Predictive geologic mapping using lake sediment geochemistry in the Melville Peninsula, Nunavut

GEM

E.C. Grunsky[1], D. Corrigan[1], U. Mueller[2], G.F. Bonham-Carter[1]

Predictive Geologic Mapping

Introduction

Government geological surveys and mineral exploration companies collect large amounts of geochemical data that are used in search for mineral commodities or environmental studies. These surveys consist of many thousands of samples (observations) with as many as 50 elements determined for each. Because the nature of the data is compositional, they must be treated according to the protocols established by Aitchison (1986). This contribution details an approach based on the application of the alr, clr and ilr transforms for process discovery and validation.

This poster highlights the value of multi-element geochemical data as an aid to regional geological mapping and potential base- and precious metal deposits through the collection and evaluation of regional geochemical survey data in the Melville Peninsula area of Nunavut, Canada (Figure 1). Figure 2 shows a generalized geological map and a corresponding digital elevation model for the area.

The geology of the area is described by Corrigan et al. (2011). The Melville Peninsula is located in the north-central Rea Craton of the western Churchill Province (wCP). The wCP is a collage of polymetamorphic and polydeformed Archean cratons unconformably overlain by Paleoproterozoic supracrustal sequences and intruded by various intra-plate Proterozoic magmatic suites. It is characterized by widespread tectonothermal reactivation related to the assembly of the supercontinent Nuna during the interval 1.95 - 1.80 Ga.

The data used in this study have been published in the Geological Survey of Canada Open File 6269 (Day et al., 2009) based on earlier studies by Hornbook et al. (1978a, 1978b). Details on the sampling methodology and analytical protocols are documented in Open File 6269.

The results presented here are from a campaign to re-analyze sample pulps using modern analytical methods including Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Instrumental Neutron Activation Analysis (INAA). In cases where elements have been analyzed using two or more methods, the elements were evaluated in terms of detection limit suitability and visual examination of the correlation of the element with each method. This included the evaluation of the degree of censoring (values less than lower limit of detection[<lld]).

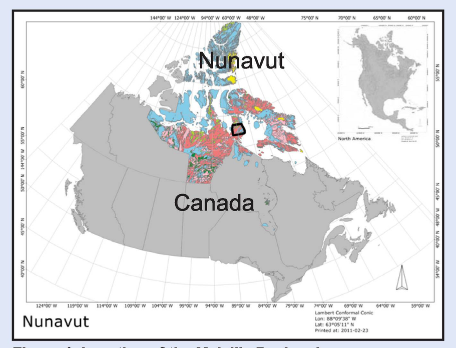

Figure 1. Location of the Melville Peninsula area.

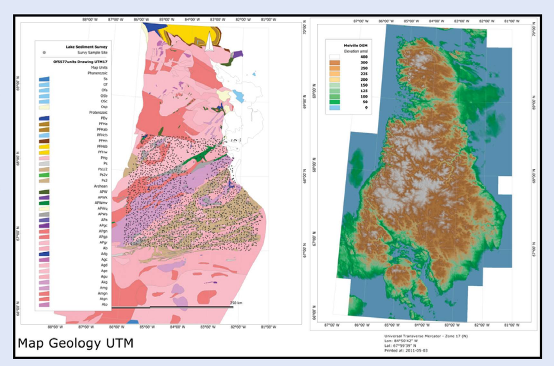

Figure 2. Generalized geology of the Melville Peninsula area based on Skulski (Open File 5577, in prep.). A map of the digital elevation model is shown on the right. Lake sediment survey sites are shown as grey dots.

Dealing with Censored Geochemical Data

One of the primary purposes of geochemical data analysis is the recognition of geochemical/geological processes. Processes are recognized by a continuum of variable responses and the relative increase/decrease of these variables. The presence of censored data generally do not affect the results of a process recognition investigation. In the compositional data analysis framework, the problem of censored data was recognized early (Aitchison, 1986). Martin-Fernandez et al. (2003) discusses various replacement options based on the nature of the censored data. Recognizing the difference between missing values and censored data is crucial in deciding how a replacement value, if any, should be estimated. More recently, Hron et al. (2010) describes a method based on neural networks that provides estimates of replacement values. This methodology, which is implemented within the R package (R-project, 2011) as **robCompositions** is used in this presentation.

Figure 3 shows quantile-quantile plots for the elements (raw data) prior to adjustments for censored values. The lower right plot of Figure 3 shows a quantile-quantile plot for Sb obtained by ICP-MS analysis. The black points represent the data as reported with a significant amount of censoring. The red points show the imputed values determined from the robCompositions package with the impKNNa function (Atichison distance option). These results are considered as reasonable estimates for replacement values.

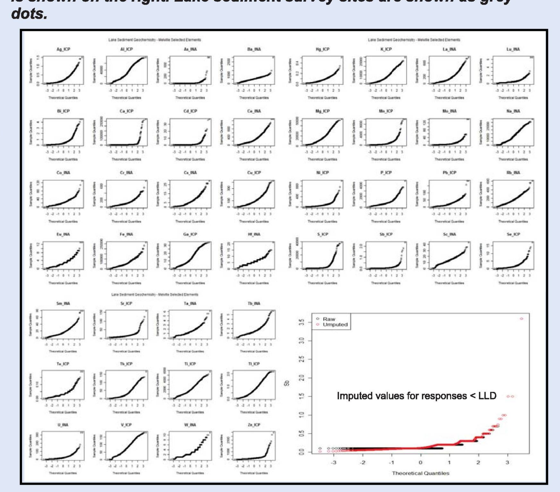

Figure 3. Q-Q plots of the elements. These plots are used to identify outliers and the degree of censoring. The lower right figure shows imputed values for Sb where the reported value < lower limit of detection (LLD). 1258 values out of 1981 values were reported at < LLD of 0.1 ppm. The imputed average replacement value is 0.15 ppm.

Multivariate Association of Geochemical Data

A useful tool for process discovery in geochemical data is principal component analysis (PCA). PCA provides a summary of multi-element relationships in a form of linear combinations of the variables (elements) based on measures of association (correlation). The application of the logcentred (clr) transform to geochemical data provides the advantage of preserving all of the variables although the covariance matrix is singular. This singularity is not a problem for the interpretation of the components. Figure 4 is a biplot of the first two components that account for 42% of the total variability). The figure also show a "screeplot" (upper right) for the eigenvalues obtained from the 46 elements and 2199 lake sediment samples. The score for each observation is depicted by a symbol and colour of the underlying geology of the sample site. Component 1 shows positive scores associated with relative enrichment of Ga-Al-K-Mg-Ti that represent feldspars and mafic minerals (including clays) that are resistant to weathering in an area where there was little ice movement and represents a "felsenmeer" of broken rock. This area occurs in the upper elevations of the Melville Peninsula (Figure 5). Negative PC1 scores show a relative enrichment of Ca-Mg and represents the carbonate platform that occurs at the lower elevations of the peninsula near the coastline. The second component shows a relative enrichment of rare earth elements (PC2>0) that are typically associated with granitoid rocks and Bi-Zn-Ni-Cs-Cd-As-Sb-Te (PC2<0) that are associated with shales and fine grained sediments. The positive PC2 scores occur in the western part of the peninsula (Figure 6) and are associated with Archean-age granitoid rocks (Apg, Amgn). The negative PC2 scores occur in an area where the geology is known as the Penrhyn group (PS1,2,3,PH), composed of black shales and sediments.

Table 1 Number of Lake Sediment Sites over each lithologic unit			
Lithologic Unit	No. of Sites	Lithologic Unit	No. of Sites
Ab	1	APWk	2
Adg	3	APWmv	21
Agd	273	APWs	71
Agu	416	PHg	73
Akg	396	Ps1/2	285
Amg	3	Ps2v	2
Amgn	193	Ps3	298
APgn	8		

Units chosen for classification

Table 2
Linear Discriminant Analysis Accuracy

	Agd	Agu	Akg	Amgn	APWs	PHg	Ps1/2	Ps3
Prior Probabilites	0.14	0.21	0.2	0.1	0.04	0.04	0.14	0.15
Agd	82.05	1.10	2.56	9.52	4.76	0.00	0.00	0.00
Agu	6.97	55.53	14.18	1.68	1.92	0.72	13.70	5.29
Akg	3.28	9.60	75.76	6.06	4.29	0.00	0.76	0.25
Amgn	29.02	3.11	23.32	39.90	4.66	0.00	0.00	0.00
APWs	11.27	16.90	21.13	5.63	42.25	0.00	0.00	2.82
PHg	0.00	15.07	6.85	0.00	0.00	31.51	17.81	28.77
Ps1/2	1.40	17.89	0.70	0.00	0.00	4.56	57.19	18.25
Ps3	0.00	17.79	0.34	0.00	0.00	4.03	21.81	56.04
Overall Accuracy			60.60					

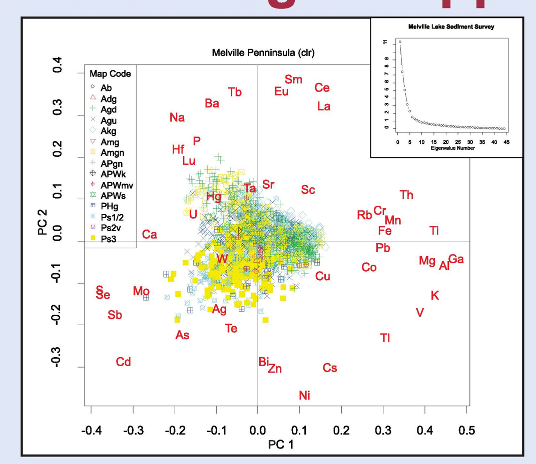

Figure 4. Biplot of the first two principal components based on a clr transform. The legend lists the lithologies associated with the lake sediment sample collection site. A screeplot of the eigenvalues is shown in the upper right.

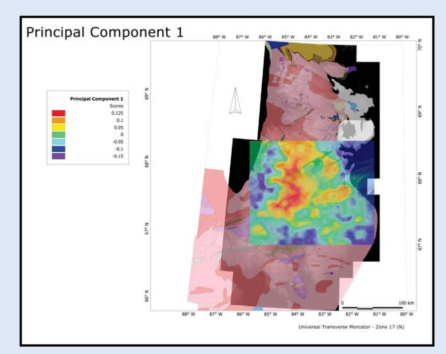

Figure 5. Map showing the first component plotted over the regional geology and digital elevation model for the area. Positive scores are associated with the upper elevations of the digital elevation mode representing a relative increase in Al-Ga-K-V-Mg-Pb-Co-Th-Mn-Fe-Cr-Rb. This combination of elements is interpreted to represent the concentration of mineralogies associated with the homogenization of regional bedrock lithologies resulting from glaciation.

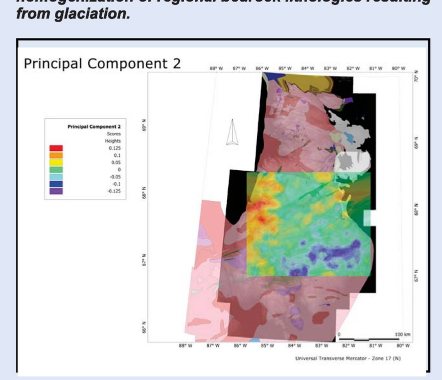

Figure 6. Map showing the second principal component plotted over the regional geology. The patterns observed on this map are more representative of local bedrock lithologies.

Classification and Prediction (ALR or ILR)

This set of data provides an opportunity to test the ability of the lake sediment geochemistry to predict the underlying geology. Each sample site was coded with a geological map unit code. The data were parsed and 8 classes were extracted for which there were a sufficient number of sites for each class (see Table 1). Using the full range of 46 elements both alr and ilr transforms were carried out on the data. In the case of the alr transform, Ce was chosen as the divisor. Standard sequential balances were calculated for the ilr transform. Regardless of the transform, the application of a linear discriminant analysis (**LDA**) in R resulted in similar results. The LDA was carried out using a cross-validation procedure of repeated sampling (20 times) to determine an average accuracy of the classification, which is shown in Table 2. A neural network prediction technique was also applied to the data and yielded similar results. Figure 7 shows a plot of the linear discriminant scores on the first two linear discriminant axes. The observations are shown by colour and symbol to reflect the underlying geology associated at each point. The figure shows three distinct groups:

1) PHg-PS1/2-PS3-Agu that represents the sediments of the Penrhyn group and grainitoid rocks with a similar chemistry (Agu).

2) Akg-Agu-Agd are massive granitoid units that reflect felsic igneous processes.

3) Agd-Amgn-APWs represent gneisses and sediments that are distinctly more clastic in nature than those in Group 1 above.

There is considerable overlap between these groups and the accuracy matrix of Table 2 shows that the overall accuracy is poor for the PHg and Amgn classes. The overall accuracy is 60%. However, given that the lake sediment geochemistry represents a certain degree of homogenization of the surrounding country rock into the lake basins and the terrain is variable in its depositional and erosional history, the results of this classification are considered acceptable as a tool for regional remote predictive mapping in areas where the geology is less well-known.

The resulting predictive classified observations were imaged the **gstat** package in R and plotted based on the posterior probability determined in the application of the LDA. The interpolated maps, by kriging, are shown in Figure 8. The patterns displayed in these images closely correspond with the observed geology.

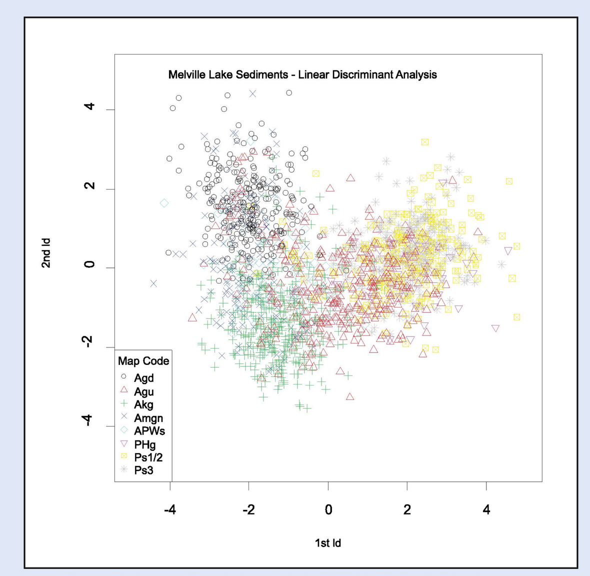

Figure 7. Linear discriminant plot of LD1-LD2 for 2199 lake sediment samples. Lithologic distinction is evident in this plot.

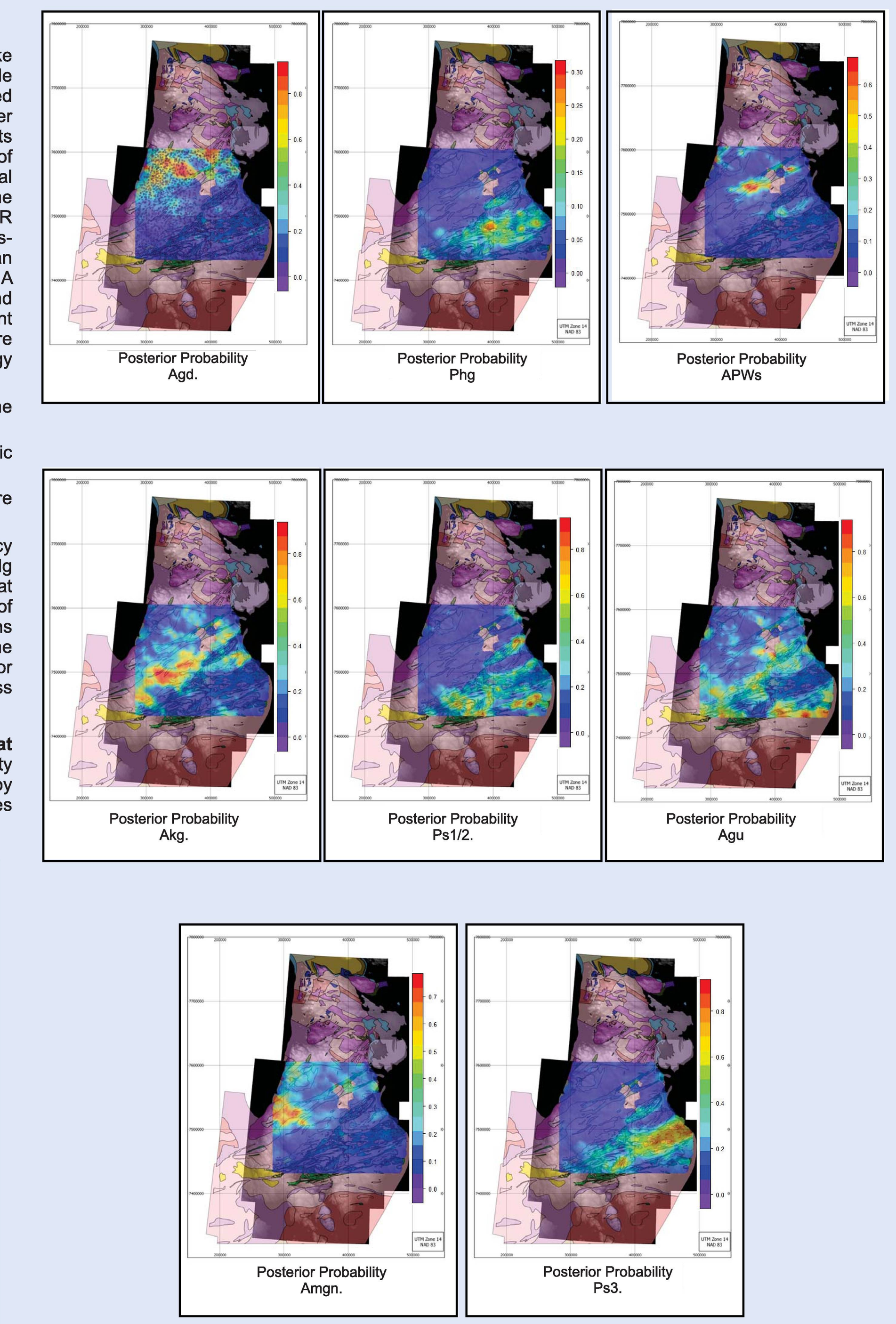

Figure 8. Interpolated images of posterior probabilities for 8 major lithologies in the Melville Peninsula. These images correspond closely with the mapped lithologies shown in Figure 2.

Spatial Analysis

Variogram maps of raw data

The raw data were analysed to ascertain the spatial structure of the distribution of the 16 elements. Experimental semivariogram maps were calculated using 30 lags with a lag value of 5km and a tolerance of 4 sectors using 36 sectors in total. A summary of the variography is shown in Table 3. All elements, except Sr, exhibit anisotropy, the directions differ, as does the degree of anisotropy present.

In addition to consideration of the entire region, the study region was subdivided into a SE and NW part with a view to ascertain whether there are any differences in the spatial continuity of the data in the two subregions compared to the entire region. The directions of greatest continuity differ markedly in the two subregions and for the majority of elements the spatial features of the NW region are those that determine the overall spatial continuity. Exceptions are As, La, Mo, Sr, U and Zn, where the Penrhyn group dominates the spatial behaviour. Variogram maps of Fe and Cu are shown for three regions over the lake sediment survey area. Figures 10a and 10d show variograms for Fe and Cu for the entire survey area. Figures 10b and 10e show variogram maps for the northwest part of the area (dominantly intrusive domain) and Figures 10c and 10f show variogram maps for the southeast (Penrhyn Group) part of the survey area. The circular map shows the variability of variance as a function of angle and lag distance. The accompanying semi-variogram line drawings highlight the differences/similarities for 4 specific directions and a range of lag distances. These figures demonstrate the spatial variability that exists due to the differences and spatial orientation of the lithologies. The study of the variography demonstrates the significant differences that exist within regional geologic domains.

Spatial Multivariate Analysis using the MAF transformation of the Lake sediment data

MAF (Minimum/maximum Autocorrelation Factors) (Switzer and Green, 1984) is in essence carried out by first calculating the principal component analysis of the given data, and using the matrix that diagonalises the correlation matrix of the given data to transform the data to new attributes which are uncorrelated at separation distance 0. This step is succeeded by a rescaling. Following this an experimental covariance matrix derived from the transformed data at a separation distance greater than 0 is used to reduce correlation between pairs of attributes at other distances. This is achieved by diagonalising the chosen covariance matrix. The resulting factors are approximately spatially uncorrelated and so any subsequent modelling can be performed on the individual factors, thus alleviating the burden of inferring a joint model of spatial dependence for the data. The covariance matrix is usually chosen at a separation distance which lies between 0 and the range of autocorrelation for the given data, and a choice of a separation distance close to the sampling distance is common, but not mandatory. The resulting factors are ordered in such a way that the first factor exhibits the highest autocorrelation and with increasing index the strength of autocorrelation decreases with the last few factors often exhibiting pure nugget

For the lake sediment data the 16 standardised variables (Al, As, Ca, Co, Cr, Cu, Ga, La, Mo, Ni, Sb, Sc, Sr, Th, U, Zn) were considered and MAF-transformed using a separation distance 15 km. Experimental semivariogram maps indicate that factors 1 to 10 exhibit anisotropy of varying strength and direction of greatest continuity, factors 11-16 are isotropic with very short ranges of autocorrelation. The most pronounced anisotropy is present in factor 1 which exhibits zonal anisotropy. A summary of the MAF features is shown in Table 4. Because the MAF factors are made up of combinations of the original variables, the individual factors can be used to identify geological features and thereby assist the interpretation of the observed distribution of elements. Figure 11a highlights the difference between the Archean granitoid and supracrustal Penrhyn Group. Figure 11b (MAF 2), highlights the Archean granitoids that are associated with the supracustal rocks in the Penrhyn and Woodburn Lake areas. Figure 11c (MAF3) highlights the supracrustal rocks of the Penrhyn Group and the

Table 3 - Variography interpretation for selected elements from the lake sediement geochemical data.

	Raw Data		
	Melville	Melville NW	Melville Penrhyn
Al	N0	N10	Isotropic in short range
As	N45	No clear	N45
Co	N0	N0	N60
Cr	Short N0	N0	N90
Cu	N80	N0	N45-N60
Ga	N5/N10	N10	N80
La	N45	N30	N60
Mo	N50	Detection	N45
Ni	N10	N0	nugget
Sb	Isotropic	Isotropic	N45
Sc	N10	N10	N90
Th	N60	N20	N80
U	N60	Isotropic	Not clear
Zn	N10-N20	N5	nugget

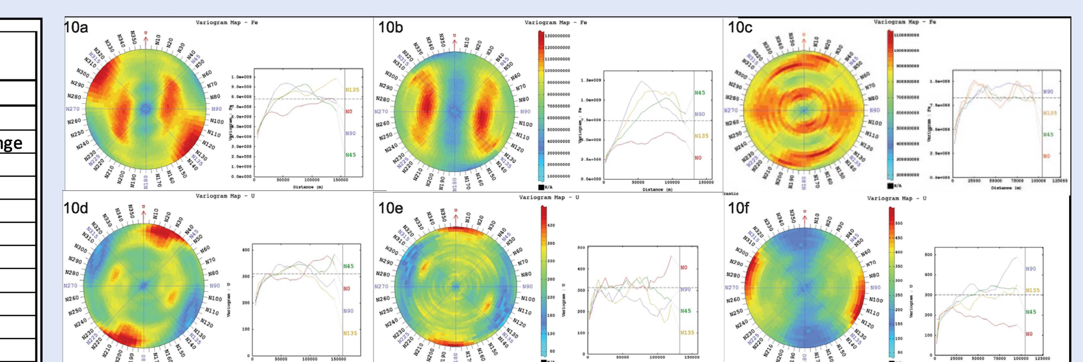

Figure 10. Variogram maps of Fe and Cu for three regions over the lake sediment survey area. Figures 10a and 10d show variogram maps for Fe and Cu for the entire survey area. Figures 10b and 10e show variogram maps for the northwest part of the area (dominantly intrusive domain) and Figures 10c and 10f show variogram maps for the southeast (Penrhyn Group) part of the survey area.

Table 4 - Spatial characteristics of the first 10 MA Factors

MAF	Anisotropy Direction	Type
F1	N60	zonal
F2	N40	geometric
F3	N50	geometric
F4	N150	geometric
F5	Isotropic	Isotropic
F6	N10	geometric
F7	N20	geometric
F8	Isotropic	Isotropic
F9	Isotropic	Isotropic
F10	N60	geometric

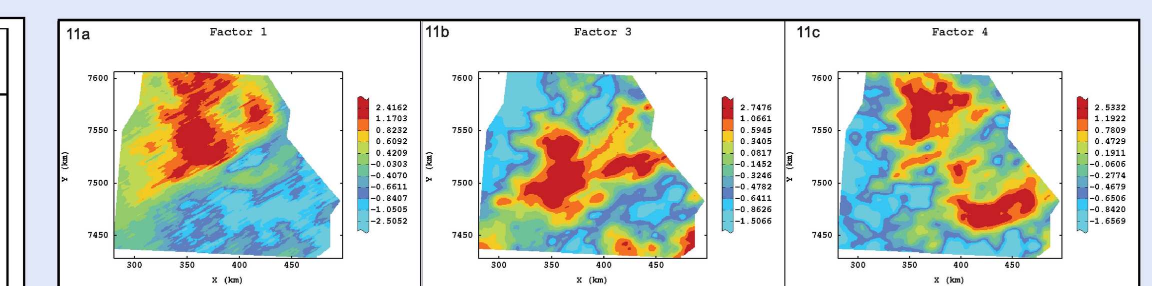

Figure 11. Kriged MAF maps of Factors 1, 3 and 4 that highlight specific spatially uncorrelated multielement patterns based on the entire lake sediment survey area. Figure 11a, MAF 1, highlights the same pattern of relative Ga-Al-K-Mg-Ti enrichment as PC1 (Figure 5). Factor 3 (Figure 11b) shows a pattern the Archean gneisses that occur throughout the central part of the survey area. Figure 11c (MAF 4), shows features that correspond the Penrhyn, Prince Albert and Woodburn Lake Groups. It also includes the area of missing samples. This suggests that further refinement in the kriging parameters is required.

Predicting Potential Mineralization

Estimating Base Metal, REE and U/Th Mineral Potential

Estimating the potential of various types of mineralization can be done through the application of regression techniques. Principal component analysis is a useful technique that identifies significant trends in the data, within the first few components and these trends are typically associated with underlying lithologies (Grunsky, 2010). Figure 8 shows the principal mineral commodity showings in Melville Peninsula.
The regression of, rare earth elements, (Ce, Eu, La, Lu, Sm, Tb) [REE] U, Cu, and Zn against the first five principal components derived from the lake sediment survey data provides a potentially useful means for separating background values of these elements from values that may be associated with rare events (under-sampled). Residual values (reported - predicted) of these elements are shown in Figures9a,b,c. A regression was not performed on Au values and the reported analytical values are shown on the map in Figure 9d.

Elevated residual U values, as shown in Figure 9a, occur within then southwestern part of the Penrhyn Group and the Archean granodioritic rocks in the northwest part of the lake sediment survey region. Residual REE elements (Figure 9b) show elevated values in the Archean granodioritic rocks in the northwest part of the lake sediment survey area and also in the Archean granitoid rocks within and adjacent to the Penrhyn Group sediments in the south and south-central parts of the area. Elevated residual Zn values (Figure 9c) occur within the Penrhyn Group in the southern part of the

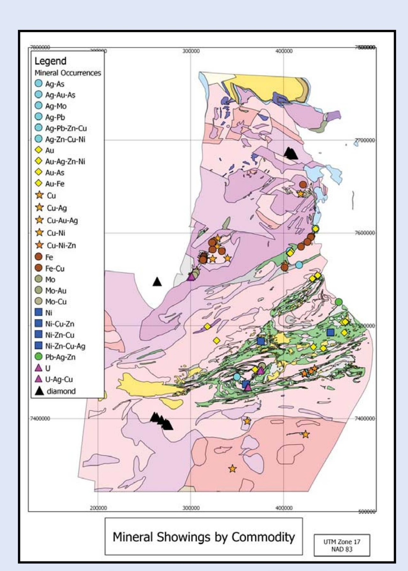

Figure 8. Principal mineral commodities in the Melville Peninsula area.

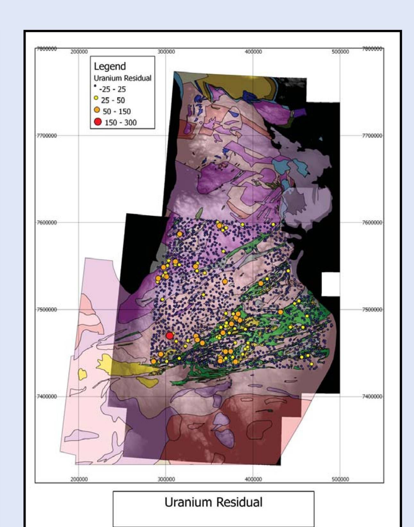

Figure 9a. Calculated residual U values based on regression of U with the first five principal components.

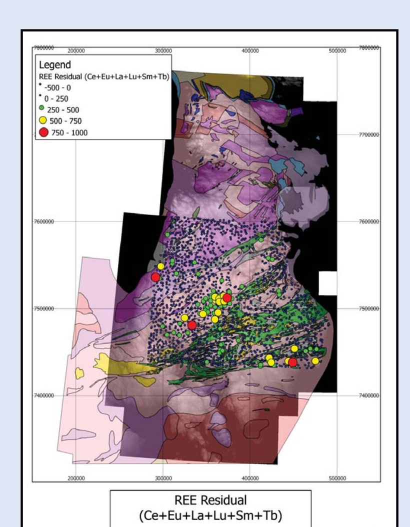

Figure 9b. Calculated residual REE values based on regression of REE with the first five principal

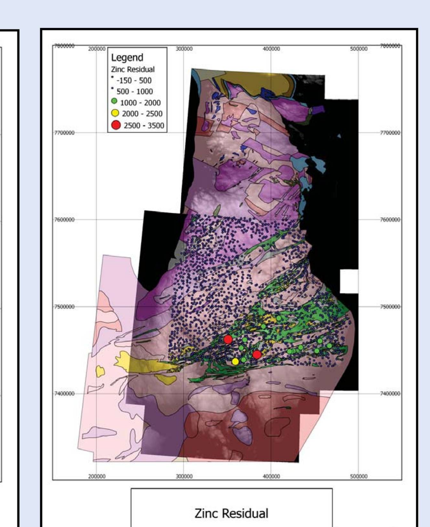

Figure 9c. Calculated residual Zn values based on regression of Zn with the first five principal

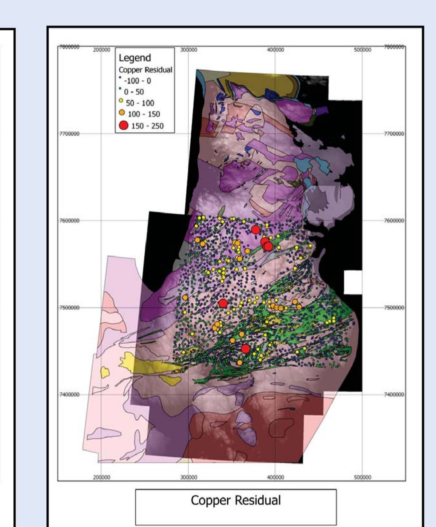

Figure 9d. Calculated residual Cu values based on regression of Cu with the first five principal

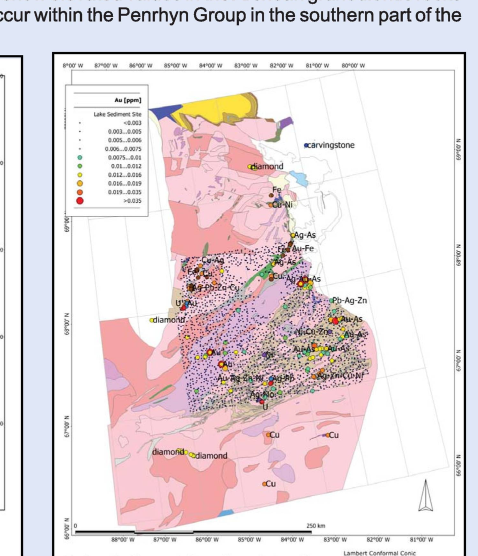

Figure 9e. Reported Au analytical values from the lake sediment survey.

Conclusions

This poster summarizes the procedures used to evaluate a geochemical survey dataset to enable regional geochemical interpretation. Through the application of data adjustment procedures, compositional data analysis, and the application of statistically based classification procedures, we can successfully interpret compositional (geochemical) data for the purposes of regional geological mapping.

The use of discriminant procedures through the use of cross-validation procedures and compositional data adjustments allows for an effective means of classifying lake sediment geochemistry based on selected geological knowledge.

Areas of potential mineralization can be recognized through the application of regression based techniques applied to the principal components derived from the clr-transform. Further refinement is possible through experimentation in the choice of the number of principal components required to isolate rare-event observations that are potentially associated with mineralization.

Multivariate spatial analysis is being explored and expanded to test its use as a mechanism for a better understanding of the spatial characteristics of multi-element geochemistry and its potential role in carrying out predictive geological mapping and the identification of potential precious, base metal, uranium and rare earth element mineralization.

Recommended citation
Grunsky, E.C., Corrigan, D., Mueller, U., and Bonham-Carter, G.F., 2012. Predictive geologic mapping using lake sediment geochemistry in the Melville Peninsula, Nunavut; Geological Survey of Canada, Open File 7171, 1 sheet. doi:4095/291901

Acknowledgements
Geological base map: updated from Skulski (Open File 5577, in prep.)

References

Aitchison, J., 1986. The statistical analysis of compositional data. Chapman and Hall, New York. 416p.

Cameron, E. M.; Coker, W. B.; Ellwood, D.J., 1981.. National geochemical reconnaissance, Melville peninsula, Northwest Territories [46n,p; 47a(s1/2) and 47b(s1/2)]; Geological Survey of Canada, Open File 738, 23 pages, doi:10.4095/129717

Corrigan, D., L. Nadeau, N. Wodicka, S. Erdmann, G. Machado, M. Houle, and T. Tremblay. 2011. GEM-Minerals Project on Melville Peninsula, Nunavut: Update and Highlights. Cordilleran Roundup, Vancouver, January 2011

Day, S.J.A.; McCurdy, M.W.; Friske, P W B; McNeil, R J; Hornbrook, E. H. W.; Lynch, J.J., 2009. Regional lake sediment and water geochemical data, Melville Peninsula, Nunavut (parts of NTS 046N, O, P, 047A and B); Geological Survey of Canada, Open File 6269, 12 pages

Grunsky,E.C., 2010. The interpretation of geochemical survey data; Geochemistry, Exploration, Environment and Analysis 10(1), p. 27-74.

Hornbrook, E.H.W.; Lynch, J.J. 1978. Geological Survey of Canada, Open File 522A, 1978., doi:10.4095/129511

Hornbrook, E.H.W.; Lynch, J.J., 1978, Geological Survey of Canada, Open File 522B, 1978; 14 sheets, doi:10.4095/129512

Hron, K. , Templ, M. , Filzmoser, P, 2010. Imputation of missing values for compositional data using classical and robust methods, Computational Statistics and Data Analysis, 54 (12), pp. 3095-3107.

NGR, 1977. National geochemical reconnaissance release NGR 33A-1977, regional lake sediment and water geochemical reconnaissance data, Melville Peninsula, Northwest Territories

NGR, 1977. National geochemical reconnaissance release NGR 33B-1977, regional lake sediment and water geochemical reconnaissance data, Melville Peninsula, Northwest Territories

Martin-Fernandez, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V., 2003. Dealing with Zeros and Missing Values in Compositional data Sets Using Nonparametric Imputation, Mathematical Geology, 35(3), 253-278.

R development core team, 2011. R: A language and environment for statistical computing. Vienna, http://www.r-project.org

Switzer, P., Green, A., 1984. Min/max autocorrelation factors for multivariate spatial imaging, Technical report no. 6, Department of Statistics, Stanford University, California, 14p..

[1] Geological Survey of Canada, 601 Booth St., Ottawa, ON K1A 0E8 Canada (email: egrunsky@nrcan.gc.ca)

[2] School of Engineering, Edith Cowan University, 270 Joondalup Drive, JOONDALUP, Western Australia 6027 (email: u.mueller@ecu.edu.au)

doi:10.4095/291901

This publication is available for free download through GEOSCAN (http://geoscan.ess.nrcan.gc.ca/).

Canada