Extracting, visualising and interpreting structure in geochemical data through compositional data analysis (CoDA) The Closure Problem

Eric Grunsky

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

egrunsky@gmail.com

2021-November-18

“Closure” – What is it?

- Geochemical analyses are typically reported as a “part” of a composition (weight %, ppm, ppb, g/t, mg/kg).
- all values are relative and sum to a constant (100%, 1000000 ppm,).
- If one value changes, then, by definition, all other values must change to maintain the constant sum.
- Thus, the variables (oxides, elements) are not independent.

“Closure” – Implications for Statistical Methods

- Statistical methods assume that the variables are independent. Since geochemical data variables are not independent, standard statistical methods are not valid.
- Statistical methods are based on values ranging from -∞ to +∞ whereas compositional data are constrained from 0 to a constant value [the “simplex”].

Effects of Closure on Values & Ratios

Element/Oxide		Normalized	Normalized
SiO2 [Wt%]	54.11	54.45	46.67
Al2O3 [Wt%]	12.36	12.44	10.66
Fe2O3 [Wt%]	4.43	4.46	12.96
FeO [Wt%]	9.53	9.59	8.22
MgO [Wt%]	3.81	3.83	3.29
CaO [Wt%]	3.25	3.27	2.80
Na2O [Wt%]	3.73	3.75	3.22
K2O [Wt%]	0.18	0.18	0.16
TiO2 [Wt%]	1.52	1.53	1.31
P2O5 [Wt%]	0.34	0.34	0.29
MnO [Wt%]	0.19	0.19	0.16
CO2 [Wt%]	1.29	1.30	6.04
S [Wt%]	0.62	0.62	0.53
H2O+ [Wt%]	3.74	3.76	3.45
H2O- [Wt%]	0.27	0.27	0.23
	99.37	100.00	100.00
Alkali Alteration (Na2O+K2O)/Al2O3	0.3163	0.3163	0.3163

Ratios don't change!

Adding CO2
To the composition changes the relative values but not the ratios.

Correlation Coefficients
Subcompositional Incoherence

Correlation Coefficients Based on 6 Elements - closed

	SiO2	TiO2	Al2O3	FeO	MgO	CaO
SiO2	1.00	-0.66	-0.68	**-0.23**	0.64	**-0.22**
TiO2		1.00	0.44	0.09	-0.44	0.12
Al2O3			1.00	-0.40	-0.21	0.50
FeO				1.00	**-0.55**	-0.73
MgO					1.00	**-0.11**
CaO						1.00

Correlation Coefficients Based on 4 Elements - closed

	SiO2	FeO	MgO	CaO
SiO2	1.00	**-0.64**	0.66	**0.04**
FeO			**-0.63**	-0.70
MgO			1.00	**-0.09**
CaO				1.00

Same data but different correlation coefficients

Compositional Data – Logratios

Additive Logratio (alr) [Aitchison (1983)]

$y_i = \log(x_i/x_D)$ $(i = 1, \ldots, D\text{-}1)$

where x_D = a compositional component of choice

Centred Logratio (clr) [Aitchison (1983)]

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

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

Isometric Logratio (ilr) [Egozcue et al. (2003)]

Combinations of elements that represent "balances" that result in an orthonormal space.

$$y_i = \sqrt{\frac{i}{i+1}} \ln \frac{g(x_1, \ldots, x_i)}{(x_{i+1})} \quad (i = 1, \ldots, D\text{-}1)$$

Combinations of elements that represent "balances" that result in an orthonormal space. i.e.

Logratio analysis is based on research by John Aitchison, Vera Pawlowsky, Juan Jose Egozcue and the University of Girona, Spain.

What About Trace Elements in Compositional Data?

- Because trace elements are << than 5% of a total composition, many assume that the effect of closure is negligible.
- The assumption must be tested before it can be accepted.
- If the assumption is tested using log-ratio analysis, then why not stay in the log-ratio space?

Historical use of ratios

- Historically rock classifications were petrographically (mineralogically) based.
- Stoichiometry is an essential part of assessing mineralogy.
- Ratios were introduced to look at the relative values of elements in minerals, thereby reducing the influence of elements not involved in the development of the mineral.

Cross, Iddings, Pirsson & Washington, 1903, Quantitative Classification of Igneous Rocks, Univ. of Chicago Press, 286p.

Pearce Element Ratios [PER]

- 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 $i = 1, \ldots, D\text{-}1$)
- The analysis of y_i ($i = 1, \ldots, D\text{-}1$) will define compositional variation related to stoichiometric processes.

Pearce, T.H. A contribution to the theory of variation diagrams. Contr. Mineral. and Petrol. 19, 142–157 (1968). https://doi.org/10.1007/BF00635485

Variation Array = Correlation/Covariance

T defined by:

Variability of the logratio

- $t_{ij} = \mathrm{var}\{\log(x_i/x_j)\}$ (i=1,…,d; j=i+1,…,D)

and the mean, E, is expressed as:

Mean of the logratio

- $x_{ij} = E\{\log(x_i/x_j)\}$ (i=1,…,d; j=i+1,…,D)

Variation Array

Variance logratios

	SiO2	Al2O3	Fe2O3	FeO	MgO	CaO	Na2O	K2O	TiO2	P2O5	MnO	CO2	S	H2Op
SiO2	0.00	0.05	0.45	0.41	0.76	0.64	1.00	2.22	0.31	0.49	0.32	1.88	1.83	0.47
Al2O3	1.33	0.00	0.30	0.24	0.51	0.43	0.96	2.43	0.15	0.36	0.20	1.99	1.91	0.30
Fe2O3	3.63	2.30	0.00	0.26	0.41	0.51	1.42	3.21	0.23	0.47	0.30	2.74	2.13	0.30
FeO	2.62	1.30	-1.01	0.00	0.15	0.37	1.27	3.20	0.11	0.31	0.11	2.22	2.16	0.13
MgO	2.86	1.54	-0.77	0.24	0.00	0.46	1.52	3.81	0.25	0.43	0.30	2.76	2.55	0.21
CaO	2.50	1.17	-1.13	-0.13	-0.37	0.00	1.46	3.68	0.31	0.57	0.33	2.37	2.74	0.47
Na2O	3.09	1.77	-0.53	0.47	0.23	0.60	0.00	3.54	1.11	1.31	1.27	3.03	3.46	1.52
K2O	5.00	3.67	1.37	2.37	2.13	2.50	1.90	0.00	3.13	3.20	3.01	3.98	3.83	3.21
TiO2	4.33	3.01	0.70	1.71	1.47	1.84	1.24	-0.66	0.00	0.30	0.15	2.24	2.15	0.21
P2O5	6.35	5.02	2.72	3.73	3.49	3.85	3.26	1.35	2.02	0.00	0.37	2.42	2.22	0.40
MnO	6.43	5.10	2.80	3.80	3.56	3.93	3.33	1.43	2.09	0.08	0.00	1.95	2.16	0.23
CO2	4.62	3.29	0.99	2.00	1.75	2.12	1.52	-0.38	0.29	-1.73	-1.81	0.00	3.39	2.26
S	7.76	6.44	4.13	5.14	4.90	5.26	4.67	2.76	3.43	1.41	1.34	3.14	0.00	2.22
H2Op	3.21	1.88	-0.42	0.59	0.34	0.71	0.11	-1.79	-1.13	-3.14	-3.22	-1.41	-4.55	0.00

Mean logratios

Consequences of Closure

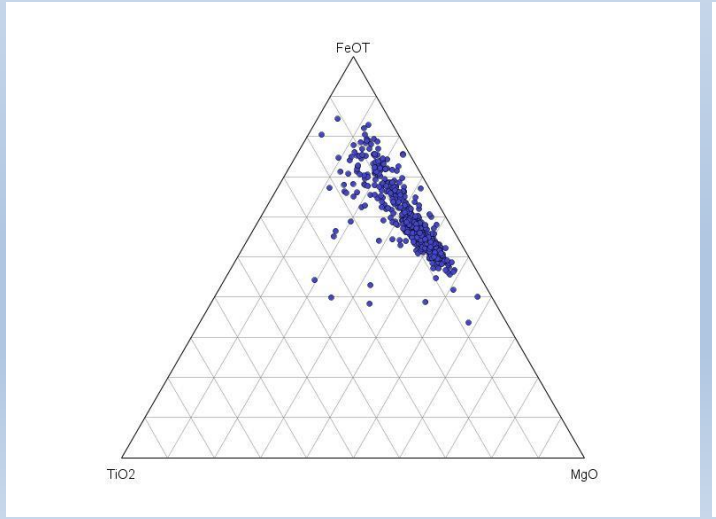
- Closing data results in the possibility of spurious correlations when comparing elements in chemical compositions. - Karl Pearson, 1897!

- With the exception of individual mineral analyses, once a composition is expressed as in constant sum form (%, ppm), it is not possible to recover the original molecular contributions without end-member modelling and many assumptions.

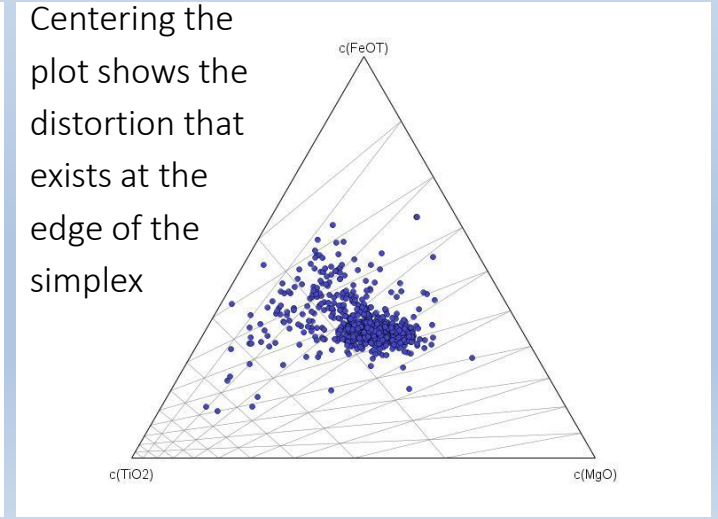
Distortion in the Simplex

- Compositional data, expressed as a part of a sum are >0 and constrained to sum to a constant. It is a “closed” system.
- This constraint is mathematically and geometrically defined as a “simplex”.

Ti-Fe-Mg Ternary Plot



Centering the plot shows the distortion that exists at the edge of the simplex



Natural Laws

- Geochemistry is a proxy for mineralogy / species in solution.
- “Natural laws” govern the relationships which are based on stoichiometry.
- The associations of elements are not random, but governed by atomic forces that result in ordered crystalline structures.
- How can we reconcile stoichiometry in a compositional framework?

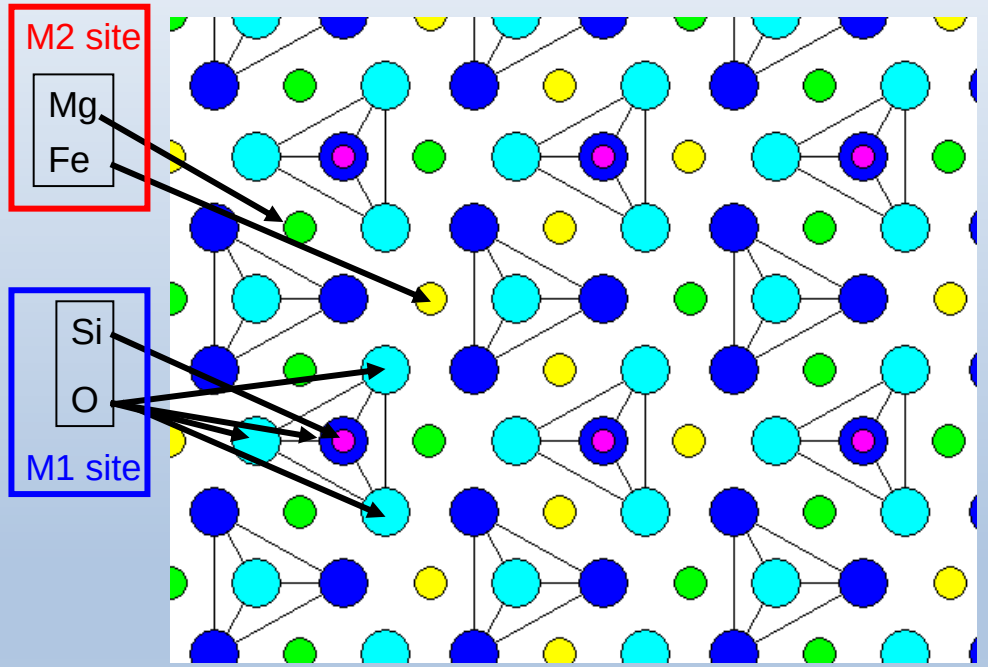
Olivine $(Mg,Fe)_2$ $[SiO_4]$ Crystal Structure

Blue/Cyan – Oxygen
Green/Yellow – Mg/Fe
Magenta – Si

Silica with 4 oxygen form SiO_4 tetrahedra with a charge of $^{-2}$, and binds with Mg-Fe-Mn cations with charges of $^{+2}$

Ti has the same ionic size/charge as Si and can also substitute.

Crystal defects can allow (e.g. Al^{+3}) other similar-sized cations to enter the structure.



Source: http://www.uwgb.edu/DutchS/petrolgy/Olivine-Structure.HTM

[Grunsky et al., 2008, CoDaWork08]

Creating a Structural Formula Based on Stoichiometry

Olivine = $(Mg,Fe)_2[SiO_4]$

For each $[SiO_4]$ tetrahedron, there are 2 cations of (Mg-Fe)

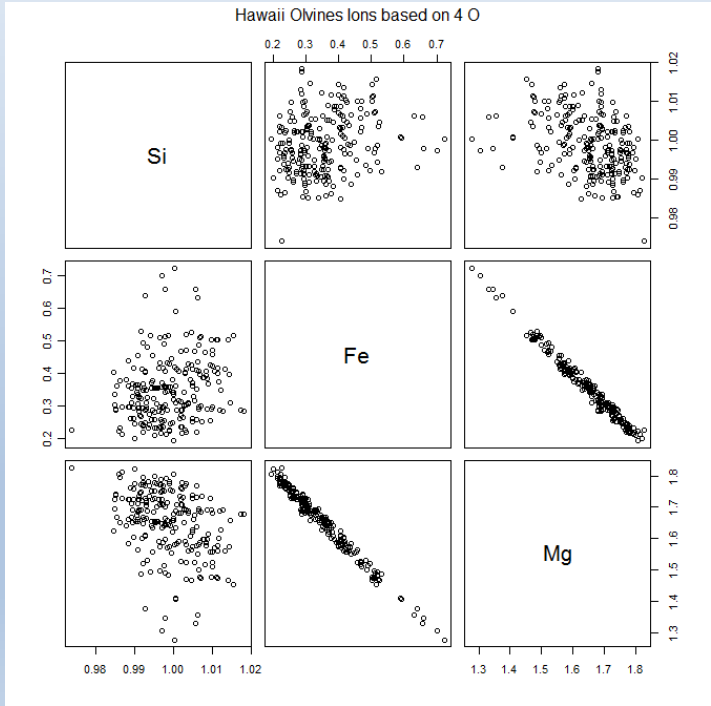
	B	C	D	E	F	G	H	
	wt% oxides	mol. weight	mol. prop. of oxides	No. Oxygen / cation	atomic proportions of oxygen from each molecule	No. anions based on 4 oxygen	No. of ions in formula	
SiO_2	41.07	60.08	0.68	2	1.37	1.96	0.98	
TiO_2	0.05	79.90	0.00	2	0.00	0.00	0.00	0.99
Al_2O3	0.56	101.96	0.01	3	0.02	0.02	0.01	
Fe_2O_3	0.65	159.69	0.00	3	0.01	0.02	0.01	
FeO	3.78	71.85	0.05	1	0.05	0.08	0.08	2.00
MnO	0.23	70.94	0.00	1	0.00	0.00	0.00	
MgO	54.06	40.31	1.34	1	1.34	1.92	1.92	
CaO	0	56.08	0.00	1	0.00	0.00	0.00	
Na_2O	0	61.98	0.00	1	0.00	0.00	0.00	
H_2O^+	0.05	18.02	0.00	1	0.00	0.00	0.00	
Total	100.45		B/C		D*E	F*T	F/E	
Total	Total of atomic proporations			2.796697				
T	4 ox/total of atomic proportions			1.430259				

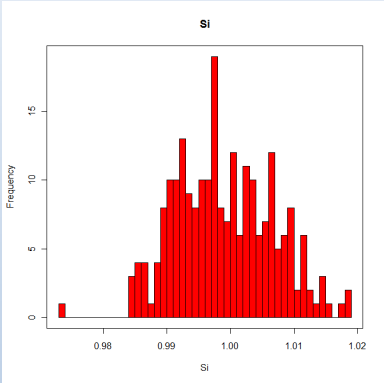
M1 site

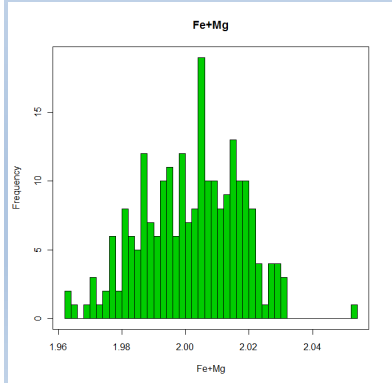
M2 site

Hawaii Olivines

$(Mg,Fe)_2 [SiO_4]$

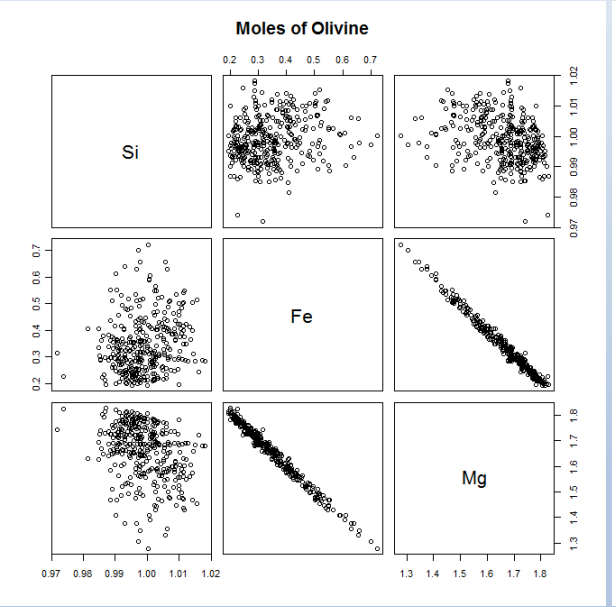


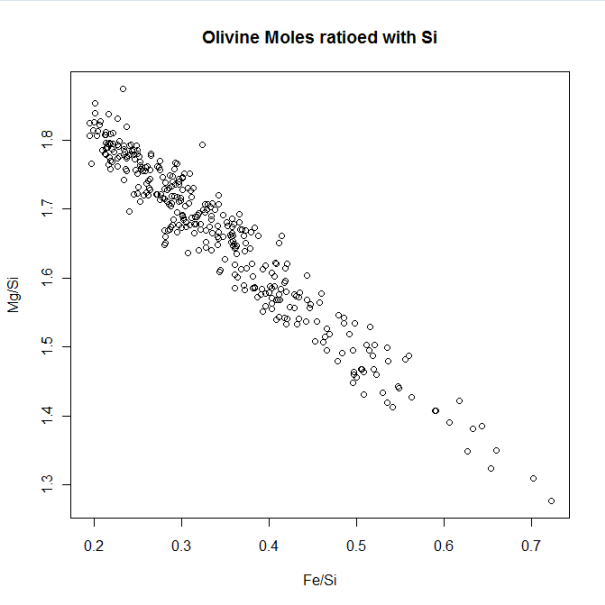




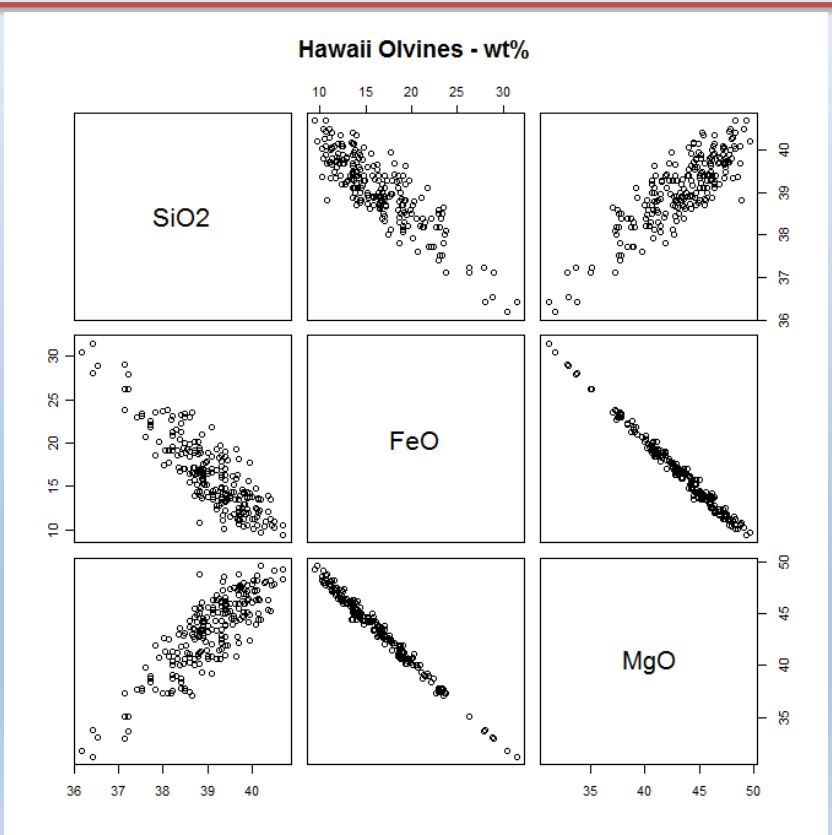
Olivine Moles & Molar Ratios

Is the correlation spurious or stoichiometrically contstrained?





Olivine Moles & Molar Ratios



Negative bias Si vs. Fe introduced with closure

Exploratory and Modelled Approaches

- Mineral stoichiometry controls the relative abundance of the elements.
- Geochemical data are **compositional and require the use of logratios** to overcome the problem of closure that interferes with the application of statistical methods (alr, clr, ilr transforms).
- **Process Discovery (Exploratory Approach) [Unsupervised]** - metrics [coordinates] used to discover geochemical processes reflected through elemental associations and geospatial patterns.
- **Process Validation/Prediction (Modelled Approach) [Supervised]** – methods for validation/prediction of processes through the use of classification methods.

The Challenges in Evaluating Geochemical Data

Different - • methods of digestion • limits of detection • instrumentation	→	Level the data where appropriate
Censoring samples < | > detection limit	→	Remove or impute elements
Missing values and zeros	→	Delete elements or compute replacement values depending on objectives.
Constant sum (closure) problem	→	Application of ratios and logratios
Adequate Spatial Sample Design	→	Properly designed spatial sampling scheme & Geostatistical evaluation

Is there a “**complete**” composition?

- Can a “complete” composition exist?
- Limitations on machine sensitivity prevent the realization of a complete composition.
- Is the value of a component reported as zero, really a zero or is it present but not detected?
- Therefore all compositions are sub-compositions.
- Sub-compositions are useful if it can be demonstrated that specific components define a unique process and the other variables only add noise or define processes that not of interest.
- Amalgamations (combining elements) can offer an advantage by reducing collinearity.

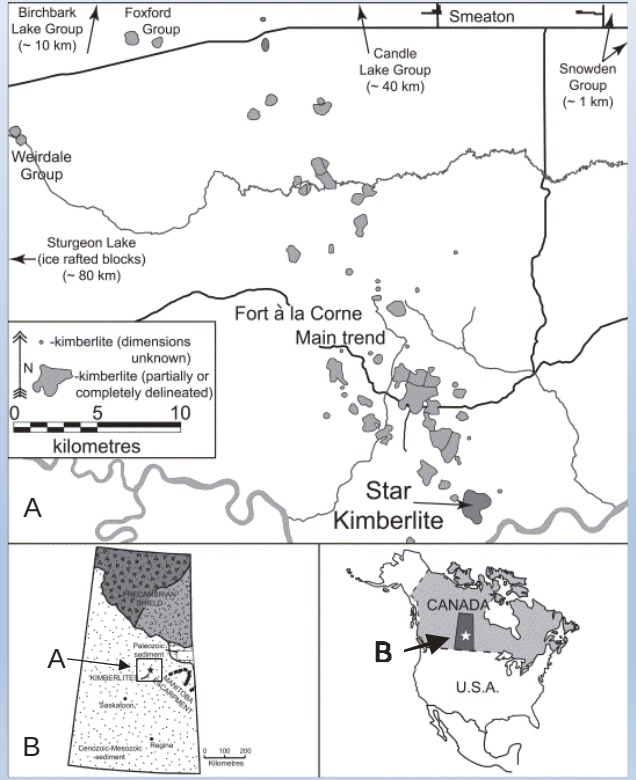
Kimberlite Classification using Lithogeochemistry Star Kimberlite Data

with Bruce Kjarsgaard
Geological Survey of Canada

Local/Camp Scale < 1:50,000

Exploration scale studies and detailed geologic mapping.

Star Kimberlite – Fort a la Corne - Saskatchewan

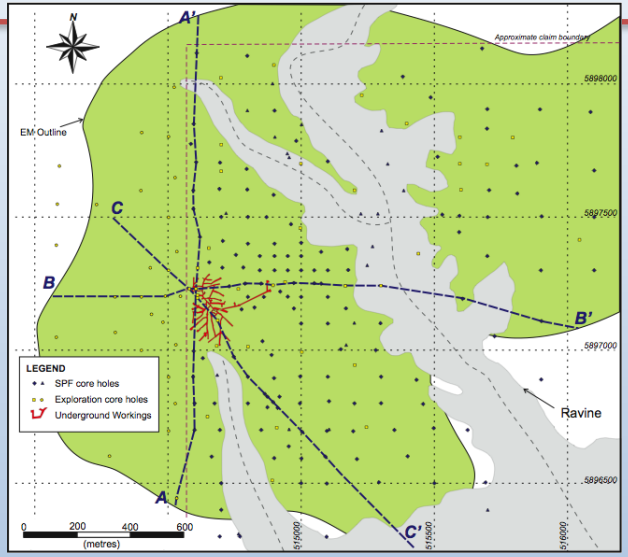


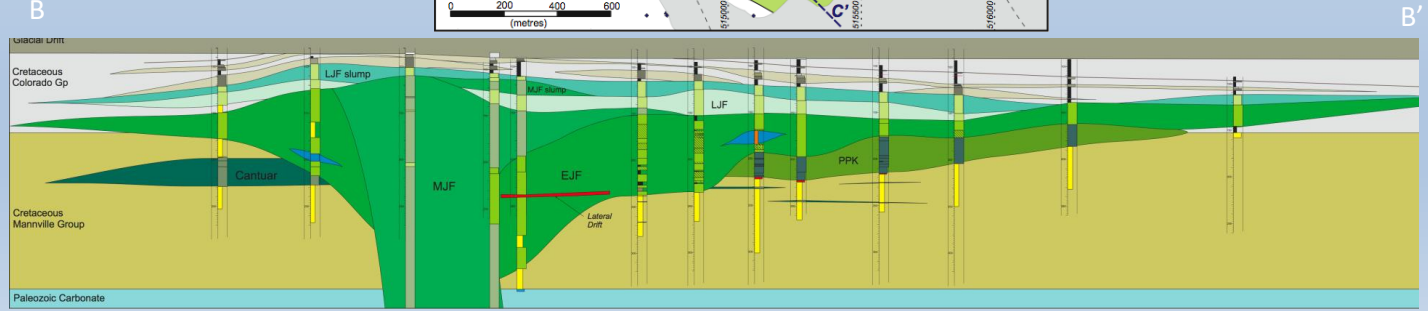
- Lithogeochemical sampling program of drill core from a series of kimberlite eruptions.
- Kimberlite mineralogy varies from olivine bearing magmas to fractionated magmas contaminated by crust.
- 283 Kimberlites analyzed the following oxides/elements converted to cation values :

Si, Ti, Al, Fe, Mg, Ca, Na, K, P, Rb, Nb, Zr, Th, V, Cr, Co, Ni, La, Er, Yb, Y, Ga

Grunsky, E.C., and Kjarsgaard, B.A., 2008. Classification of eruptive phases of the Star Kimberlite, Saskatchewan, Canada based on statistical treatment of whole-rock geochemical analyses; Applied Geochemistry. v. 23 (12), p. 3321-3336. (ESS Contribution # 20080330), 10.1016/j.apgeochem.2008.04.027.

Star Kimberlite- Cross Section





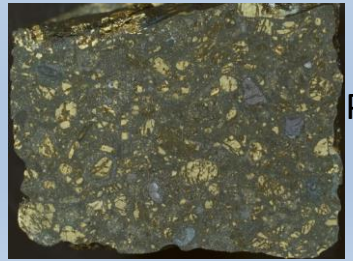
West - East cross section

Kimberlite Phases

Early Joli Fou

Mid Joli Fou

Late Joli Fou



Pense

Cantuar