APPLICATION OF GEOGRAPHICALLY WEIGHTED REGRESSION ANALYSIS TO LAKE-SEDIMENT DATA FROM AN AREA OF THE CANADIAN SHIELD IN SASKATCHEWAN AND ALBERTA

Nadia Yavorskaya [1], Stephen Amor[2]

[1] 450 Bonner Av., Winnipeg, MB, Canada. E-mail: gisofnadia@gisgeochem.com

[2] Geological Survey of Newfoundland and Labrador, P.O. Box 8700, St. John's, NL, Canada, A1A 1W5. Email: stephenamor@gov.nl.ca

Introduction

The National Geochemical Reconnaissance (NGR) program conducted by the Geological Survey of Canada (GSC) generated a regional geochemical database, comprising analyses of lake sediments and waters, covering a significant portion of the Canadian Shield. Analyses for more than 40 elements, including a 27-element INAA suite, have been available for more than two decades for NGR lake-sediment samples from some map sheets of Canada's National Topographic System (NTS). An extensive list of references to this work is provided by Amor (2014). Archived NGR samples from Labrador have more recently been re-analyzed by ICP-OES, after multi-acid digestion, for a broad suite of elements (McConnell & Finch, 2012). An examination of the expanded Labrador lake-sediment database has shown an element association that is believed to be related to the relative contents of inorganic clastic and organic material in the lake sediments (Amor, 2014). It was also shown that with the availability of the analyses of a large number of diagnostic elements, the clastic component can be modelled statistically and, more importantly, that deviations from it can be quantified in the form of residuals. Some of these residuals define known bedrock and surficial geological features, and provide possible indications of Ba and Li mineralization that were not apparent in the raw data (Amor, 2014).

A similar investigation of the lake-sediment data has been carried out in an area of northern Saskatchewan and north-eastern Alberta, to test whether similar inter-correlations of the clastic-sediment association exist in other parts of the Canadian Shield. The area covered by the dataset is the largest outside of Labrador, with almost continuous coverage, where samples were also subjected to additional INAA analysis (Figure 1).

Because the samples were not subjected to additional ICP-ES analysis, the elements suitable for regression treatment in the Alberta and Saskatchewan data are less numerous than those in the Labrador data. However, the data do include analyses for Ba, Br, F, Cs, Hf, Na, Rb and Sc, which appear to represent the clastic component in the Labrador study, and LOI, which represents its organic counterpart.

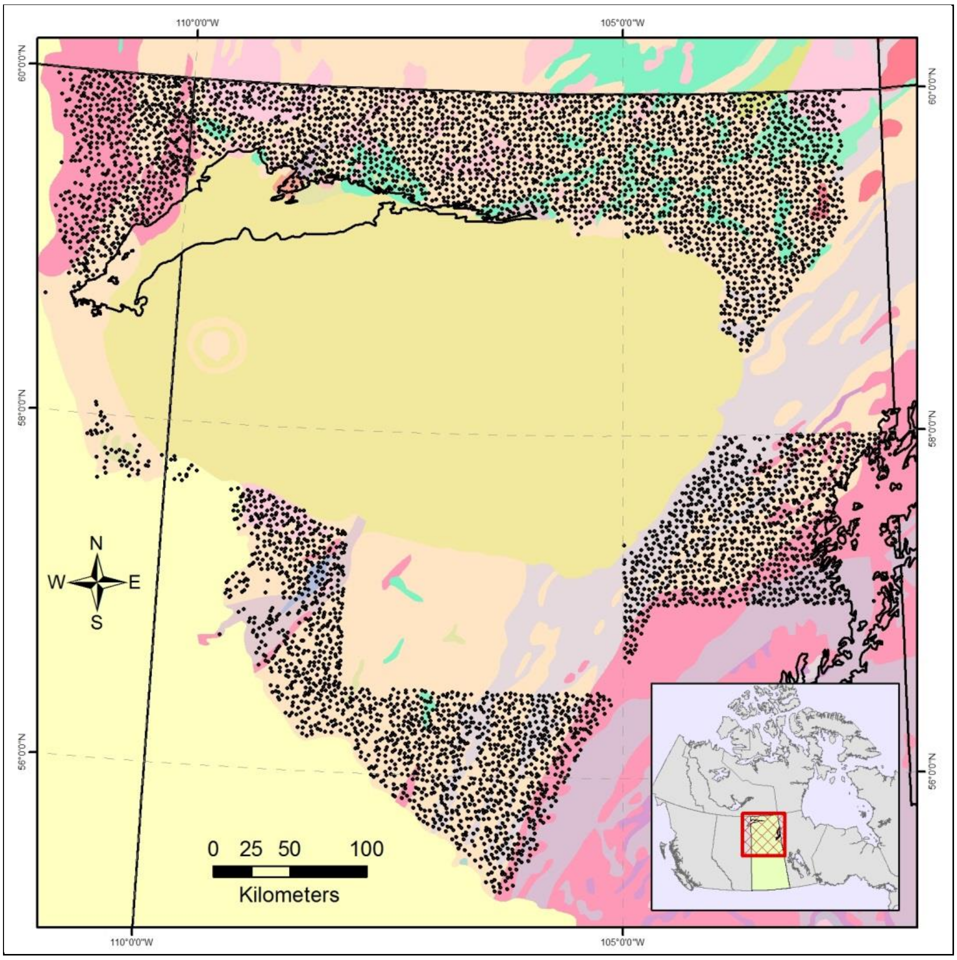

Figure 1. NGR lake-sediment samples analysed by INAA in Alberta and Saskatchewan

Methodology

In preparation of the data for multiple regression, values of Ba were subjected to a log-log transformation, Cs, F and Na to log transformation and Br, Hf, Rb and Sc to square-root transformation, in order to create data whose distributions were closer to normal. Transformation of the LOI values was unnecessary.

The content of potentially economic elements in the lake sediments was modelled with statistically significant elements from the clastic-organic association, together with a group of elements reflecting deposit types, by using multiple regression. The aim of this modelling is to remove the component of each sample's content that is attributable to the organic/clastic content and then relate what is left (the residuals) to geology and possible mineralization.

It is possible that a regression equation derived from samples, some of which may be separated by a distance of up to 500 km, may not be equally applicable over the entire area of coverage. This issue was addressed by dividing the data into a series of subareas (Figure 2) and creating a separate battery of equations derived by ordinary least-squares (OLS) regression.

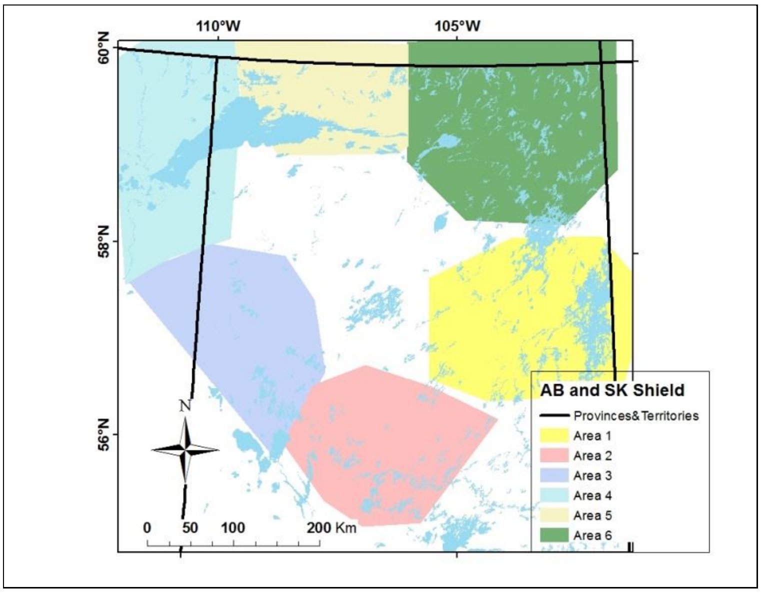

Figure 2: Subareas of the study area

The results of these regression exercises are summarized in the form of R^2 of regression in Table 1 and it appears that in the case of several elements, notably Ba, Br and Sc, better fits can be obtained from more locally restricted data sets.

Table 1: R^2 of multiple ordinary least squares (OLS) regression in five subareas of the study area, and the study area as a whole (in Area 3, several of the input variables are too severely censored for regression to be performed).

	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6	All
Ba	0.72	0.75		0.79	0.80	0.86	0.59
Br	0.55	0.49		0.36	0.36	0.45	0.37
Cs	0.70	0.64			0.58	0.69	0.60
F				0.60	0.57	0.60	0.61
Hf	0.78	0.79			0.74	0.77	0.68
LOI	0.68	0.76		0.71	0.77	0.65	0.70
Na	0.84	0.88		0.67	0.83	0.89	0.81
Rb	0.92	0.89			0.78	0.86	0.86
Sc	0.55	0.31		0.67	0.54	0.68	0.47

In order to avoid creating over-generalized regression equations, an option of Geographically Weighted Regression (GWR), available in ArcGIS 10.1, was applied. This computer-intensive method creates a regression equation (and regression residuals) for each sample, using a dataset consisting only of the sample's neighbours. Consequently, the results reflect local, rather than global, element associations and deviations from them. The procedure for choosing the elements for GWR analysis was as follows:

1. OLS was run on all of the data.
2. Elements with a Variance Inflation Factor (VIF) greater than 7.5, or marked in the output report file as insignificant, were excluded.
3. OLS was re-run on the elements remaining after Step 2.

After a few iterations, only significant elements, with VIF less than 7.5 remain. Table 2 shows the example of OLS results for transformed data with Ni as the dependent variable (Ni-models). The “Explanatory variables” column indicates the elements used in each test model.

Table 2. Summary of OLS Results with Ni as the dependent variable

OLS model	Adj.R^2	VIF>7.5	insignificant	Explanatory variables for tests
Ni model 1	0.75	Ce,La, Rb,Sm	Mn,V,Zn	As, Ba, Br, Cd, Ce, Co, Cr, Cs, Cu, Fe, Hf, Hg, La, LOI, Mn, Mo, Na, Rb, Sc, Sm, Th, U, V, Zn
Ni model 2	0.74	Ce,Rb, Sm	Rb,Th	As, Ba, Br, Ce, Co, Cr, Cs, Cu, Fe, Hf, LOI, Mo, Na, Rb, Sc, Sm, Th, U
Ni model 3	0.73	_	Sc,U	As, Br, Co, Cr, Cs, Cu, Fe, Hf, LOI, Mo, Na, Sc, U
Ni model 4	0.73	_	_	As, Br, Co, Cr, Cs, Cu, Fe, Hf, LOI, Mo, Na

.

Geographically Weighted Regression was subsequently applied to those elements that were left after the OLS runs. Results for each run are summarized, in the form of element residuals and C-coefficients, in associated shape and DBF files. The C-coefficients define the relationship between each explanatory variable and the dependent variable.

Results and discussion

Clastic-organic model

It was determined that there is a strong intercorrelation between Rb-Na-Hf-Ba-Cs-Sc in the transformed analytical data. This result is similar to that of the Labrador study, which determined Ba, F, Cs, Hf, Na, Rb and Sc to be related to the content of inorganic clastic material in the lake sediments while an association of LOI and Br was interpreted to represent the chemically dispersed (mostly organic) content of the sediments (Amor, 2014).

In order to define which suite of elements provided the best model for the clastic and organic components in the Saskatchewan data, several OLS analyses were conducted with Ba as dependent variable. Table 2 shows the results. The

Adjusted R-Squared (Adj.R^2) and the Akaike's Information Criterion (AICc) are the measures of the performance of each model. The Adjusted R-Squared value indicates how much of the variation in Ba has been explained by the model, while the best model is the one with the lowest AICc value. The lowest AICc value and the highest Adj R^2 number were returned for Ba-model 4 (Ba/ Cs/ Na/ Rb/ Sc/ LOI). Cs can be omitted from the model as it does not show strong relationship in the model.

The final suite that is responsible for clastic-organic components, as determined by OLS analysis, is the Ba/ Na/ Rb/ Sc/ LOI suite. These elements can then be included in a GWR analysis to compensate the influence of clastic and organic components.

Table 3. Summary of OLS Results with Ba as the dependent variable

OLS model	AICc	Adj. R^2	High coefficients	VIF>7.5	insignificant	Explanatory variables
Ba-model 1	10902	0.56	Fe, Mn, Na, Rb, Sc	Rb, Ce, Co ,Cs, Fe, Hf, La, LOI, Sc, Sm, Th	Ce, La, Sm	All elements
Ba-model 2	11318	0.63	Na, Rb, Sc	Ce, Cr, Cs, Hf, La, Th	Ce, La, Sm	All elements except Fe, Mn
Ba-model 3	7162	0.73	Na, Rb, Sc	-	Hf	Cs, Na, Hf, Rb, Sc, LOI
Ba-model 4	7160	0.73	Na, Rb, Sc	-	-	Cs, Na, Rb, Sc, LOI

Economic element model

The Ni regression model is discussed as an example of the Geographically Weighted Regression analysis. The suitable predictor elements for the Ni model were determined as described in the Methodology section and are shown in the last row of Table 2 (Ni-model 4). This suite includes some of the elements that are strongly correlated with Ni (Co, Cr, Cu, Cs) as well as the “clastic” component determined from the GWR analysis (LOI, Na, Rb, Sc).

The calculated Ni residuals have been evaluated in terms of their relationship to known Ni occurrences and mafic-ultramafic hosted mineralization (Fig. 4). GWR residuals for the Ni-model show a closer spatial relationship to the known nickel occurrences than do the raw Ni values. The element coefficients from a table in a GWR shapefile show how the relationship between each explanatory variable and the dependent variable changes across the study area. The “Hot Spot Analysis with Rendering” tool in ArcGIS 10.1 allows the identification of areas where the element coefficients exhibit a random spatial pattern. The results of applying this tool to the coefficients of Co in Model 5 indicate that the latter are not randomly distributed.

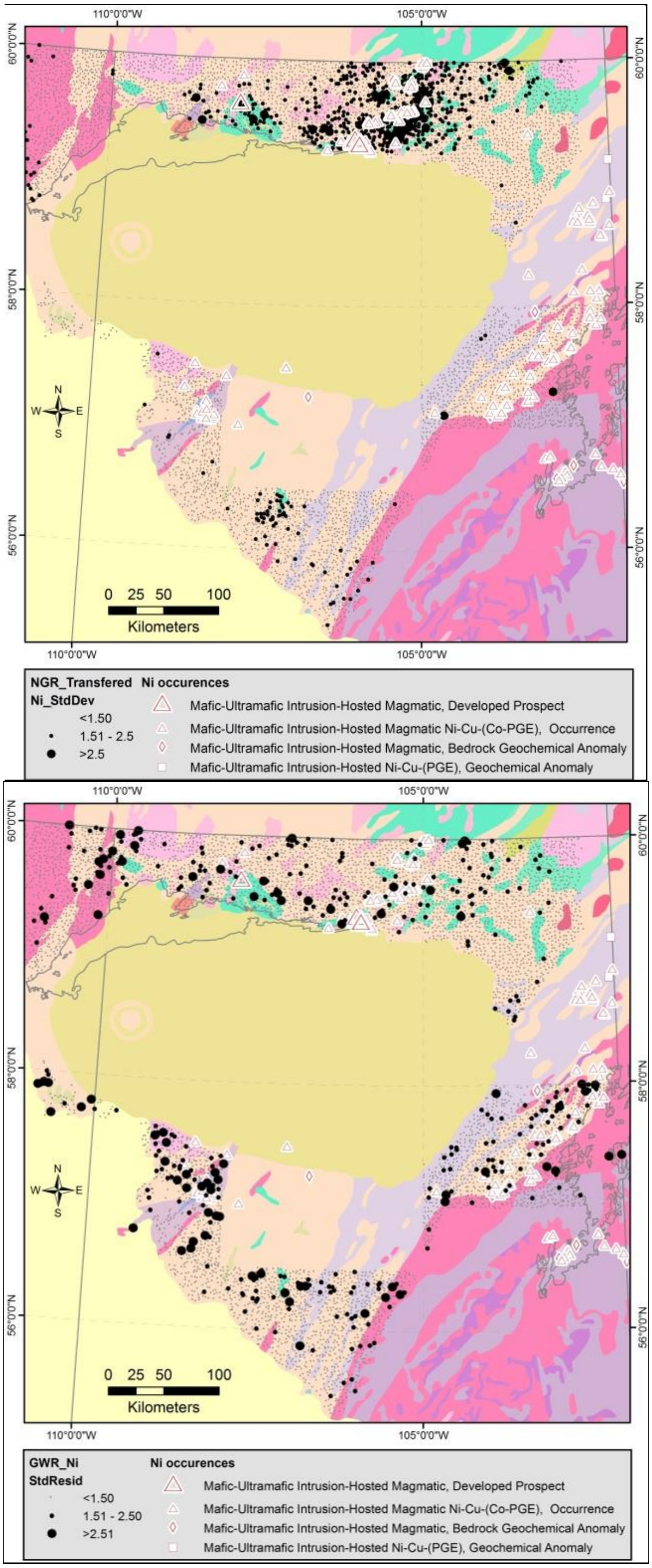

Figure 4. Distribution of Ni in transformed analytical data (top) and GWR residuals for Ni model 5 (bottom).

The locations in yellow in Figure 5 show areas where Co is a strong predictor for Ni in this model, which suggests that those areas are of interest for Ni-Co mineralization.

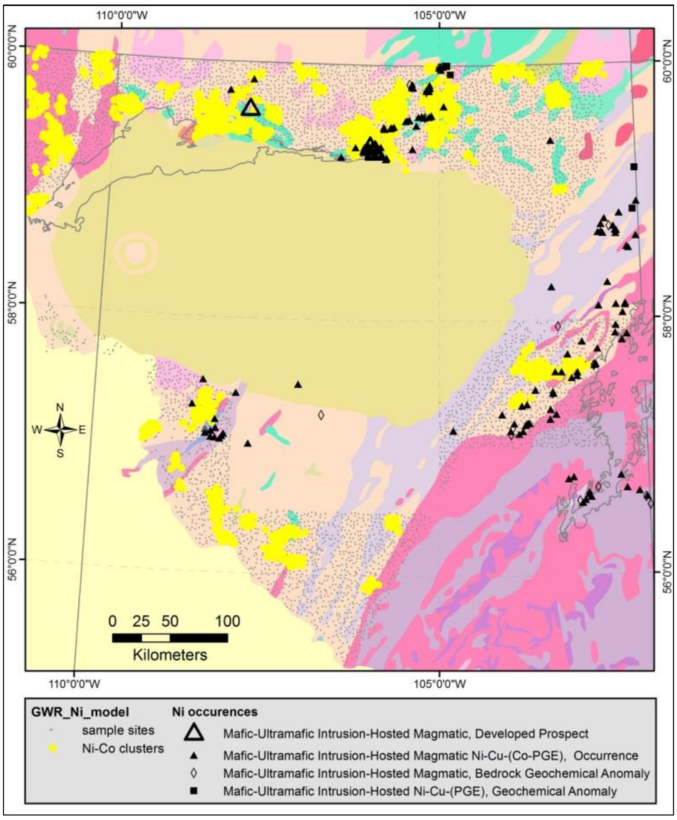

Figure 5. Results of the Hot Spot Analysis for the Co coefficients in Model 5

In many cases GWR residuals show a closer spatial relationship to mineral occurrences than do raw values of the same elements; and the GWR method has potential advantages in locating previously unknown Ni mineralization.

References

AMOR, S.D. 2014. Geochemical quantification of the clastic component of Labrador lake sediments, and applications to exploration. Government of Newfoundland and Labrador, Department of Natural Resources, Geological Survey, Open File LAB/1625.

MCCONNELL, J.W. & FINCH, C. 2012. New ICP-ES geochemical data for regional Labrador lake-sediment and lake-water surveys. Government of Newfoundland and Labrador, Department of Natural Resources, Geological Survey, Open File LAB/1602.