

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

Process Prediction and Predictive Mapping

Eric Grunsky

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

egrunsky@gmail.com

2021-November-18

Evaluating Geochemical Data Using Multivariate Methods and Predictive Mapping

The Process of Predictive Mapping

Process Discovery

The use of empirical methods for identifying structure in data and forms the basis/justification to build or test models:

- Adjust data for censoring/missing values.
- Transform data to the logratio space (alr/clr/ilr).
- The use of different metrics (PCA, t-SNE, ICA, UMAP + others) to observe/discover patterns in the data.
- Discovery of processes through empirical analysis (principal component analysis, logratio analysis, multidimensional scaling, cluster analysis).
- Determine suitable classes for predictive mapping (e.g. lithologic units).
- Tag classes to sample sites where available using GIS.

The Process of Predictive Mapping

Process Validation

The use of modelled methods for process confirmation:

- Analysis of variance (AOV) to determine which elements or principal components give maximum separation of the classes.
- Methods of classification used for prediction (LDA, QDA, NN, LR, RF + others)
- Classification to determine measures of posterior probability or typicality from which a probability of class membership is assigned to each site. Other methods can be used (e.g. Random Forests).
- Spatial analysis to calculate semi-variograms and subsequent kriging (interpolation) to produce predictive maps for each class.

Calculate:

- accuracy of prediction for each class and overall accuracy.
- **precision** the fraction of true positives/[true + false positives].
- **recall** the fraction of true positives/false negatives.
- Estimates of accuracy are derived from Process Validation and based on the confusion matrix predicted classes vs. actual classes.

Kimberlite Classification using Lithogeochemistry

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.
- •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

Kimberlite Phases



Process Discovery Kimberlites – Logcentred PCA



Kimberlite Fractionation Trends [stoichiometric control]





ΣPC1-2 = 66% ΣPC1-4 = 80% ΣPC1-7 = 90%Overall variation Signal PC 1-7 Noise PC 8-22

Kimberlite Suite Linear Discriminant Analysis Process Validation



Accuracy/Confusion Matrix

	eJF	mJF	IJF	Pense	Cantuar
eJF	90.90	0.00	0.00	0.00	9.10
mJF	0.00	96.78	0.00	2.58	0.65
IJF	0.00	3.58	85.71	10.71	0.00
Pense	0.00	2.50	0.00	97.50	0.00
Cantuar	3.45	10.35	0.00	0.00	86.21

•The use of LDA enables the classification, and prediction of kimberlite phases that are relatively rich in diamonds.

Predictive Lithologic Mapping from the Tellus data, Northern Ireland

Tellus Survey "A" and "S" Sample Sites



Surface soil/ A sample (5-20 cm depth)

6862 sites

XRF analysis for major oxides and trace elements on pressed powder pellets (xrf) (6783 sites).

ICP (OES/MS) analysis by aqua regia (ar) digestion (6768 sites) Fire assay for Au, Pd, Pt by ICP-MS

Deep soil/ S sample (35-50 cm depth)

6867 sites

ICP-OES/MS analysis by aqua regia (ar) digestion (6847 sites) ICP-OES/MS analysis by 4 acid (4a) digestion (6859 sites) Fire assay for Au, Pd, Pt by ICP-MS



"A" zone – 5-20 cm depth

"S" zone – 35-50 cm depth

Predictive Mapping Age Brackets



Age Brackets - Sample Sites



Principal Component Analysis



Colour/Symbol by Age Bracket



Amalgamation - Fit-for-purpose

- There are cases where groups of elements are highly associated (correlated), such as rare earth elements (REE) where elements combine in specific lattice sites in minerals.
- These elements can create difficulty in linear algebraic methods where collinearity can result in non-unique solutions.
- Grouping these elements (amalgamating) reduces the risk of collinearity and simplifies the identification of processes.

Amalgamation

- Amalgamation of elements:
- LREE La, Ce, Pr, Nd
- HREE Lu, Yb, Er, Ho, Dy, Tb
- Mafic Fe, Sc, V, Cr
- Feldspar (Cations)- Ba, Na, K,
- Feldspar (Anions) Al, Ga

PCA Screeplots



PCA Biplots



Amalgamated [clr transform]



Analysis of Variance Elements (clr) & Principal Components



Many elements (>20) required for a useful discrimination between classes (Age Brackets)

Only 8 PCs required for a useful discrimination between classes (Age Brackets)



Linear Discriminant Analysis





Not a significant difference

Classification Accuracy

Counts											
predicted											
true	CzOl	CzPl	Mes	NeoP	Pg	Pl	PlCr	PlDv	PlOr	PlSi	
CzOl	11	42	37	6	0	0	56	0	2	0	
CzPl	13	1448	74	54	0	9	71	0	21	1	
Mes	4	69	57	10	0	0	132	0	58	0	
NeoP	6	7	6	736	1	17	189	1	49	1	
Pg	0	11	1	5	72	1	10	0	12	12	
Pl	0	5	4	90	0	10	34	2	25	0	
PlCr	8	8	32	216	1	15	1070	6	172	6	
PlDv	1	2	1	95	0	0	35	27	142	1	
PlOr	0	1	33	33	40	0	54	5	1117	28	
PlSi	0	0	0	1	1	0	2	0	16	148	
error rate = 30.93 %											
predict	ed				%	/ D					
-	CzOl (CzPl	Mes N	leoP	Pg	Pl H	PlCr P	lDv P	olor P	lSi	
CzOl	7.1	27.3 2	24.0	3.9	0.0	0.0 3	36.4	0.0	1.3	0.0	
CzPl	0.8	85.6	4.4	3.2	0.0	0.5	4.2	0.0	1.2	0.1	
Mes	1.2 2	20.9 1	.7.3	3.0	0.0	0.0 4	10.0	0.0 1	7.6	0.0	
NeoP	0.6	0.7	0.6 7	2.7	0.1	1.7 1	18.7	0.1	4.8	0.1	
Pq	0.0	8.9	0.8	4.0 5	8.1	0.8	8.1	0.0	9.7	9.7	
Pĺ	0.0	2.9	2.4 5	2.9	0.0	5.9 2	20.0	1.2 1	4.7	0.0	
PlCr	0.5	0.5	2.1 1	4.1	0.1	1.0	59.8	0.4 1	1.2	0.4	
PlDv	0.3	0.7	0.3 3	1.2	0.0	0.0	1.5	8.9 4	6.7	0.3	
PlOr	0.0	0.1	2.5	2.5	3.1	0.0	4.1	0.4 8	5.2	2.1	
PlSi	0.0	0.0	0.0	0.6	0.6	0.0	1.2	0.0	9.5 8	8.1	

- A multivariate distance (Mahalanobis Distance MD) is measure from a given observation (sample site) to the centroid of each Age Bracket class. Distances are based on the principal components used for the classification.
- The predicted class is assigned based on the shortest MD.
- Posterior probabilities are forced fits. An observation must belong to one class.



Tellus Soil A (XRF) Pg Paleogene Posterior Probability · 0.0 - 0.1 200000 250000 300000 350000 · 0.01 - 0.2 0.2 - 0.3 0.3 - 0.4 • 0.4 - 0.5 0.5 - 0.6 0.6 - 0.7 • 0.7 - 0.8 • 0.8 - 0.9 • 0.9 - 1.0 Age Bracket Palaeogene Cenozoic, Palaeogene Palaeocene Cenozoic, Palaeogene Oligocene Mesozoic, Cretaceous Jurassic Triassic Palaeozoic, Carboniferous Mississ. Pennsyl. Palaeozoic, Ordivician Silurian Permian 20 30 40 50 km Caledonian (Silurian Devonian) Palaeozoic, Carboniferous Devonian Lower Palaeozoic 200000 250000 300000 350000 Neoproterozoic

Cenozoic Basalts



Lacustrine; Clay, sand & lignite

Continental redbed facies sandstone & mudstone



Basalt, trachyte, syenite & tuff + sediments



Appinite Suite, Granite, granodiorite



Basalt, andesite + sediments



Paleozoic Wackes/Conglomerate/Redbeds/Mudstone



Paleozoic Granitoid/Gabbroic Rocks



Neoproterozoic Pelite/Semi-pelite/Psammite



Predictive Lithologic Mapping from the Tellus data, Northwest Ireland

Gallagher, V., Grunsky, E., Fitzsimons, M., Browne, M., Lilburn, S. and Symons, J. (2021), Tellus Border and West Stream Water Data Analysis and Interpretation. Geological Survey Ireland report.

Evaluation of Tellus Waters Geochemistry Northwest Ireland

- Geological Survey of Ireland carried out a geochemical survey of surface waters.
- 6,739 stream water sites were sampled and studied.
- Intent to examine the usefulness of surface waters to reflect natural and anthropogenic influences.

Predicting Bedrock Lithologies from Tellus Waters Geochemistry



Principal Component Analysis



Tellus Waters Geochemistry - RockType



Random Forests Prediction of Rock Type

PC1 500 400 Mean Decrease in Gini 300 PC5 PC2 PC7 200 PC9 POPOR 100 0 10 20 30 40 50 Order

Random Forest Classification RockType

	inter noon type						Countr								
	amphiholito	anninito	folyolo	abbro	granito	grouwacko	limostono	mafuel	ognoice	polito	ngnoiss	quartzito	rodbod	candistano	clate
amphiholite	872	appinte	leivoic	gabbio	1/	gieywacke 13	00			78	pgrieiss	quartzite	neubeu	Sanustone	Siate
anninite	072	0	0	0	14	13	0		, 0) 0	/0	0	0	0	0	
felvolc	2	0	0	0	1	4	0		, 0) 0	1	0	0	0	0	
gabbro	20	0	0	0	7	3	10		, 0) 0	7	0	0	0	1	
granite	60	0	0	0	318	13	41	, () 0	35	0	2	0	7	
greywacke	19	0	0	0	7	748	93		0	4	0	- 0	0	21	4
limestone	23	0	0	0	, ,	37	1735		, c	2	0	0	1	63	(
mafyolc	3	0	0	0	0	10	18		, 0) 0	0	0	0	0	5	
ogneiss	4	n	0	0	5	10	10) N	5	0	0	0	0	6
pelite	242	0	0	0	16	29	61	(0 0	489	0	3	0	14	(
ogneiss	35	0	0	0	0	1	20) () 0	7	6	0	0	2	(
quartzite	148	0	0	0	10	- 8	31	() 0	66	0	55	0	7	(
edbed	18	0	0	0	1	5	36	; (0 0	17	0	0	23	22	(
sandstone	69	0	0	0	0	22	338		0 0	22	0	0	0	501	(
slate	5	0	0	0	2	34	4) 0	2	0	0	0	1	-
							Accuracy %								
	amphibolite	appinite	felvolc	gabbro	granite	greywacke	limestone	mafvolo	ogneiss	pelite	pgneiss	quartzite	redbed	sandstone	slate
amphibolite	80.2	0	0	0	1.3	1.2	9.1	. () 0	7.2	0	0.5	0	0.6	C
appinite	80	0	0	0	20	0	0) (0 0	0	0	0	0	0	(
felvolc	25	0	0	0	12.5	50	0) (0 0	12.5	0	0	0	0	(
gabbro	41.7	0	0	0	14.6	6.2	20.8	; () 0	14.6	0	0	0	2.1	(
granite	12.6	0	0	0	66.8	2.7	8.6	i (0 0	7.4	0	0.4	0	1.5	(
greywacke	2.1	0	0	0	0.8	83.8	10.4	. (0 0	0.4	0	0	0	2.4	0.1
limestone	1.2	0	0	0	0.1	2	93.1	. () 0	0.1	0	0	0.1	3.4	(
mafvolc	8.3	0	0	0	0	27.8	50) () 0	0	0	0	0	13.9	(
ogneiss	25	0	0	0	31.2	6.2	6.2) 0	31.2	0	0	0	0	(
pelite	28.3	0	0	0	1.9	3.4	7.1	. 0) 0	57.3	0	0.4	0	1.6	(
pgneiss	49.3	0	0	0	0	1.4	28.2	. () 0	9.9	8.5	0	0	2.8	(
quartzite	45.5	0	0	0	3.1	2.5	9.5	; () 0	20.3	0	16.9	0	2.2	(
redbed	14.8	0	0	0	0.8	4.1	29.5	, (0 0	13.9	0	0	18.9	18	(
sandstone	7.2	0	0	0	0	2.3	35.5	i (0 0	2.3	0	0	0	52.6	(
slate	9.8	0	0	0	3.9	66.7	7.8	, r) 0	39	0	0	0	2	5.9

Overall Accuracy

69.73

Random Forests - Prediction





Random Forest Predictive Map



Random Forest Predictive Map





From surface sediment geochemical surveys to prediction of major crustal blocks

U Mueller¹, E C Grunsky², P de Caritat³ and J Coombes¹

¹Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia ²Department of Earth and Environmental Sciences, University of Waterloo, Canada N2L 3G1 ³Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia

National Geochemical Survey Australia (with Patrice de Caritat, Geoscience Australia)

- National Geochemical Survey of Australia
- Geochemical composition (major, minor and trace elements) was determined on sediment samples collected as part of the National Geochemical Survey of Australia (NGSA) between 2006 and 2011 by Geoscience Australia and all States/NT geological surveys (www.ga.gov.au/ngsa).
- The NGSA collected catchment outlet (similar to floodplain/overbank) sediment samples from two depths (0-10 cm, and ~60-80 cm) in the sediment profile at 1186 sites across Australia.
- Collected samples were sieved to two grain-size fractions (<2 mm, and <75 μm) and analysed geochemically, resulting in four separate data sets (Caritat & Cooper, 2016).
- Whole-rock compositional data for 50 elements were used in this study.

Aim

- Australia is composed of crustal blocks reflecting architecture of the continent
- National Geochemical Survey of Australia analysed surface regolith over continent
- The objective of this study is to use surface regolith compositional data for spatial prediction of membership to the major crustal blocks across the Australian continent
- Previous work has shown that the geochemistry of surficial materials (lake sediments, glacial till) can be used to define "type" compositions for the lithologies with which they represent based on statistical classification methods.

Surface Geology Map Australia



Major Crustal Blocks of Australia

- Classically interpreted from deep seismic reflection data and potential field geophysics.
- Fundamental geological units forming a continent,
- Boundaries often are crustalor lithospheric-scale discontinuities that may act as melt and fluid conduits in a mineral system.



Methods

- Treat the geochemical data as compositions and apply logratio transforms.
- Compute Minimum/Maximum Autocorrelation Factors (Switzer and Green, 1984).
- Apply analysis of variance to MA factors based on MCB classification.
- Apply linear discriminant analysis to validate and provide a classification framework of the MCBs using non- robust estimates of covariance.
- Soft indicator kriging to interpolate posterior probabilities derived from LDA.
- Derive most likely crustal block and measures of uncertainty.

Classification Accuracy

MCB	BOS c/g E	3OS f/g 1	ſOS c∕g⊺	FOS f/g	_
MCB01	0.81	0.85	0.87	0.82	S.
MCB02	0.71	0.72	0.74	0.80	7
MCB03	0.67	0.69	0.67	0.74	
MCB04	0.57	0.69	0.72	0.69	S-02
MCB05	0.41	0.24	0.42	0.44	
MCB06	0.55	0.54	0.40	0.45	
MCB07	0.42	0.53	0.53	0.56	30%
MCB08	0.62	0.72	0.69	0.69	
MCB10	0.42	0.44	0.46	0.63	S.
MCB11	0.32	0.32	0.30	0.44	9
MCB12	0.48	0.48	0.39	0.30	_
MCB13	0.18	0.29	0.21	0.26	S-0
MCB14	0.83	0.89	0.94	0.89	
MCB15	0.72	0.64	0.60	0.48	
MCB16	0.19	0.26	0.23	0.19	20°S
MCB18	0.60	0.65	0.60	0.70	
MCB19	0.16	0.12	0.04	0.16	
MCB20	0.32	0.21	0.26	0.47	30°5
MCB21	0.56	0.44	0.49	0.62	
MCB22	0.10	0.15	0.15	0.25	ş
MCB23	0.33	0.43	0.29	0.24	40%
MCB24	0.14	0.14	0.29	0.36	



Posterior Probabilities, MCB04, MCB02, TOS c/g (left) and BOS f/g (right)

Measure of Uncertainty

- From the estimated posterior probabilities the following measures of spatial uncertainty about the allocation of a location to a specific MCB were computed:
- Local classification of uncertainty: $\varphi(x) = 1 max_m(p_m(x))$
- Local entropy, given by $H(x) = -\sum_{i=1, p_i(x)\neq 0}^{K} p_i(x) \ln(p_i(x))$
- *K* denotes the number of classes
- The local classification uncertainty attains values from 0 to 1 and is close to 0 when class membership is reasonably certain. Similarly, the local entropy provides a measure of local randomness. The maximum value is given by ln *K*, which is attained when all K MCBs are equiprobable, and its minimum value is 0, when the MCB is certain.

Measures of Uncertainty ($\boldsymbol{\varphi}$)



Measures of Uncertainty (Entropy)



In Conclusion ...

- Geochemical survey data are well suited for identifying processes associated with mineralogy, weathering, transport.
- The use of multi-element geochemistry enables the recognition of processes through combinations of elements that reflect mineral stoichiometry. Soils, stream sediments, lithogeochemistry, laterite, etc. all have distinct mineral associations and are reflected in the inter-element associations.
- The geospatial association of geochemical survey sites with measures of underlying geology and surficial processes record the influence of igneous, sedimentary, mineralizing processes and the subsequent interaction with the biosphere (ecosystem) and climate.
- The use of advanced analytics and machine learning methods enhance the ability to "Discover" and "Validate/Predict" processes.
- The results demonstrated from predictive mapping confirm the capacity of geochemical data to test new hypotheses from which new geological/geochemical process maps can be created
- The establishment of training sets (specific lithologies, alteration, soils, ecosystems, landforms, climate) can assist in the study and prediction in areas where there is a lack of information.
- Overlap between classes (e.g. lithologies) is expected and the use of posterior probabilities can identify the degree of distinctiveness and overlap.
- The use of advanced analytics and machine learning methods provide an objective, repeatable and defensible framework from which processes can be identified, predicted and validated.

Some References

• Gallagher, V., Grunsky, E.C., Fitzsimmons, M.M., Browne, M.A., Lilburn, S., Symons, J., in press. Tellus Regional Surface Water Geochemistry: environmental and mineral exploration applications, Geochemistry: Exploration, Environment, Analysis.

• Greenacre, M., Grunsky, E., Bacon-Shone, J. 2021. A comparison of isometric and amalgamation logratio balances in compositional data analysis, Computers & Geosciences, 148, https://doi.org/10.1016/j.cageo.2020.104621

• Grunsky, E.C., Arne, D., 2020. Mineral-Resource Prediction Using Advanced Data Analytics and Machine Learning of the QUEST-South Stream-sediment Geochemical Data, Southwestern British Columbia, Canada, Geochemistry: Exploration, Environment, Analysis, https://doi.org/10.1144/geochem2020-054

Grunsky, E.C., de Caritat, P., 2019. State-of-the-Art Analysis of Geochemical Data for Mineral Exploration, Geochemistry, Exploration Environment Analysis, Special Issue from Exploration 17, October, 2017, Toronto, Canada, DOI 10.1144/geochem2019-031

• McKinley, J.M., Grunsky, E.C., Mueller, UA., 2017. Environmental monitoring and peat assessment using a multivariate analysis of regional-scale geochemical data, Mathematical Geosciences, V. 50, (2), pp, 235-246. doi.org/10.1007/s11004-017-9686-x.

• Grunsky, E.C., de Caritat, P., Mueller, U.A., 2017. Using surface regolith geochemistry to map the major crustal blocks of the Australian continent, Gondwana Research, 46, 227-239, http://dx.doi.org/10.1016/j.gr.2017.02.011.

• McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., de Caritat, P., Filzmoser, P., van den Boogaart, K.G., Tolosana-Delgada, R., 2016. The single component geochemical map: Fact or fiction?, Journal of Geochemical Exploration, Volume 162, March 2016, Pages 16-28, ISSN 0375-6742, http://dx.doi.org/10.1016/j.gexplo.2015.12.005.

• Harris, J. R., & Grunsky, E. C., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Computers & Geosciences, 80, 9-25. 10.1016/j.cageo.2015.03.013.

• Harris, J.R., Grunsky, E.C., Behnia, P., Corrigan, D. 2015. Data- and knowledge-driven mineral prospectivity maps for Canada's North. Ore Geology Reviews, 71, pp. 788-803. Doi: http://dx.doi.org/10.1016/j.oregeorev.2015.01.004.

• Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. Journal of Geochemical Exploration, 141: 15-41 doi 10.1016/j.gexplo.2013.07.013.

• Montreuil, J-f., Corriveau, L., Grunsky, E.C., 2013. Compositional data analysis of hydrothermal alteration in IOCG systems, Great Bear magmatic zone, Canada: To each alteration types its own signature, Geochemistry: Exploration, Environment, Analysis vol. 13, no. 4, 2013; p. 229-247, doi:10.1144/geochem2011-101.

• Eppinger, R.G., Fey, D.L., Giles, S.A., Grunsky, E.C., Kelley, K.D., Minsley, B.J., Munk, L, Smith, S.M., 2013. Summary of Exploration Geochemical and Mineralogical Studies at the Giant Pebble Porphyry Cu-Au-Mo Deposit, Alaska, USA, Economic Geology. V.108, 495-527. ESS Contribution #20120332.

• Grunsky, E.C., 2013. Predicting Archean Volcanogenic Massive Sulfide Deposit Potential from Lithogeochemistry: Application to the Abitibi Greenstone Belt, Geochemistry: Exploration, Environment, Analysis vol. 13, 2013; p. 317-336, doi:10.1144/geochem2012-176.

• Grunsky, E.C., Drew, L.D., Sutphin, D.M., 2010. Process recognition in multi-element soil and stream-sediment geochemistry, Applied Geochemistry, 24(8), pp 1602-1616, 10.1016/j.apgeochem.2009.04.024.

• Grunsky, E.C., 2010. The interpretation of geochemical survey data; Geochemistry, Exploration, Environment Analysis 10(1), p. 27-74, doi: 10.1144/1467-7873/0

• 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.

Thanks for you attention.

Eric Grunsky, Ph.D., P.Geo.(ON) Ottawa ON CANADA email: <u>egrunsky@gmail.com</u>

Geoscience Data Analytics Adjunct Professor Dept. Earth & Environmental Sciences, Univ. Waterloo