Developments in Geochemical Data Processing and Presentation:

Some Useful Tools for Evaluating Multi-element Geochemical Data

Eric Grunsky
Geological Survey of Canada

International Geochemical Exploration Symposium
Dublin, Ireland
03-Sep-2003

Natural Resources Canada Ressources naturelles Canada

Canada

Geochemical Survey Data

- This presentation is concerned with the interpretation of multi-element geochemical survey data.
- Geochemical survey data are collected over the geographic domain from a variety of geologic media.
- The intent of these surveys are to characterize the background and identify atypical observations that associated with:
  - potential mineralization; and
  - environmental hazards.

Natural Resources Canada Ressources naturelles Canada

Canada

Geochemical Space

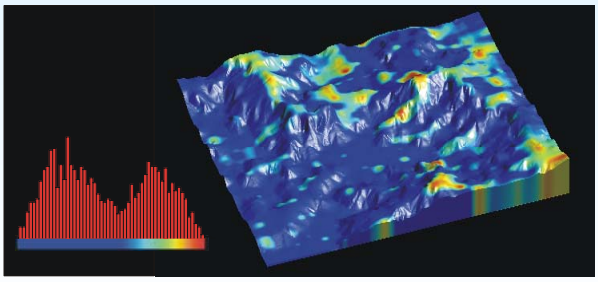
- Numerous graphical methods can characterize the variables. (Q-Q plots, box & whisker, scatter plots, dot plots, histograms, scatter plot matrices)
- Data analysis, statistics and data visualization assist in extracting structure from the data.
- Inter-element relationships of geochemical data reflect mineralogy and infer geological processes.

Natural Resources Canada Ressources naturelles Canada

Canada

Geographical Space

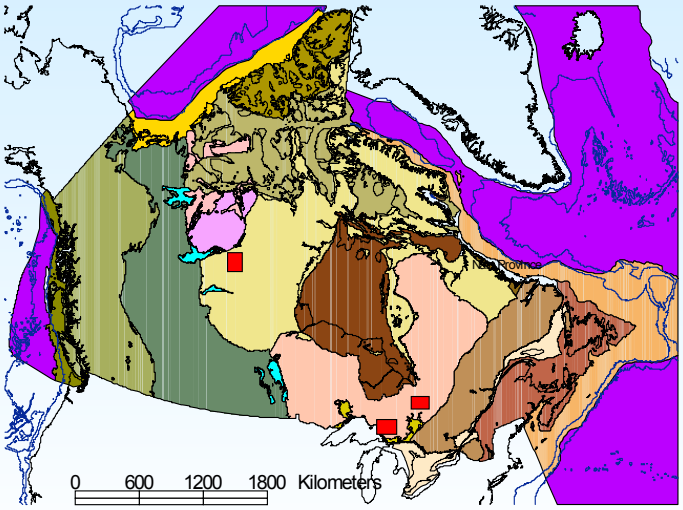
- Humans relate visually to the spatial representation of concepts.
- Interpretation is enhanced when symbols or imagery are represented in the geospatial domain.



Natural Resources Canada Ressources naturelles Canada

Canada

Examples of Exploratory Data Analysis in Canada



Natural Resources Canada Ressources naturelles Canada

Canada

Goals of Geochemical Data Analysis

- Detect inter-element associations and data structures that correspond to spatially-based geologic processes.
- Isolate atypical observations or groups of observations that are potentially identified with processes of interest (mineral deposits, hazardous environments)
- Interpret multi-element associations in the context of mineralogy and/or solute speciation.

Approaches to Investigate Geochemical Data

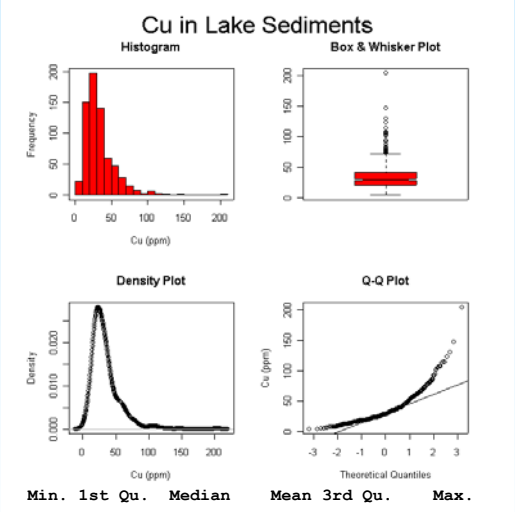
- Exploratory Approach
  - investigate and characterize data
- Modelled Approach
  - test groups of samples against known parameters and characteristics associated with a target (mineralization or hazard)
  - MANOVA, Discriminant methods, Allocation / Typicality

Natural Resources Canada Ressources naturelles Canada

Canada

Univariate Exploratory Approach

- Histogram
- Box plot
- Density plot
- Q-Q plot
- Order Statistics



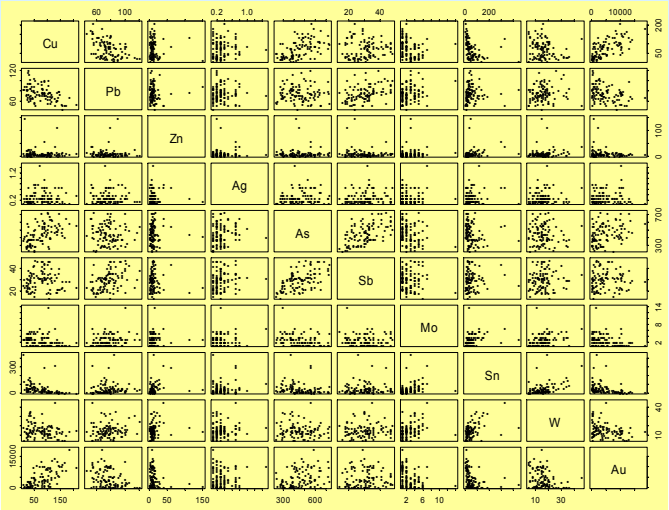

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4.00   20.00   29.00   34.34   41.00  204.00
```

Natural Resources Canada Ressources naturelles Canada

Canada

Multivariate Exploratory Approach
Scatterplot Matrix

Golden Grove, Western Australia



Natural Resources Canada
Ressources naturelles Canada

Canada

Multivariate Exploratory Approach

- Principal components analysis
- D^2 (χ^2) plots
- Empirical Indices
  - Weighted Sums

Natural Resources Canada Ressources naturelles Canada

Canada

Special Problems

- Censoring - samples < detection limit, problems in estimating means & variances,
- Non-normal distributions that hamper accurate statistical measures,
- Missing values,
- Different limits of detection and instrumentation – levelling,
- Constant sum (closure) problem.

Natural Resources Canada Ressources naturelles Canada

Canada

Non-normal Distributions: Transformation of Data

- Standard statistical procedures assume a normal distribtuion.
- Most geochemical data distributions are non-normal and positively skewed - mixtures of populations.
- Data may be transformed before statistical methods are used.
- Transforming the data minimizes the effects of outliers.
- Commonly used method is Box-Cox Power Transformation

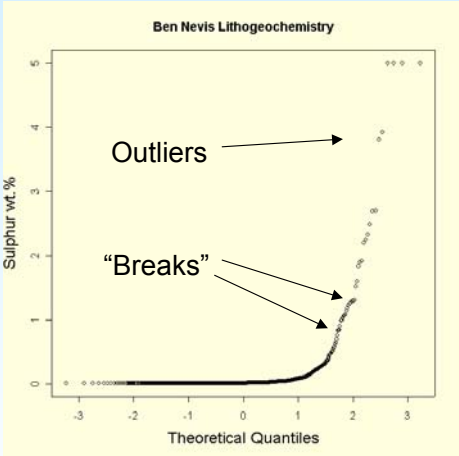
$$y = (x^{\lambda} - 1)/\lambda \text{ for } \lambda>0$$

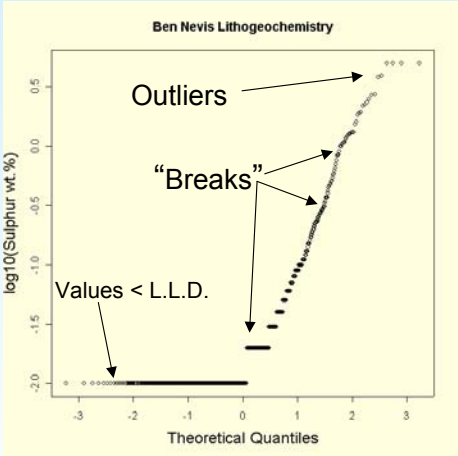
$$y = \ln(x) \text{ for } \lambda=0$$

Natural Resources Canada Ressources naturelles Canada

Canada

Q-Q Plots & Transformations





Natural Resources Canada Ressources naturelles Canada

Canada

Geochemical Definitions

Threshold - outer limit of background variation.

- Background variation determined from orientation studies over area.
- Assumes adequate representation of background exists.
- Do not use mean ± 2 standard deviations.
- Use percentiles (e.g., 95th or 98th percentile)

Anomaly - a value that is beyond the threshold.

Target Population - samples within and surrounding an area of interest.

- characteristic geochemistry of the deposit or hazard
- each mineral deposit type/hazard should have its own target population.

Background Population - samples from a regional survey.

- lithological variation and other regional geochemical effects.
- not related to mineralization, hazard or other atypical event.

Natural Resources Canada Ressources naturelles Canada

Canada

Robust Estimation

Estimates of the means, variances and covariances, etc. in which the presence of outliers is down-weighted or eliminated.

Univariate: Trimmed Means, L-Estimates, Adaptive Trimmed Means, Huber W-Estimates, Dominant Cluster Mode

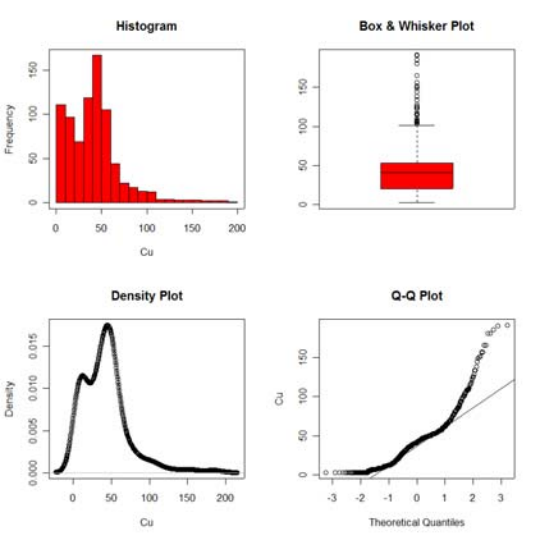
Multivariate: Minimum Volume Ellipsoid, Minimum Covariance Determinant

Any choice of robust estimate is better than none at all.

Robust Estimation

Lake Sediment Geochemistry

Method		n
Mean	56	825
Median	42	825
Trimmed Mean	40	783
Trimmed Median	40	783
Huber M-estimator	40	825
Min.Vol. Ellipsoid	38	825
Min. Cov. Det.	NA	800
Min. Cov. Det.	39	540



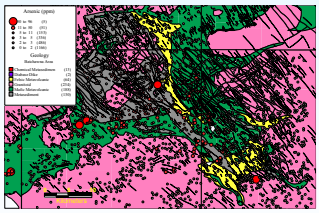
MVE & MCD based on: S, Ba, Co, Cr, Cu, Li, Ni, Pb, Zn, Sr, V, Y, Zr

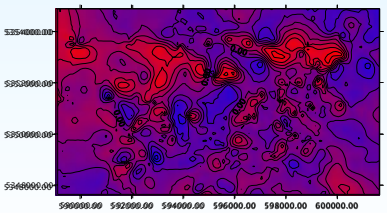
Natural Resources Canada Ressources naturelles Canada

Canada

Spatial Presentation of Data

- Bubble Plots - Presentation of values based on size of symbols.
- Contouring - Lines of equal value drawn from interpolated or triangulated data.
- Imaging - Raster or gridded data converted to pixels and assigned integer values for image display systems.
- Imaging with 3D rendering (usually with Digital Elevation Data)
- Spatial Continuity - adequate sample density?

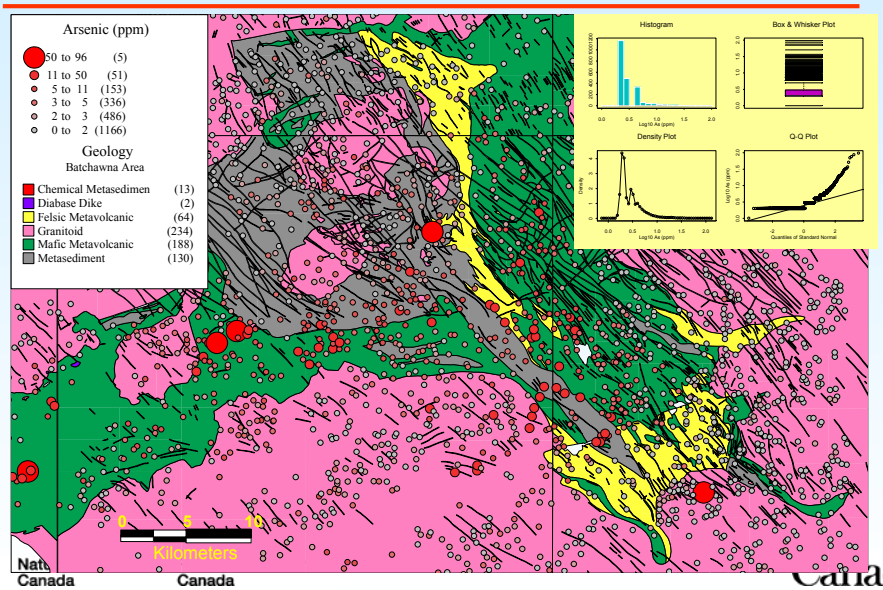




Natural Resources Canada Ressources naturelles Canada

Canada

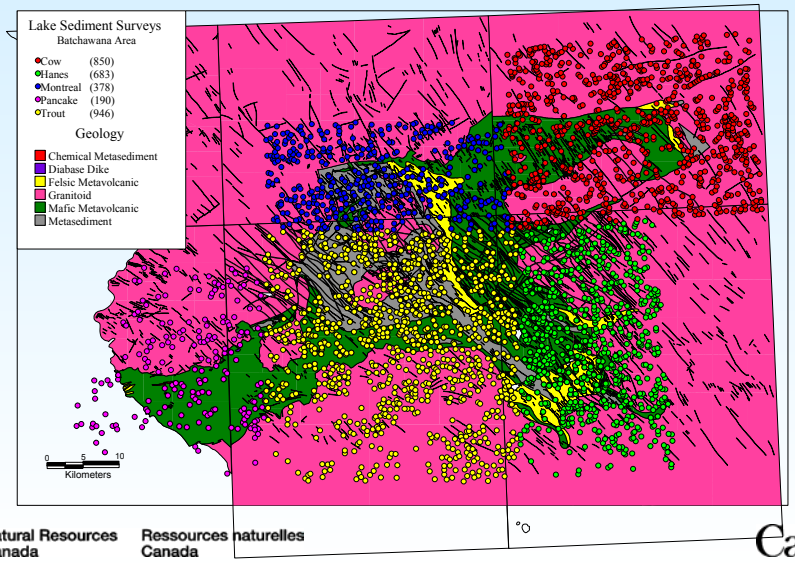
Bubble Plots



Canada Canada Canada

(The Art of) Levelling Lake Sediment Surveys

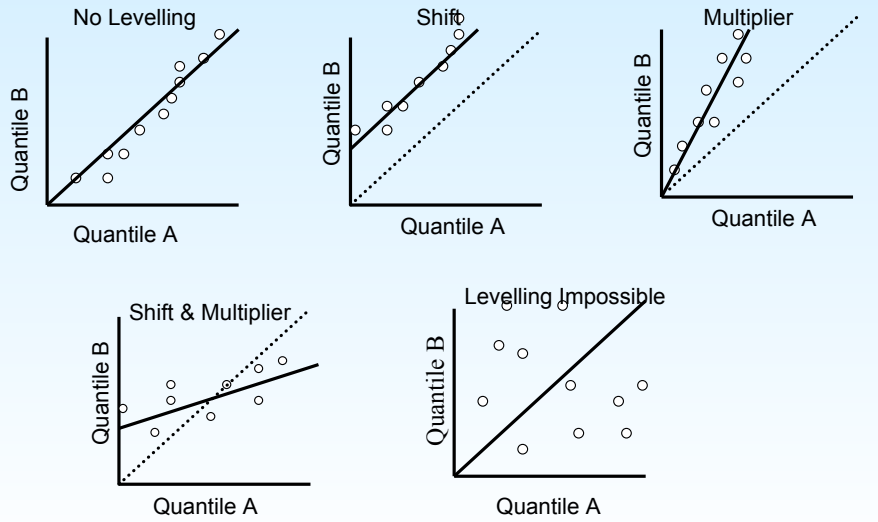
Batchawana, Ontario



Natural Resources Canada Ressources naturelles Canada

Canada

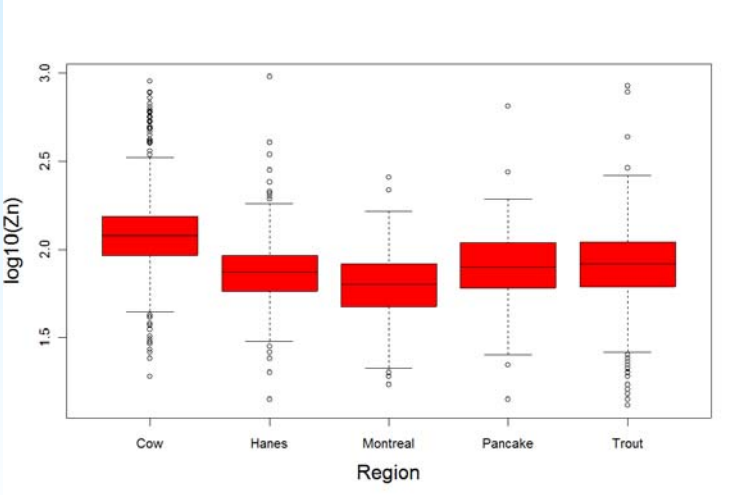
Parametric Levelling Scenarios



Natural Resources Canada Ressources naturelles Canada

Canada

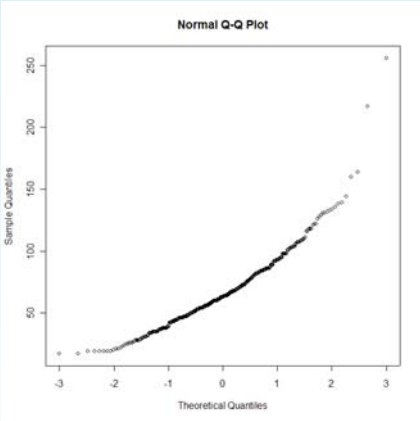
Batchawana Lake Sediments (Zn)



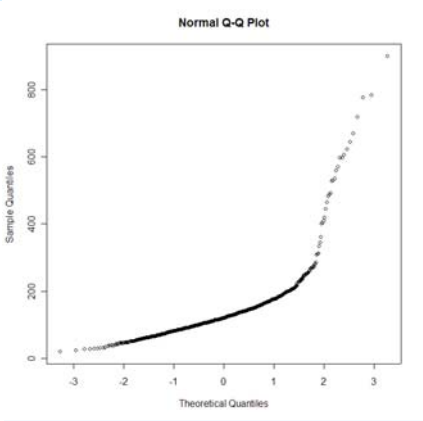
Natural Resources Canada Ressources naturelles Canada

Canada

Zn Q-Q Plots



Montreal River

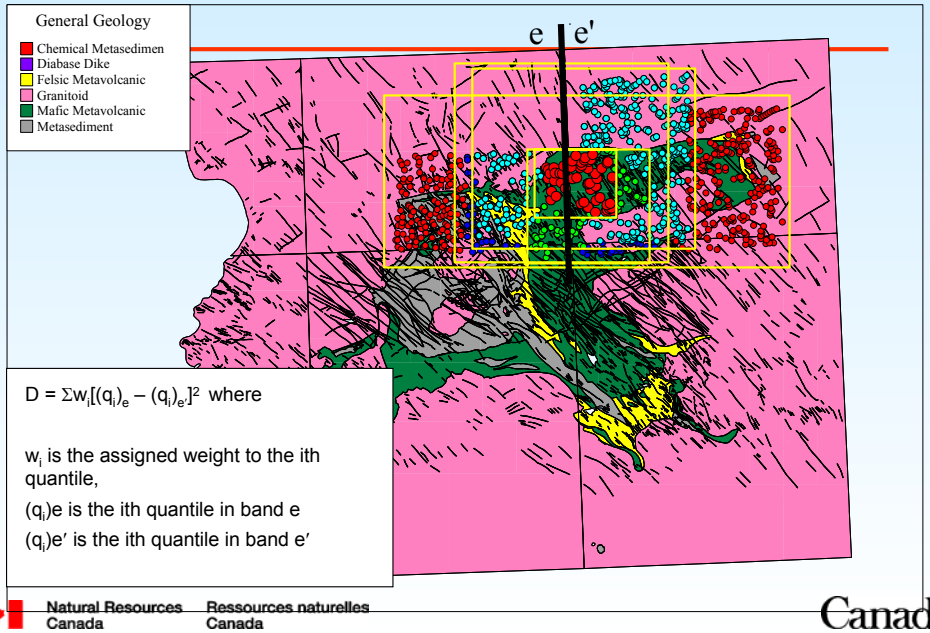


Cow River

Natural Resources Canada Ressources naturelles Canada

Canada

Band Selection for Quantile Regression



$D = \Sigma w_i[(q_i)_e - (q_i)_{e'}]^2$ where

w_i is the assigned weight to the ith quantile,

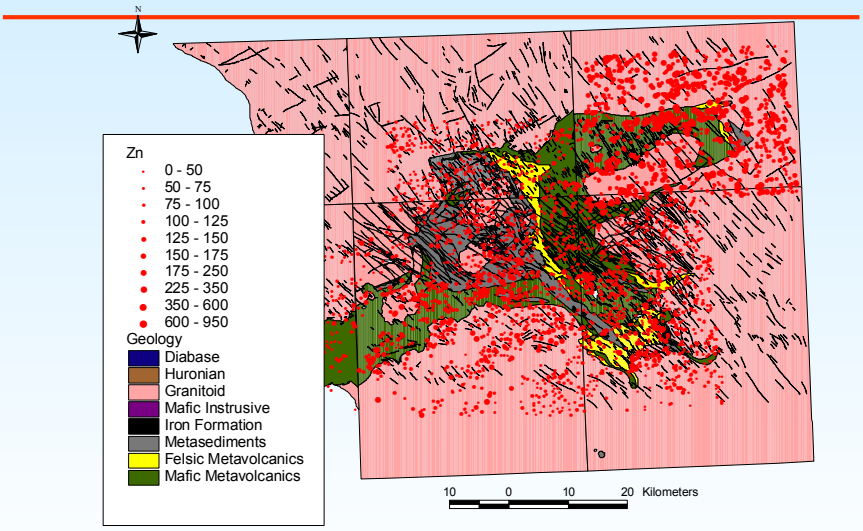
$(q_i)e$ is the ith quantile in band e

$(q_i)e'$ is the ith quantile in band e′

Natural Resources Canada Ressources naturelles Canada

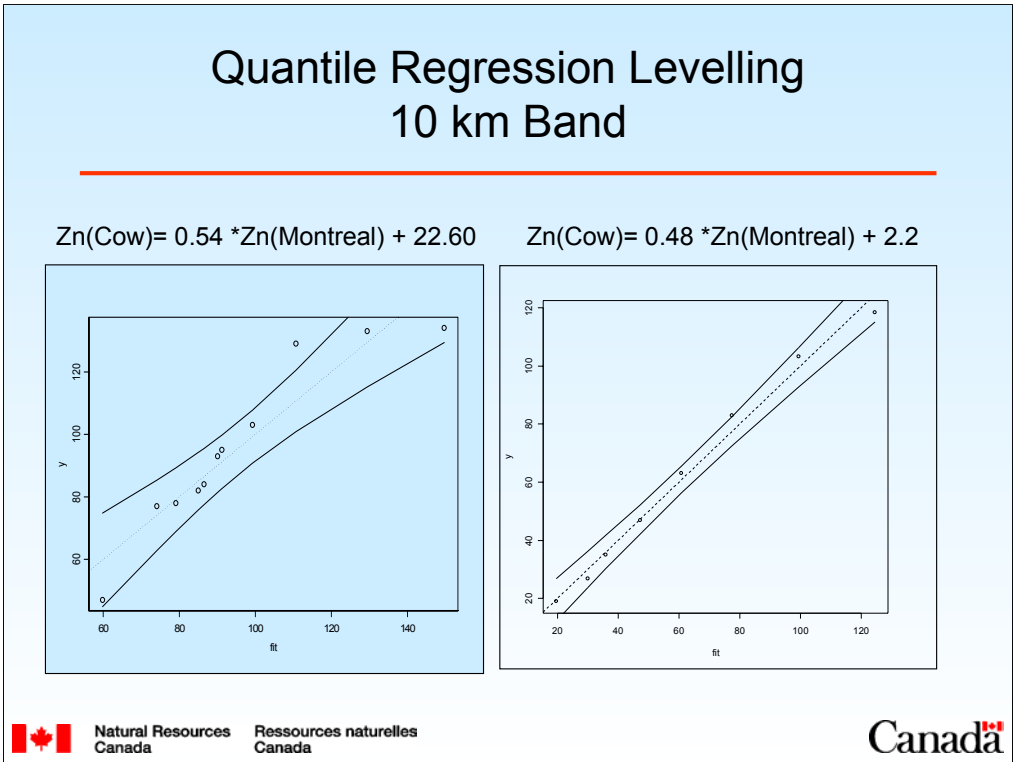
Canada

Zn Before Levelling



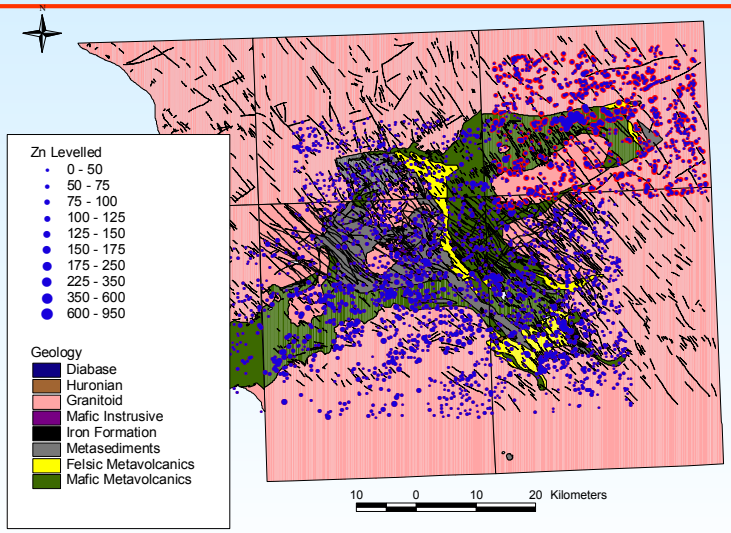
Natural Resources Canada Ressources naturelles Canada

Canada



weights favour quantile pairs at or near the median of the distribution and are based on the ordinates of a normal distribution (weight for the median value (50th percentile) = 0.399).

Zn - after levelling



Natural Resources Canada Ressources naturelles Canada

Canada

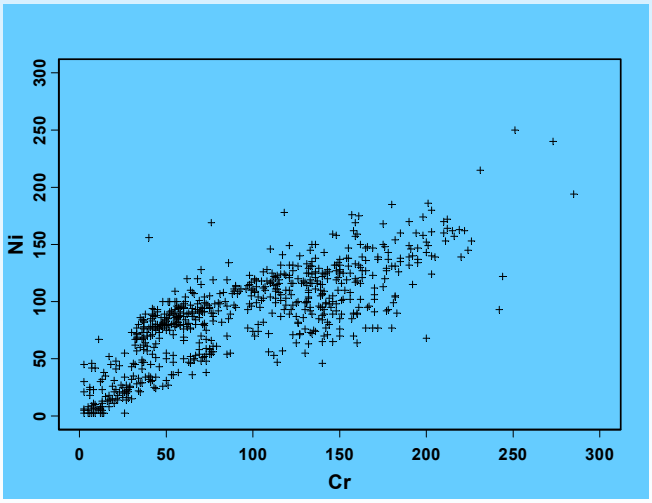
Principal Components Analysis

Soil and Lithogeochemistry Examples

Natural Resources Canada Ressources naturelles Canada

Canada

Metavolcanics & Mafic Intrusions



Natural Resources Canada Ressources naturelles Canada

Canada

A simple example of PCA

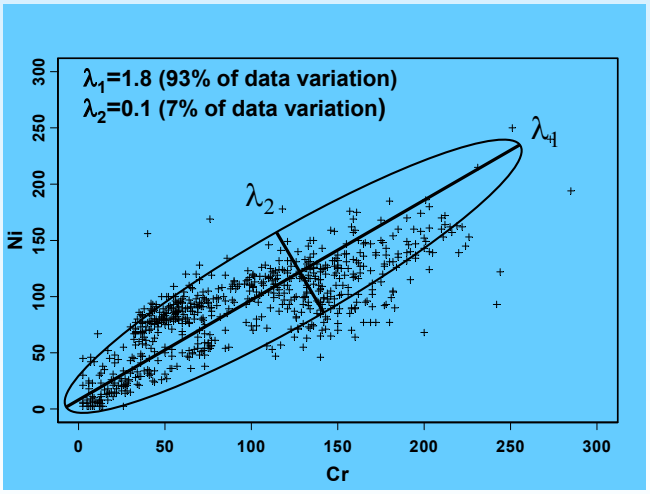
Mafic volcanics and intrusions from a lithogeochemical sampling program from the Blake River Group, Ontario.

Several distinct lithotypes including calc-alkalic rhyolites, dacite, andeisites and basalst together with tholeiitic mafic instrusions.

At least three groups are obvious but there is a linear trend that clearly displays the positive correlation between Cr and Ni.

Ni substitutes for Cr in lattice sites in orthosilicates.

Metavolcanics & Mafic Intrusions
Principal Components



Natural Resources Canada Ressources naturelles Canada

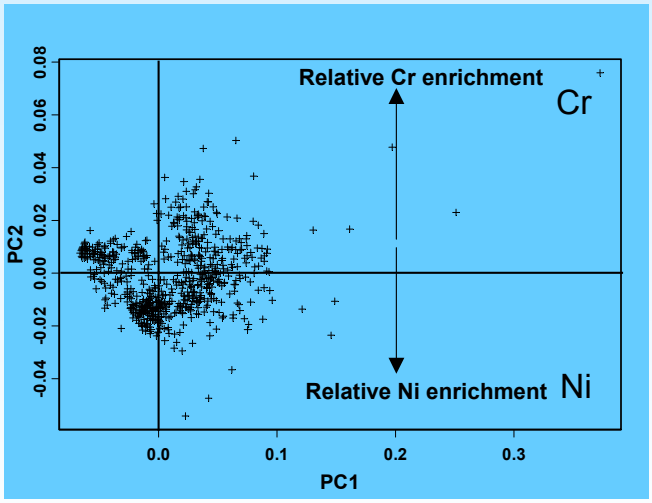
Canada

Plots of component scores describe the relationships of the samples and the elements with respect to each other.

Components are ordered from most to least significant.

More significant components reflect obvious geochemical processes whereas the less significant components may be related to processes such as mineralization.

PCA - Metavolcanics & Mafic Intrusions



Natural Resources Canada Ressources naturelles Canada

Canada

By multiplying the eigenvalue x eigenvector x sstandardized values of individual samples (observations), a principal component score can be calculated.

Similarly, scores of the elements can be computed. When scaled, both the samples and elements can be plotted on the same diagram.

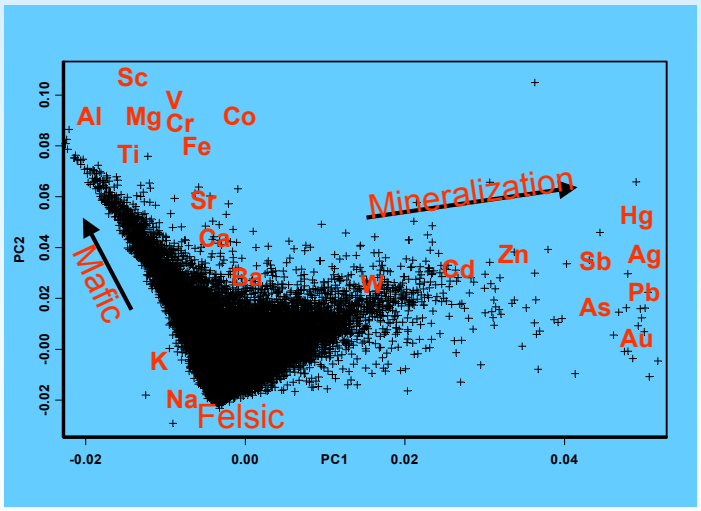
This enables the detection of element and sample associations.

All samples show a corresponding positive relationship between Cr and Ni (C1)

C2 shows that there is also an inverse relationship between Cr and Ni. This inverse association may or may not be significant. Statistical tests can be carried out to test this.

Typically in exploratory geochemical investigations, statistical tests are not carried out. In fact, many of the lesser components which may not be statistically significant provide information on processes that are underpresesented in the sample population and as a result have the appearance of being outliers.

Campo Morado PCA - Soil Geochemistry

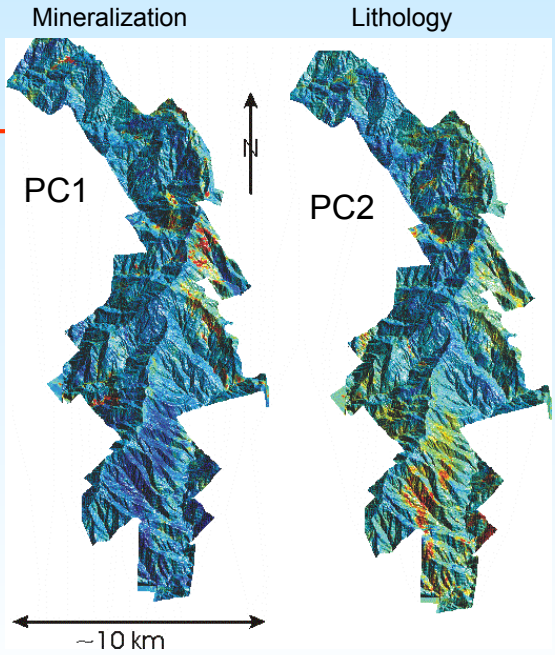


Natural Resources Canada Ressources naturelles Canada

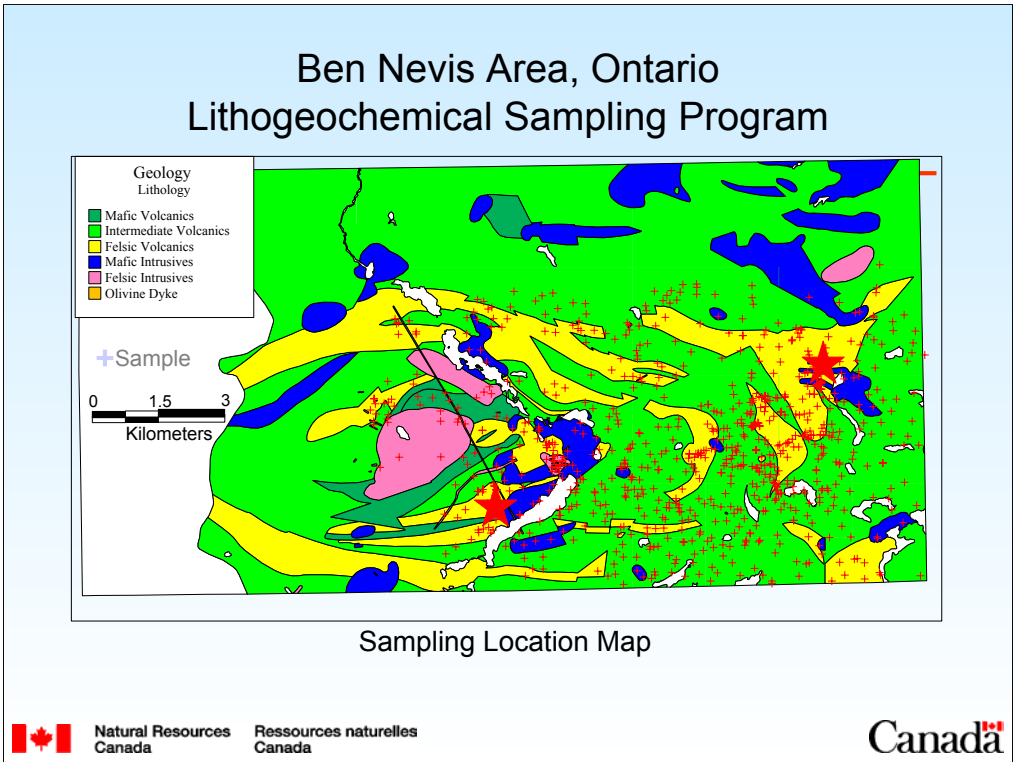
Canada

PC's Draped over DEM

- Distinction between mineralization at source from hydromorphic dispersion.
- Stratigraphic interpretation assisted by topography.



Natural Resources Canada
Ressources naturelles Canada
Canada



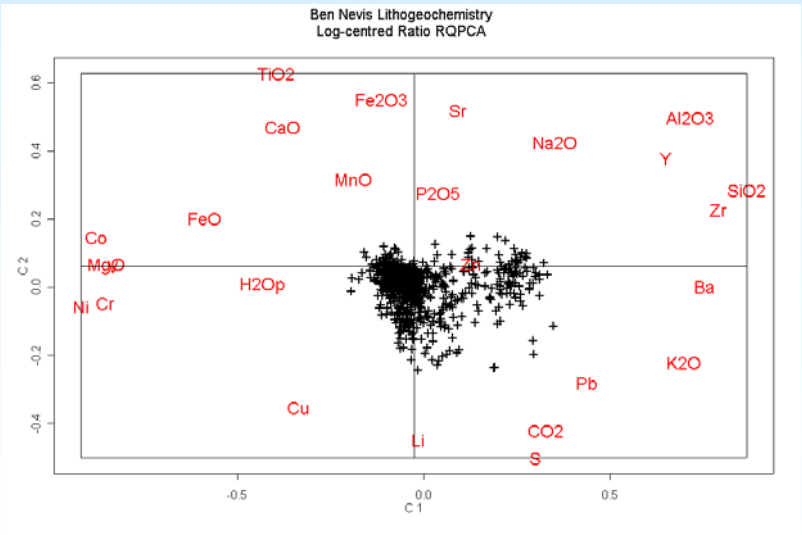
Lithogeochemical sampling program was undertaken in 1979 to examine the nature of alteration associated with two significant zones of mineralization.

825 samples taken and analyzed for:

SiO_2, Al_2O_3, Fe_2O_3, FeO, MgO, CaO, Na_2O, K_2O, TiO_2, P_2O_5, MnO,

CO_2, S, H_2O^+, Ba, Co, Cr, Cu, Li, Ni, Pb, Zn, Sr, V, Y, Zr

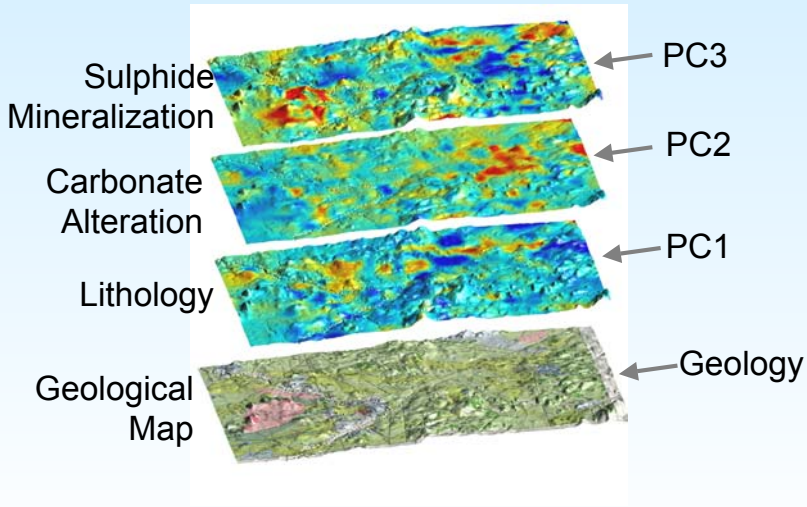
Principal Components Analysis



Natural Resources Canada Ressources naturelles Canada

Canada

Ben Nevis area, Ontario
Integration of PCA with Topography



Natural Resources Canada Ressources naturelles Canada

Canada

Mahalanobis Distance Plots

Measuring Multivariate Atypicality

Natural Resources Canada Ressources naturelles Canada

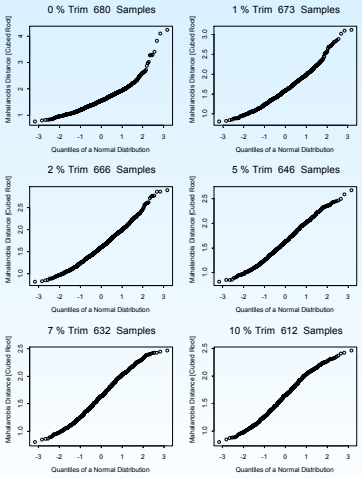
Canada

Mahalanobis Scores

Plots of Lake Sediment Survey
Cu, Zn, As, Sb, W

$$\mathbf{D}_i^2 = (\mathbf{x}_i - \mathbf{x})' \, \mathbf{S}^{-1} \, (\mathbf{x}_i - \mathbf{x})$$

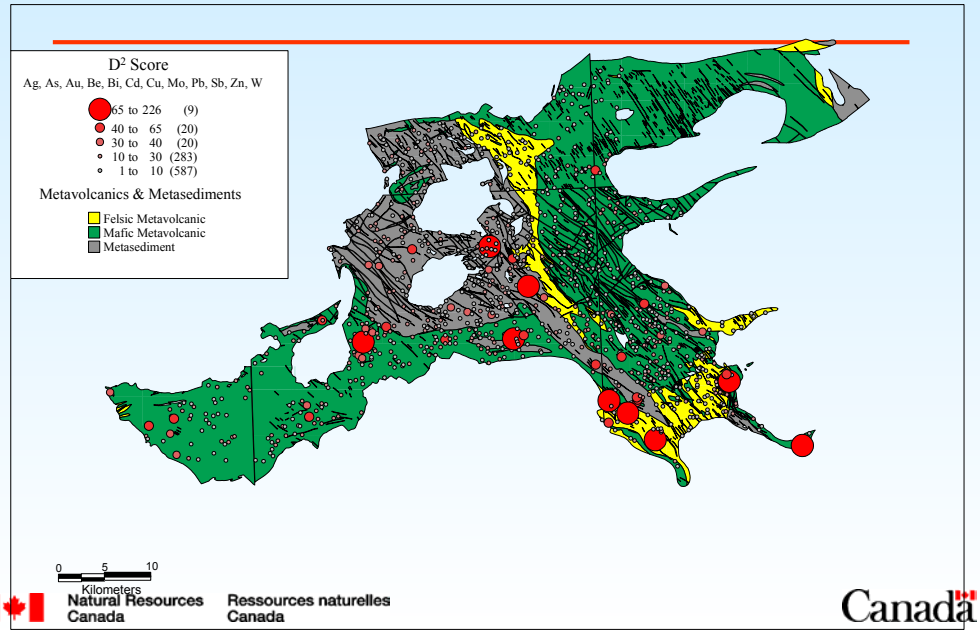
where **S** is the covariance matrix and **x** is the mean vector for the data. The estimates of **S** and **x** can be obtained from robust procedures which enhance the presence of outliers.



Natural Resources Canada Ressources naturelles Canada

Canada

D^2 Scores


D2 Score
Ag, As, Au, Be, Bi, Cd, Cu, Mo, Pb, Sb, Zn, W
65 to 226 (9)
40 to 65 (20)
30 to 40 (20)
10 to 30 (283)
1 to 10 (587)
Metavolcanics & Metasediments
Felsic Metavolcanic
Mafic Metavolcanic
Metasediment
0 5 10
Kilometers
Natural Resources Canada
Ressources naturelles Canada
Canada

Empirical Indices

Weighted Sums

Natural Resources Canada Ressources naturelles Canada

Canada

Weighted Sums Technique

1 Subjective Relative Importance – r

Weight - w

$$w_j = r_j \Big/ \sum_{j=1}^{m} Abs(r_j)$$

Normalize weights

2

$$a_j = w_j \Big/ \sqrt{\sum_{j=1}^{m} w_j^2}$$

3 Determine normal scores based on robust estimates of mean and standard deviation

$$z_{ij} = \left(x_{ij} - \bar{x}_j^* \right) \Big/ s_j^*$$

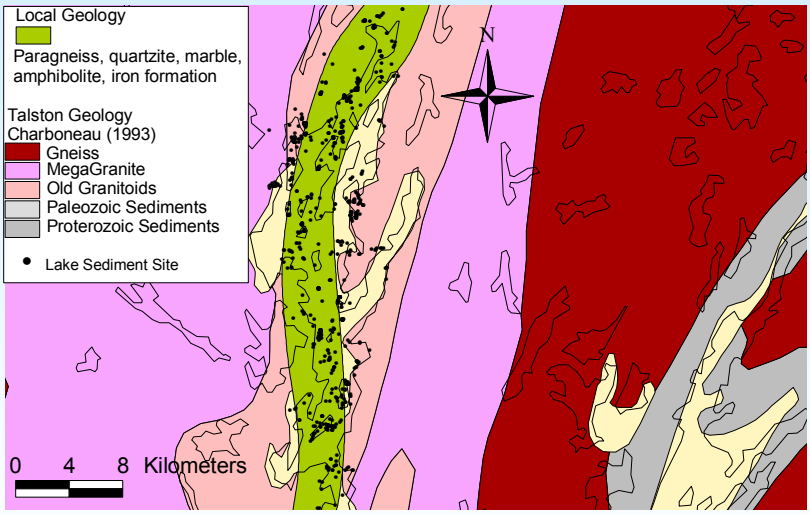
4 Compute weighted sum

$$ws_i = \sum_{j=1}^{m} a_j \cdot z_{ij}$$

Natural Resources Canada Ressources naturelles Canada

Canada

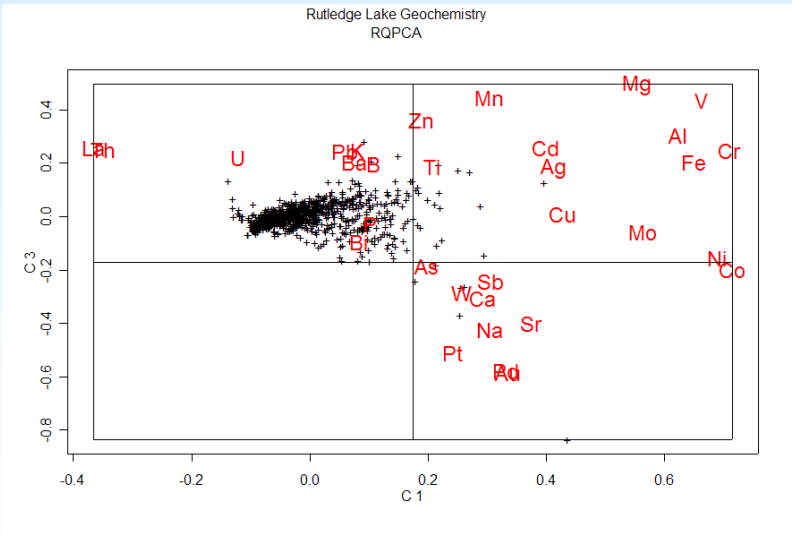
Rutledge Lake, NWT
Lake Sediment Geochemistry



Natural Resources Canada Ressources naturelles Canada

Canada

PCA - Lake Sediments



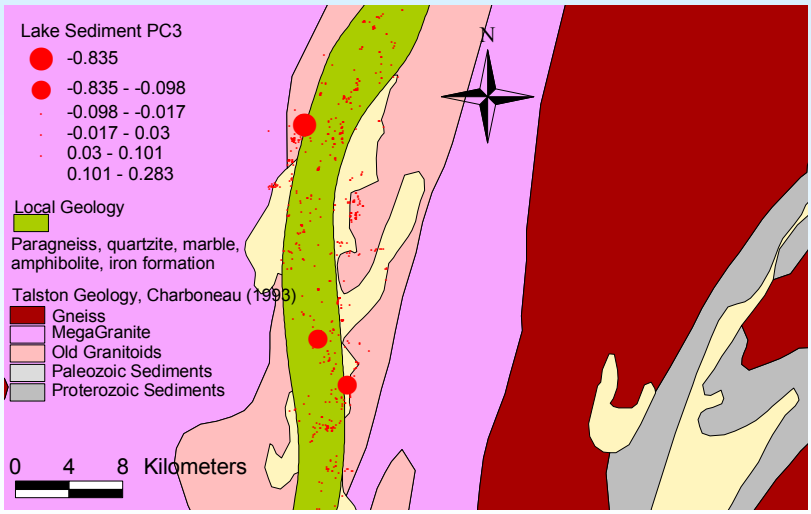
Natural Resources Canada Ressources naturelles Canada

Canada

Rutledge Lake - PC 3

32 Trace Elements

Au Pt Pd Cu Pb Zn Ag Ni Co Mn Fe As U Mo Th Sr Cd Sb Bi V Ca P La Cr Mg Ba Ti B Al Na K W



Natural Resources Canada Ressources naturelles Canada

Canada

Weighted Sums – Rutledge Lake

PC 3 & PC 4 exhibit strong associations of Pt, Pd & Au with Na, Sr, Ca, W, As, Sb

Sb – highly censored.

Weights derived from R-mode loadings

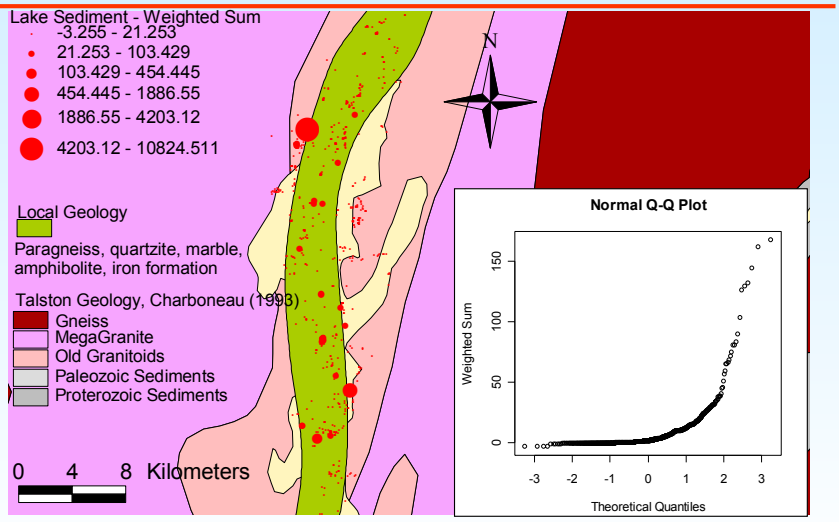
Au	Pt	Pd	Sr	Na	Ca	W	Sb	As
0.31	0.28	0.3	0.2	0.2	0.18	0.14	0.24	0.06

Natural Resources Canada Ressources naturelles Canada

Canada

Weighted Sums - Rutledge Lake

Pt, Pd, Au, Na, Sr, Ca, W, As, Sb



Natural Resources Canada Ressources naturelles Canada

Canada

Sequence of Geochemical Data Analysis - 1

Mineral exploration and hazards identification, threshold selection and anomaly recognition is assisted by the following:

- Q-Q plots, Histograms, Density and Box-Whisker plots
- Scatterplots and scatterplot matrices
- Trimming of outliers
- Adjusting for censored data
- Data transformations
- Survey Levelling

Natural Resources Canada Ressources naturelles Canada

Canada

Sequence of Geochemical Data Analysis - 2

- Robust estimates for outlier detection.
- Application of dimension reducing techniques (PCA).
- Discovery of structures/trends (PCA, Cluster analysis).
- D^2 plots to isolate multivariate outliers.
- Empirical, knowledge-based, indices that are tailored to specific geochemical characteristics (Weighted Sums, NUMCHI, Scoresum).

Natural Resources Canada Ressources naturelles Canada

Canada

Statistical Tools

- Many good commercial statistical packages available
- Public domain freeware/shareware
  - "R" statistical package – http://www.r-project.org
  - Established statistical routines based on the S and S-Plus languages.
  - Very good visualization
  - Very large collaborative community
  - Continuous developments of statistical procedures and data visualization
  - Essential tool for the geochemical variable domain that can integrate with tools for the geographic domain.
  - GeoDAS – spectral and fractal analysis

Natural Resources Canada Ressources naturelles Canada

Canada

Acknowledgements

- Ontario Geological Survey
- Geological Survey of Canada
- CSIRO, Division of Exploration & Mining
- Hunter-Dickinson Inc.
- Platinum Group Metals Inc.
- IGES Organizing Committee

Natural Resources Canada Ressources naturelles Canada

Canada