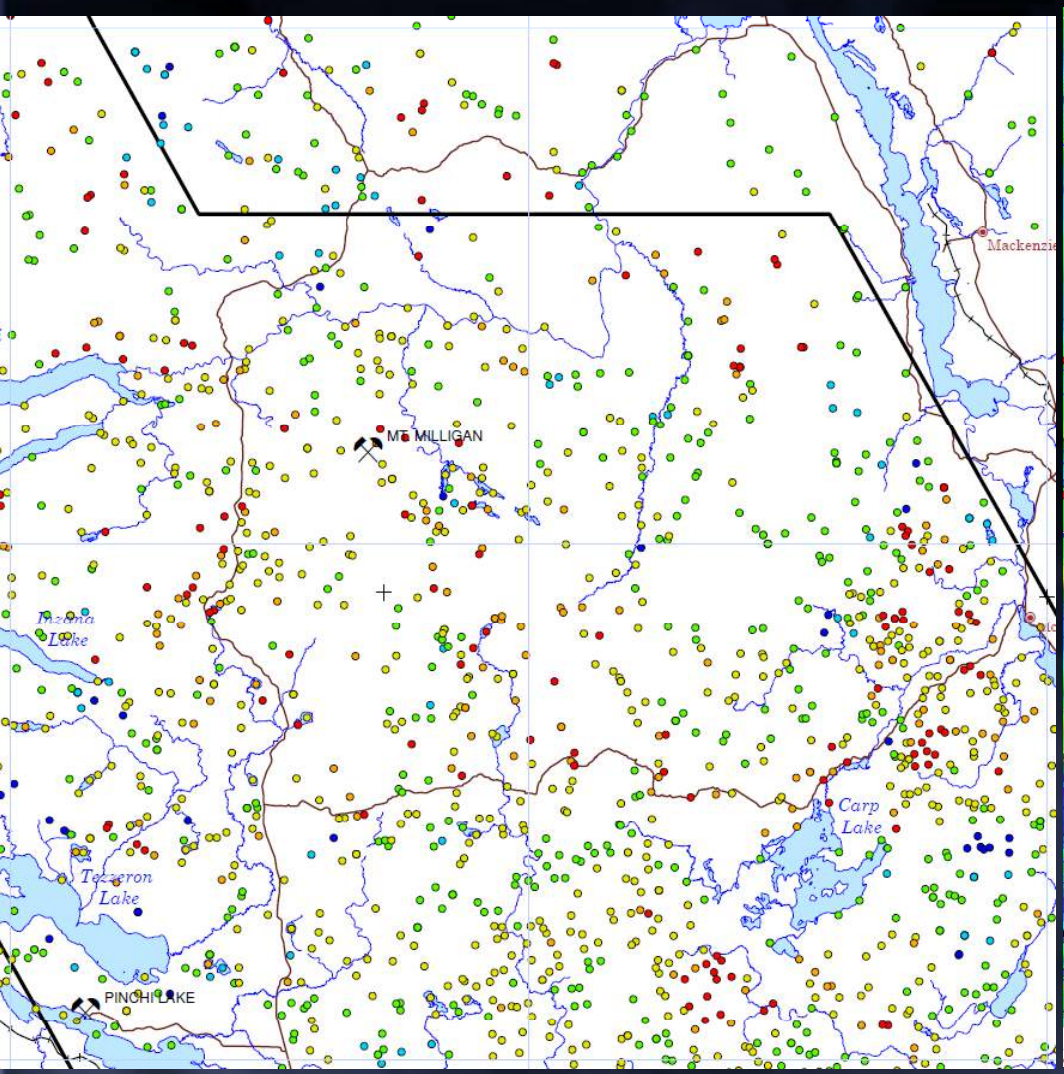
>	1.5
1 to	1.5
0 to	1
-1 to	0
-1.5 to	-1
-10 to	-1.5

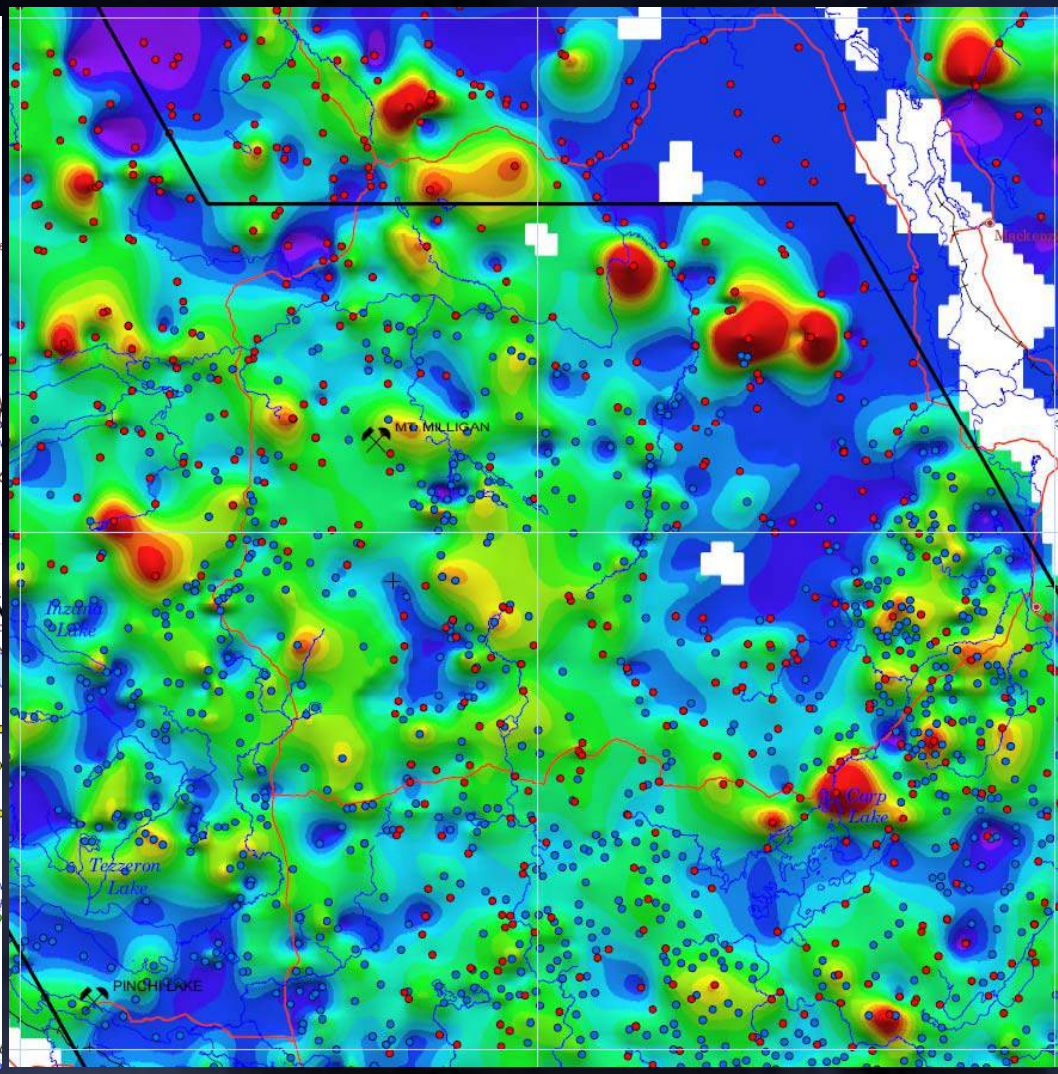
values above cluster norm

values below cluster norm

Geoscience BC

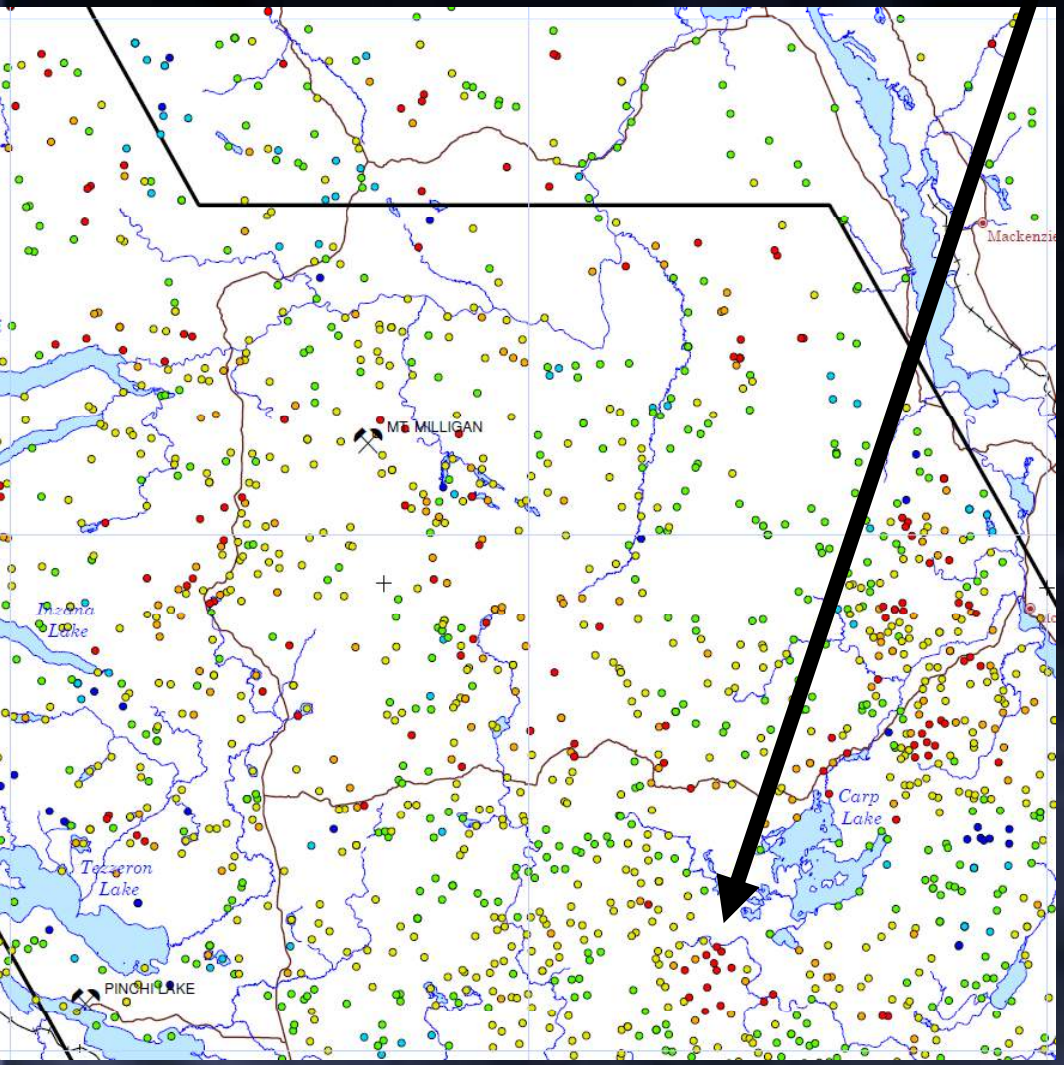
Gold:
A comparison of cluster normalized anomalism with the original levelled gold values



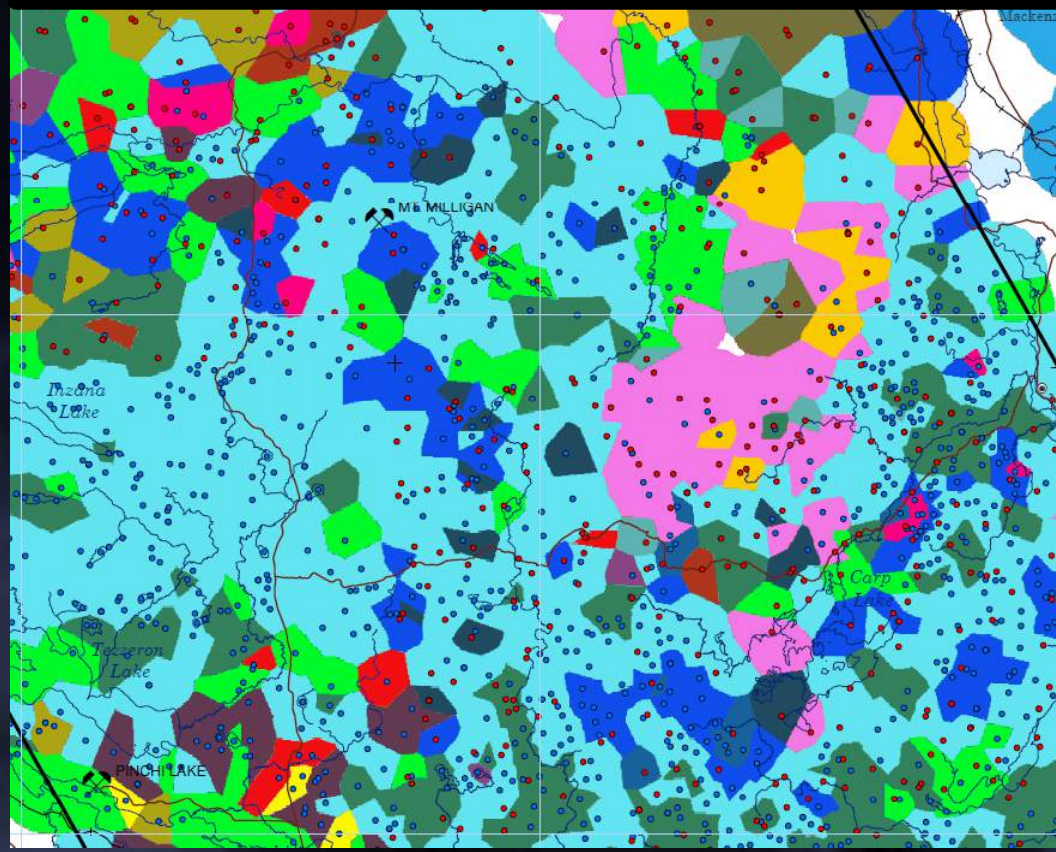


Geoscience BC

Gold:
A comparison of cluster normalized anomalism with K-means cluster classes



Note that samples normalized by their cluster averages identify areas of anomalous sample sites that cross K-means cluster boundaries. Differences between stream and lake samples are minimized.



Geoscience BC

Make your own maps using the digital data!

K-Means cluster number of BMU data point reports to

Q-Error of data point

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Easting	Northing	somx	somy	BMU	Cluster	q-error	UTME_83	UTMN_83	Ag_lg	Al_lg	As_lg	Au_lg	Ba_lg
2	301918	5946149	26	29	904	11	4.29	301918	5946149	-0.89	0.18	1.57	1.37	2.93
3	302058	5943340	25	28	868	11	12.08	302058	5943340	-0.77	0.13	2.14	2.86	2.93
4	302631	5945658	26	28	903	12	4.53	302631	5945658	-0.89	0.16	1.55	1.48	2.89
5	302635	5929077	13	11	431	17	2.04	302635	5929077	-1.09	0.05	0.48	0.31	2.81
6	302718	5967570	28	28	973	12	2.31	302718	5967570	-1.09	0.07	1.37	0.6	2.92
7	303030	5965795	27	28	938	12	3.25	303030	5965795	-1.19	0.06	1.3	1.01	2.96
8	303131	5955449	28	30	975	12	2.1	303131	5955449	-1.17	0.07	0.94	0.34	2.94
9	303131	5955449	28	30	975	12	2.1	303131	5955449	-1.17	0.07	0.94	0.34	2.94
10	303131	5955449	28	30	975	12	2.1	303131	5955449	-1.17	0.07	0.94	0.34	2.94

BMU number & SOM coordinates of data point

Map coordinates

Levelled data used as input, 42 elements Log(10) values

Geoscience BC

Caveats

- This is not magic - it is statistics.... be careful of implying too much.
- The data is stream and lake geochemistry, look at the map to see what the local situation is. Don't give up the basics of geochemical interpretation.
  - Where is the drainage coming from?
  - Is the sample contaminated by a local source?
- Is the anomaly real? Do you see it in the raw data now you have been directed to it by the SOM analysis?
  - [remember normalization makes small numbers big!]

Thank you and good luck!

Acknowledgements to

Stephen Fraser & Jane Hodgkinson, CSIRO
Colin Barnett & Peter Williams, BW Mining

Geoscience BC