



Geochemical Data Evaluation and Interpretation

Eric Grunsky
Geological Survey of Canada

*Workshop 2: Exploration Geochemistry –
Basic Principles & Concepts
Exploration 07
8-Sep-2007*



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07



Outline

- What is geochemical data?
- The Basics of how geochemical data is evaluated.
- Case Studies:
 - Levelling geochemical survey data
 - Lithology/Alteration/Mineralization signatures in metavolcanics
 - Multivariate geochemical signatures & topography in the tropics.
- Concluding Remarks





Goals of Geochemical Data Analysis

- Detect inter-element relationships of geochemical data that reflect that **mineralogy** or chemical species interactions 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).
- Pattern recognition is a key concept in data analysis.





Geochemical Data Processing & Visualization

Evaluation of Geochemical Data

- Variable Space
 - Statistics and Data visualization. Numerous graphical and statistical methods characterize the variables.
- Geographic Space
 - Geographic representation of data using Geographic Information Systems (GIS) or Image Analysis Systems
 - Geostatistical Analysis – spatial processes.





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.





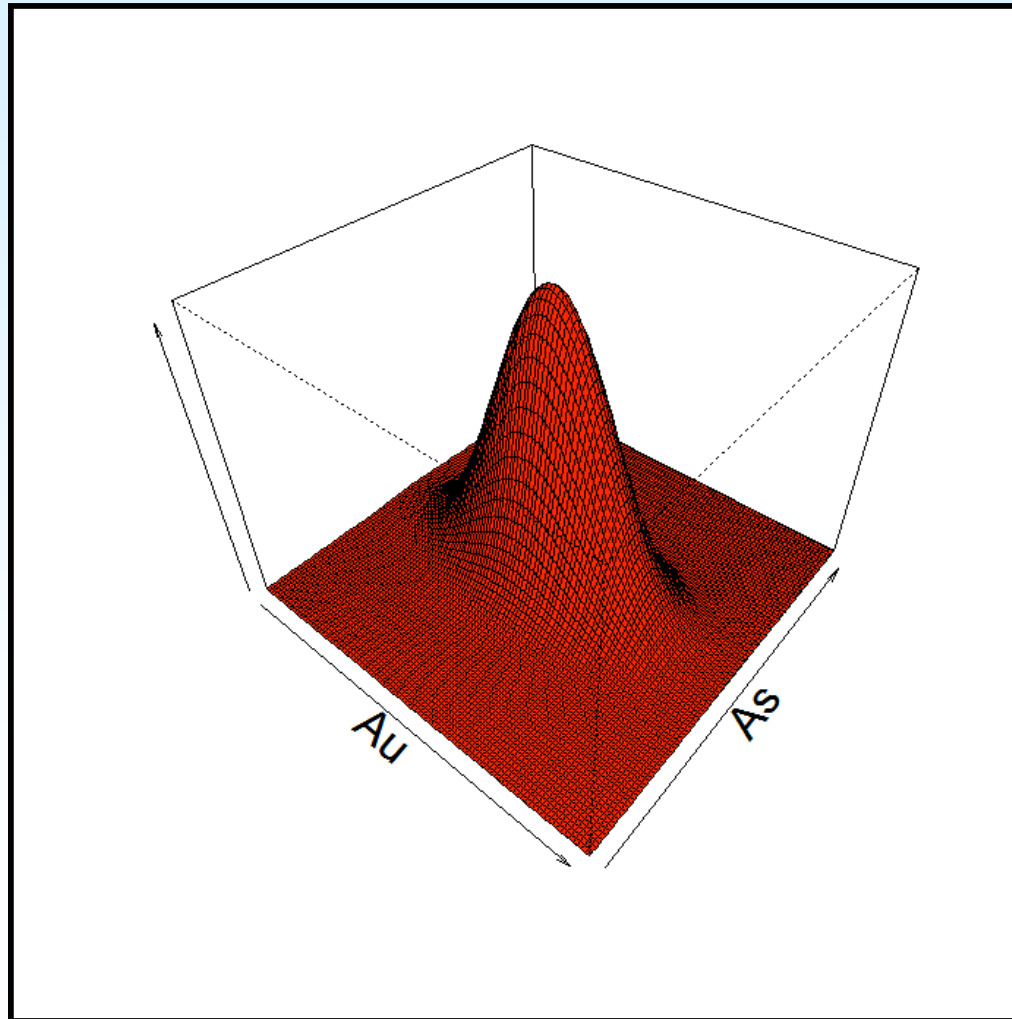
Measures of Geochemical Data

- Common practice to express major elements as weight % oxides (e.g. SiO_2).
- Trace elements commonly expressed as parts per million (ppm).
- Mixing major element oxides with elements reported without oxides is not a good idea (scale differences).
- Convert to moles and then evaluate.





Ideal Distribution of Elements



Natural Resources
Canada

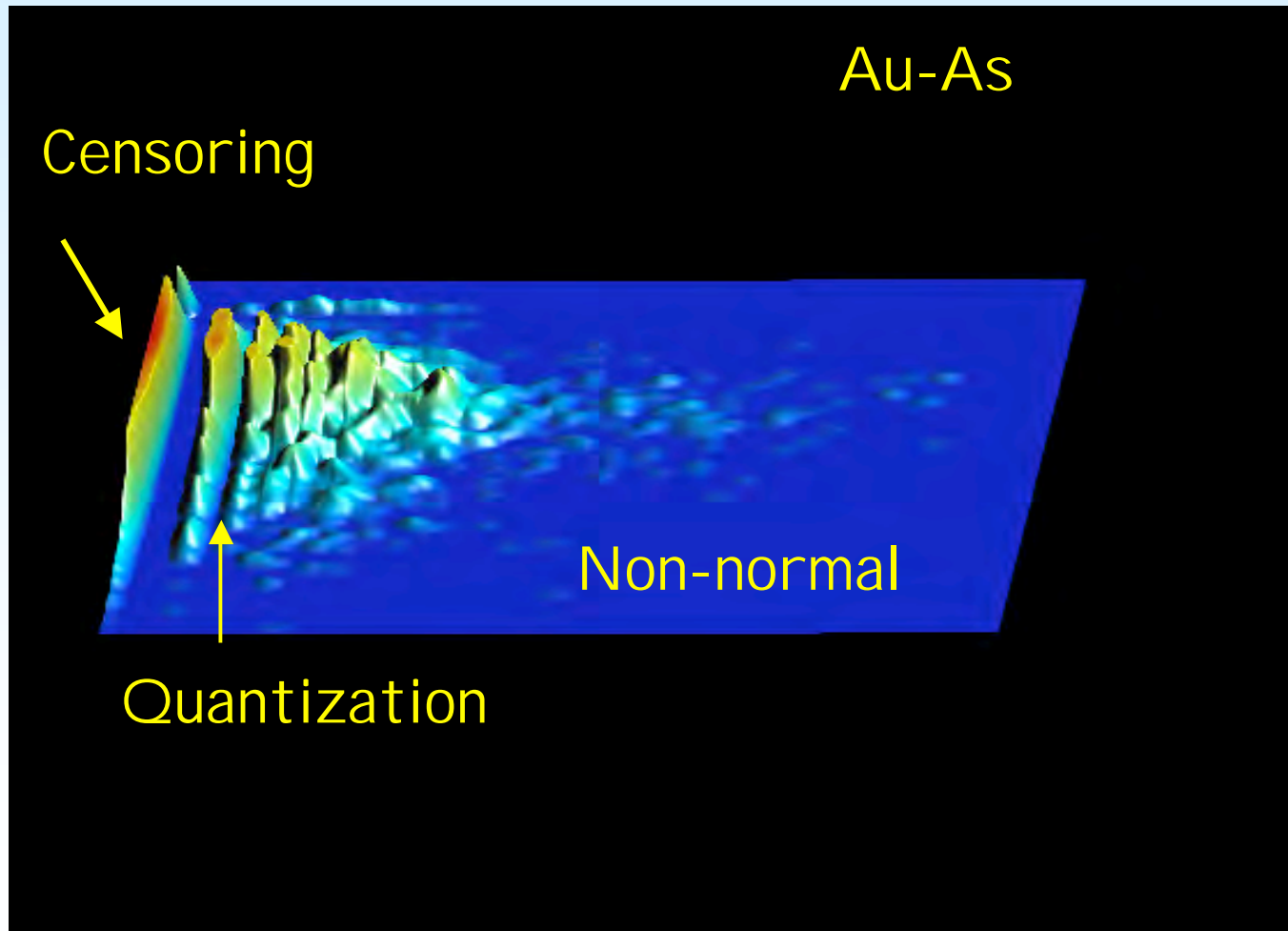
Ressources naturelles
Canada



Exploration07



The True Nature of Distributions

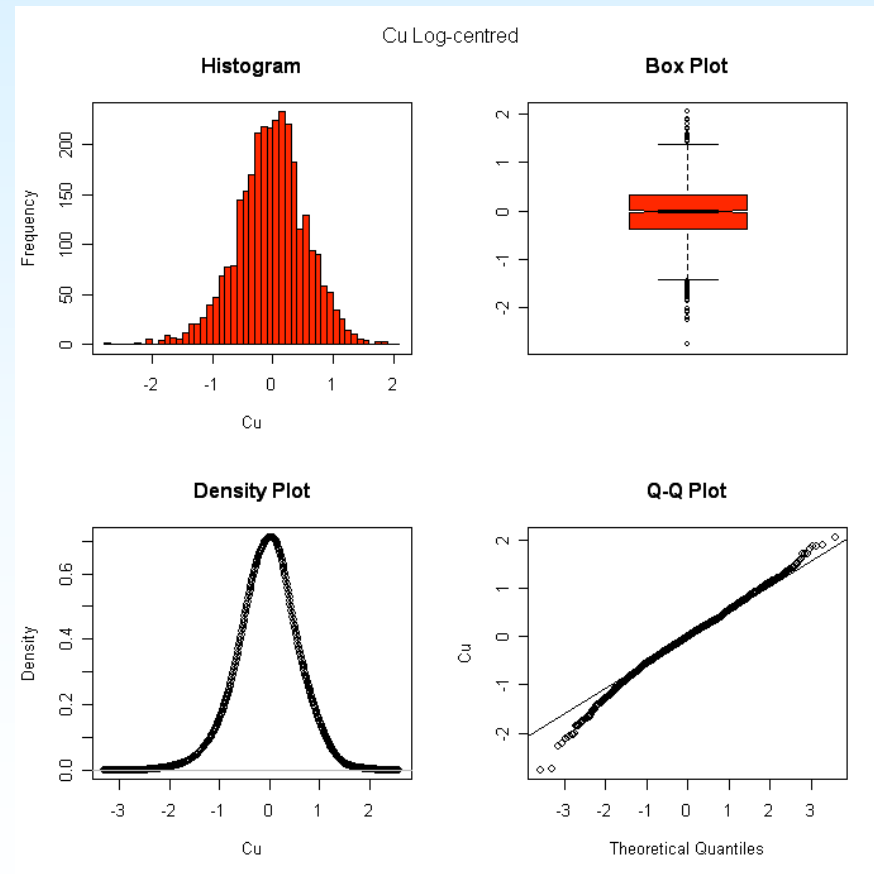




Univariate Exploratory Approach

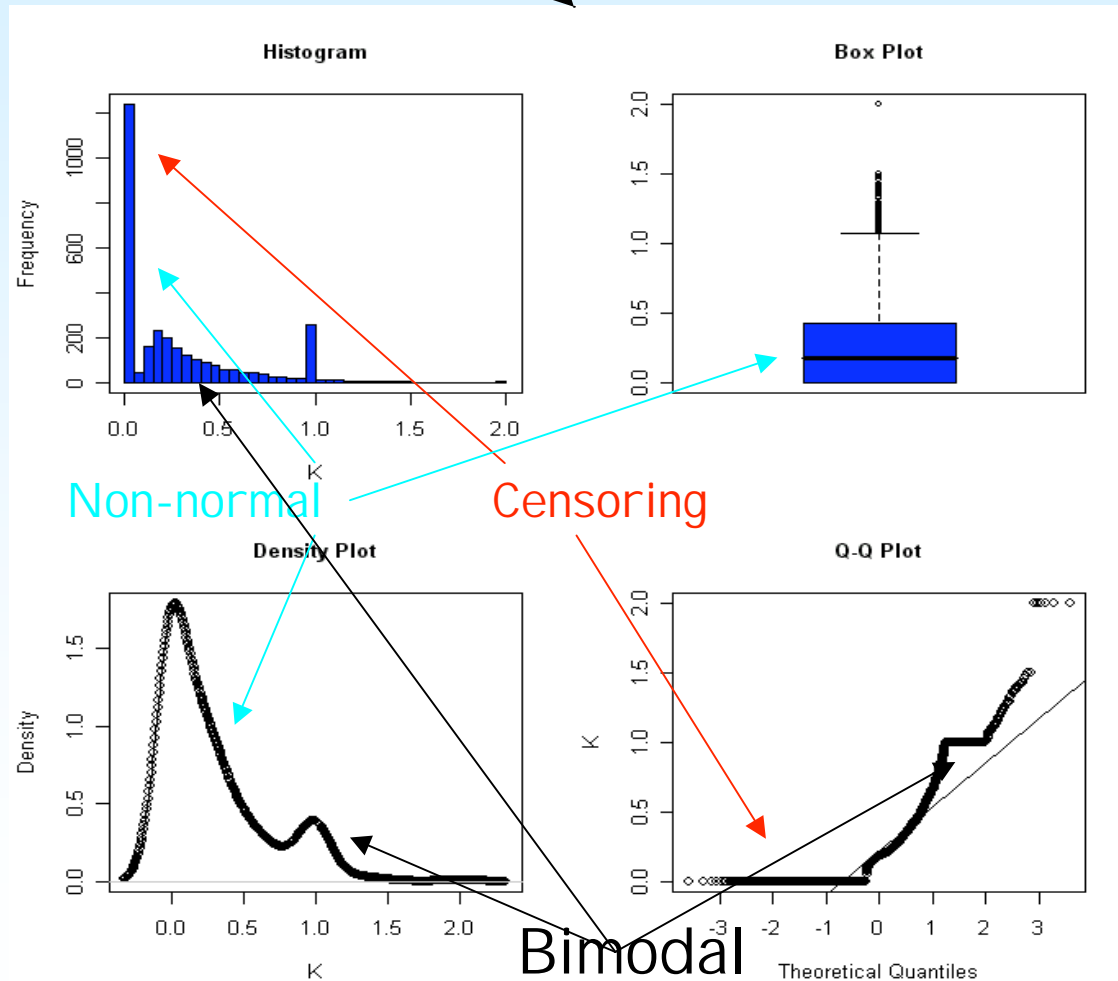
A "near perfect" distribution.

- Histograms
- Ranked data
- Q-Q plots
- Density plot
- Summary tables





Typical Distribution





Summary Table

Summary Statistics for Lake Sediments, Batchawana Area, Ontario																		
	LLD	Num Obs	Minimum	1%	5%	10%	25%	50%	Median	Mean	75%	90%	95%	99%	Maximum	Std.Dev.	MAD	Coefficient of Variation
LOI	2.96	3019	3	35.5	40.11	40.5	43.94	49.5	44	44	53.5	56.535	57.25	57.5	91.5	13.7	13.3	0.3
Ag	0.2	2900	0.2	0.2	0.2	0.2	0.2	0.3	0.5	0.7	1	1	1	1	72	1.5	0.4	2.3
Al	0.36	3047	0.4	1.385	1.955	2	2.19	2.22	2	2.5	3	3.5	3.5	4.17	8	1.2	1.4	0.5
As	0.5	3046	0.5	0.85	0.9	1	1.25	1.3	1.2	2.2	1.5	1.5	2	2	96	4	0.4	1.8
Au	1	3042	1	1	1	1	1	1.5	1	2.1	2	2	2.5	4	64	2.1	0	1
Ba	30	3047	30	132	156.5	160	160.5	175	148	167.8	178.5	195	235	295	680	85.2	71.2	0.5
Be	0.5	3047	0.5	0.5	0.5	0.5	0.5	0.75	0.5	0.8	1	1	1	1	54.1	1	0	1.3
Bi	2	3047	2	2	2	2	2	3	2	2.9	5	5	5	5	10	1.4	0	0.5
Br	1	3046	1	3.4	14.5	17.45	18.05	31	22	25.6	34.5	37.95	43	57.5	132	16.1	14.1	0.6
Ca	0.23	2685	0.2	0.71	0.805	0.87	0.915	1	1	1	1	1	1	1.08	9.1	0.4	0.1	0.4
Cd	0.2	3047	0.2	0.3	0.45	0.5	1	1	1	1	1	1.05	1.1	1.5	6	0.6	0.3	0.5
Co	1	3047	1	4	4.5	5	5.5	5.5	6	6.9	6.5	6.5	8	10.5	105	5	3	0.7
Cr	1	3047	1	18	25.5	26	31.5	32	27	31.3	32	41	41.5	47.5	328	18.2	13.3	0.6
Cu	2	3047	2	13	17	21	23.5	28	29	34.2	31.5	37.5	44	45.5	441	24.3	14.8	0.7
Fe	0.14	2649	0.1	0.4	0.45	0.965	1	1.5	1	1	1.5	1.5	1.505	1.745	15	0.7	0.3	0.7
Hf	1	3046	1	1	1.5	2	2	2.5	2	2.3	2.5	2.5	3.5	6	30	1.4	1.5	0.6
K	0.05	1809	0.1	0.19	0.265	0.33	0.37	0.425	0.3	0.5	0.56	1	1	1	2	0.3	0.3	0.7
La	1	3046	1	13	20.5	27	27	31	25	29	38	44.5	50	50.5	408	19.3	13.3	0.7
Lu	0.1	1605	0.1	0.15	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.25	1	2	0.2	0	0.7
Mg	0.04	1636	0	0.09	0.095	0.1	0.15	0.28	0.2	0.3	0.285	0.305	0.31	1	2	0.2	0.1	0.9
Mn	20	3047	20	76	89	97.5	101.5	126.5	114	159.8	134	142.5	160	309.5	3410	168	77.1	1.1
Mo	1	3047	1	1	1.5	1.5	1.5	2	2	2.3	2.5	2.5	2.5	3	84	3.2	1.5	1.4
Na	0.03	1999	0	0.17	0.355	0.44	0.52	0.935	0.5	0.7	1	1	1.055	2	4	0.5	0.5	0.8
Ni	3	3047	3	11.5	12	15	15.5	16.5	16	17.3	18	19.5	22.5	29	153	7.9	5.9	0.5
P	150	2197	150	300	515	650	825	830	820	941	970	1060	1105	2315	4700	508.6	474.4	0.5
Pb	2	3047	2	7	8	10	11	12	10	11.6	13.5	14	16	17	1340	27.3	5.9	2.4
Sb	0.1	1627	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.15	0.15	0.2	1	7	0.3	0	1.8
Sc	0.1	3046	0.1	2.6	4.5	5.35	5.4	6	5	5.2	6.35	6.5	7.5	7.5	19	2.2	1.5	0.4
Sr	12	3047	12	48	50	60.5	66	66	60	78.3	94	109.5	117	170	427	54.3	34.1	0.7
Ta	0.5	3046	0.5	0.5	0.5	0.5	0.5	2	2	1.4	2	2	2	2	3	0.7	0	0.5
Th	0.4	3044	0.4	1.9	2.4	2.5	3.25	3.3	3	3.3	3.65	4.5	5.5	8	26	1.7	1.5	0.5
Ti	0.009	1557	0	0.03	0.0495	0.0565	0.06	0.06	0.1	0.1	0.076	0.105	0.121	0.255	0.3	0	0	0.5
U	0.1	3009	0.1	1.9	2.3	2.5	2.65	2.95	2	4.2	4.1	4.5	5	18.5	195.5	7.5	1.5	1.8
V	5	3047	5	11	18.5	22.5	24	24.5	24	27.1	27.5	37	41	45.5	301	15.9	13.3	0.6
W	1	3046	1	1	1	1	1	1	1	1.7	1	1.5	1.5	2	46	1.7	0	1.1
Zn	13	3047	13	52	62.5	63.5	75.5	98	86	98.6	102.5	114	116.5	145	952	68.1	38.5	0.7





Comparative Boxplots



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07

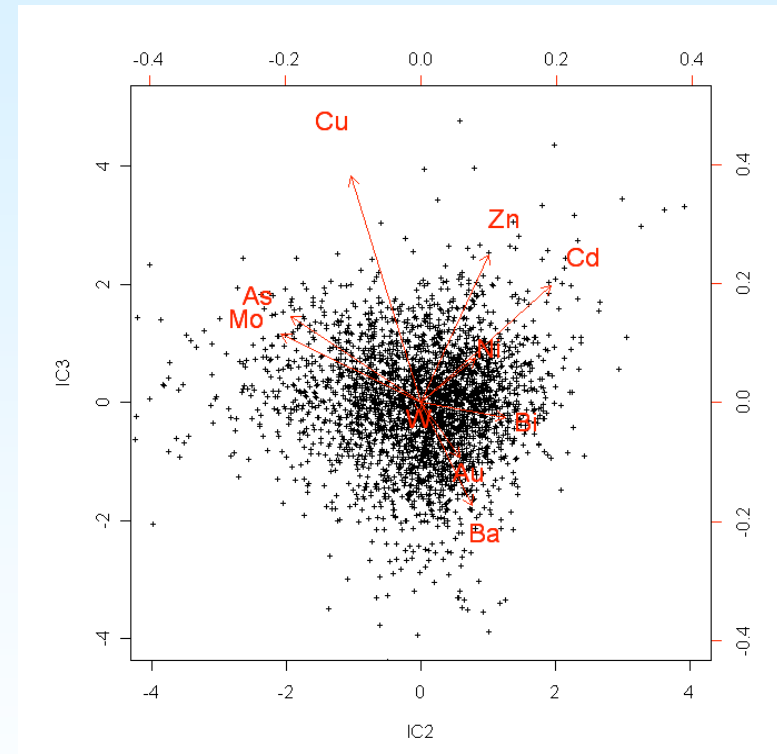


Multivariate Exploratory Approach

Multivariate

- Principal components
- Independent Components
- Multi-dimensional scaling
- Cluster analysis
- χ^2 plots
- Empirical Indices

(NUMCHI, CHI *-6X, PEG-4, Weighted Sums, SCORESUM)





Modelled Approach based on Target and Background Populations

Groups chosen from exploratory analysis &
orientation studies

- Canonical Variate Analysis
- Posterior Probability/Typicality
- Neural Networks / Self Organizing Maps /
Random Forests / Support Vector Machines





Special Problems

- Censoring - samples $<$ detection limit - replacement,
- Non-normal distributions that hamper statistical testing - transformations,
- Missing values and zeros - replacements,
- Different limits of detection and instrumentation - levelling,
- Constant sum (closure) problem - Logratio analysis.





Transformations

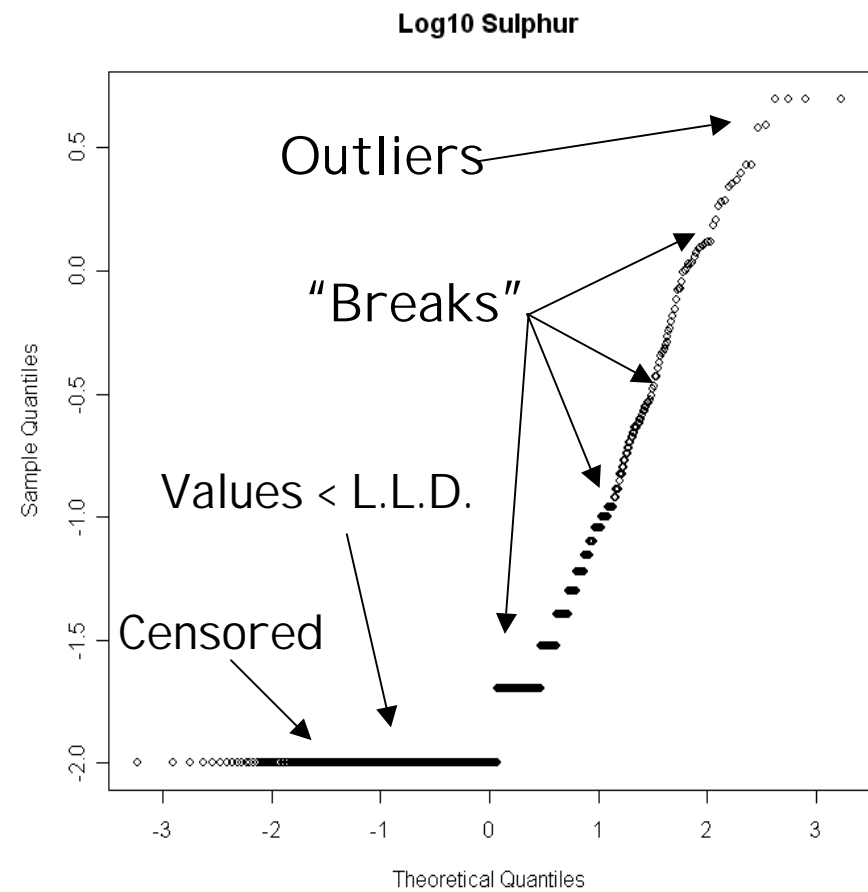
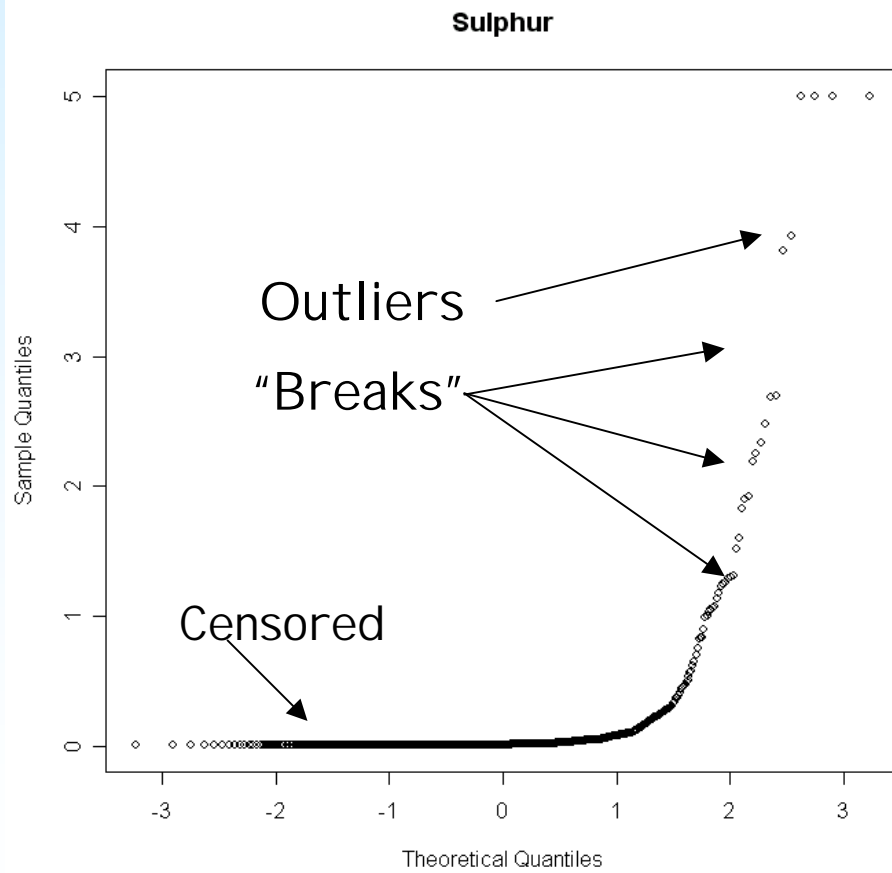
Transformations are useful to:

- Scale the data in order to view subtle features and minimize the effects of outliers,
- Apply statistical tests (i.e. log-ratios)



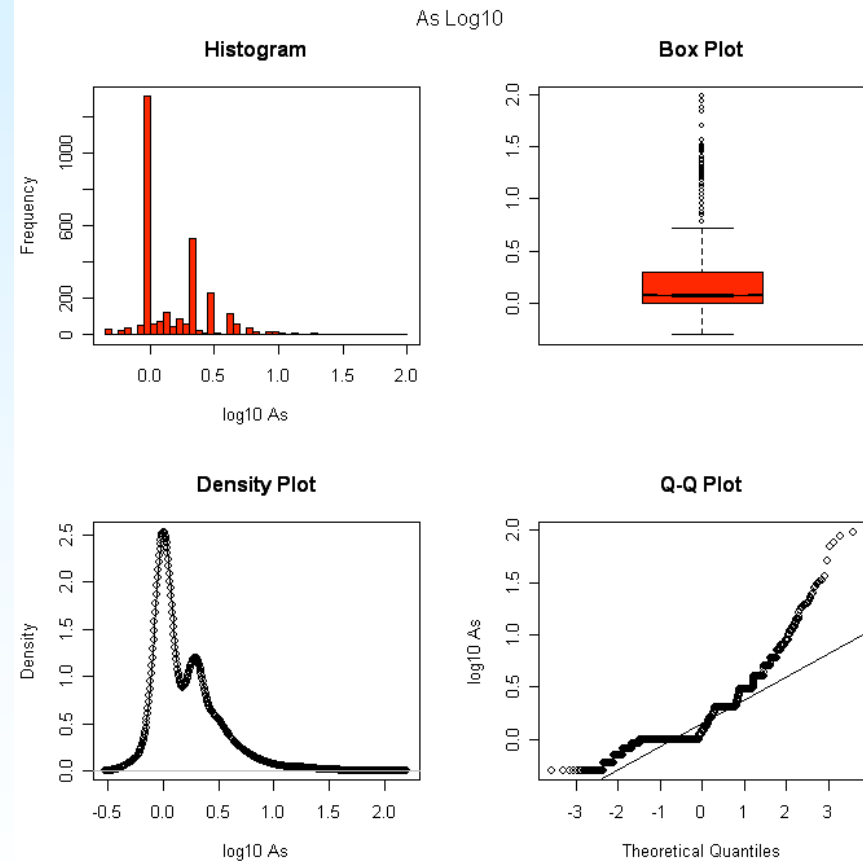
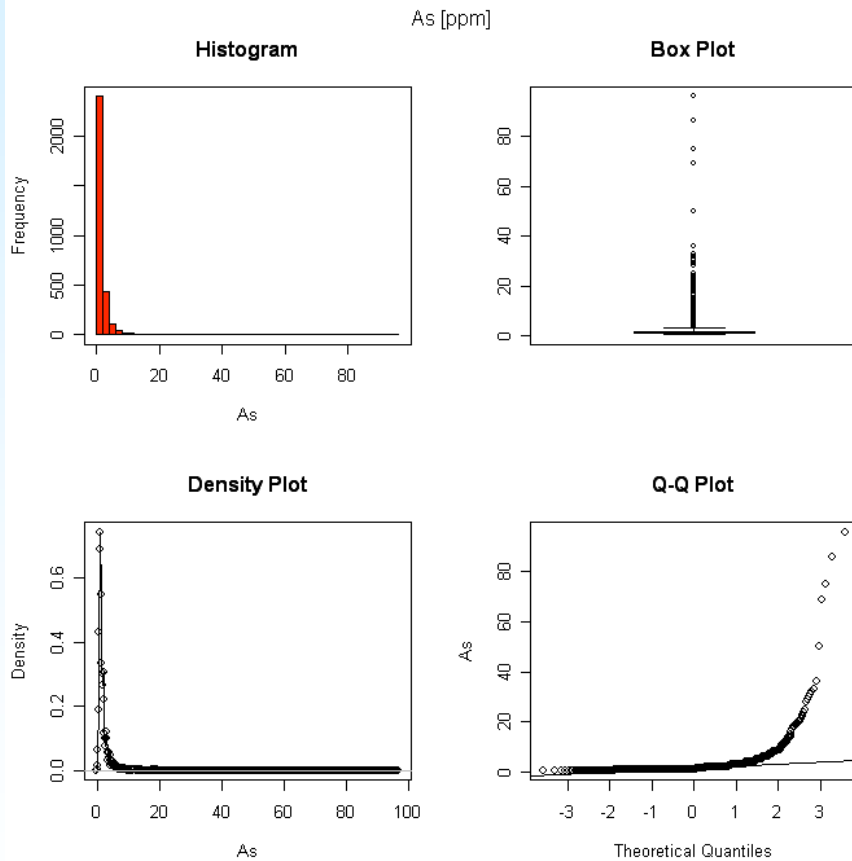


Transformations



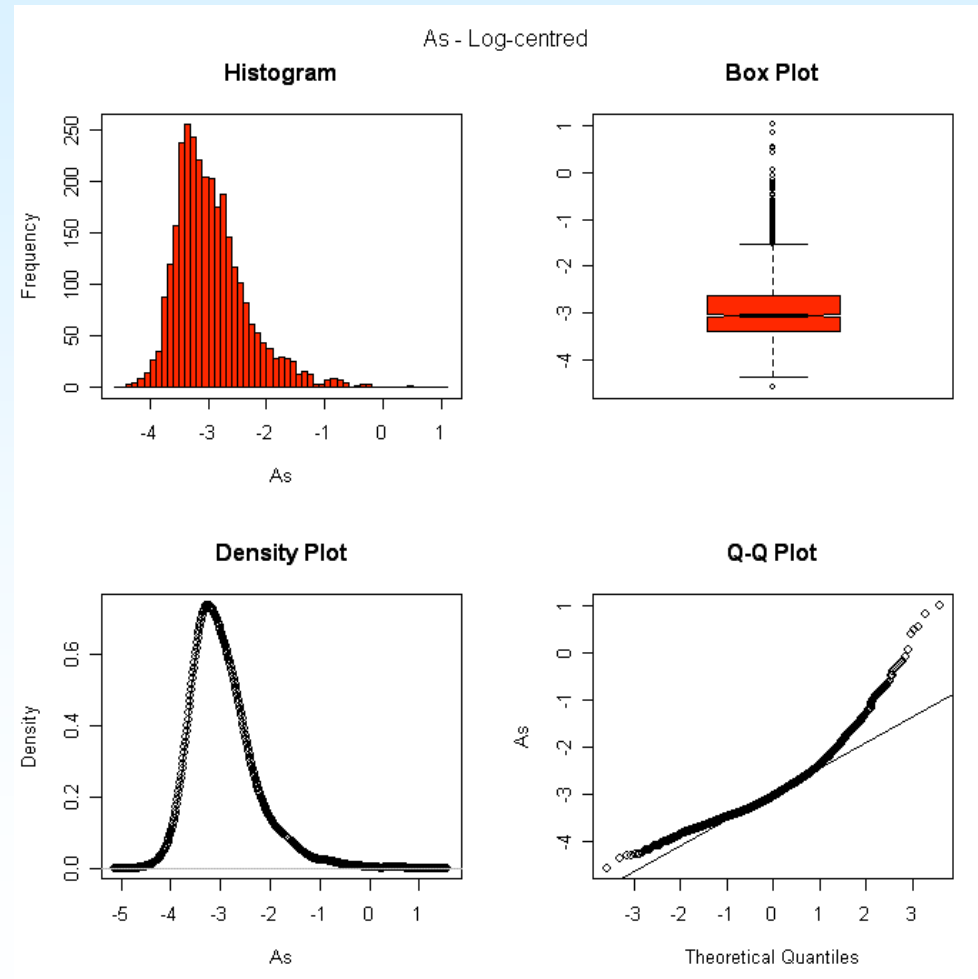


Exploratory Data Analysis & Transformations





Lake Sediments - Logcentred

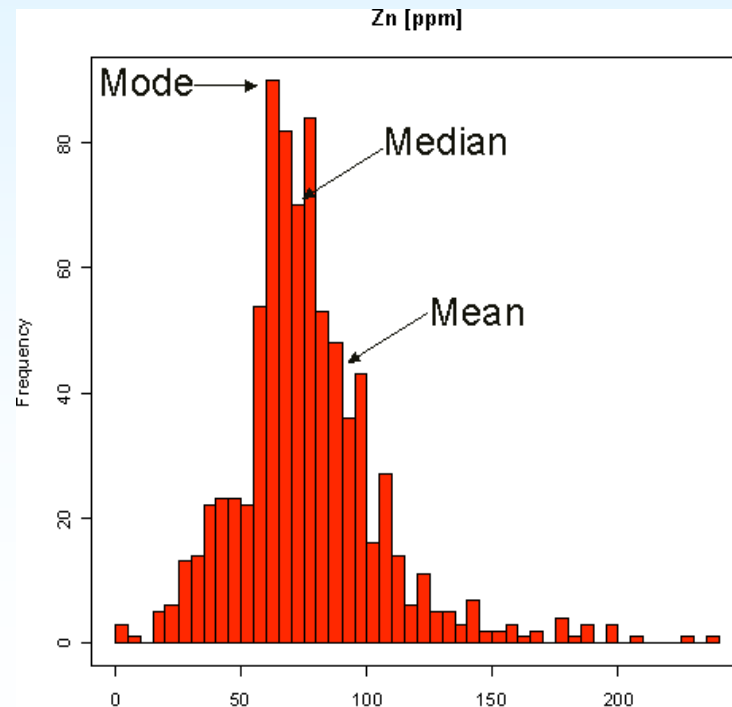




Robust Statistics

- Presence of extreme or atypical values in a sample population can have a dramatic effect on the estimation of the mean and variance.

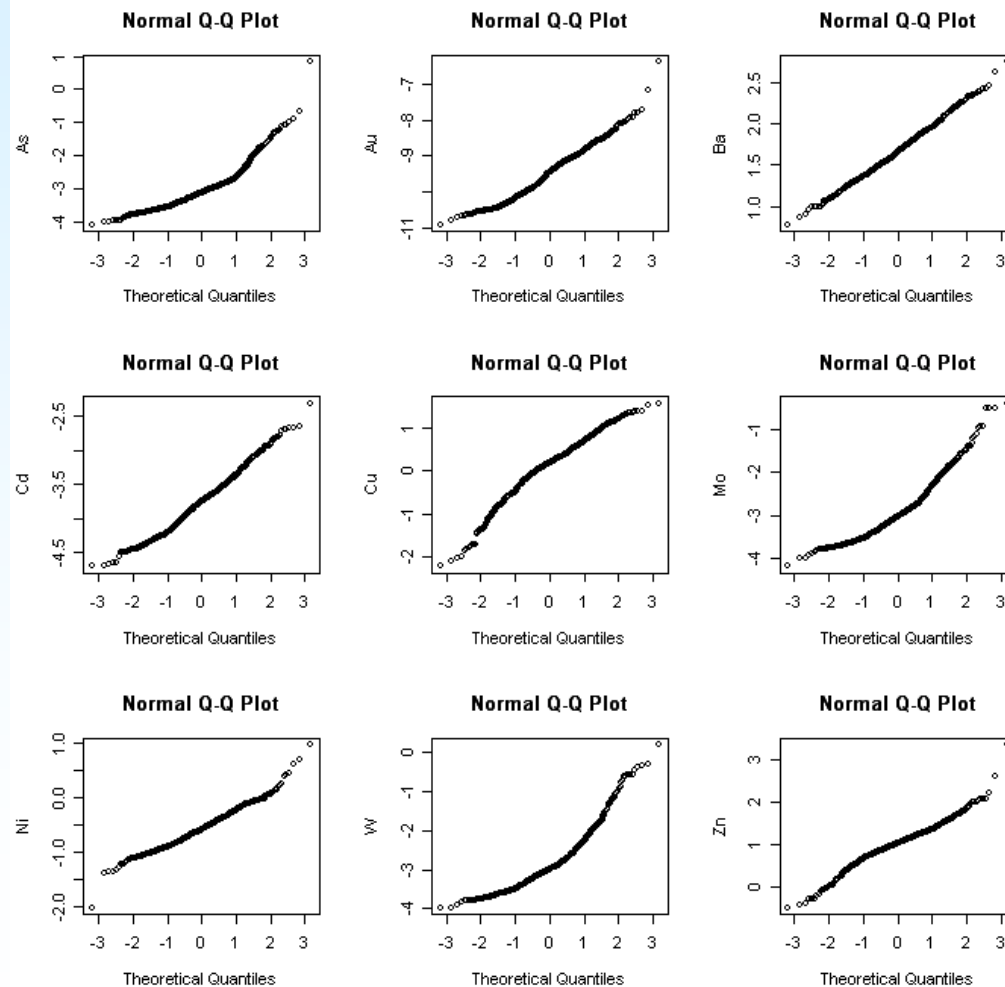
Univariate Methods	Zn
Mean	89
Robust Estimate (Huber M Estimator)	76
Univariate Median	74
Mode	65
Multivariate Methods	
Minimum Volume Ellipsoid	73
Minimum Covariance Determinant 800 observations	78
Minimum Covariance Determinant 540 observations	73





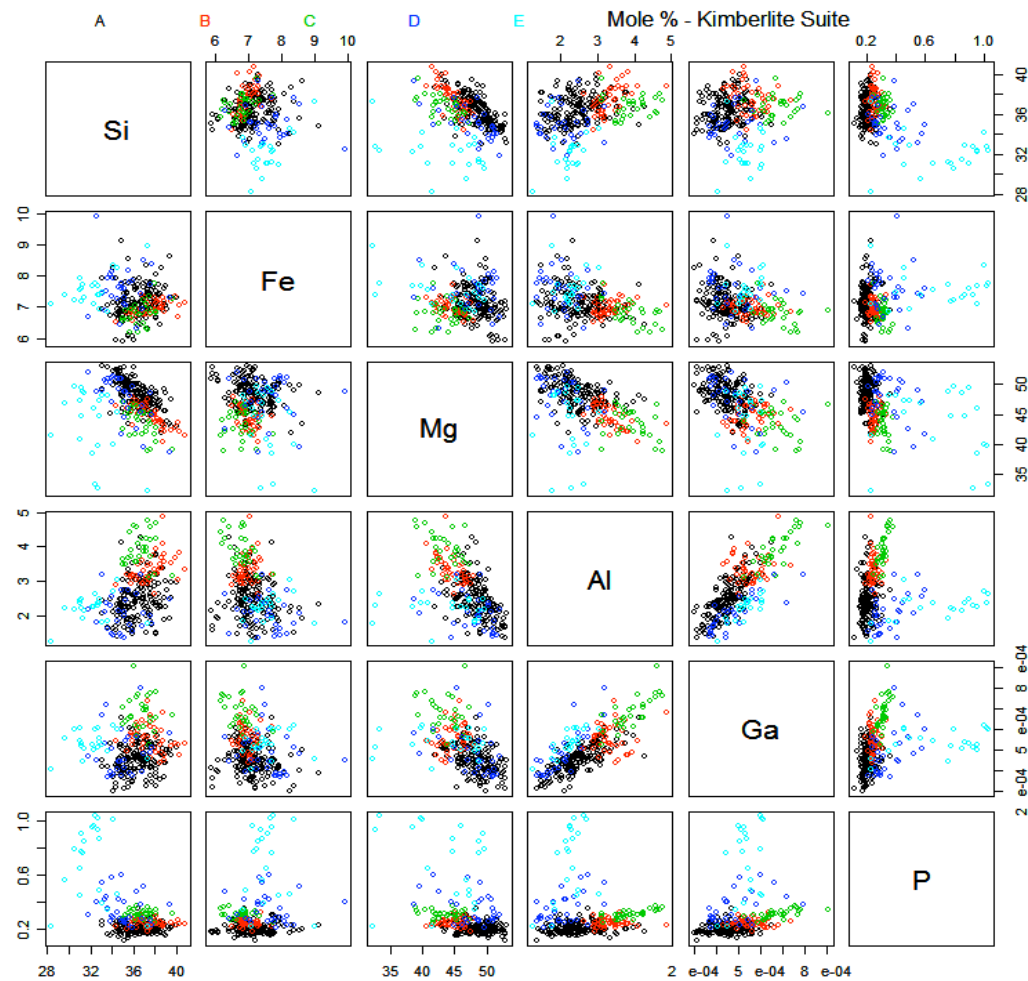
Multivariate Visualization Quantile-Quantile Plots

Batchawana Lake Sediments - Log-centred



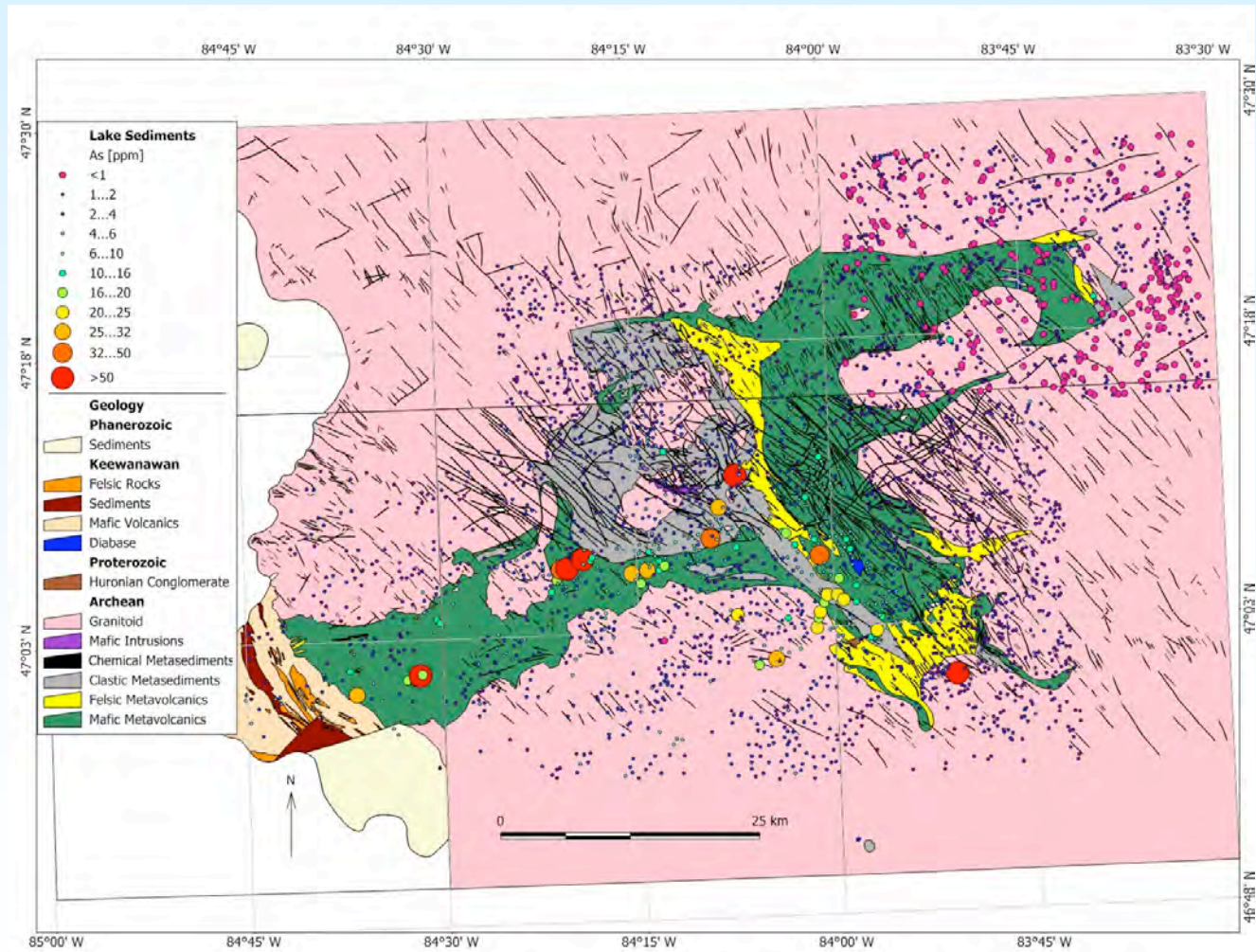


Multi-element Relationships Pairs Plots





Spatial Evaluation Bubble Plots



Natural Resources
Canada

Ressources naturelles
Canada

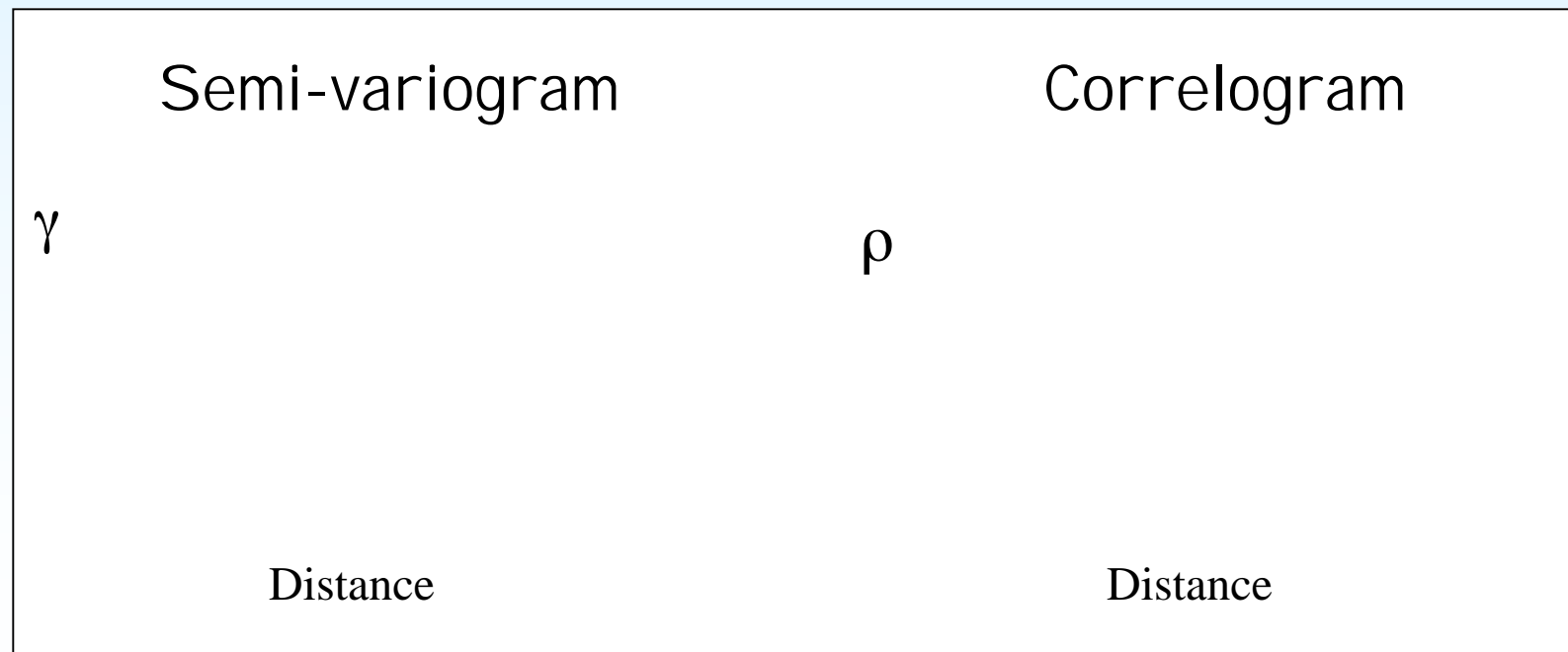


Exploration07



Adequate Support - Spatial Analysis

- Spatial Structure - Fe - Lake Sediments
- Adequate sampling density (support)

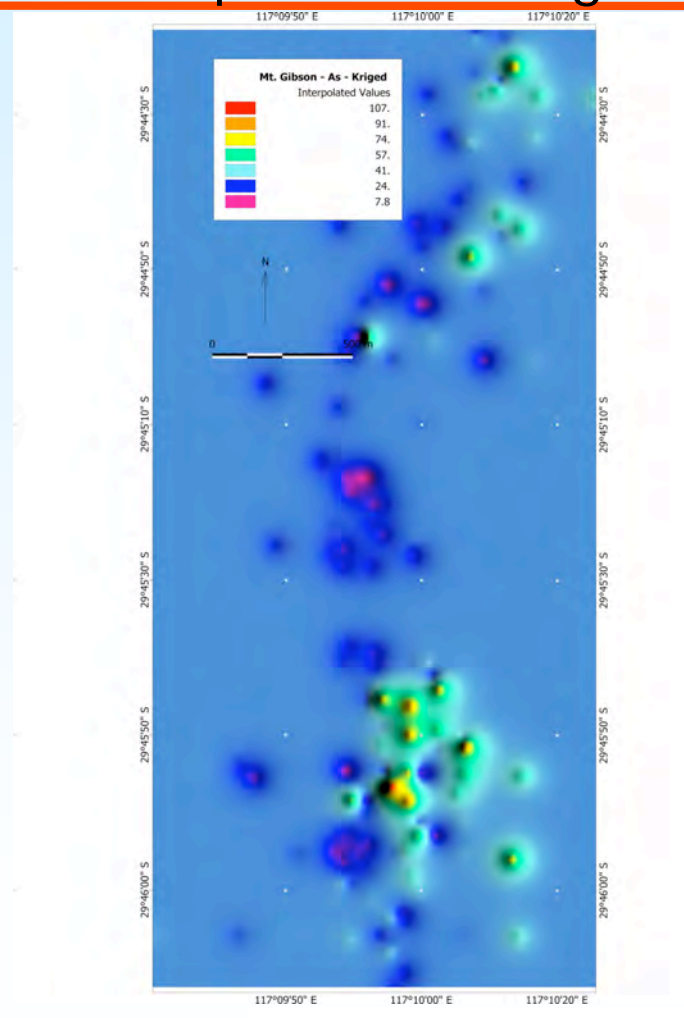
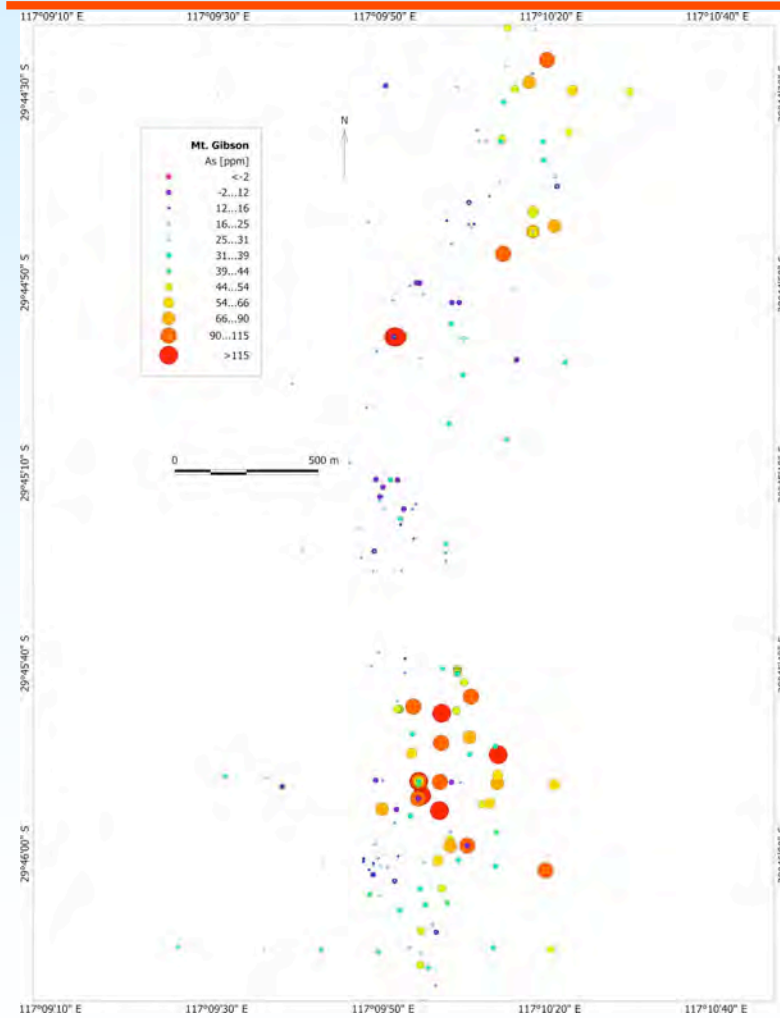




Spatial Presentation

Bubble Plot

Interpolated Image



Natural Resources
Canada

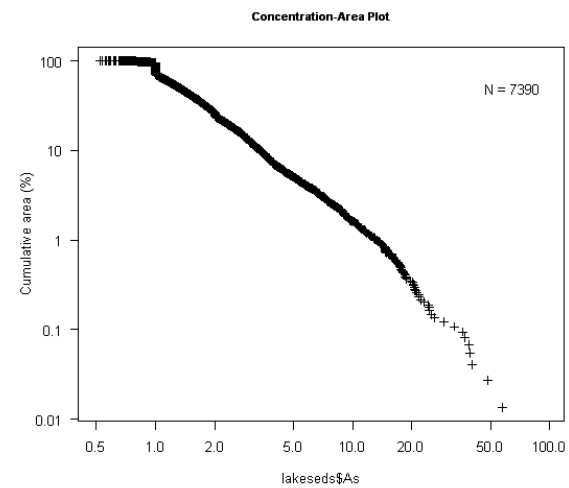
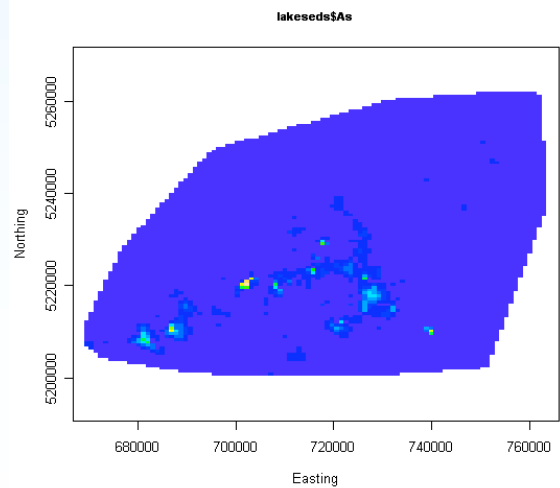
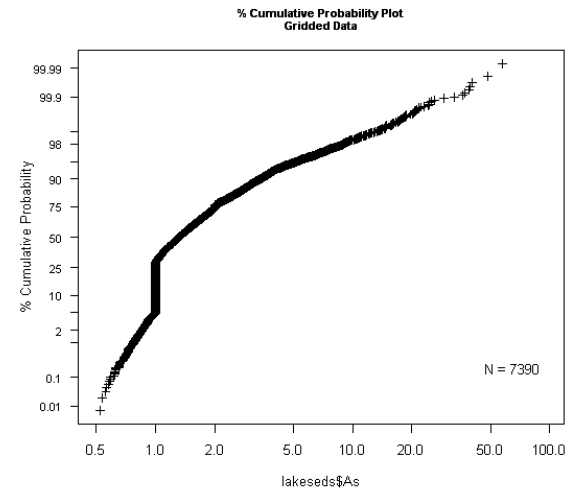
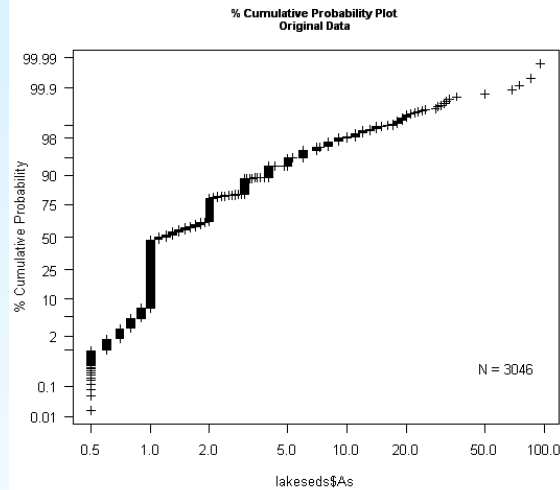
Ressources naturelles
Canada



Exploration07



Spatial Analysis and Anomaly Recognition Using Fractals





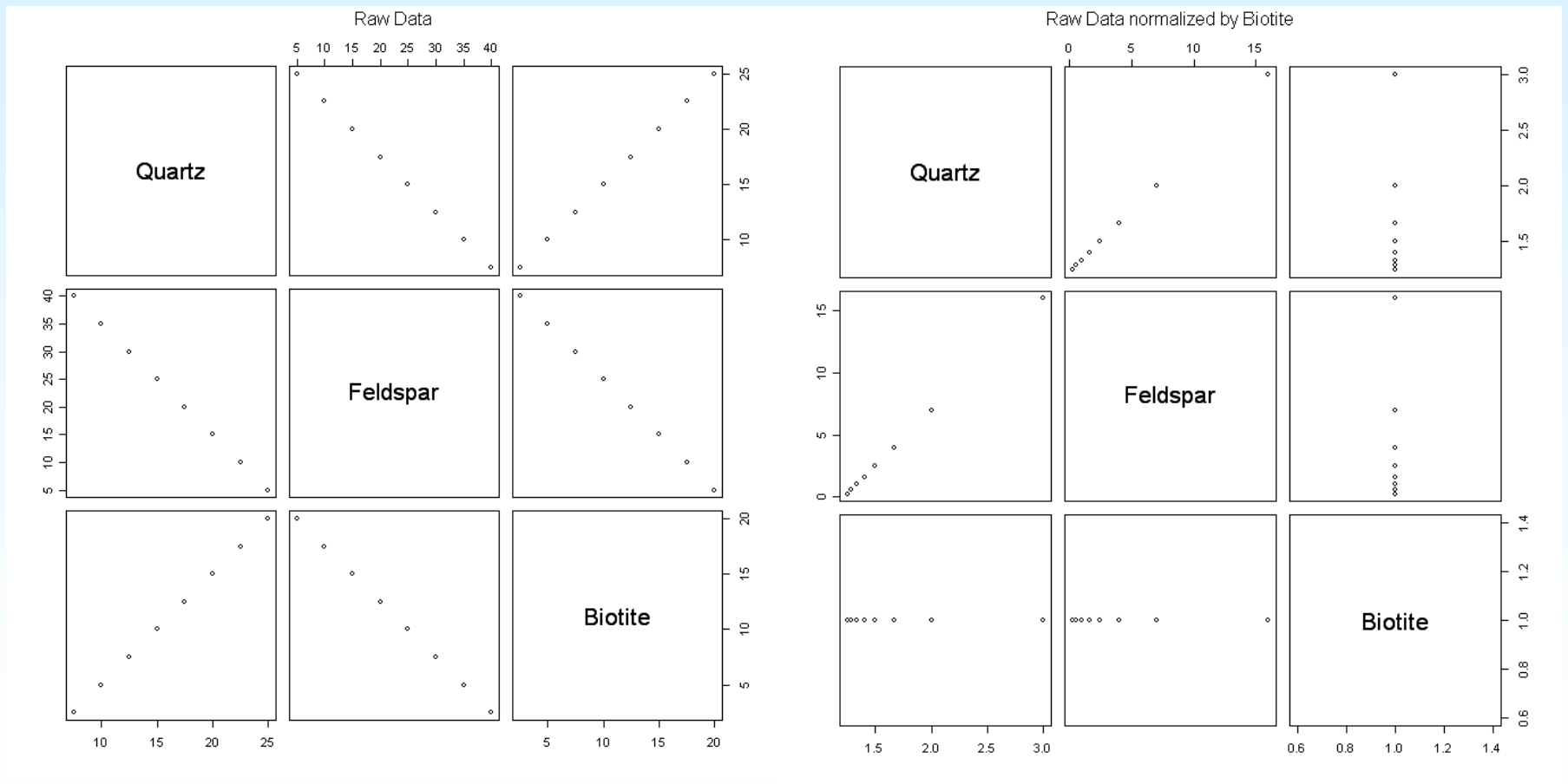
Compositional Data

Quartz	Feldspar	Biotite	Total
15	80	5	100
20	70	10	100
25	60	15	100
30	50	20	100
35	40	25	100
40	30	30	100
45	20	35	100
50	10	40	100





Compositional Relationships



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07



Compositional Data - Logratios

Aitchison (1982) described the use of logratios as a way of overcoming the problem of compositional data for statistical analysis.

Logratio

- $y_i = \log(x_i/x_D)$ ($i = 1, \dots, D-1$)

where x_D = a compositional component of choice

Centred Logratio

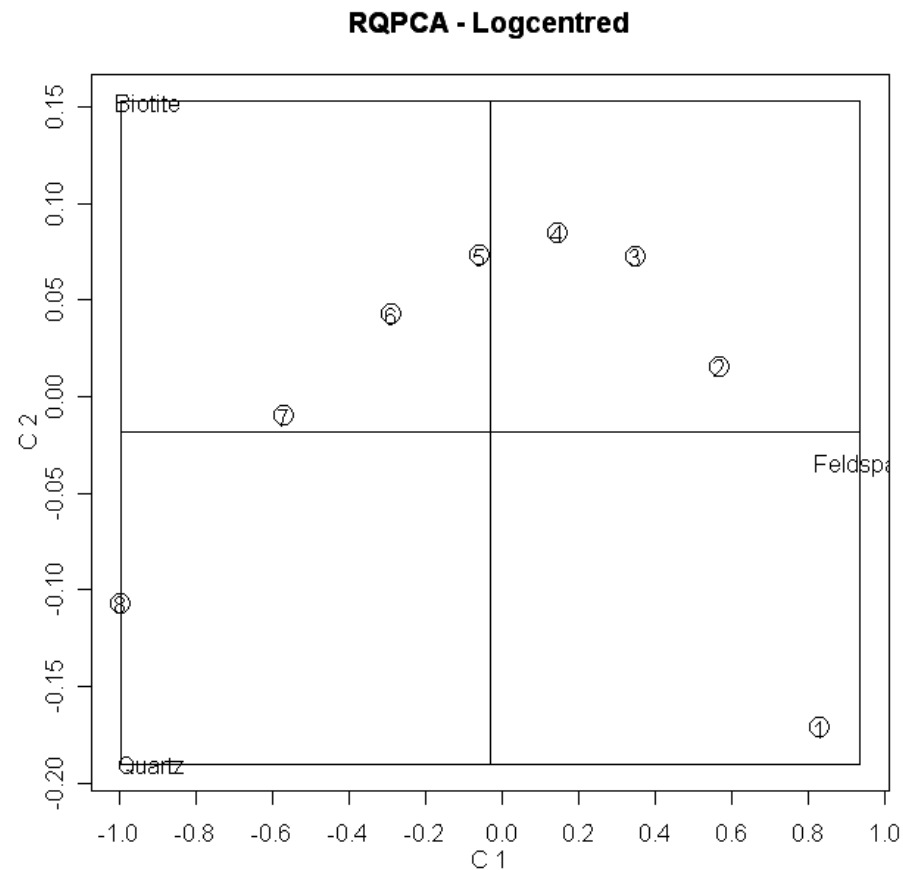
$$z_i = \log(x_i/g(x_D)) \quad (i = 1, \dots, D),$$

where $g(x_D)$ is the geometric mean of the composition





Assessing the Entire Composition with Multivariate Methods





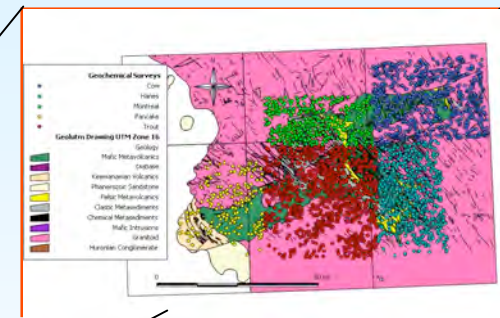
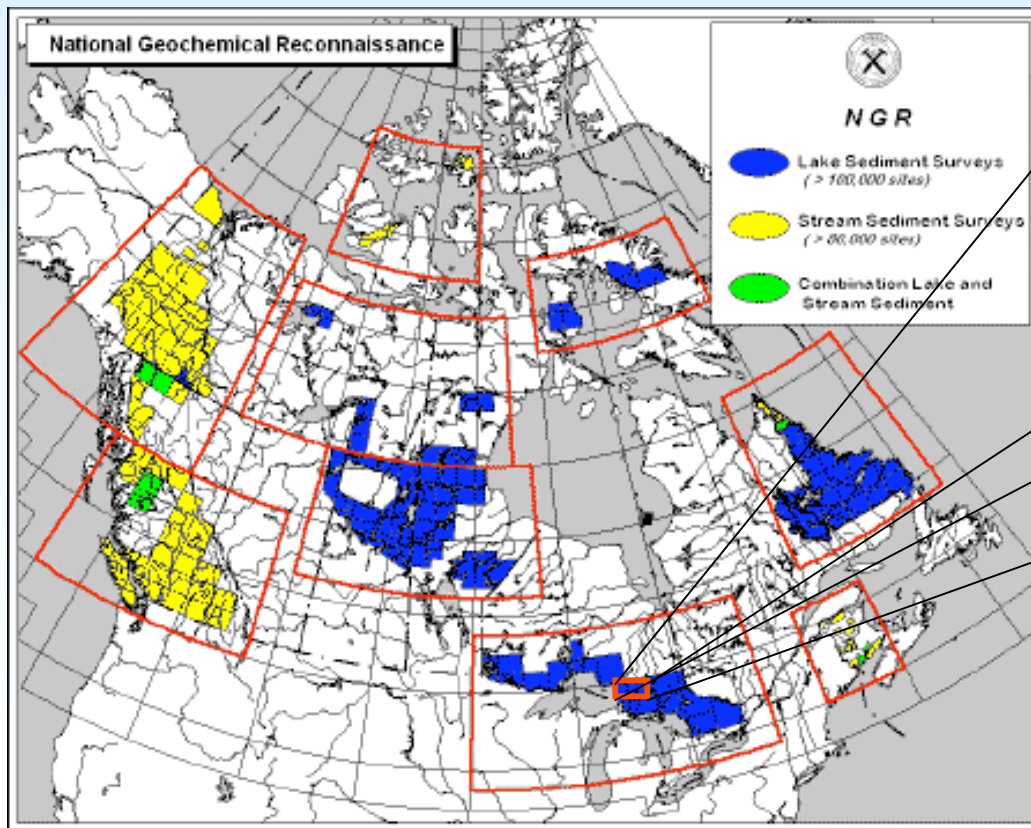
Pearce Element Ratios

- PER's are based on the "preservation" of a constituent within a magmatic system.
- It is the relative change of a constituent w.r.t another constituent that describes compositional variation.
- $y_i = x_i/x_j$ (where x_j is the preserved constituent and $i = 1, \dots, D-1$)
- The analysis of y_i ($i = 1, \dots, D-1$) is more likely to define compositional variation related to stoichiometric processes.





Levelling Geochemical Survey Data



Natural Resources
Canada

Ressources naturelles
Canada

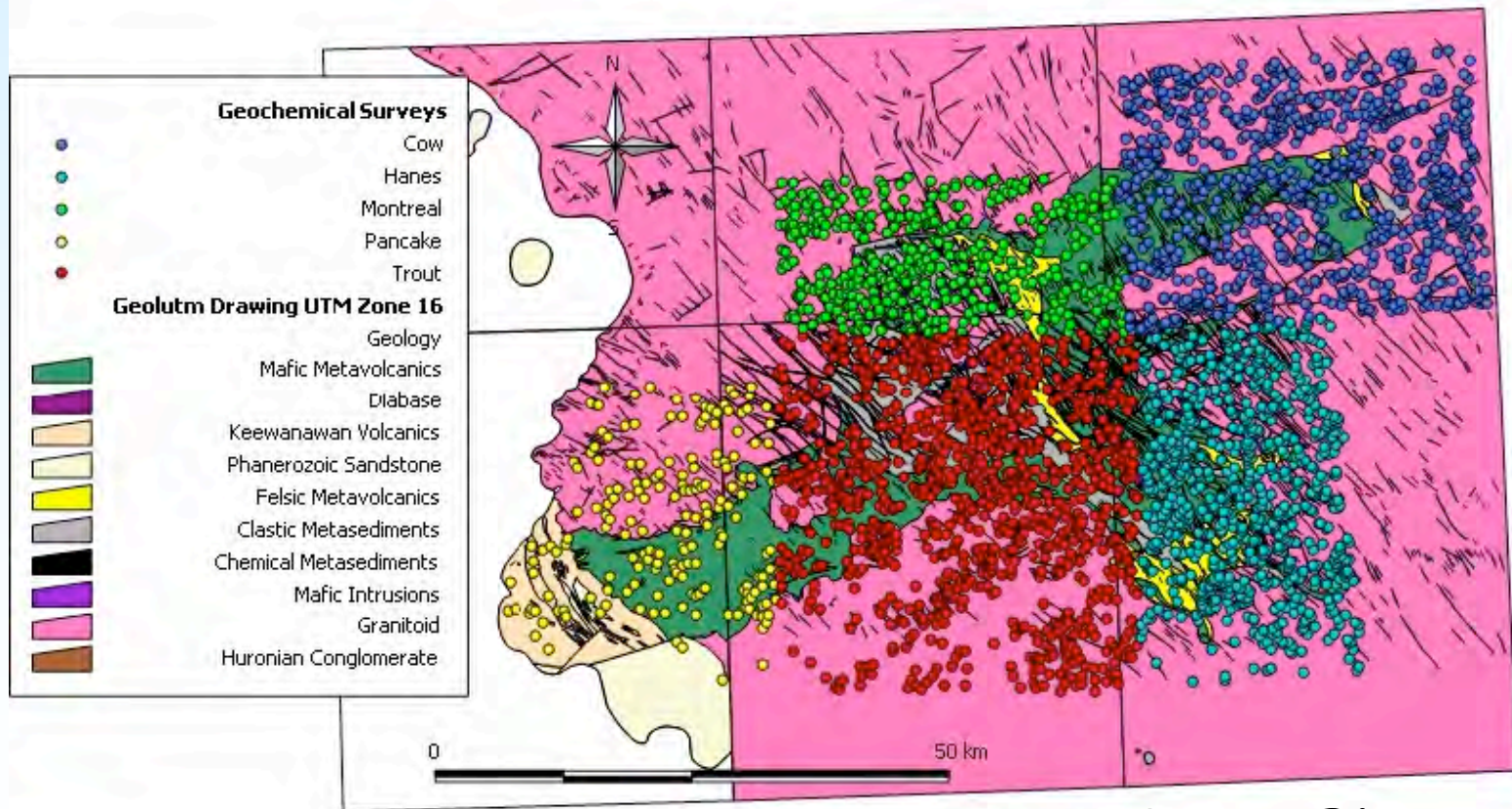


Exploration07



Levelling Geochemical Survey Data

> 35 elements Batchawana Greenstone Belt



3047 Sites



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07



Parametric Levelling Scenarios

Cr - Lake Sediment Surveys

Quantile Interval=0.05



Natural Resources
Canada

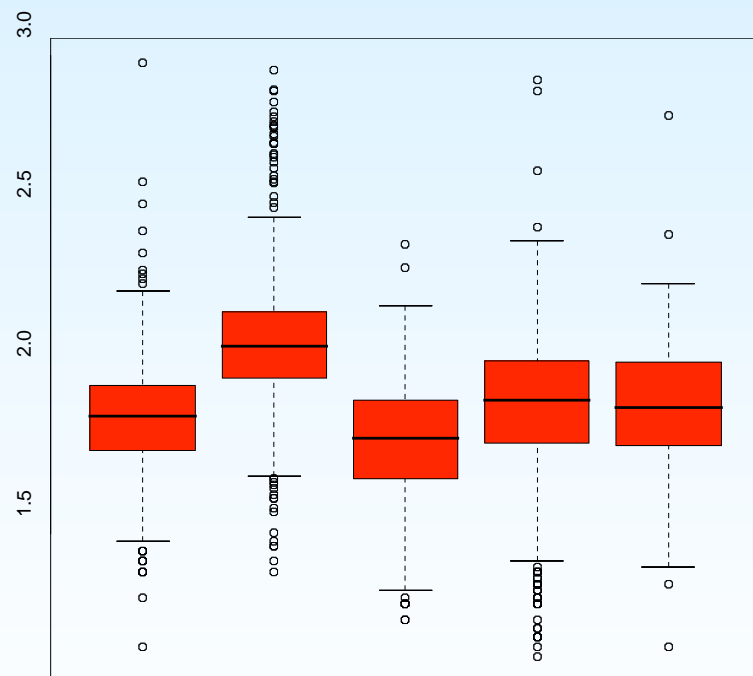
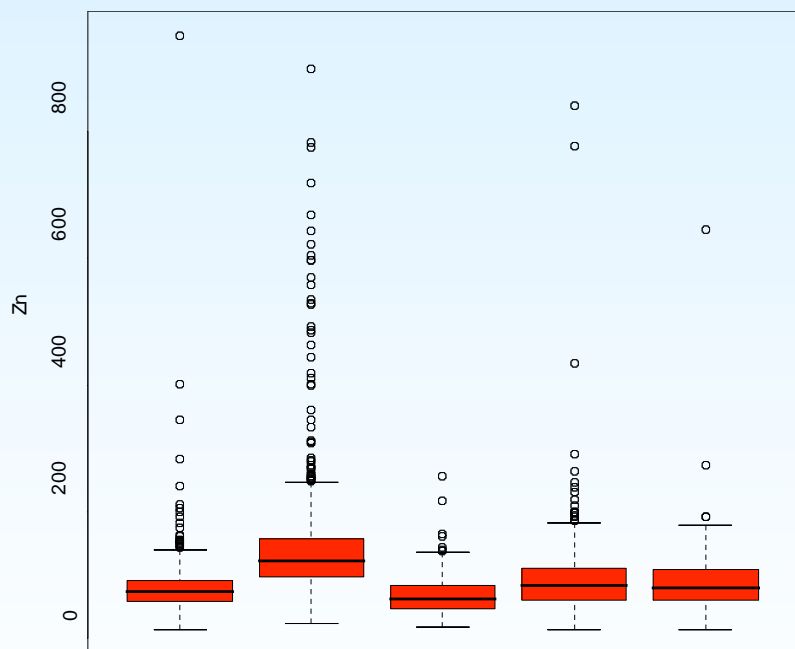
Ressources naturelles
Canada



Exploration07

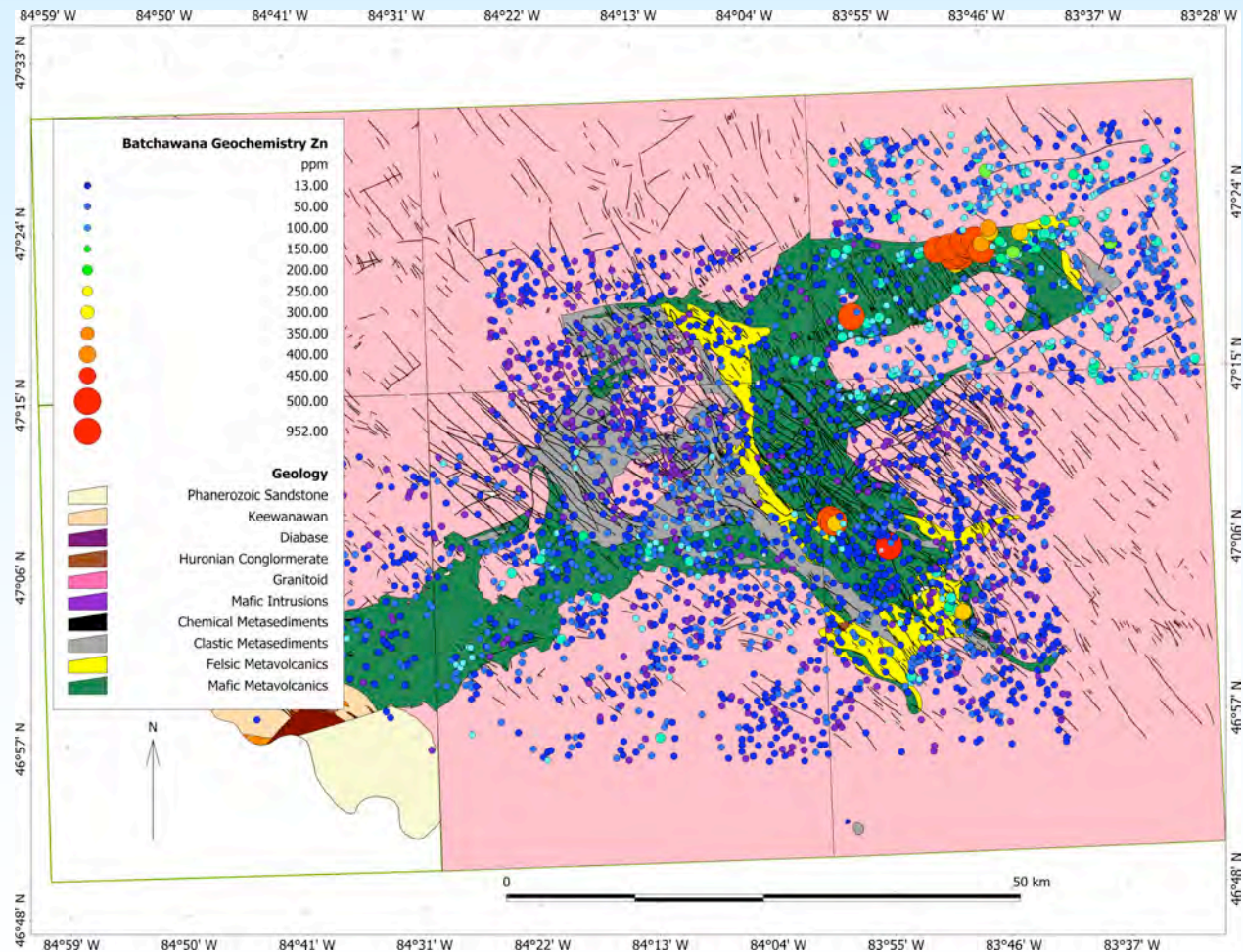


Comparative Boxplots for Levelling





Zn in Lake Sediments Batchawana Area



Natural Resources
Canada

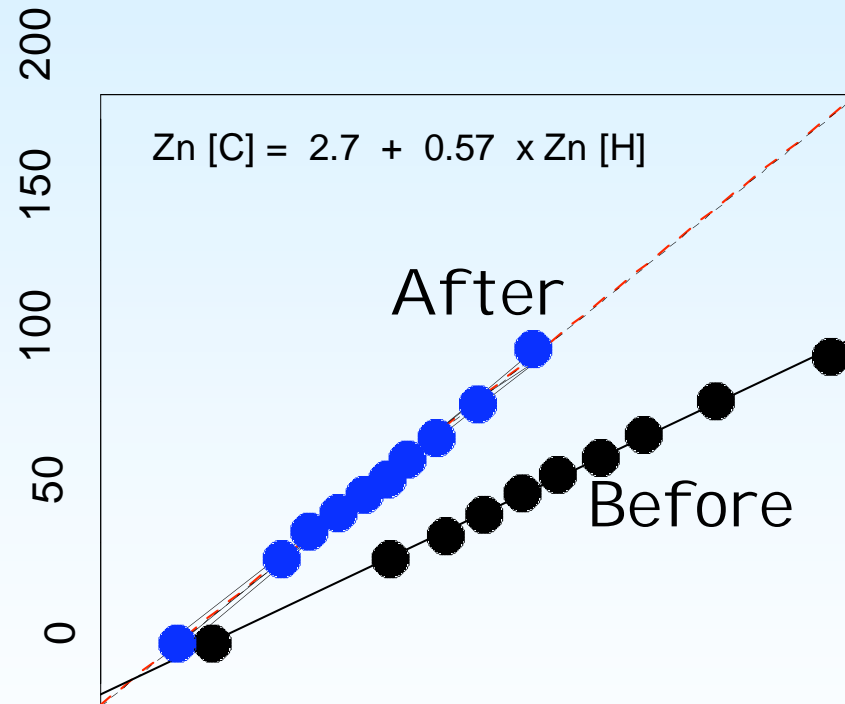
Ressources naturelles
Canada

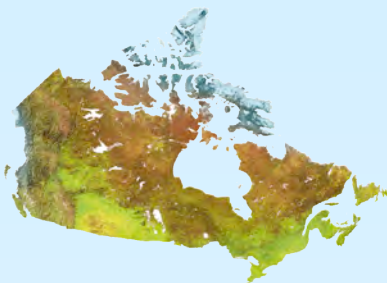


Exploration07

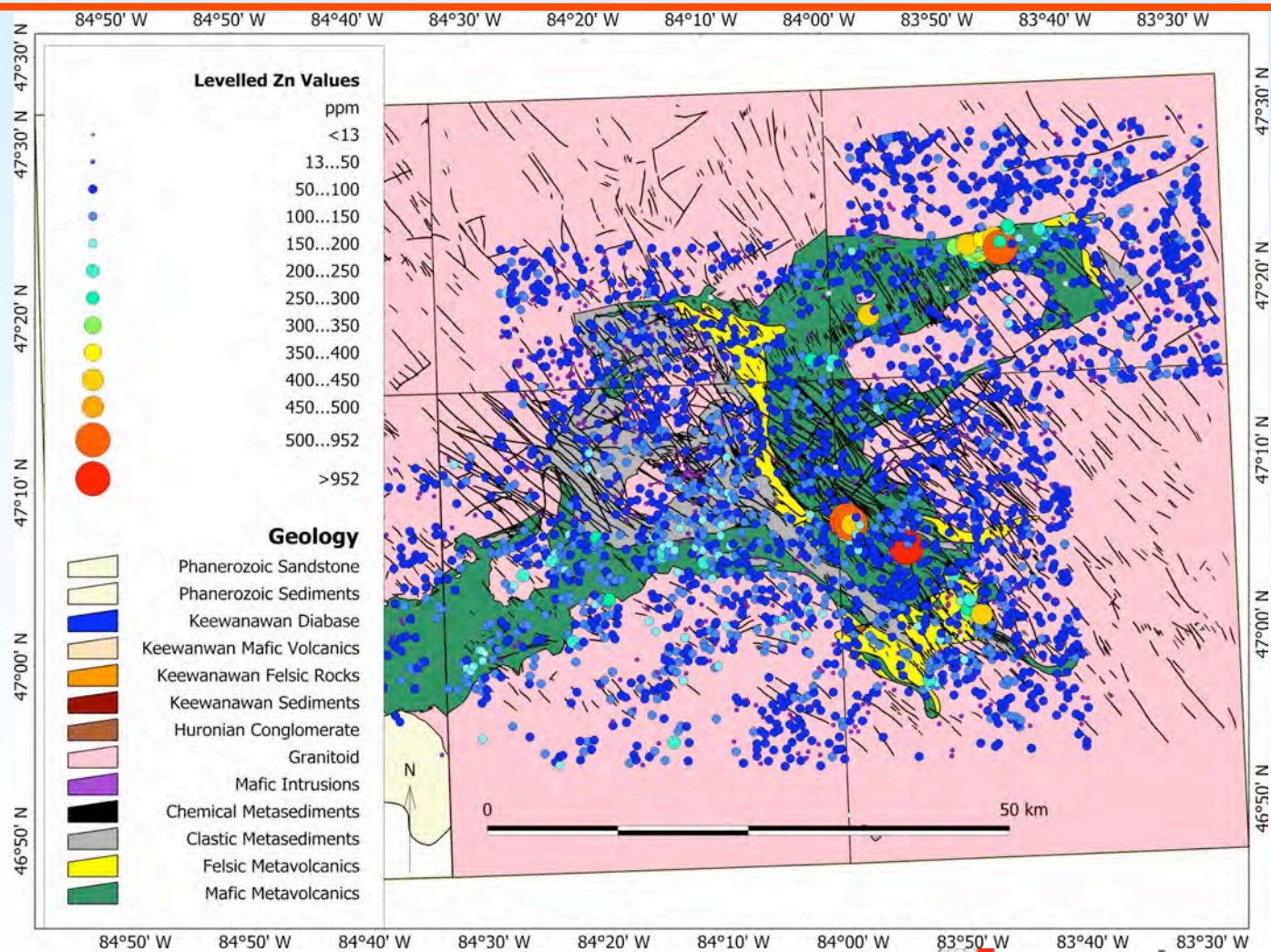


Levelling Equation Quantile Linear Regression





Zn in Lake Sediments after levelling





Multivariate Methods

- In many surveys a number of elements may be important.
 - Commodity elements
 - Pathfinder elements
 - Background characterization
- There are many methods to evaluate multi-element geochemical data.





Empirical Methods

- Weighted Sums - a choice of specific elements that define a process of interest. Weights are assigned in terms of importance
- SCORESUM - a score is given for each element that exceeds a predetermined threshold (i.e. if 6 elements exceed their respective thresholds, the SCORESUM value is 6)
- Specific Indices
- $CHI - 6 * X = As + 3.56xSb + 10xBi + 3xMo + 30xAg + 30xSn + 10xW + 3.5xSe$
- Mahalanobis Distance Plots (X^2 plots) - multivariate equivalent of a quantile-quantile plot.





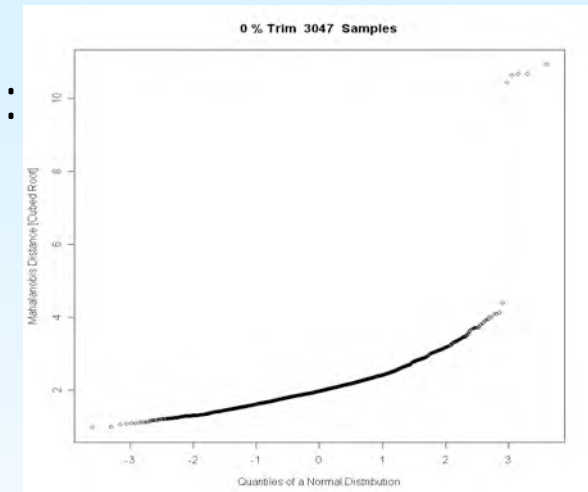
χ^2 Plot

The Mahalanobis distance is defined as:

$$D^2 = [x \quad -\bar{x}]' C^{-1} [x \quad -\bar{x}]$$

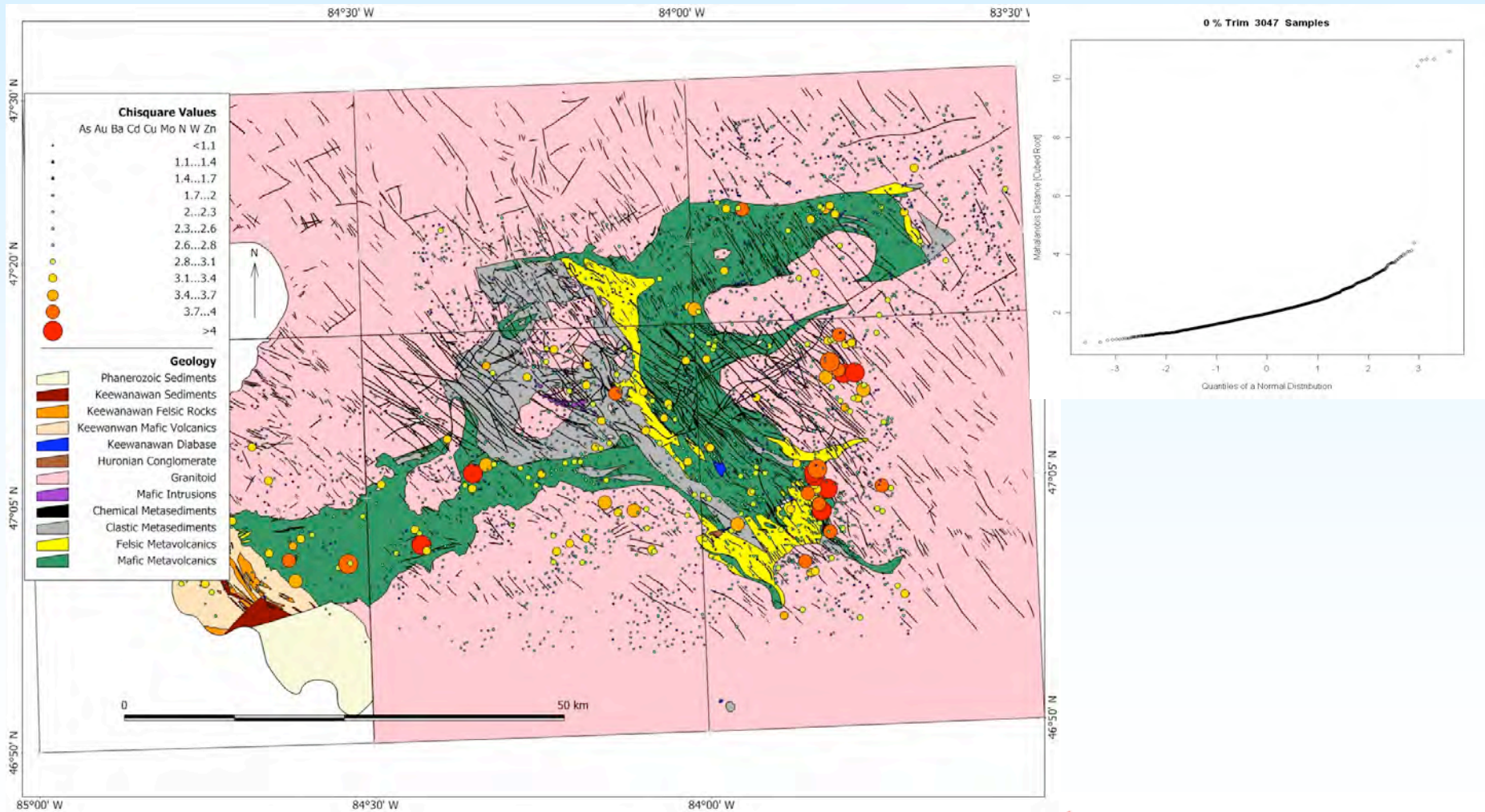
Where:

- x is a vector of variables for the observations,
- \bar{x} is a vector of the group mean,
- C^{-1} is the inverse of the covariance matrix.





Chi-square Plot



Natural Resources
Canada

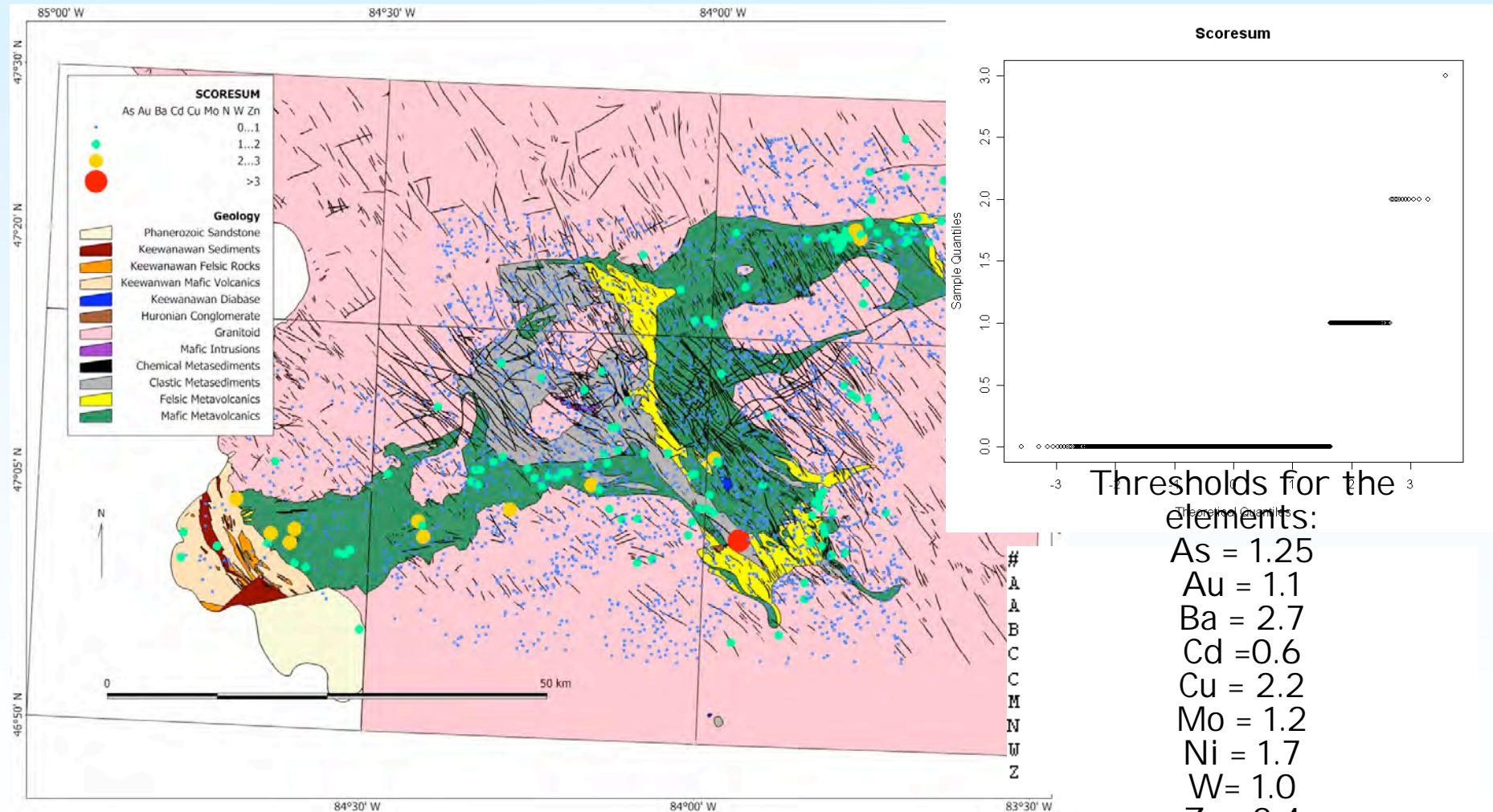
Ressources naturelles
Canada



Exploration07



SCORESUM



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07



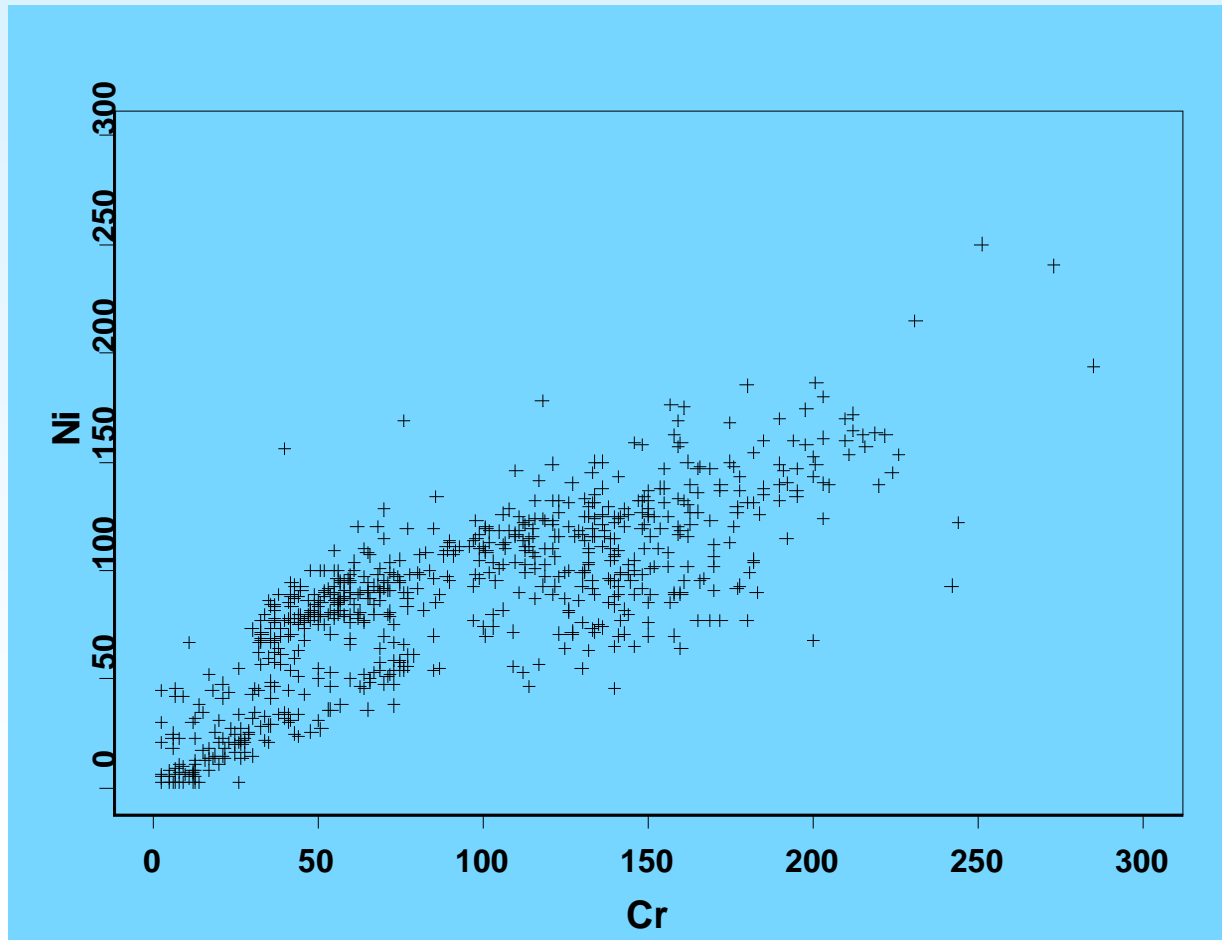
Principal Components Analysis

- A multivariate method based on the correlation/covariance of groups of elements.
- Based on correlations linear combinations of elements can be extracted that are orthogonal (independent of each other).
- Each successive component accounts for less of the overall data variation.





Metavolcanics & Mafic Intrusions



Natural Resources
Canada

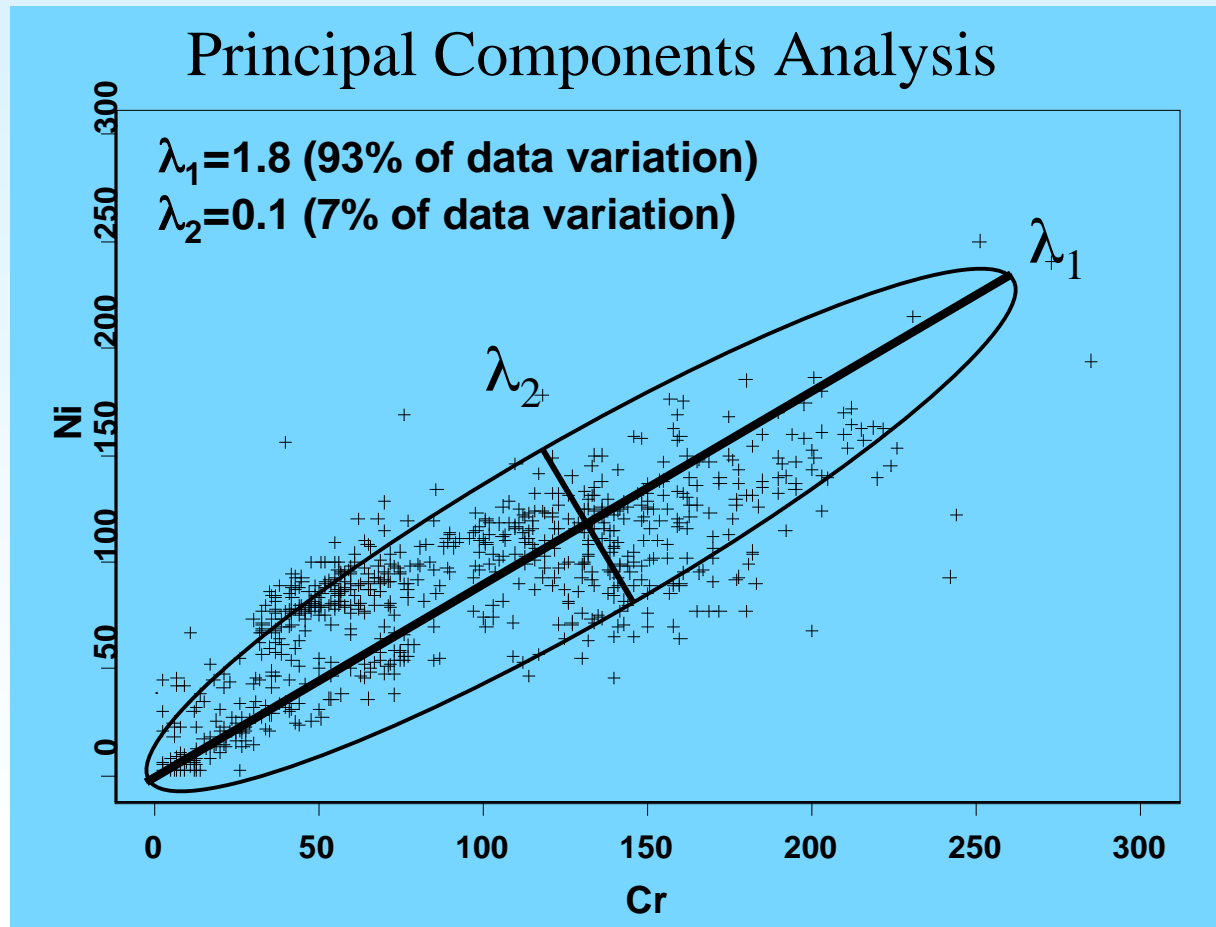
Ressources naturelles
Canada



Exploration07

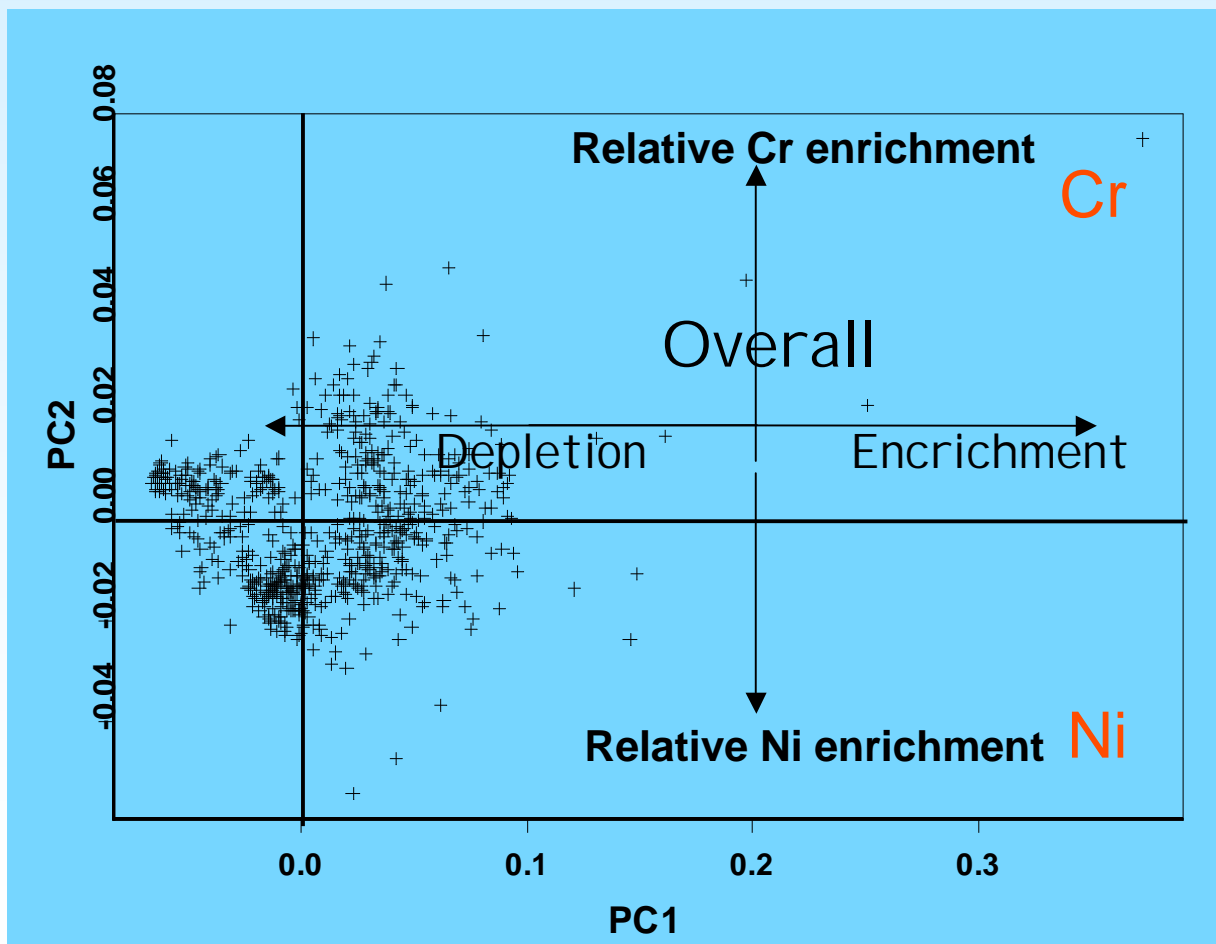


Metavolcanics & Mafic Intrusions



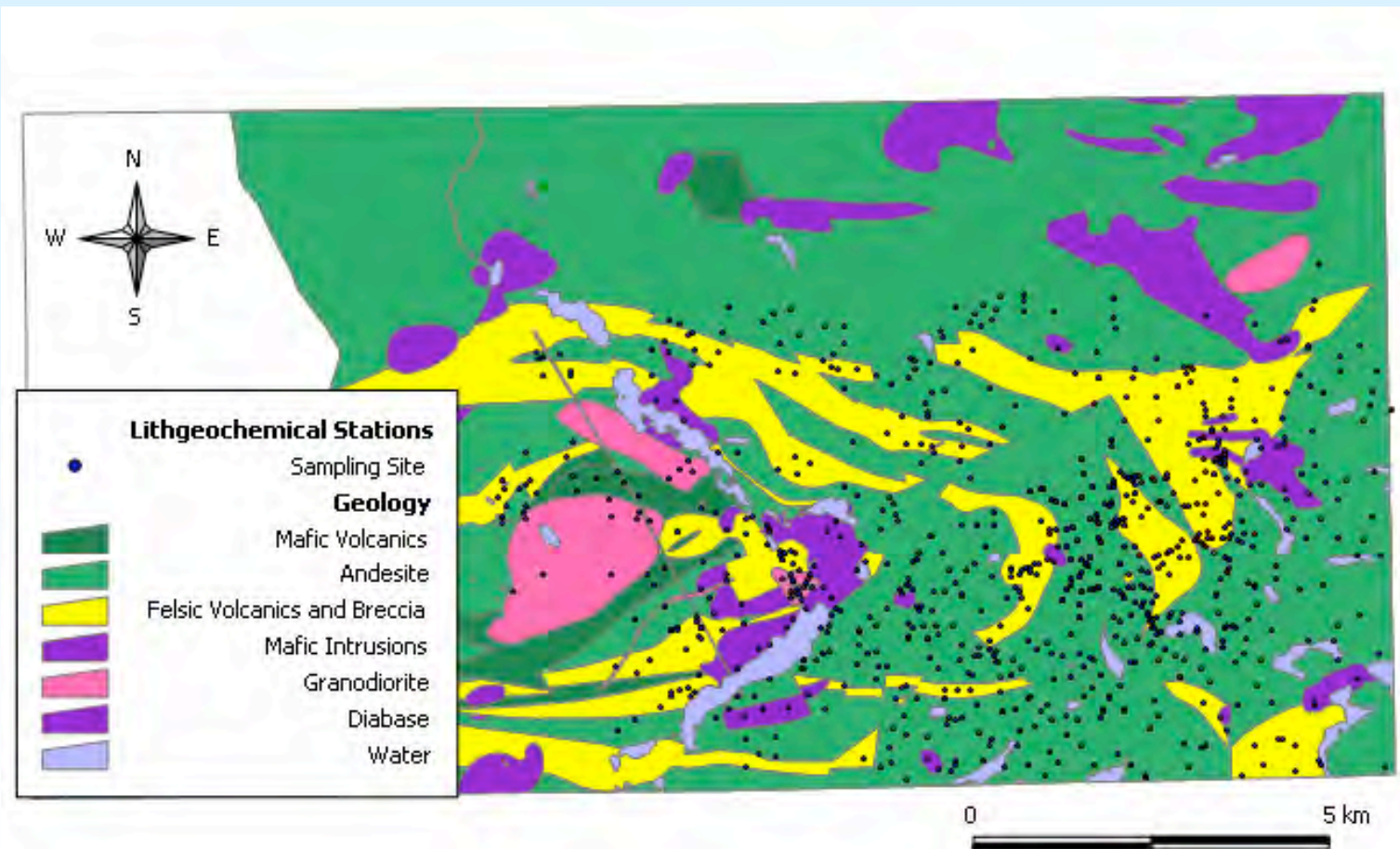


Principal Components Analysis Metavolcanics & Mafic Intrusions





Mapping Volcanic Stratigraphy, zones of alteration and mineralization



Sample Location Map – 825 sites



Natural Resources
Canada

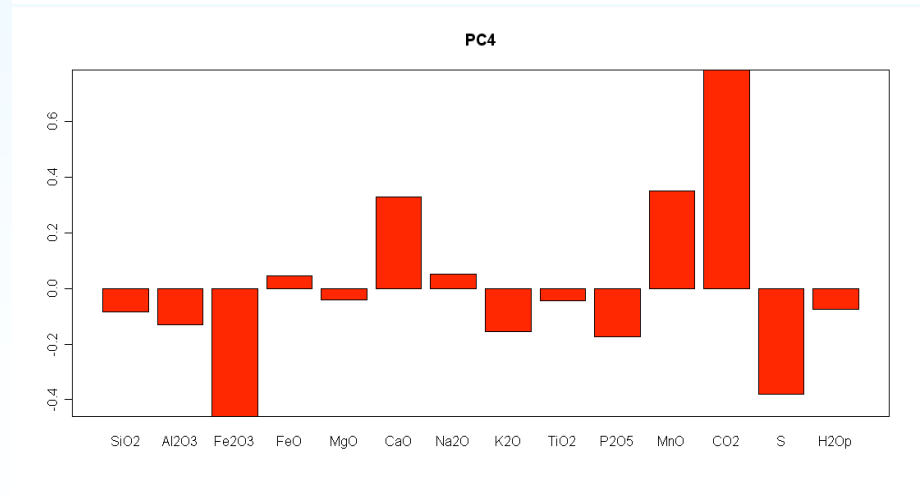
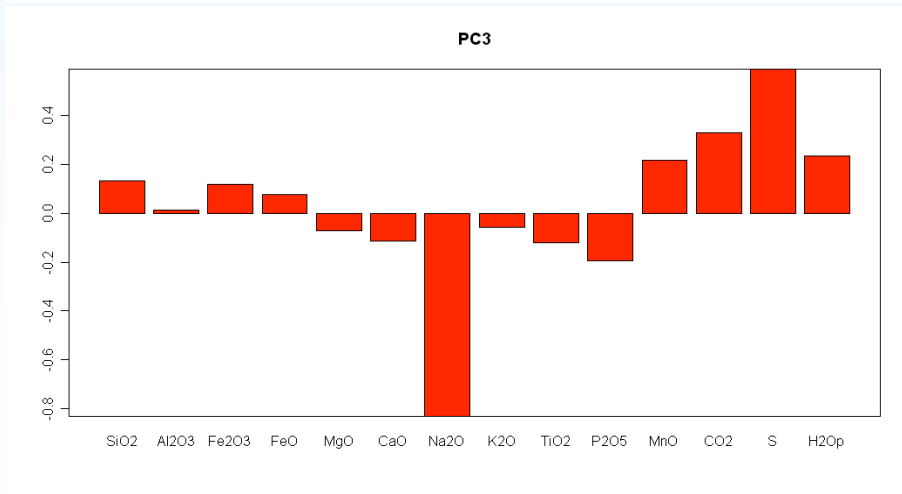
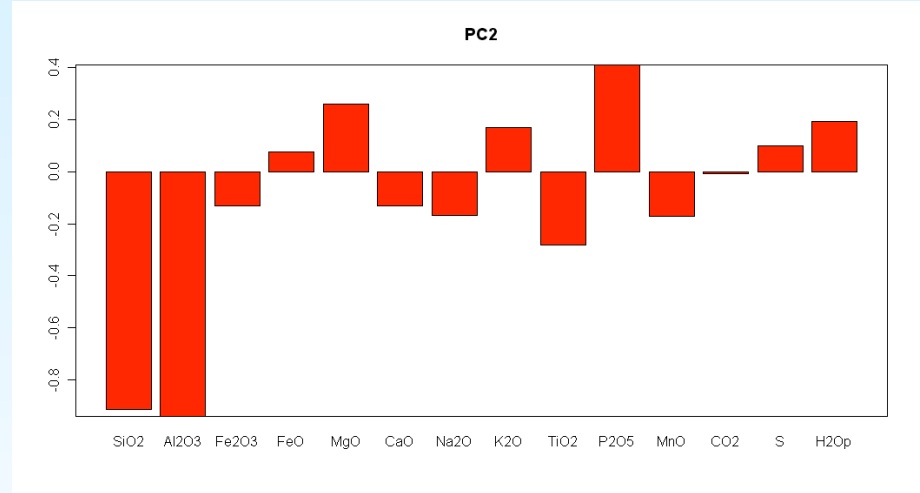
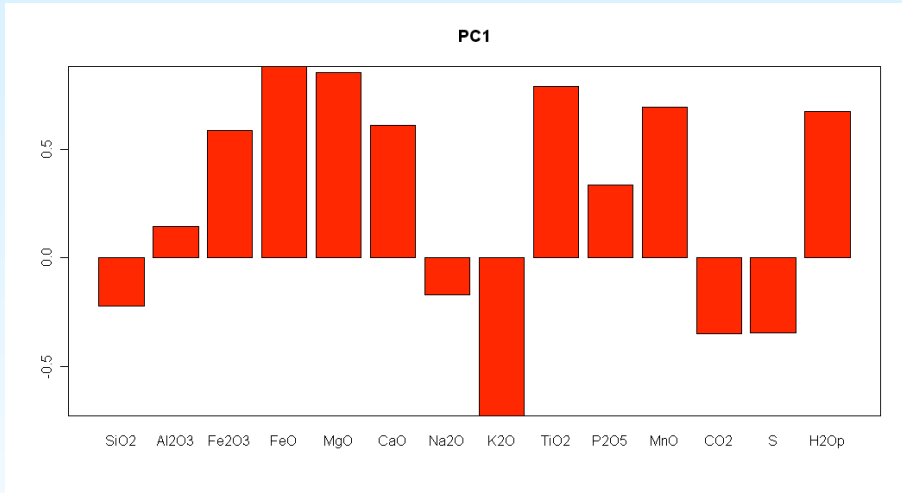
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Components Analysis



Natural Resources
Canada

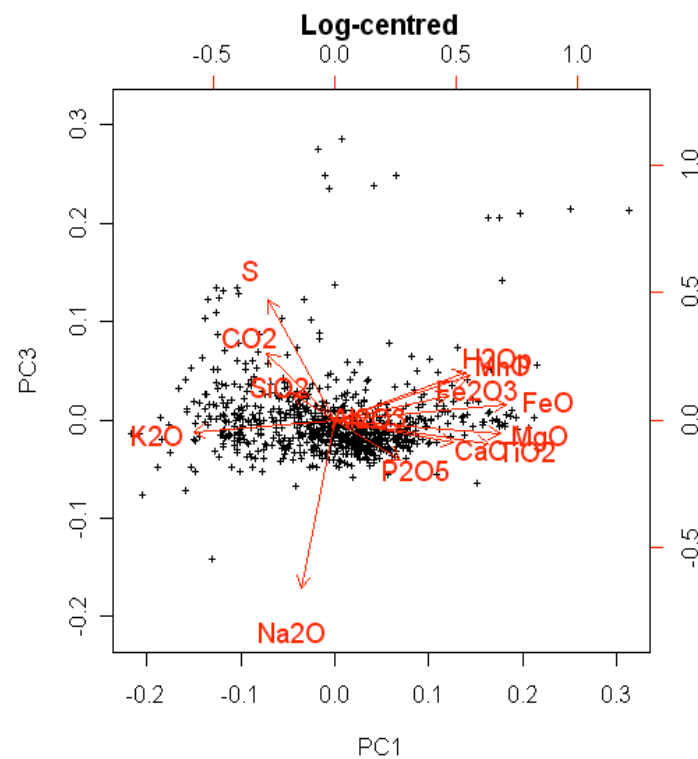
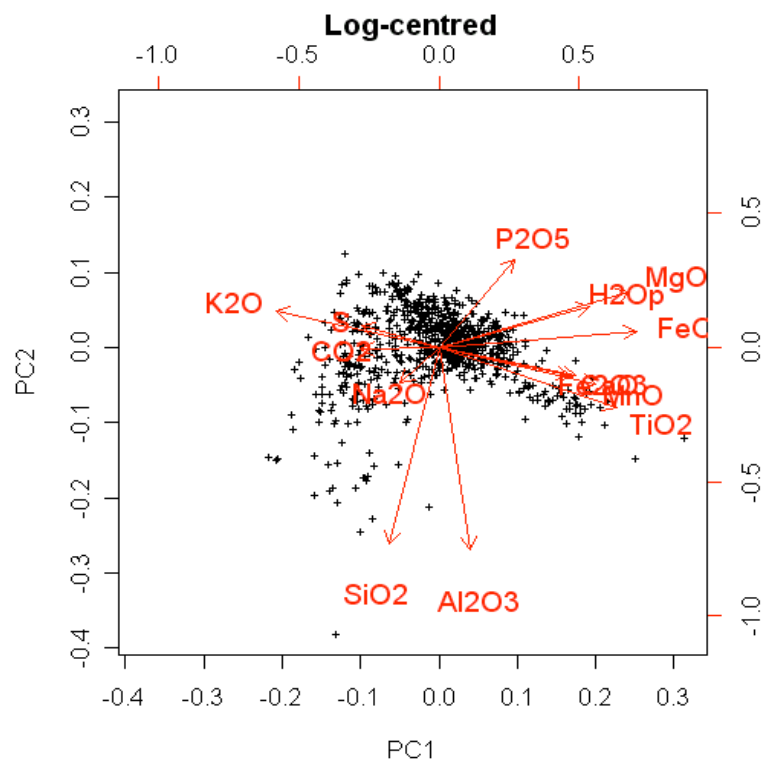
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Components Analysis



Natural Resources
Canada

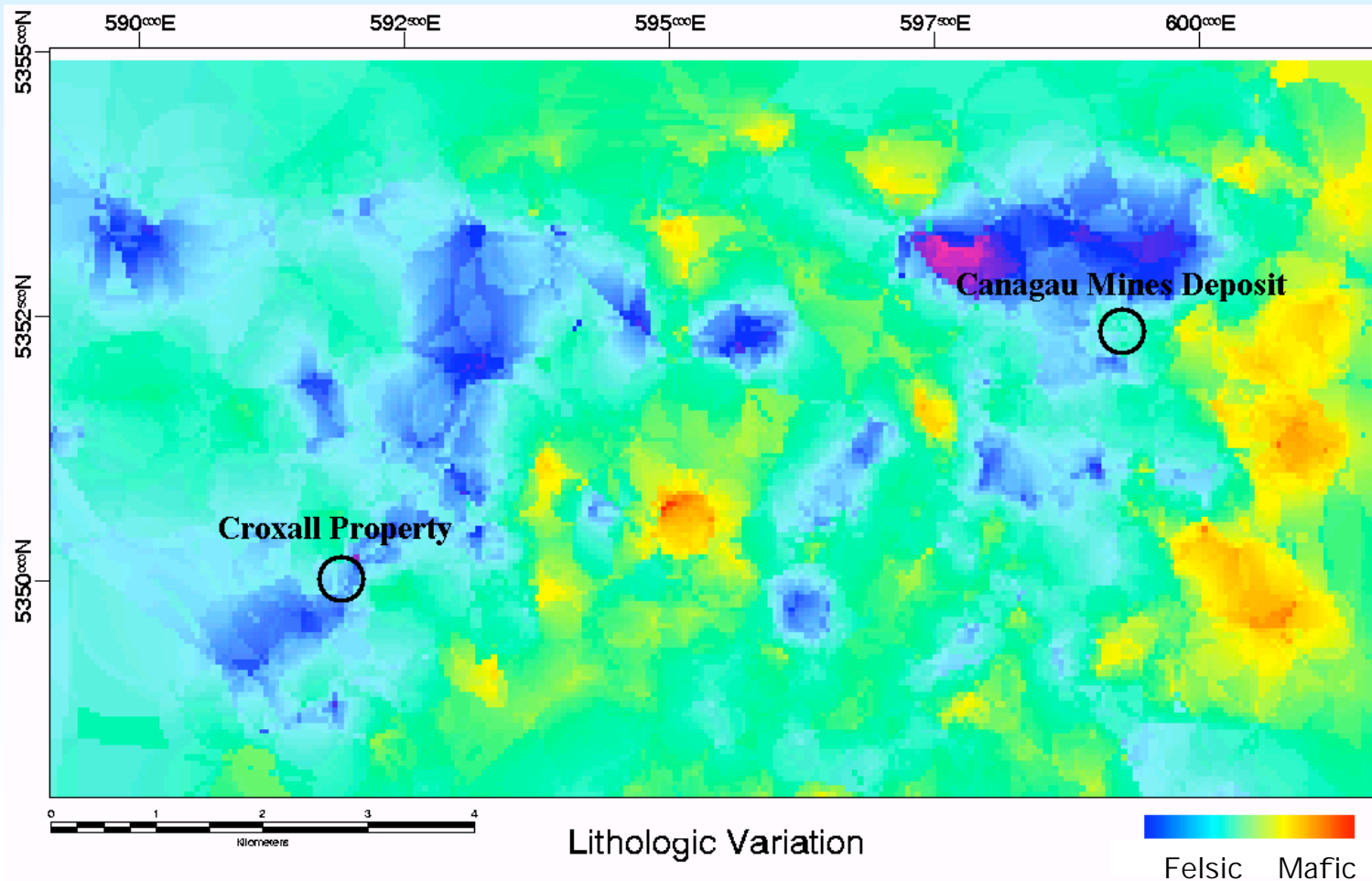
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Component 1 Log-centred Data



Natural Resources
Canada

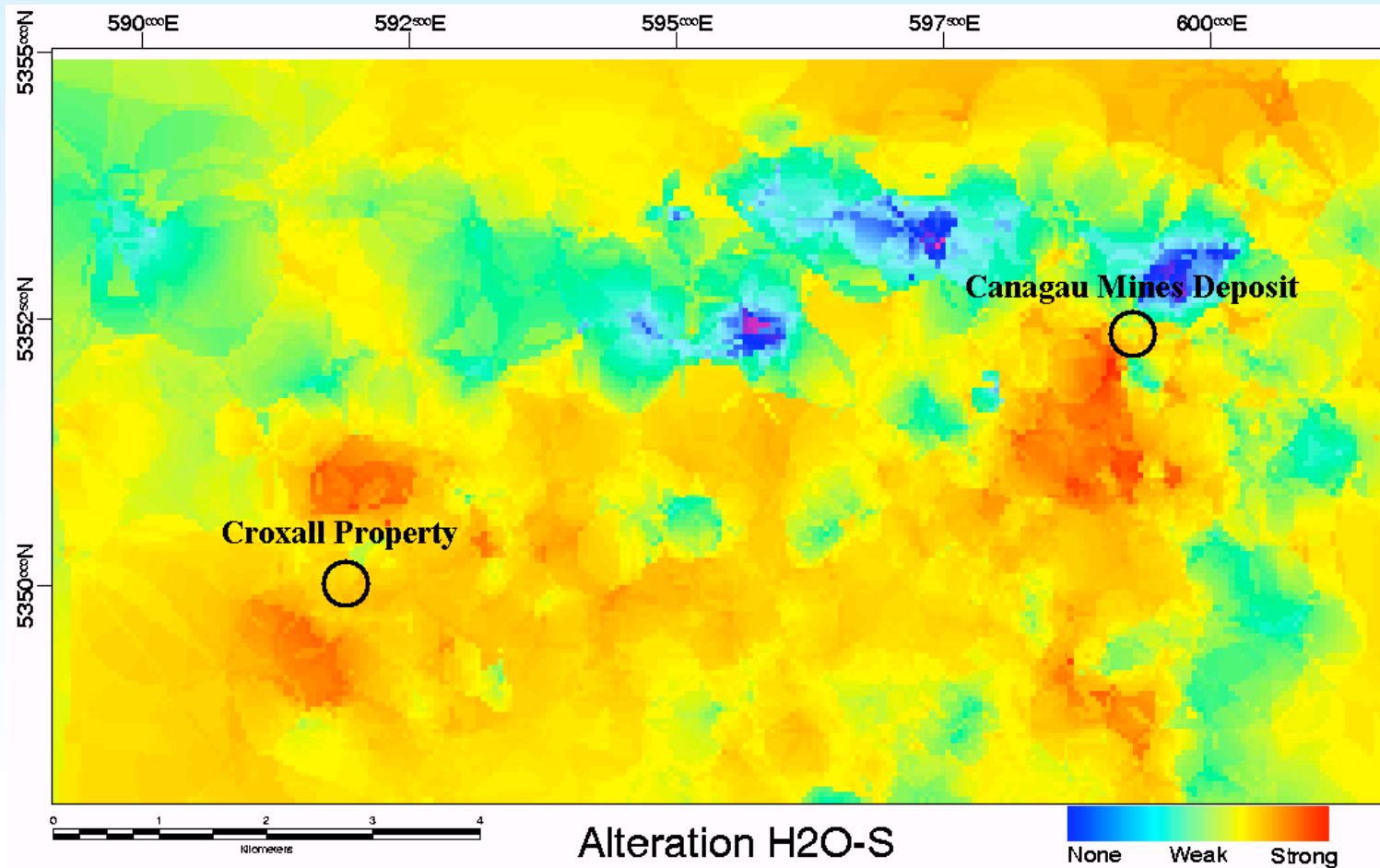
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Component 2 Log-centred Data



Natural Resources
Canada

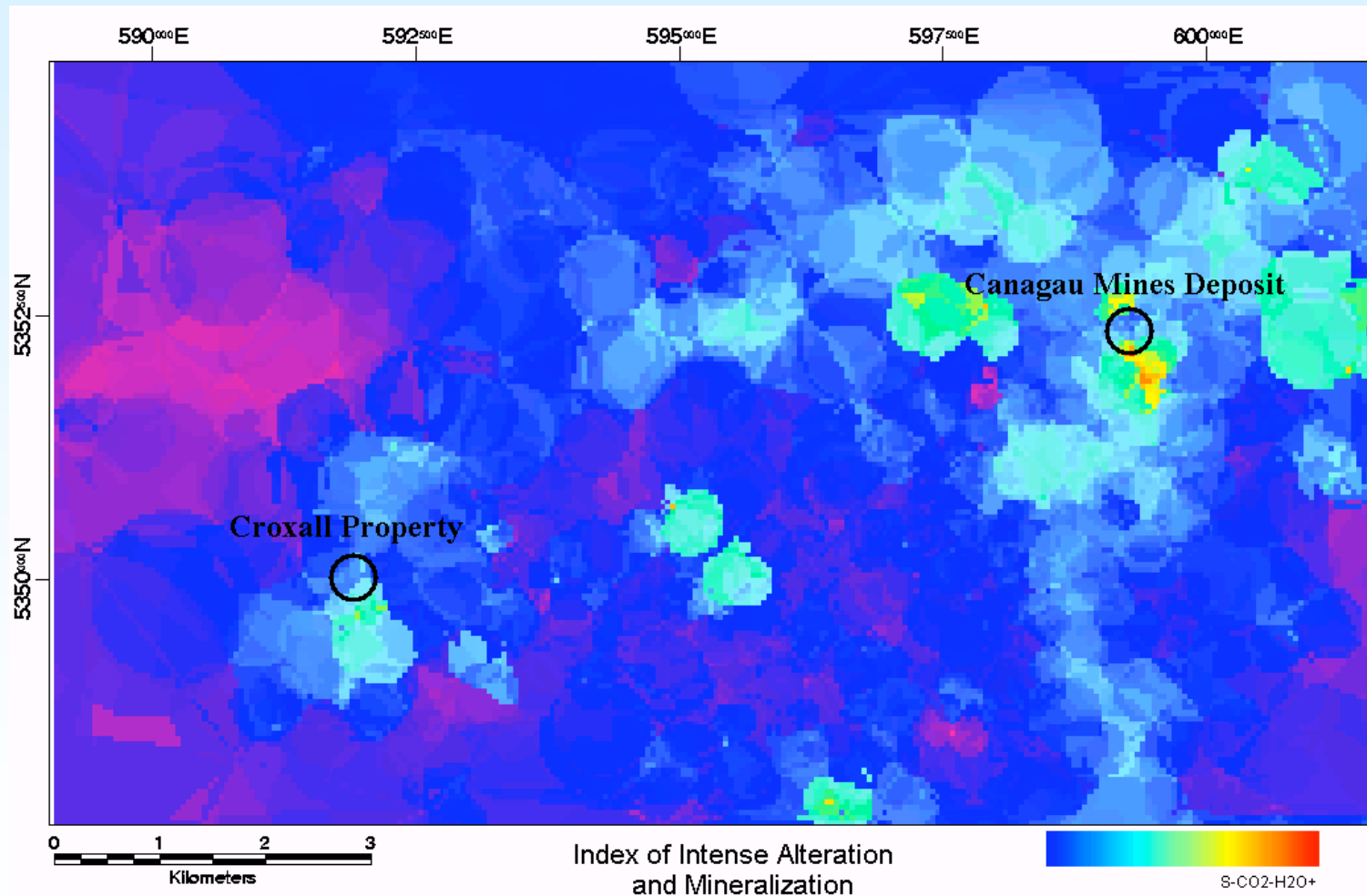
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Component 3 Log-centred Data



Natural Resources
Canada

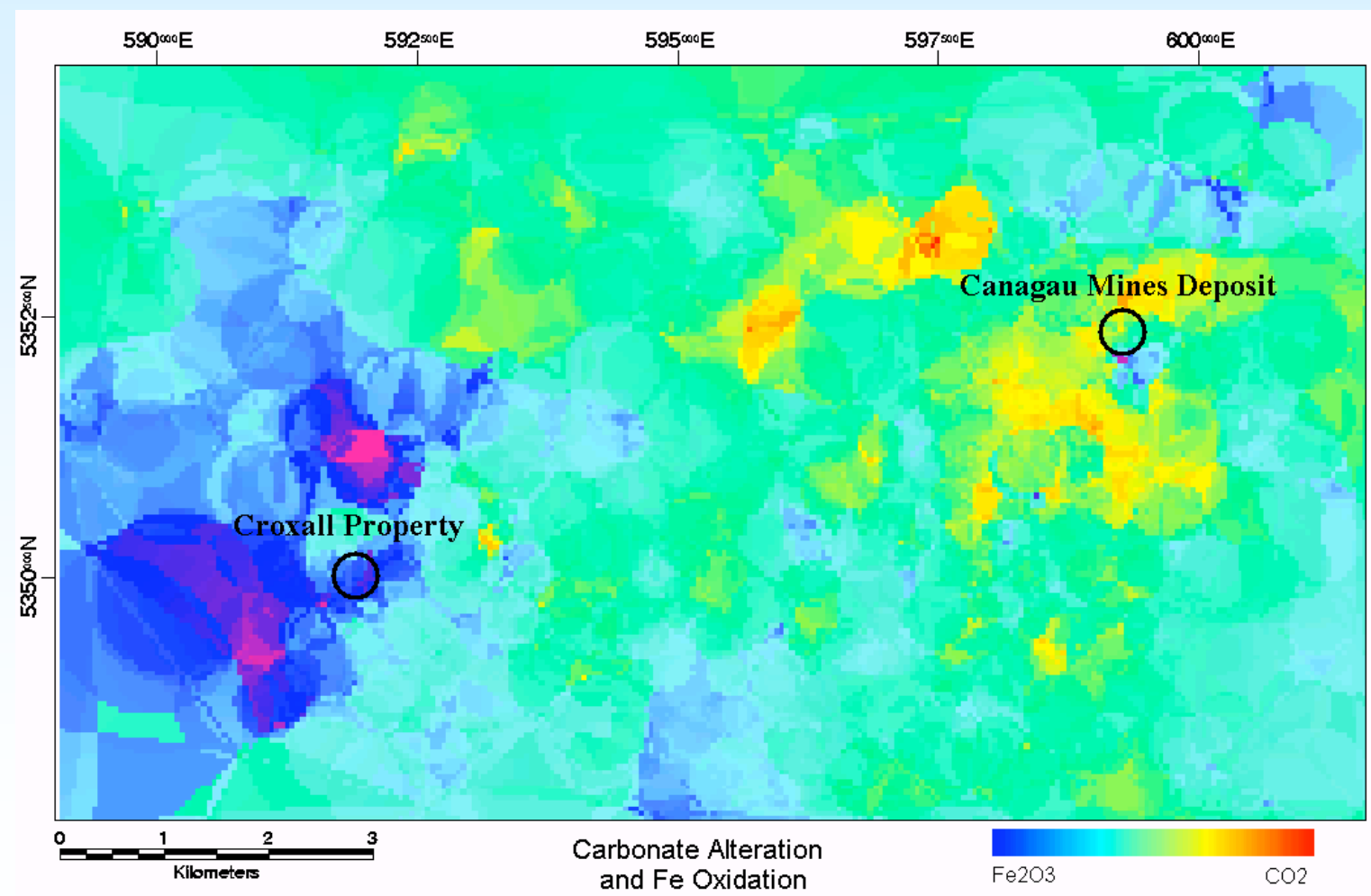
Ressources naturelles
Canada



Exploration07



Ben Nevis Metavolcanics Principal Component 4 Log-centred Data





Soil Sediment Geochemistry in a Tropical Environment



Natural Resources
Canada

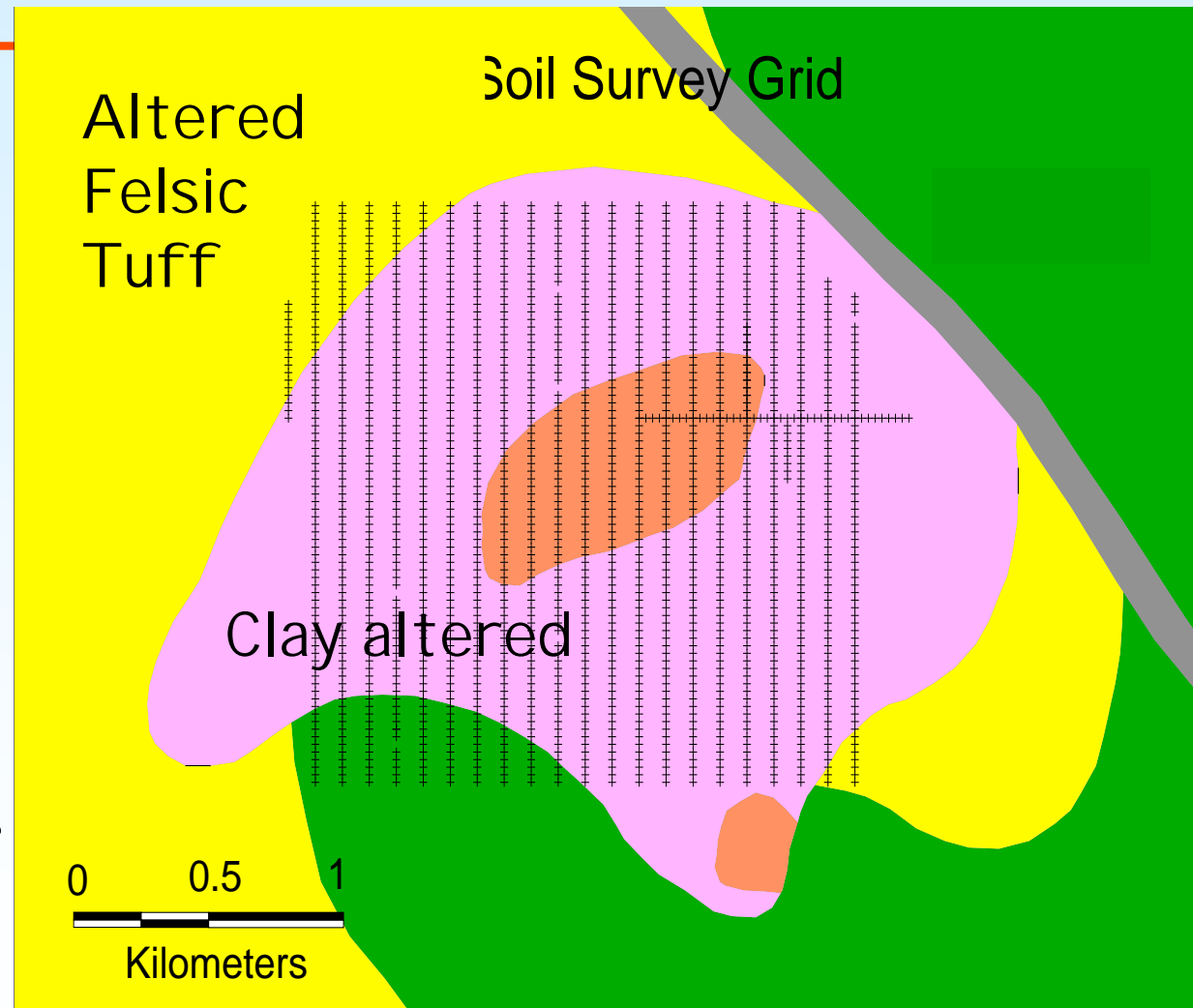
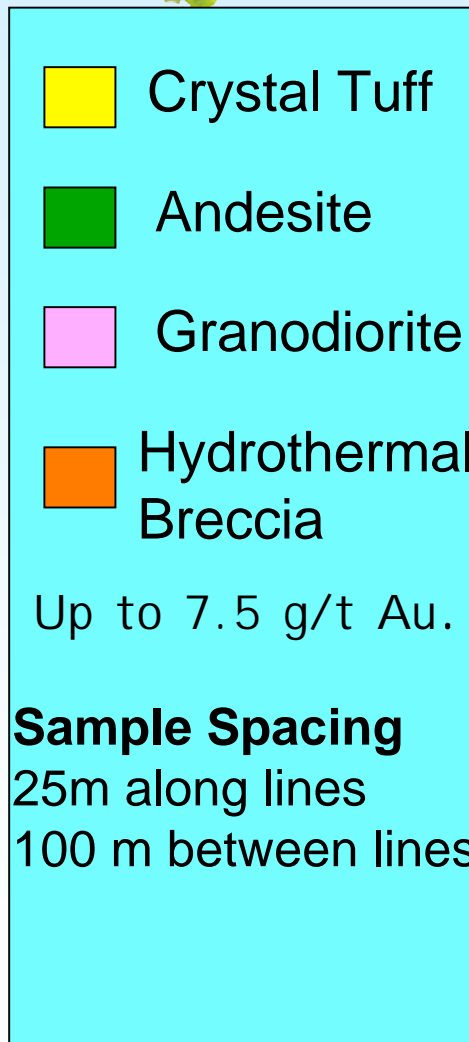
Ressources naturelles
Canada



Exploration07

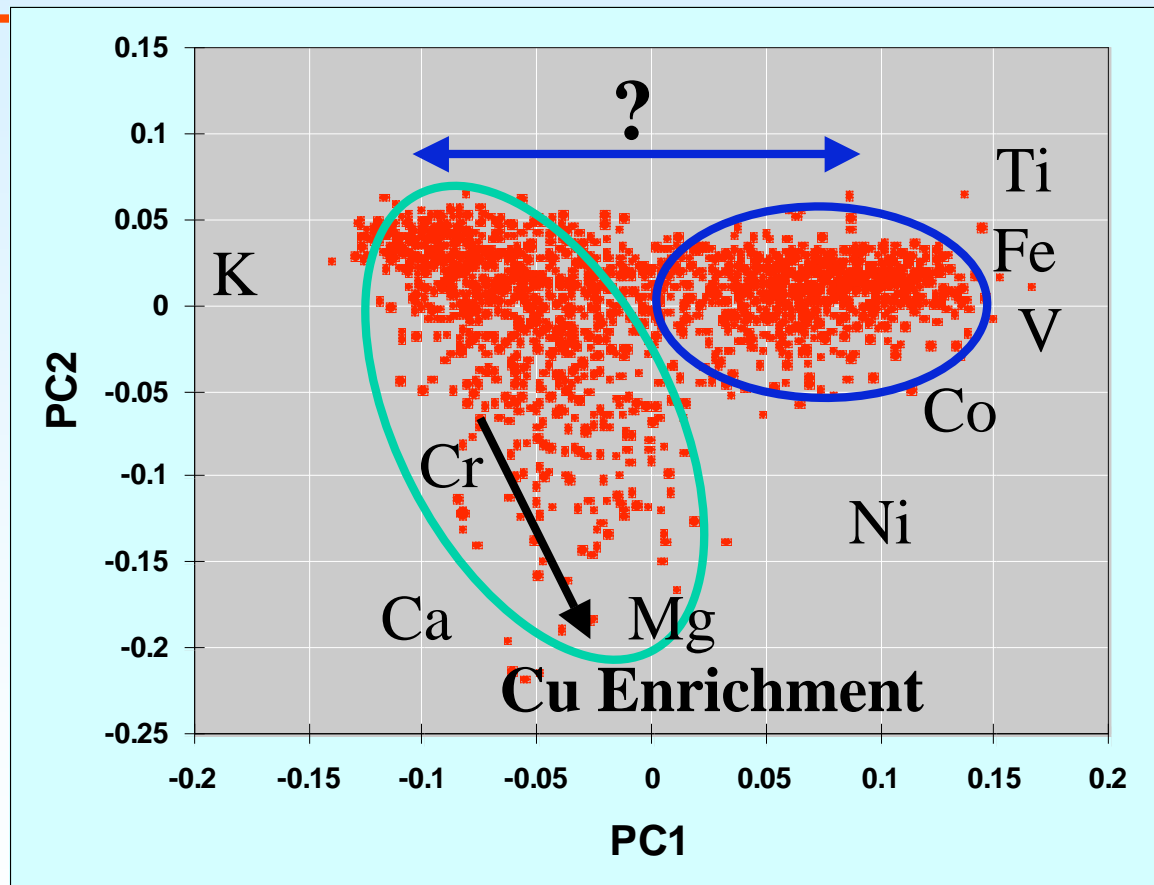
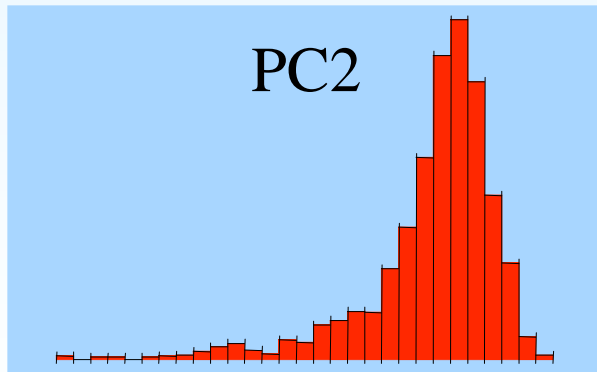
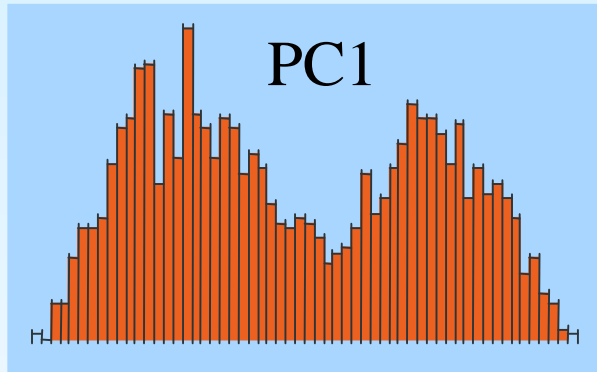


Geology of the Area





Soil Geochemistry Principal Components Analysis



Natural Resources
Canada

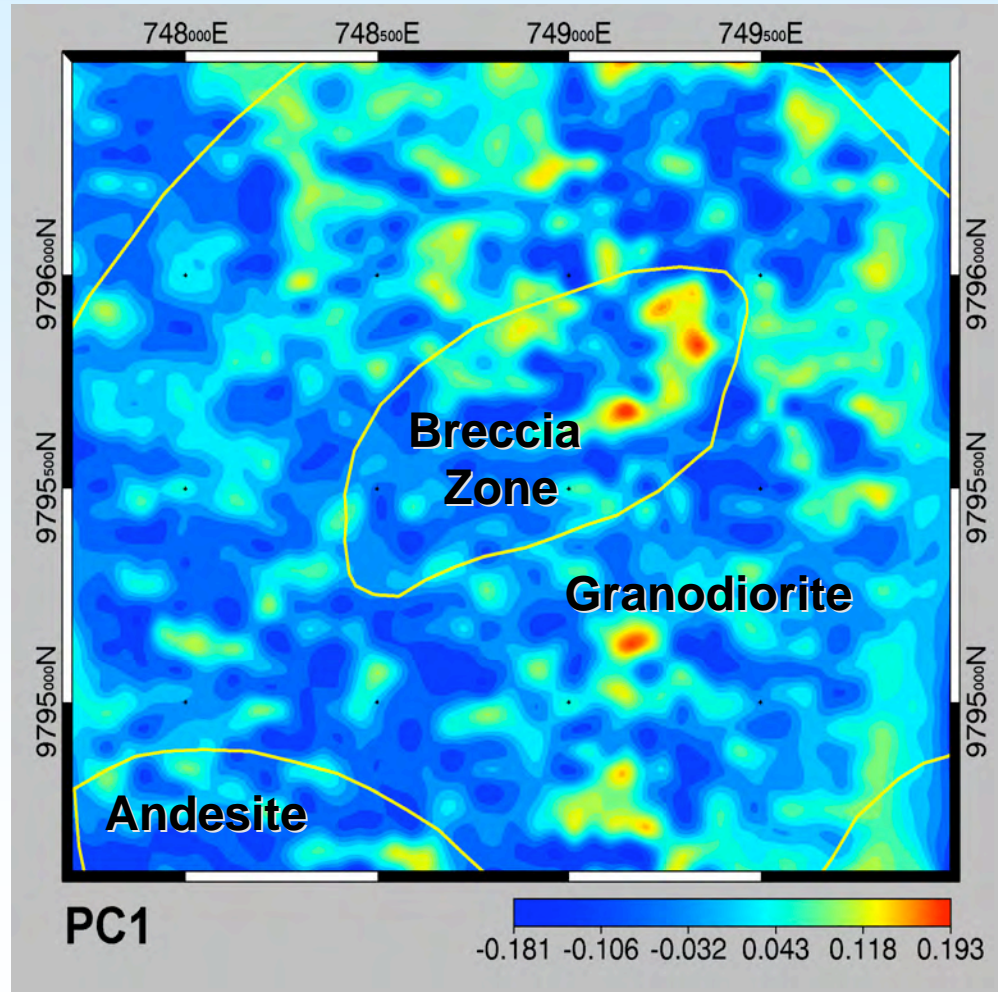
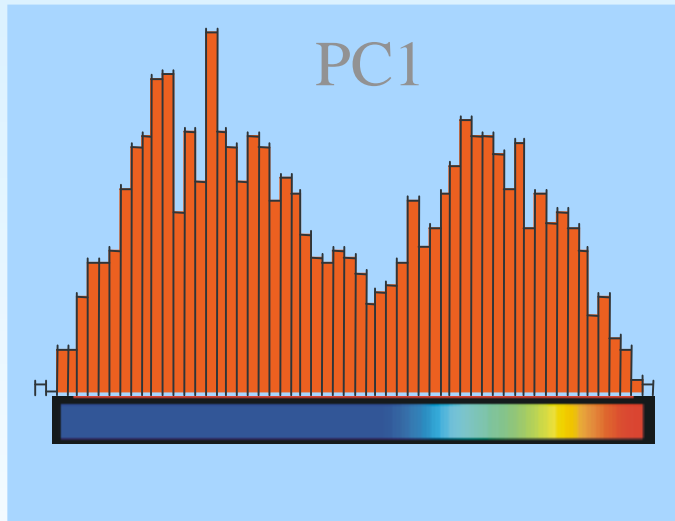
Ressources naturelles
Canada



Exploration07



PC1



Natural Resources
Canada

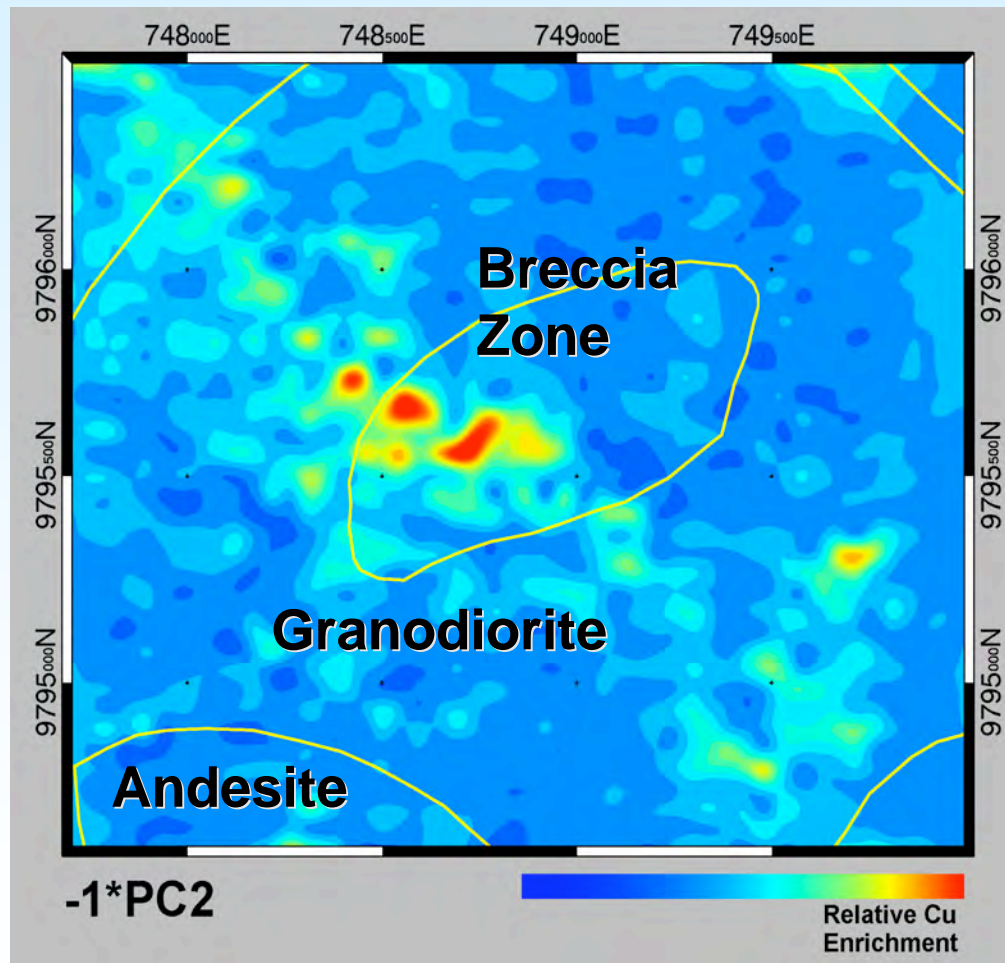
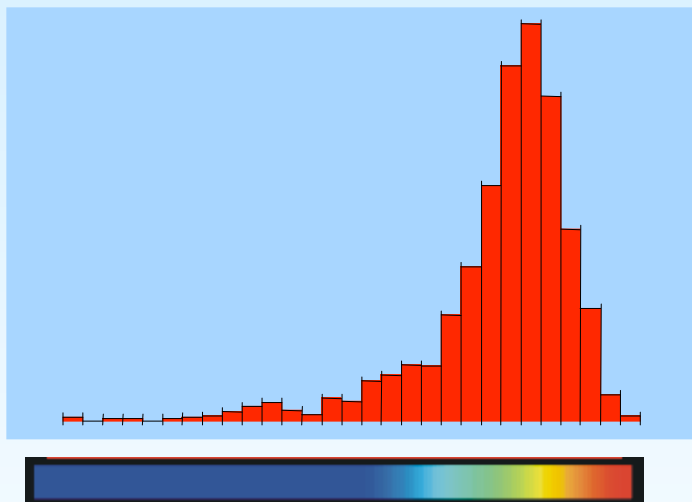
Ressources naturelles
Canada



Exploration07



PC2 - Cu Mineralization



Natural Resources
Canada

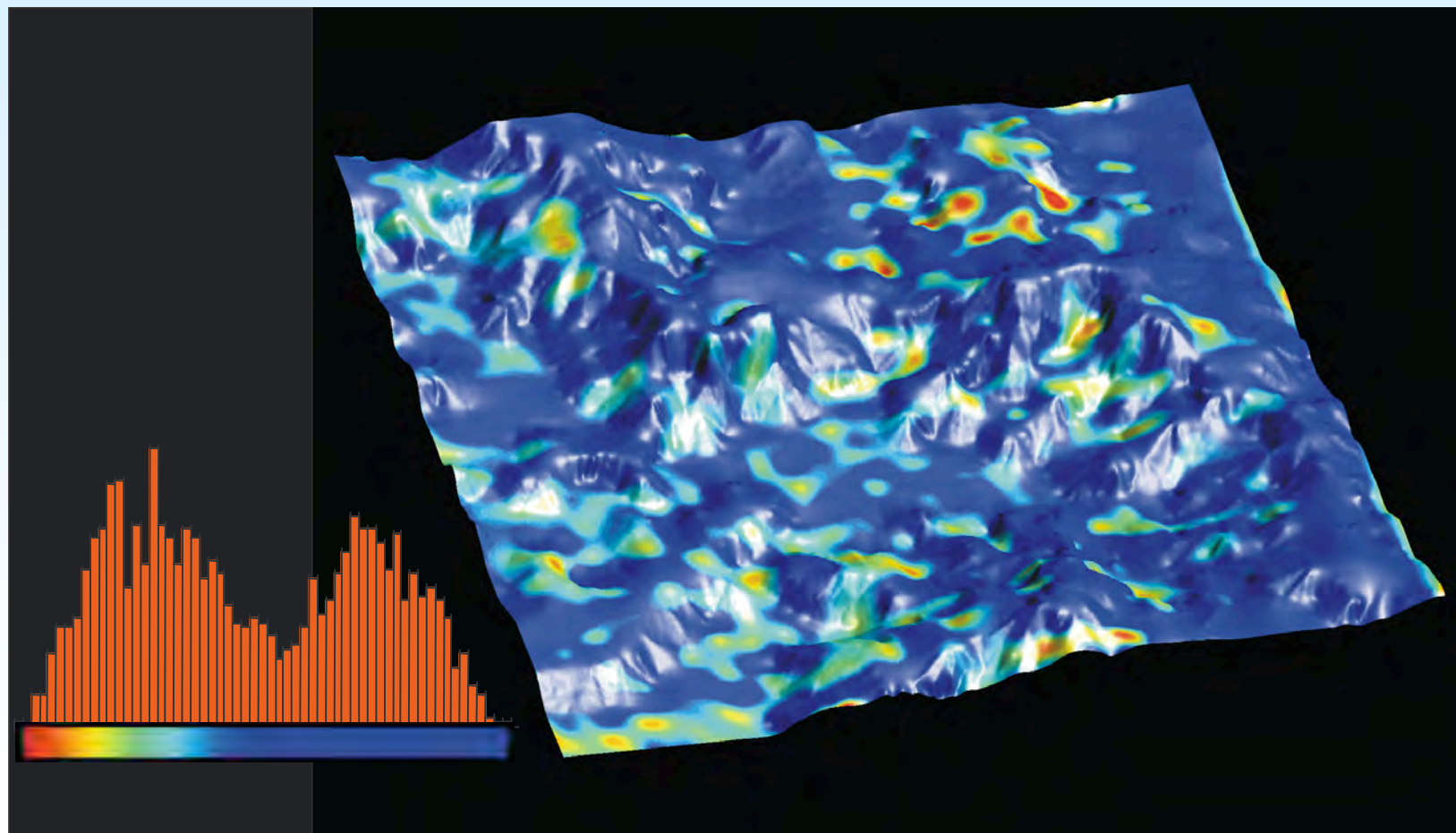
Ressources naturelles
Canada



Exploration07



Soil Geochemistry PC1 over Topography



Natural Resources
Canada

Ressources naturelles
Canada

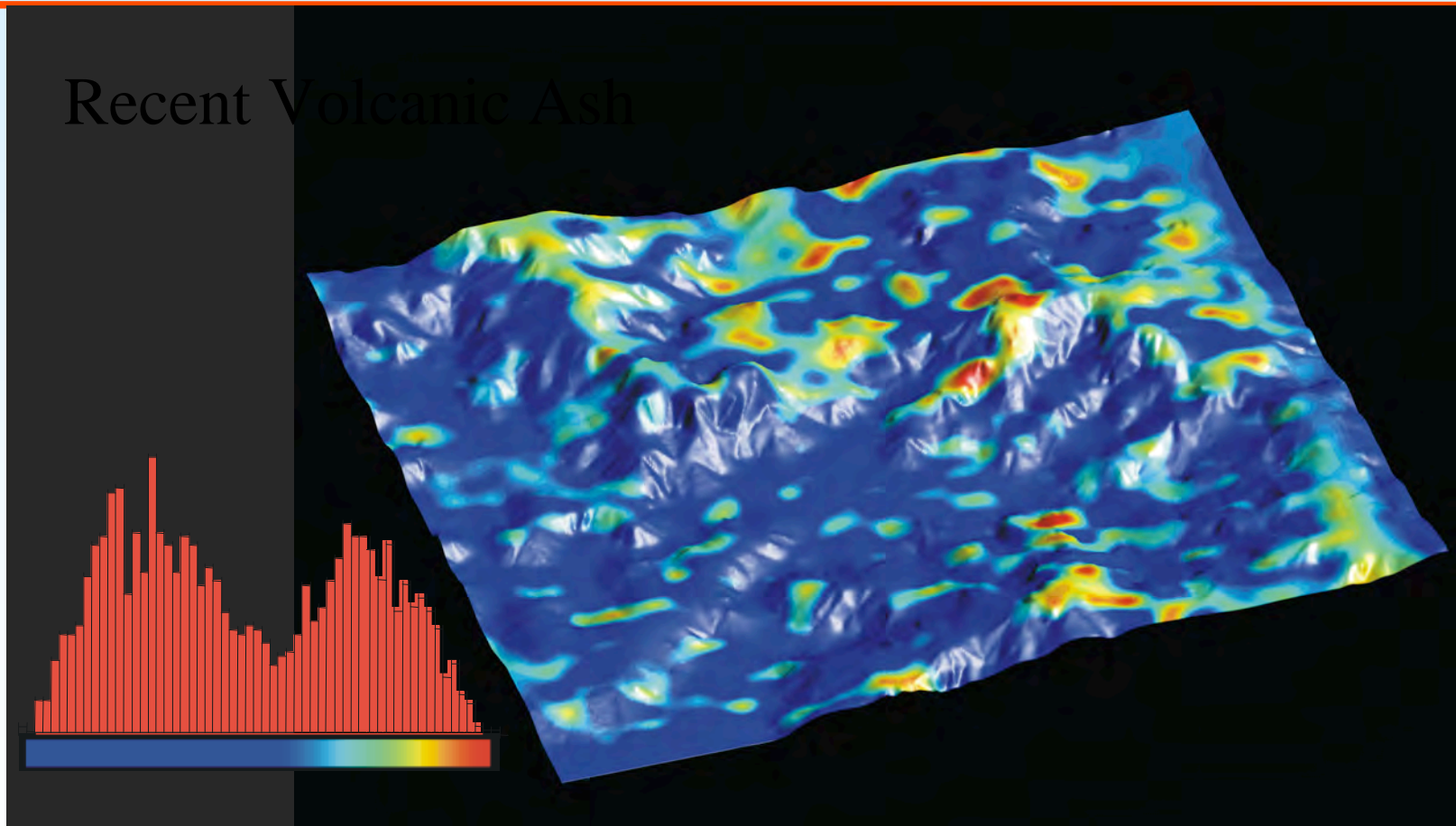


Exploration07



Soil Geochemistry PC1 over Topography

Recent Volcanic Ash



Natural Resources
Canada

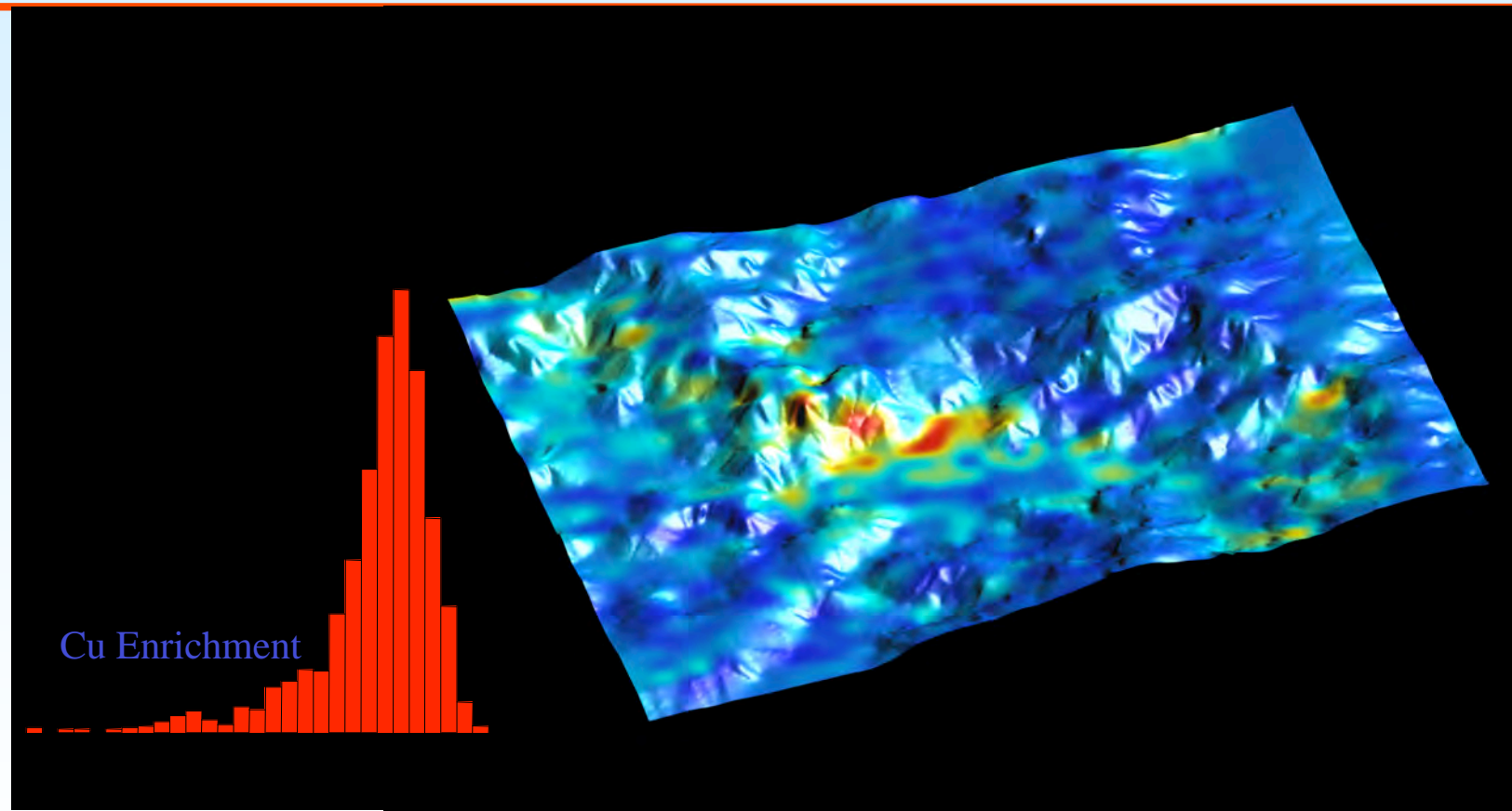
Ressources naturelles
Canada



Exploration07



Soil Geochemistry PC2 - Cu Signature



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07



Software for Geochemical Data Analysis

- Desktop statistical and mapping packages are common today.
- Many public domain and commercial packages available for geochemical data analysis.
- See Exploration 2007 proceedings.





Software for Evaluating Geochemical Data

- Evaluation of data in the spatial and variable domains are currently separate.
- GIS and Geostatistics packages are best for the spatial evaluation.
- Statistical packages best for the variable domain.





Software ...(cont'd)

- The R Project for Statistical Computing (www.r-project.org) provides a comprehensive environment for evaluating data.
- Specific packages include:
 - rgr (The GSC Applied Geochemistry EDA Package)
 - gstat (geostatistical modelling, prediction and simulation)
- Other packages (cluster, fastICA, MASS, e1071, kohonen, nnet, randomForest)





Some Comments

- To Assess/Interpret multi-element geochemistry:
 - Data analysis methods (exploratory),
 - Statistical methods (modelling),
 - Visualization in the variable and geospatial domains,
 - Geographic rendering in 2, 2.5 and 3 dimensions.





In Conclusion....

From Mess to Message

Get to know your data

There are a wealth of tools available to
investigate and evaluate.

Use them all !



Natural Resources
Canada

Ressources naturelles
Canada



Exploration07