Extracting, visualising and interpreting structure in geochemical data through compositional data analysis (CoDA) Introduction

Eric Grunsky

Department of Earth and Environmental Sciences, University of Waterloo, Ontario

egrunsky@gmail.com

2021-November-18

Goals of Geochemical Data Analysis

- Detect inter-element relationships of geochemical data that reflect mineralogy or molecular associations and describe or infer geological processes.
- Isolate atypical observations or groups of observations that are potentially identified with processes of interest (mineral deposit, hazardous environment, alteration/lithologic mapping).

Geochemical Survey Data

Geochemical surveys are conducted to provide baseline information for:

- Mineral exploration
- Geologic mapping
- Baseline values for environment/land use purposes

- Geochemical survey data are a rich source of information for geological, geochemical, environmental and climatic processes.
- More than 50 elements can be analyzed at sufficiently low detection limits.
- Geochemical data reflect processes that form or affect mineralogy.
- These data represent a multivariate data space over a two or three dimensional geographic space and time.

Geochemical Sample Media

- Choice of sample media reflects different processes.
- Method of sample preparation affects analytical results.
- Method of instrumentation affects analytical results.
- Spatial density (support) affects the ability to detect various processes.

Effective Geochemical Data Evaluation

- There is an ever-increasing number of data sets collected by both industry and government that provide geochemical data at regional and mineral camp scales.
- Modern methods of data analysis integrate both the detailed and regional data with the expectation to recognize mineralization or vectors to mineralization.
- Integrating regional and project specific geochemical data requires careful evaluation to ensure that the data are properly levelled against a common standard.
- The role of standard reference materials in geochemical quality control/assurance cannot be overstated.
- The sample design should be of sufficient density to provide information at the scale of interest.

Defining Geospatial Scales of Geochemical Surveys

Continental Scale – > 1:500,000 & < 1:1,000,000

Mapping large crustal blocks/tectonic assemblages.

Regional Scale - > 1:50,000 & < 1:500,000

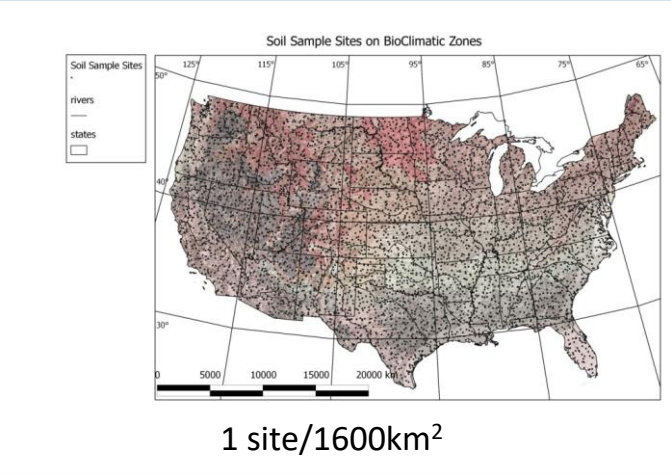
Regional geological mapping

Local/Camp Scale < 1:50,000

Exploration scale studies and detailed geologic mapping.

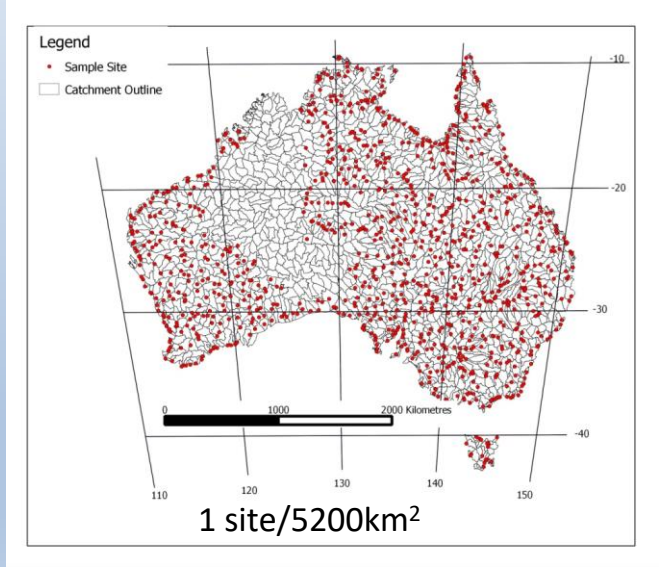
Continental Scale – > 1:500,000 & < 1:1,000,000
Mapping large crustal blocks/tectonic assemblages.

USGS Soil Survey



1 site/1600km^2

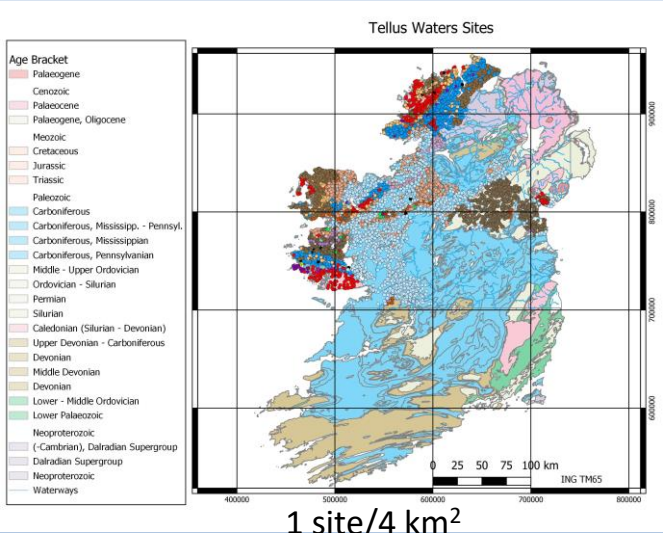
NGSA -National Geochemical Survey of Australia



1 site/5200km^2

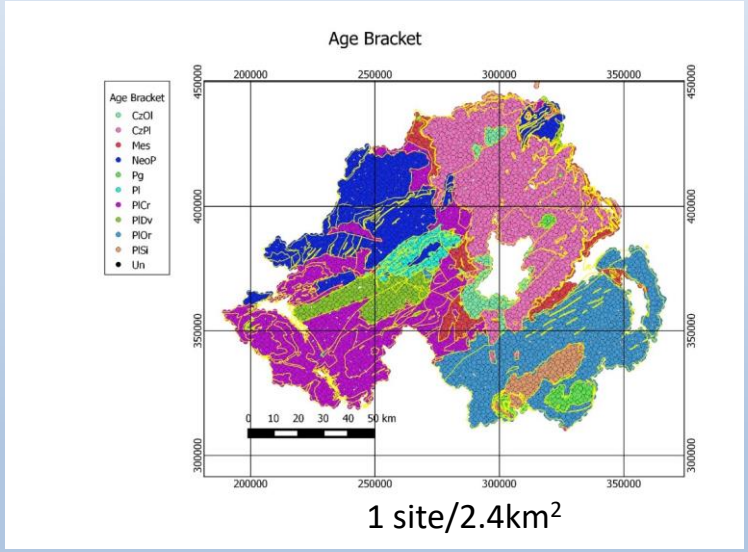
Regional scale of geochemical surveys 1:250,000

Tellus Waters Sampling Sites – NW Ireland



1 site/4 km^2

Northern Ireland Soil Sampling Sites



1 site/2.4km^2

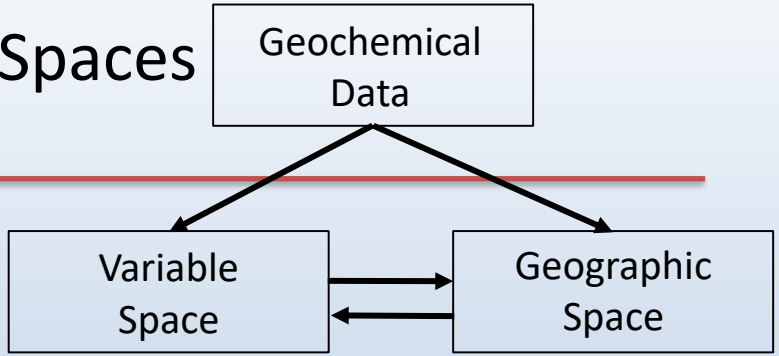
Structure in Geochemical Data

Structure in geochemical data are trends/patterns that can be described by linear and non-linear methods.

Linear processes - **stoichiometry** – the ordered arrangement of elements according to atomic forces that form minerals.

Non-linear processes – comminution, material transport and sorting (gravitational).

Geochemical Data Spaces



Variable Space – structure in the elements (stoichiometry)

- Statistics and Data visualization. Numerous graphical and statistical methods characterize and describe the variables.
- Different metrics display features that are not obvious.
- Metrics include cartesian, logratios, principal components, independent components, t-distributed stochastic neighbour embedding + many others.

Geographic Space – 2D or 3D (geospatial structure)

- Geographic representation of data using Geographic Information Systems (GIS) or Image Analysis Systems
- Geostatistical Analysis – spatial processes.
- Geographic metric can be linear (Easting/Northing), spherical (longitude/latitude), polar (angular).
- Defines "**geospatial coherence**" of processes.

Geochemical Definitions

Target Population

Samples within and surrounding a mineral deposit.

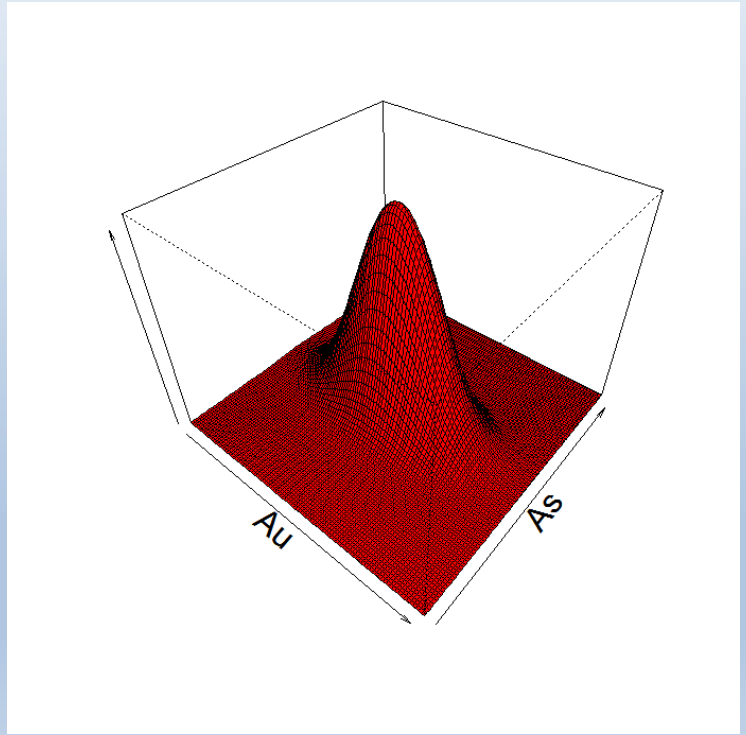
- characteristic geochemistry of the deposit.
- each mineral deposit type should have its own target population.

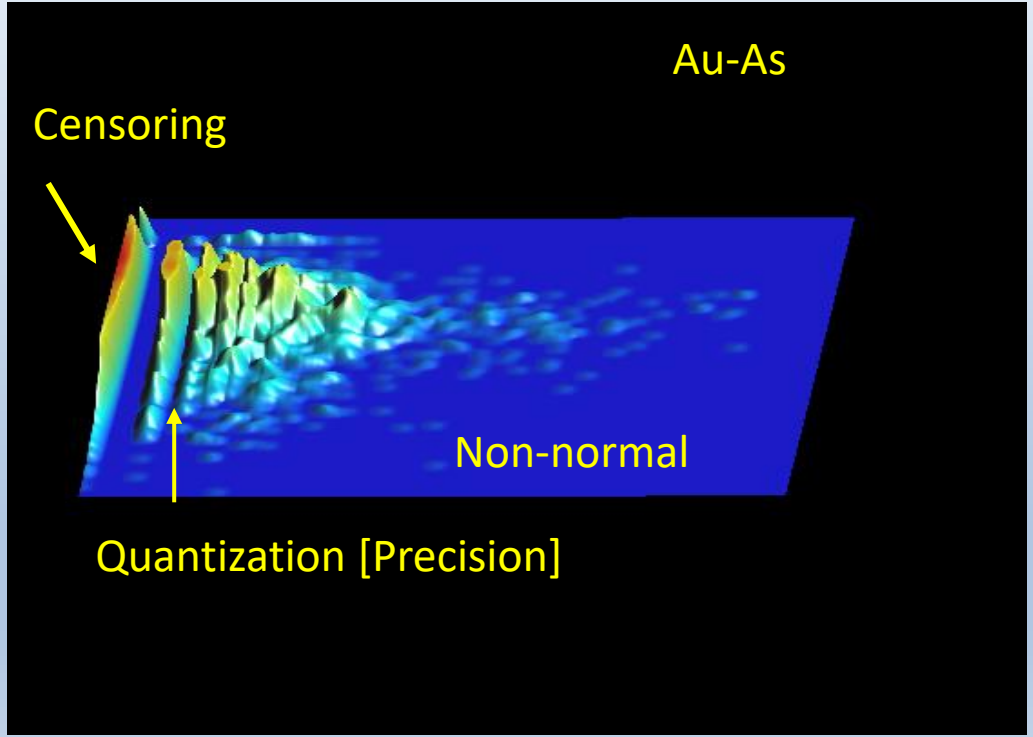
Background Population

Samples from a regional survey.

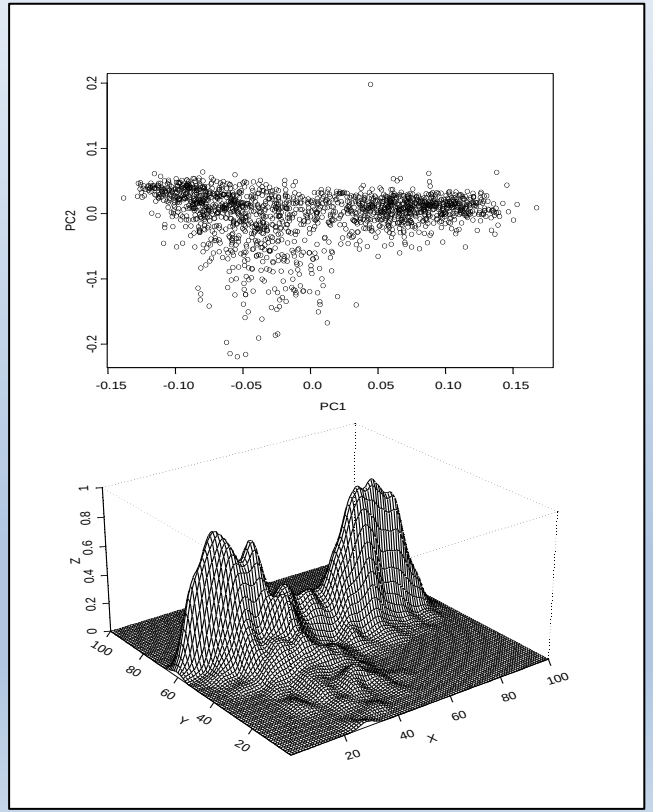
- lithological variation and other regional geochemical effects.
- not related to mineralization.

Ideal vs. Actual Covariation of Two Elements

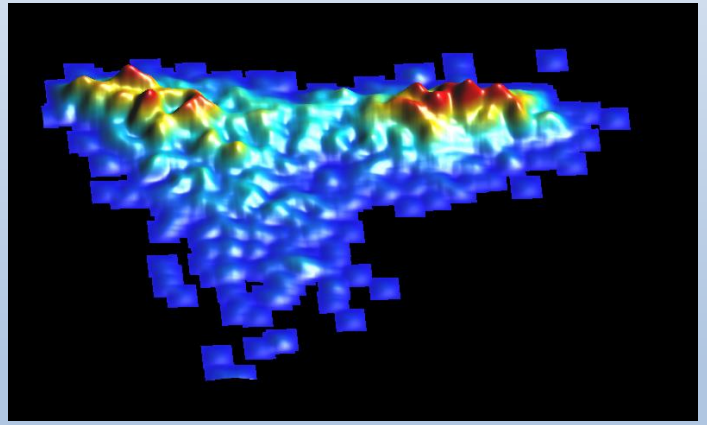




Visualizing Distributions



Densities of Bi-variate Distributions



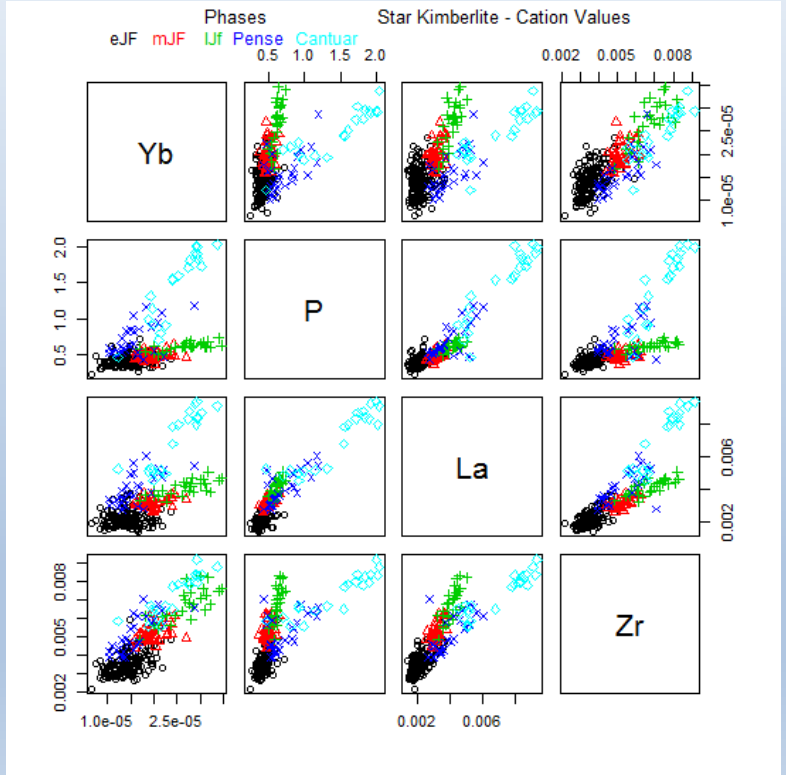
Multi-element Evaluation

- It is advantageous to examine geochemical data using a multi-element (multivariate) approach.
- Geochemistry as a proxy for mineralogy is represented by the variability of more than one element.
- Combining elements reflects the reality of mineral assemblages that comprise earth surface materials.
- Emphasis will be placed on the multivariate approach to geochemical data analysis and interpretation.

Multivariate
Exploratory Approach

Multivariate

- Scatterplot matrix
- Principal components
- Weighted Sums
- Cluster analysis
- χ^2 plots
- Empirical Indices
- Independent components



Transformation of Data

Standard statistical procedures assume a normal distribution.

- Most geochemical data distributions are non-normal and positively skewed - mixtures of populations.
- Data should be transformed before statistical methods are used.
- Outliers should be eliminated prior to transforming data.
- Commonly used method is Box-Cox Power Transformation

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

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

Transformations