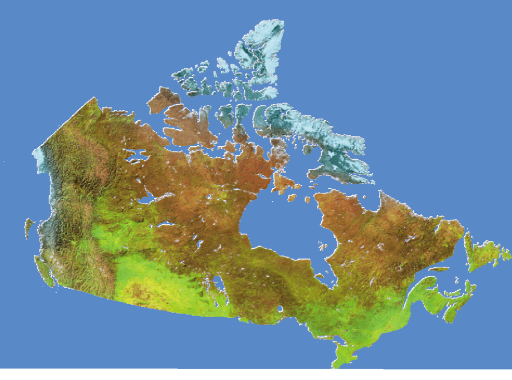
GEM

Practical aspects of compositional data analysis using regional geochemical survey data

E.C. Grunsky[1] and D. Corrigan[1]



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]).

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.

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 felspars 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.

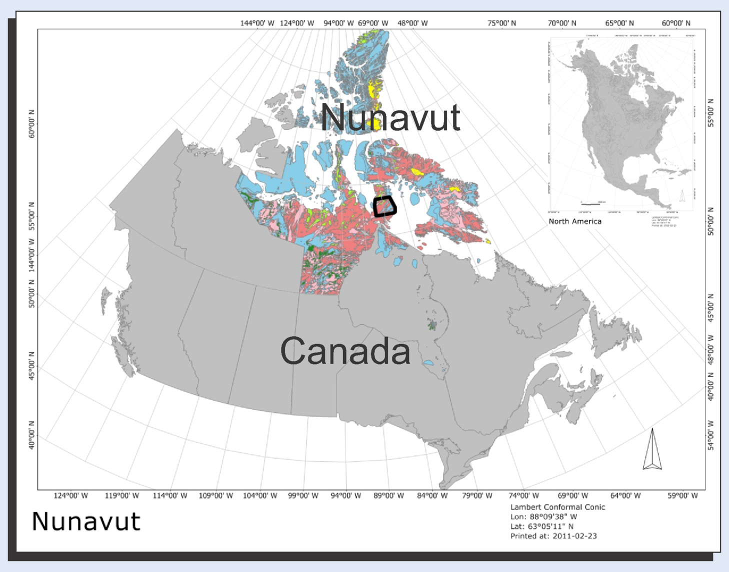


Figure 1. Location of the Melville Peninsula area.

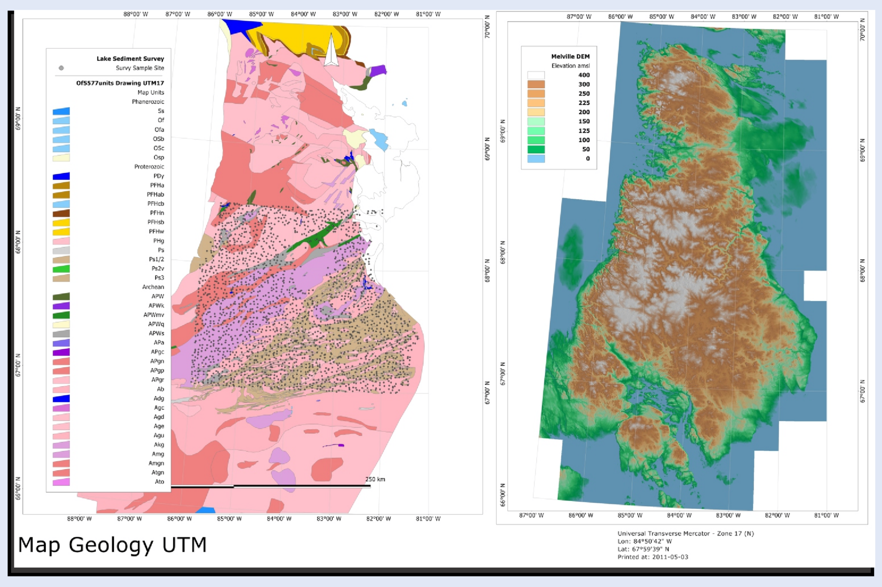


Figure 2. Generalized geology of the Melville Peninsula area based on Skulski (Open File 5577, in prep.). A map of the digital elevation model is shown on the right. Lake sediment survey sites are shown as grey dots.

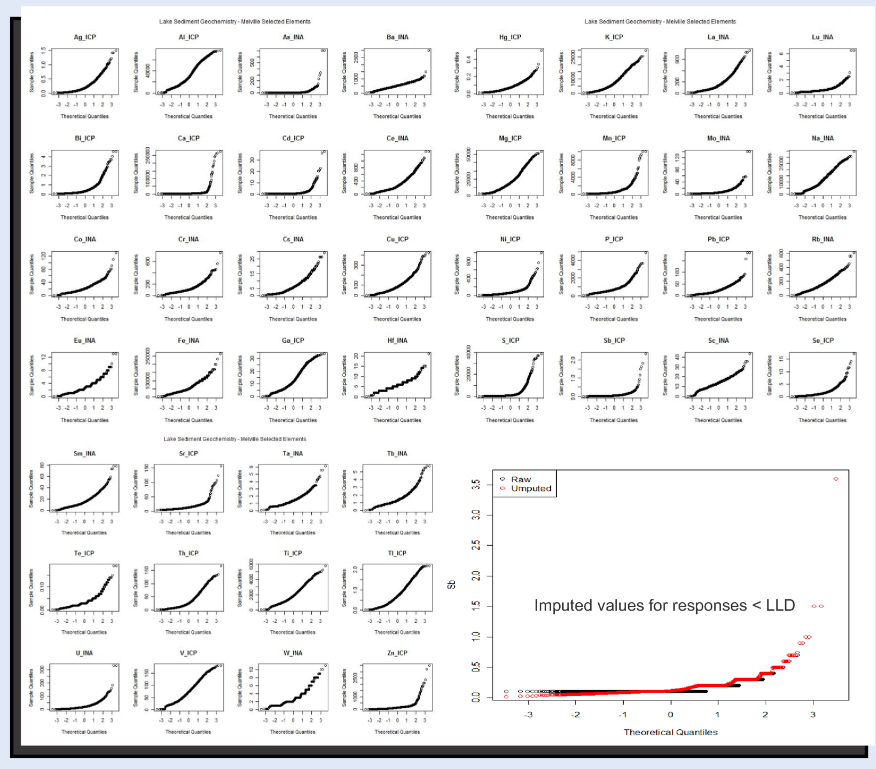


Figure 3. Q-Q plots of the elements. These plots are used to identify outliers and the degree of censoring. The lower right figure shows imputed values for Sb where the reported value < lower limit of detection (LLD). 1258 values out of 1981 values were reported at < LLD of 0.1 ppm. The imputed average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

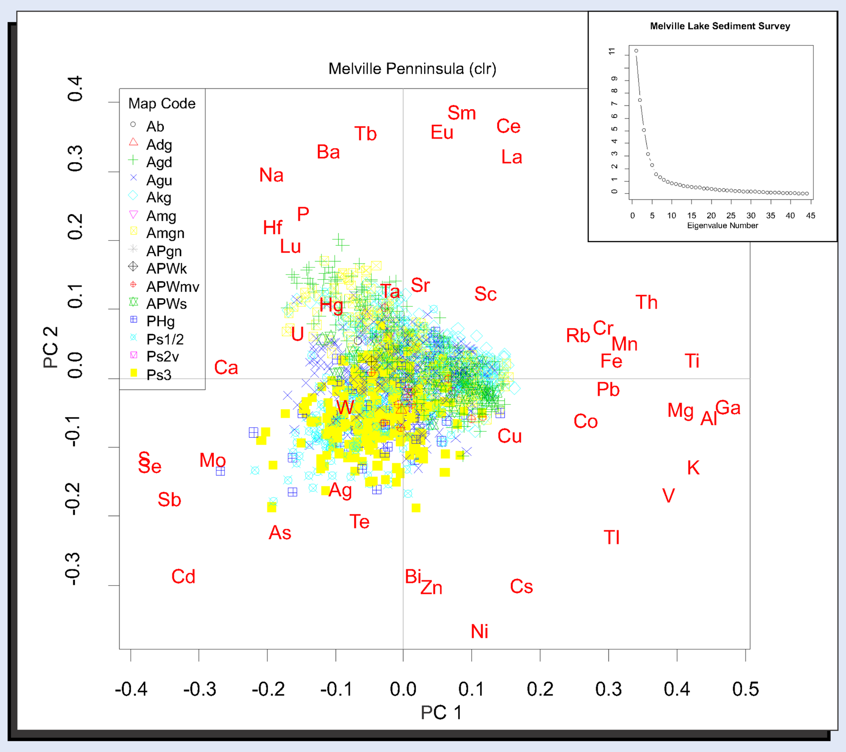


Figure 4. Biplot of the first two principal components based on a clr transform. The legend lists the lithologies associated with the lake sediment sample collection site. A screeplot of the eigenvalues is shown in the upper right.

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 represent 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. An example semi-variogram used to fit the data is shown in Figure 8. The interpolated maps, by kriging, are shown in Figure 9. The patterns displayed in these images closely correspond with the observed geology.

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.

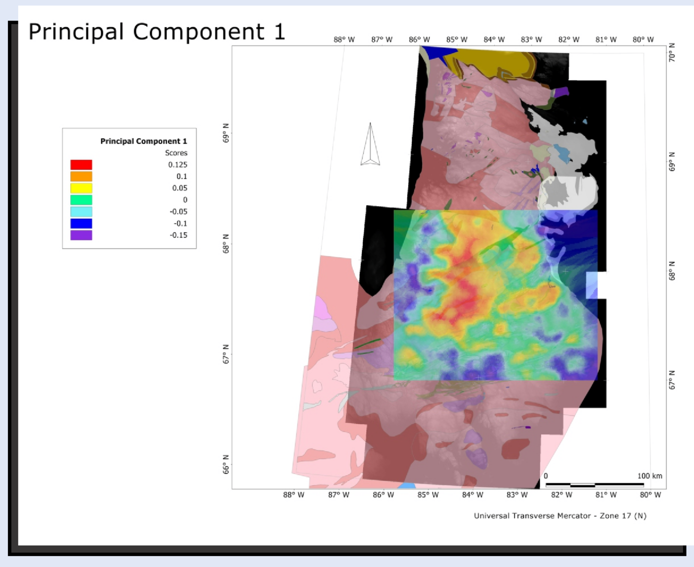


Figure 5. Map showing the first component plotted over the regional geology and digital elevation model for the area. Positive scores are associated with the upper elevations of the digital elevation mode representing a relative increase in Al-Ga-K-V-Mg-Pb-Co-Th-Mn-Fe-Cr-Rb. This combination of elements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

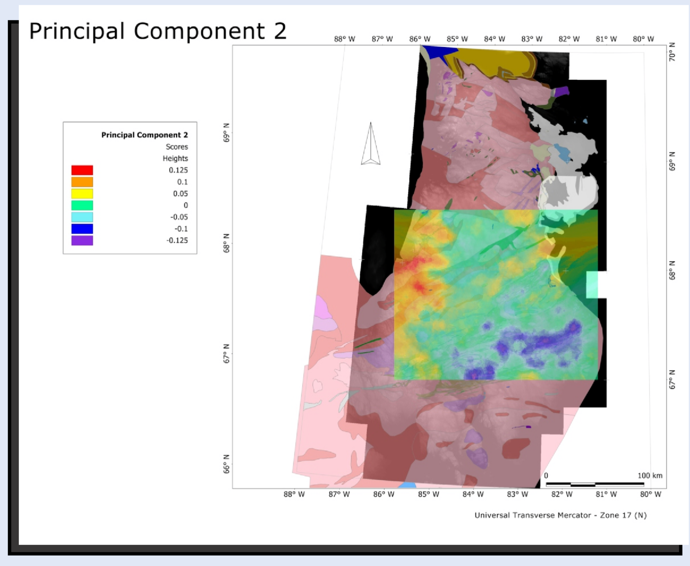


Figure 6. Map showing the second principal component plotted over the regional geology. The patterns observed on this map are more representative of local bedrock lithologies.

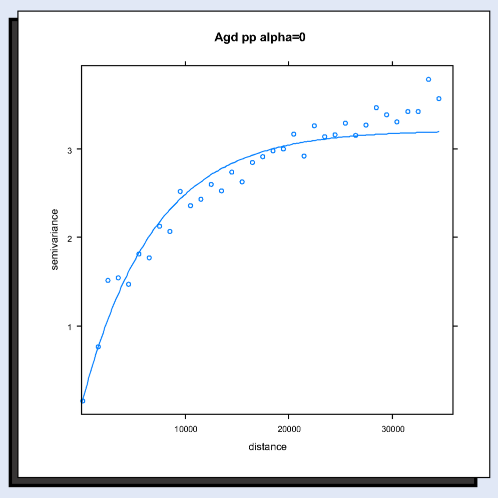


Figure 8. Fitted semi-variogram for posterior probability of lithology Agd.

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

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					

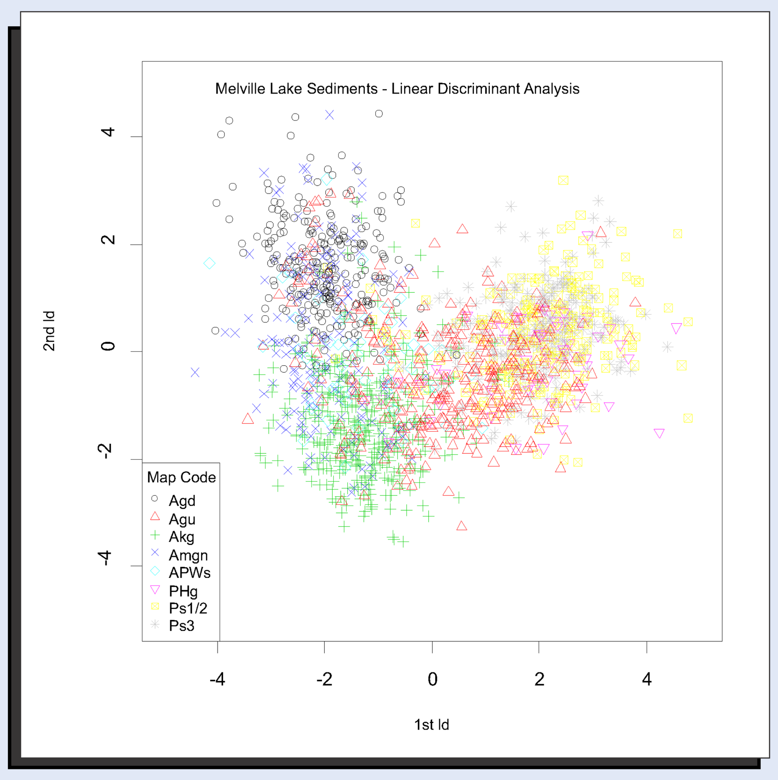


Figure 7. Linear discriminant plot of LD1-LD2 for 2199 lake sediment samples. Lithologic distinction is evident in this plot.

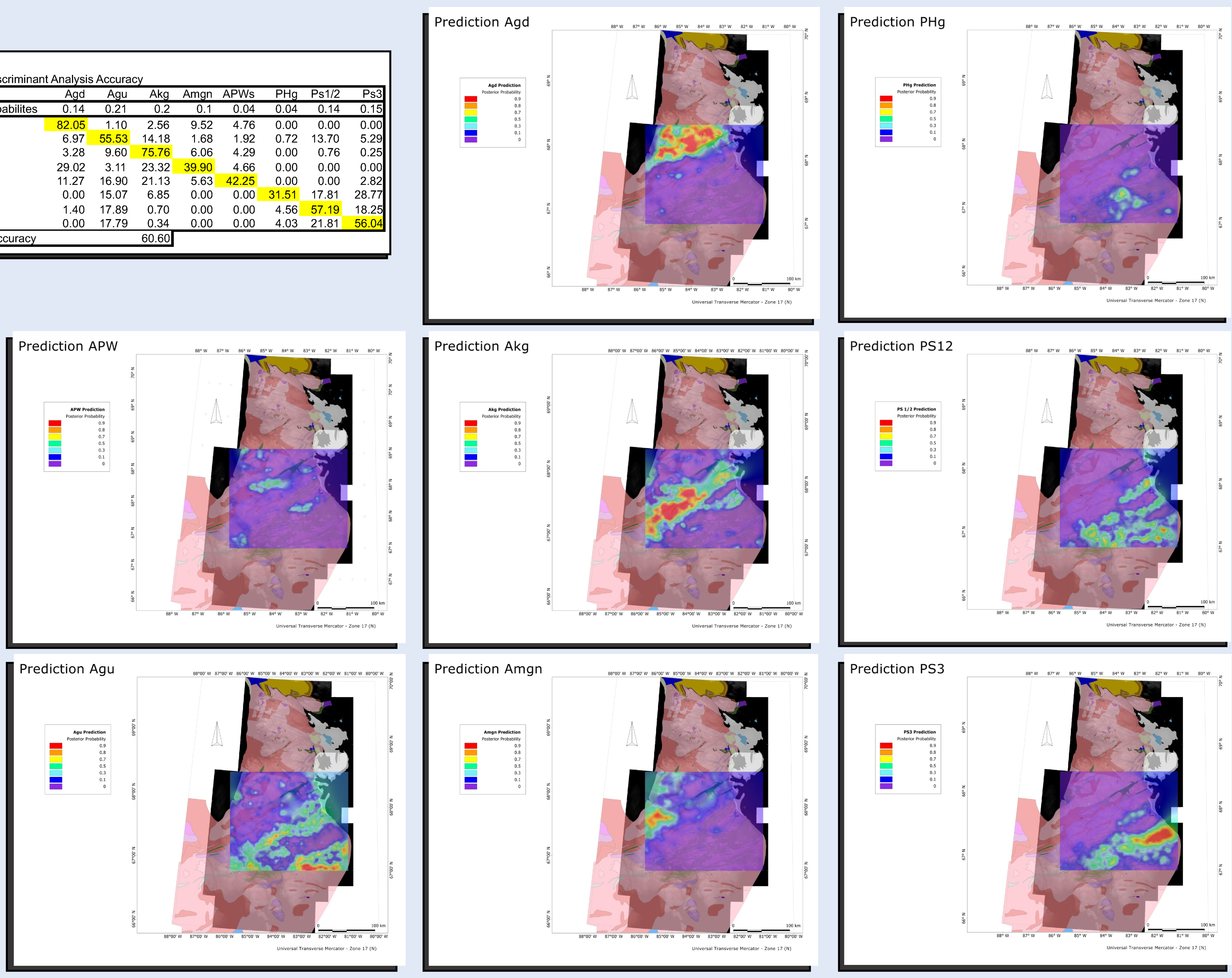


Figure 9. Interpolated images of posterior probabilities for 8 major lithologies in the Melville Peninsula. These images correspond closely with the mapped lithologies shown in Figure 2.

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

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.

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

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

Acknowledgments

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

doi:10.4095/295694

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

Recommended citation
Grunsky, E.C. and Corrigan, D., 2015. Practical aspects of compositional data analysis using regional geochemical survey data ; Geological Survey of Canada, Open File 7176, 1 poster. doi:10.4095/295694

OPEN FILE
DOSSIER PUBLIC
7176
GEOLOGICAL SURVEY OF CANADA
COMMISSION GÉOLOGIQUE DU CANADA
2015

Publications in this series have not been edited; they are released as submitted by the author.

Les publications de cette série ne sont pas révisées; elles sont publiées telles que soumises par l'auteur.

[1] Geological Survey of Canada, 601 Booth St., Ottawa, Ontario, Canada
(Eric.Grunsky@NRCan-RNCan.gc.ca)