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
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.
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.
Geographical Space

- Humans relate visually to the spatial representation of concepts.
- Interpretation is enhanced when symbols or imagery are represented in the geospatial domain.
Examples of Exploratory Data Analysis in 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
Univariate Exploratory Approach

- Histogram
- Box plot
- Density plot
- Q-Q plot
- Order Statistics
Multivariate Exploratory Approach

- Principal components analysis
- $D^2 (\chi^2)$ plots
- Empirical Indices
  - Weighted Sums
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.
Non-normal Distributions: Transformation of Data

- Standard statistical procedures assume a normal distribution.
- 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

\[
y = \frac{x^\lambda - 1}{\lambda} \quad \text{for } \lambda > 0
\]

\[
y = \ln(x) \quad \text{for } \lambda = 0
\]
Q-Q Plots & Transformations

Outliers

"Breaks"

Values < L.L.D.

Ben Nevis Lithogeochemistry

Ben Nevis Lithogeochemistry
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.
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

<table>
<thead>
<tr>
<th>Method</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>56</td>
</tr>
<tr>
<td>Median</td>
<td>42</td>
</tr>
<tr>
<td>Trimmed Mean</td>
<td>40</td>
</tr>
<tr>
<td>Trimmed Median</td>
<td>40</td>
</tr>
<tr>
<td>Huber M-estimator</td>
<td>40</td>
</tr>
<tr>
<td>Min. Vol. Ellipsoid</td>
<td>38</td>
</tr>
<tr>
<td>Min. Cov. Det.</td>
<td>NA</td>
</tr>
<tr>
<td>Min. Cov. Det.</td>
<td>39</td>
</tr>
</tbody>
</table>

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

Arsenic (ppm)
- 0 to 96 (5)
- 11 to 50 (51)
- 5 to 11 (153)
- 3 to 5 (336)
- 2 to 3 (486)
- 0 to 2 (1166)

Geology
- Batchawana Area
  - Chemical Metasediment (13)
  - Galous Dike (2)
  - Felsic Metavolcanic (64)
  - Metasediment (130)
  - Mafic Metavolcanic (188)
  - Granitoid (234)

Histogram

Box & Whisker Plot

Density Plot

Q-Q Plot

Quantiles of Standard Normal

Log10 As (ppm)
(The Art of) Levelling Lake Sediment Surveys
Batchawana, Ontario
Parametric Levelling Scenarios

No Levelling

Shift & Multiplier

Levelling Impossible

Quantile A

Quantile B

Quantile A

Quantile B

Quantile A

Quantile B

Quantile A
Batchawana Lake Sediments
(Zn)
Zn Q-Q Plots

Montreal River

Cow River
Band Selection for Quantile Regression

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

where

- \( w_i \) is the assigned weight to the \( i \)th quantile,
- \( (q_i)_e \) is the \( i \)th quantile in band \( e \)
- \( (q_i)_{e'} \) is the \( i \)th quantile in band \( e' \)
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).
Principal Components Analysis

Soil and Lithogeochemistry Examples
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, andesites and basalt together with tholeiitic mafic intrusions.

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.
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.
By multiplying the eigenvalue x eigenvector x standardized 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 underrepresented in the sample population and as a result have the appearance of being outliers.
Field Morado PCA - Soil Geochemistry

Mineralization

Mafic

Felsic

Sc, V, Cr, Co, Ti, Sr, Ca, Ba, Na, K, As, Sb, Ag, Hg, Zn, Cd, W, Pb, Au
PC’s Draped over DEM

- Distinction between mineralization at source from hydromorphic dispersion.
- Stratigraphic interpretation assisted by topography.
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$_2$O$_3$
- Fe$_2$O$_3$
- FeO
- MgO
- CaO
- Na$_2$O
- K$_2$O
- TiO$_2$
- P$_2$O$_5$
- MnO
- CO$_2$
- S
- H$_2$O$^+$
- Ba
- Co
- Cr
- Cu
- Li
- Ni
- Pb
- Zn
- Sr
- V
- Y
- Zr
Principal Components Analysis
Ben Nevis area, Ontario
Integration of PCA with Topography

Sulphide Mineralization
Carbonate Alteration
Lithology
Geological Map

PC3
PC2
PC1
Geology
Mahalanobis Distance Plots

Measuring Multivariate Atypicality
Mahalanobis Scores

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

$$D_1^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x})$$

where $S$ is the covariance matrix and $\bar{x}$ is the mean vector for the data. The estimates of $S$ and $\bar{x}$ can be obtained from robust procedures which enhance the presence of outliers.
D^2 Scores

D^2 Score
Ag, As, Ba, Be, Cd, Ca, 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

Canada Natural Resources
Ressources naturelles Canada
Empirical Indices

Weighted Sums
Weighted Sums Technique

1. Subjective Relative Importance – \( r \)
   Weight - \( w \)
   \[ w_j = \frac{r_j}{\sum_{j=1}^{m} |Abs(r_j)|} \]

   Normalize weights

2. \( a_j = \frac{w_j}{\sqrt{\sum_{j=1}^{m} w_j^2}} \)

3. Determine normal scores based on robust estimates of mean and standard deviation
   \[ z_{ij} = \left( x_{ij} - \bar{x}_j^* \right) / s_j^* \]

4. Compute weighted sum
   \[ WS_i = \sum_{j=1}^{m} a_j \cdot z_{ij} \]
Rutledge Lake, NWT
Lake Sediment Geochemistry

Local Geology
Paragneiss, quartzite, marble, amphibolite, iron formation

Talston Geology
Charbonneau (1993)
- Gneiss
- MegaGranite
- Old Granitoids
- Paleozoic Sediments
- Proterozoic Sediments
- Lake Sediment Site

Natural Resources
Canada
Ressources naturelles
Canada

Canada
PCA - Lake Sediments

Rutledge Lake Geochemistry
RQ-PCA

Lith, U, Pb, Zn, Mn, Mg, V, Al, Fe, Cr, Ti, Cd, Ag, Cu, Mo, Ni, Co, As, W, Sb, Na, Sr, Pt, Au
Rutledge Lake - PC 3
32 Trace Elements

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

Lake Sediment PC3
-0.835
-0.835 - -0.098
-0.098 - -0.017
-0.017 - 0.03
0.03 - 0.101
0.101 - 0.283

Local Geology
Paragneiss, quartzite, marble, amphibolite, iron formation

Talston Geology, Charbonneau (1993)
Gneiss
MegaGranite
Old Granitoids
Paleozoic Sediments
Proterozoic Sediments

0 4 8 Kilometers
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

<table>
<thead>
<tr>
<th></th>
<th>Au</th>
<th>Pt</th>
<th>Pd</th>
<th>Sr</th>
<th>Na</th>
<th>Ca</th>
<th>W</th>
<th>Sb</th>
<th>As</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.31</td>
<td>0.28</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.18</td>
<td>0.14</td>
<td>0.24</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Weighted Sums - Rutledge Lake
Pt, Pd, Au, Na, Sr, Ca, W, As, Sb

0 4 8 Kilometers

Lake Sediment - Weighted Sum
-3.255 - 21.253
21.253 - 103.429
103.429 - 454.445
454.445 - 1886.55
1886.55 - 4203.12
4203.12 - 10824.511

Local Geology
Paragneiss, quartzite, marble, amphibolite, iron formation
Talston Geology, Charboneau (1993)
Gneiss
MegaGranite
Old Granitoids
Paleozoic Sediments
Proterozoic Sediments

Normal Q-Q Plot

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
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).
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
Acknowledgements

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