# Some Applications of Geographically Weighted Regression to Regional Geochemical Data

Stephen Amor

January 17, 2019

# TOPICS TO BE COVERED

1. Ordinary and Geographically Weighted Regression
2. Earlier Studies
3. Datasets in this study
4. Results:

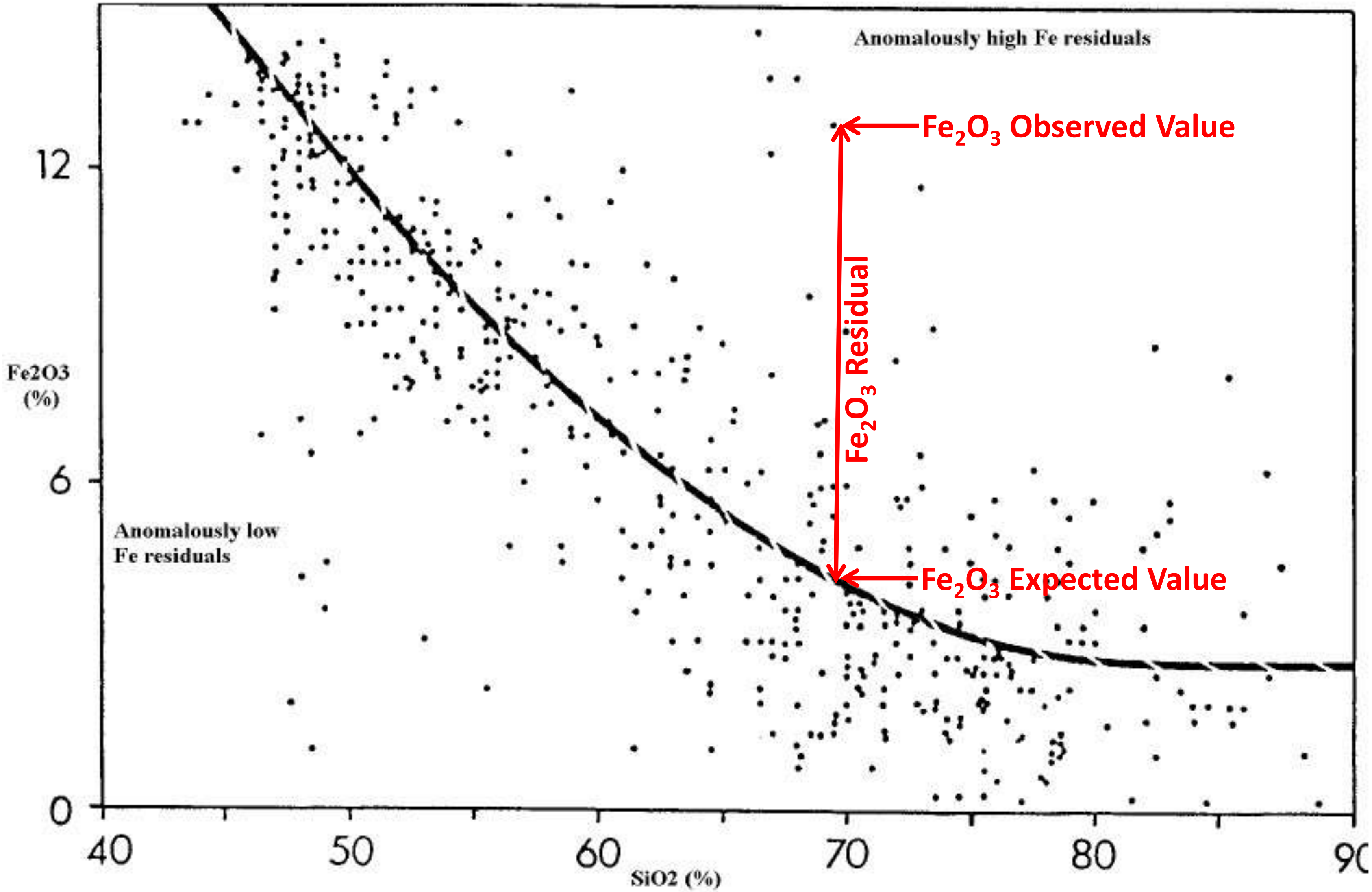
   Local $R^2$

   Residuals
5. Issues still to be addressed

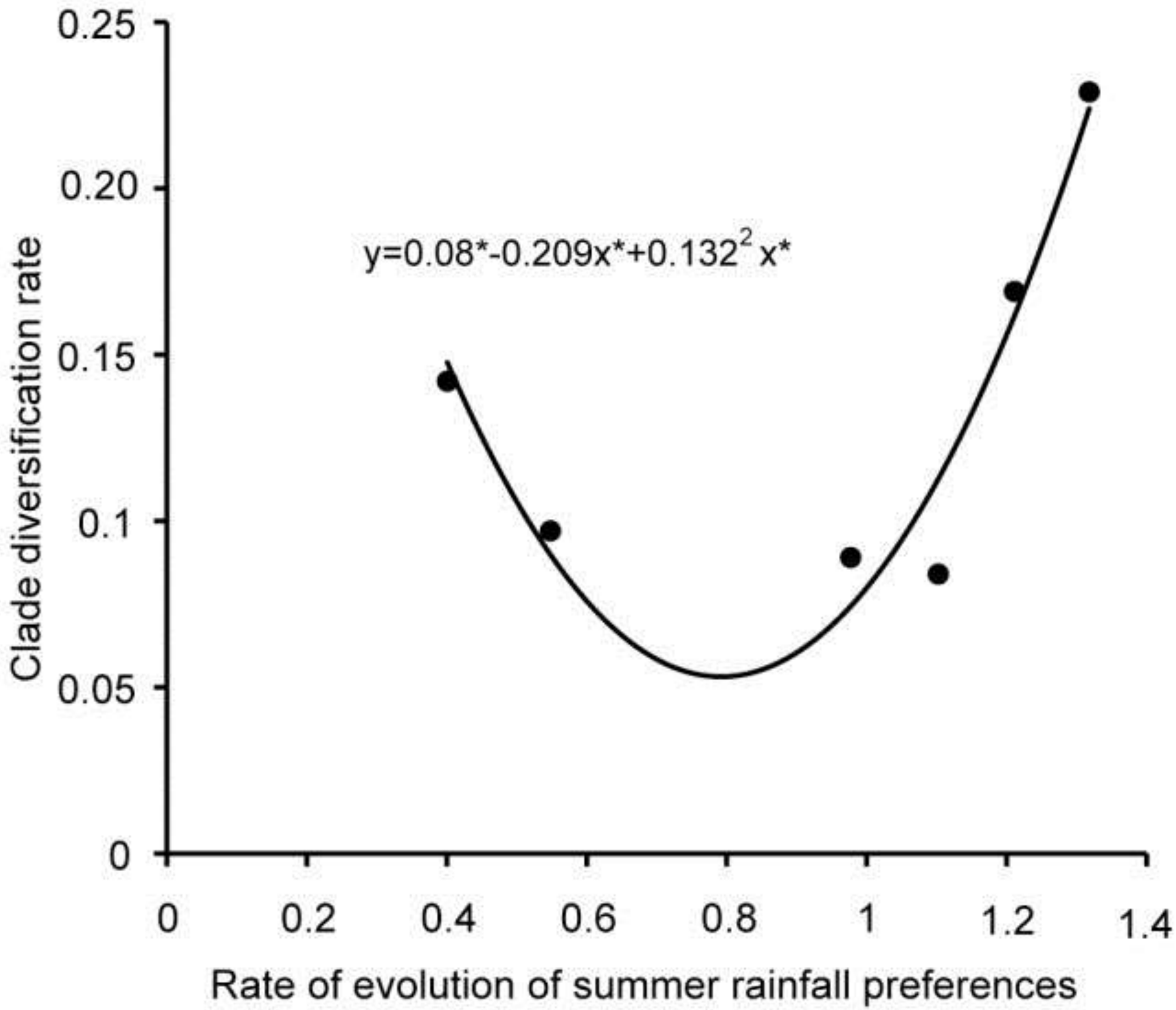
# Ordinary Least Squares (OLS) Regression

## Explanation of Residuals

# Simple (curvilinear) regression: Volcanic Rocks, NW Ontario

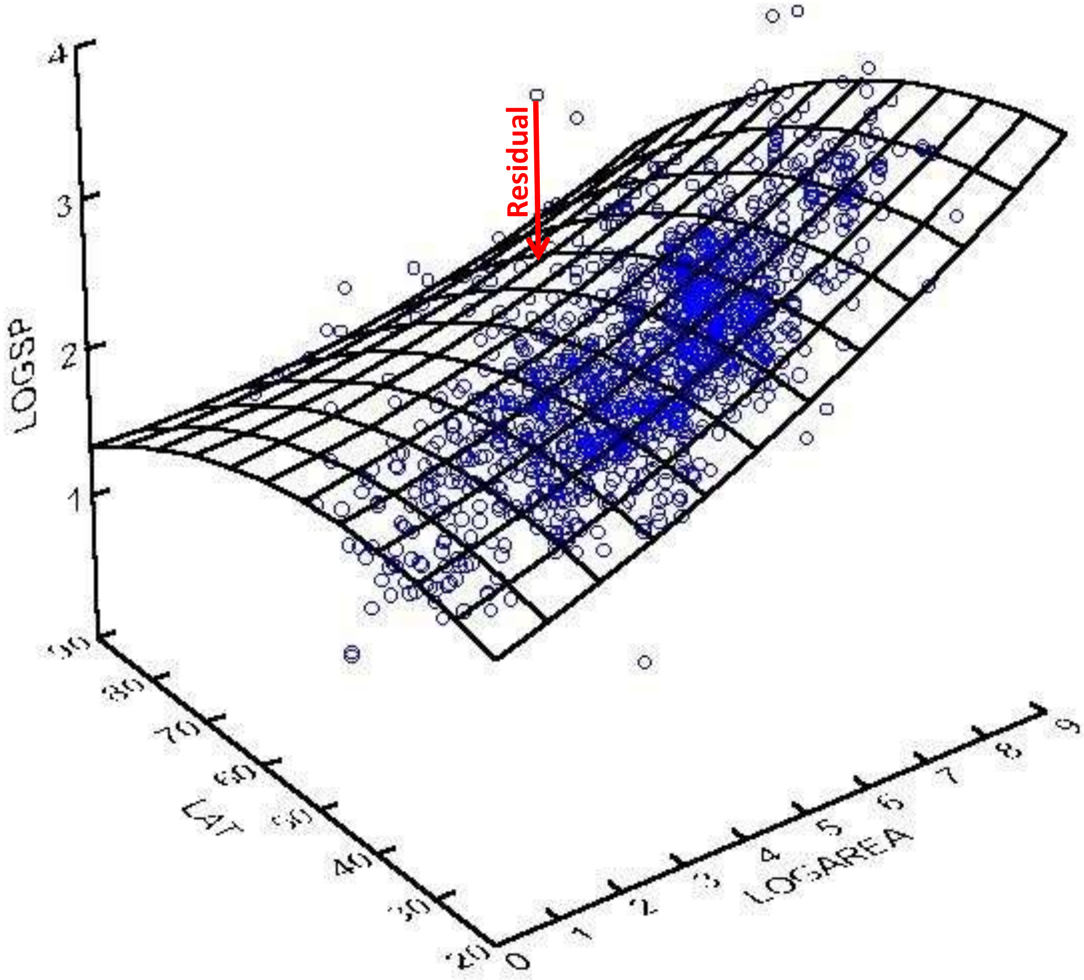


# Example of a quadratic regression equation

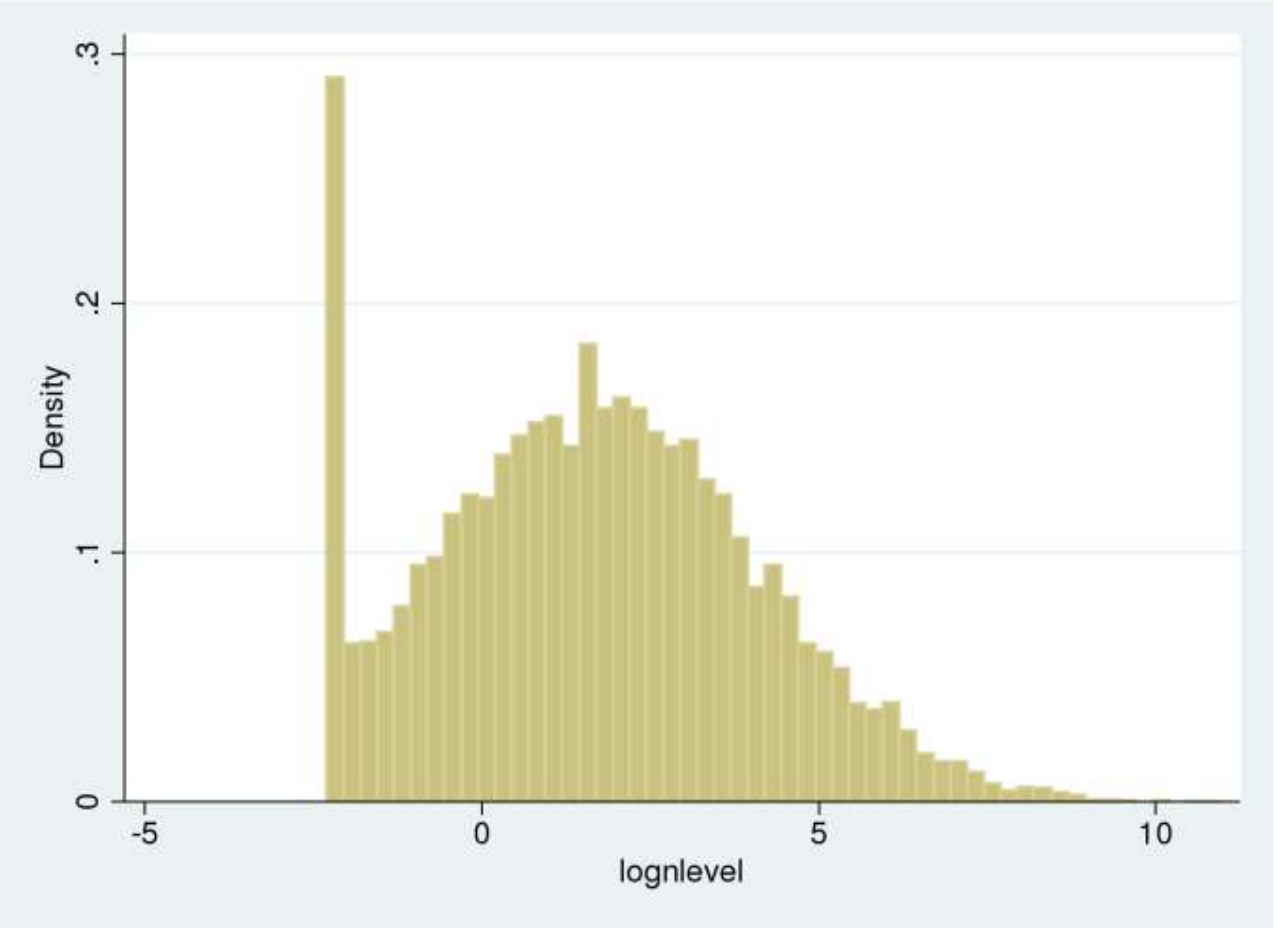


# Multiple regression

**Fitting a surface to bivariate data**



## Censoring / truncation of data – an important consideration




Analyses below the detection limit can be assigned a value of half the detection limit if there are not too many of them. How if more than 10% of the analyses of an element are below the detection limit, that element is not suitable for regression analysis and should be omitted.

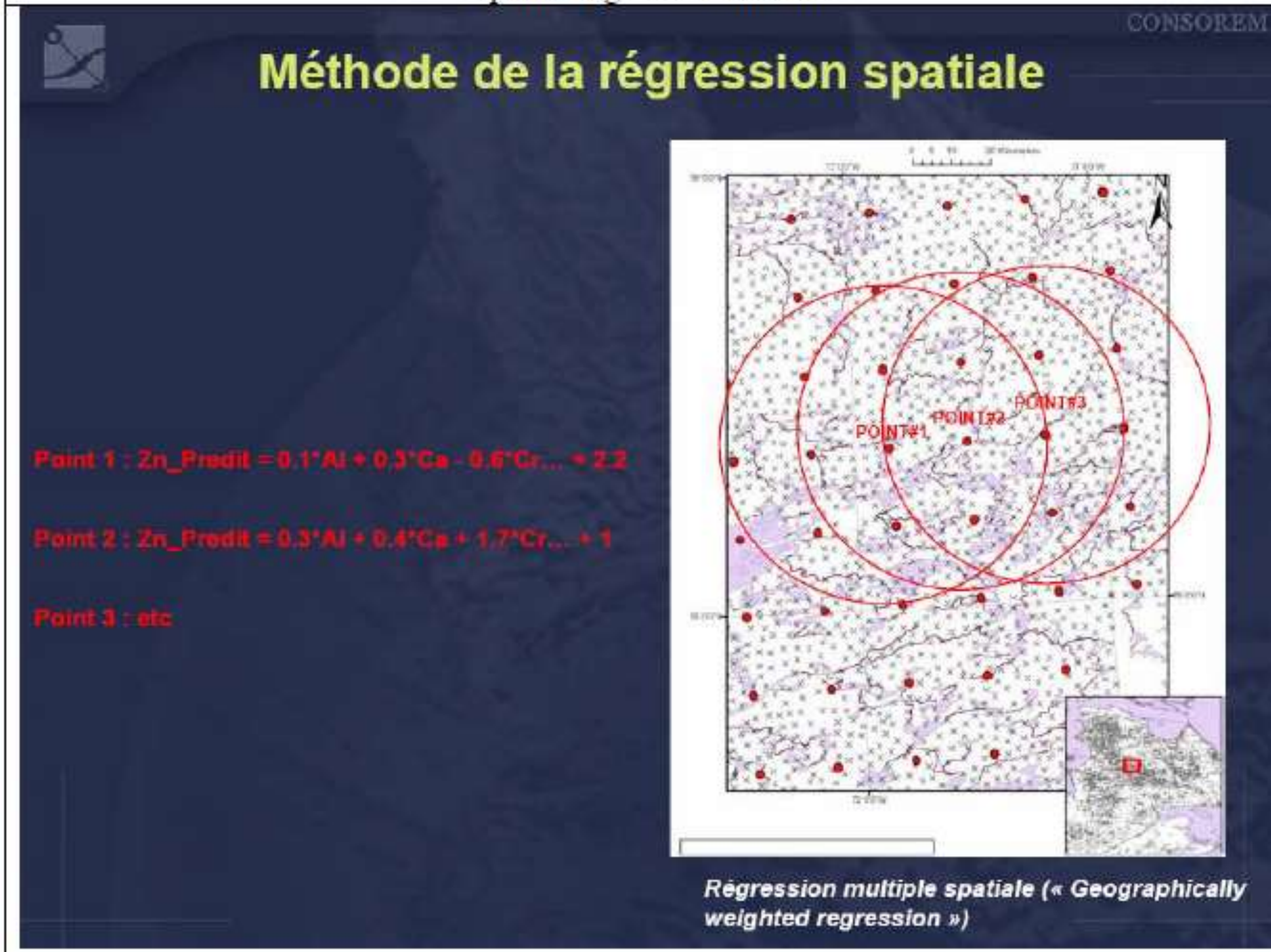
All input variables should be approximately normally distributed and log-transformation of geochemical variables .is often necessary

# Geographically Weighted Regression (GWR)

**A new equation for each point**

**(only suitable for large, areally extensive geochemical data sets)**

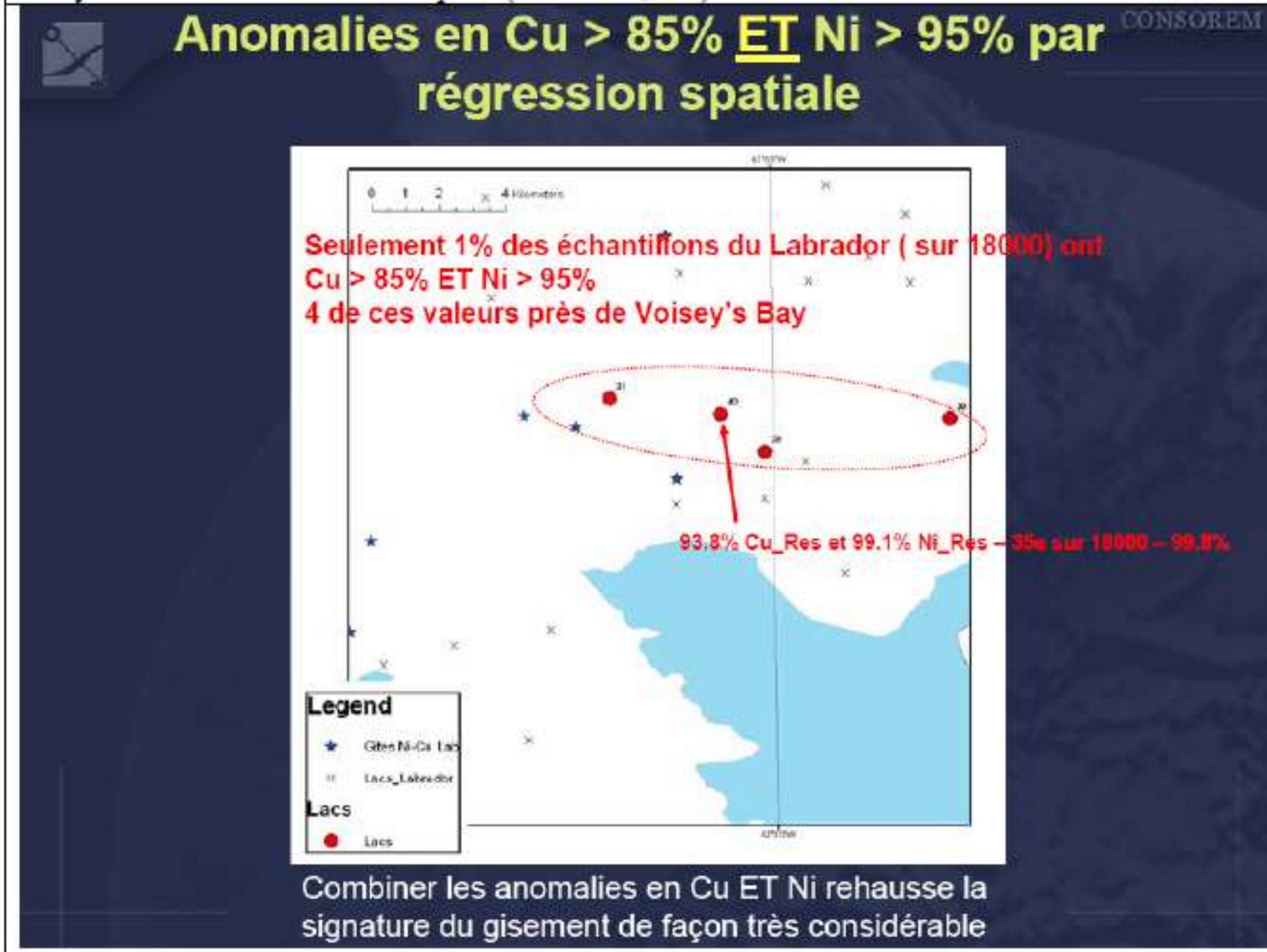
## Spatial regression method




**The “search radius” within which samples are selected for each regression can be varied**

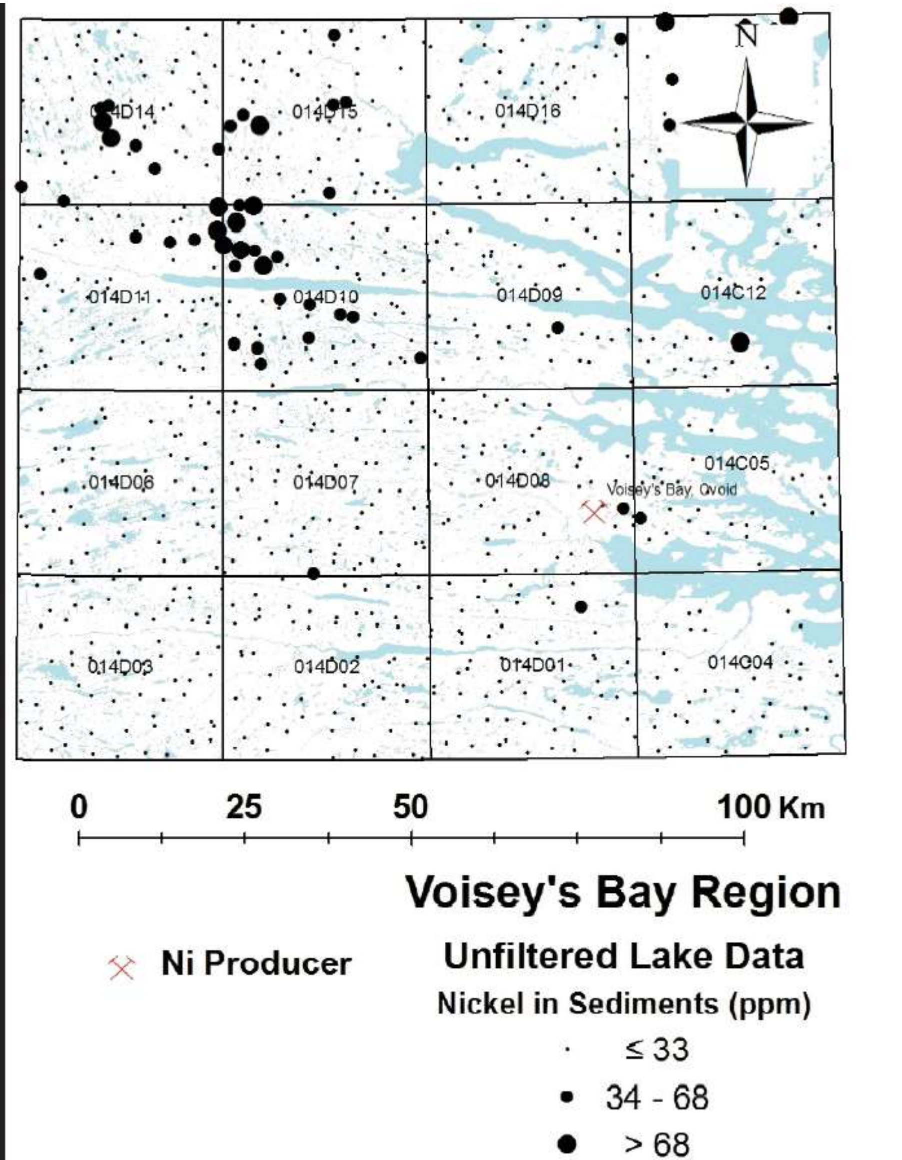
Anomalies in Cu residuals (>85-percentile) <u>and</u> Ni residuals (> 95-percentile) by spatial regression

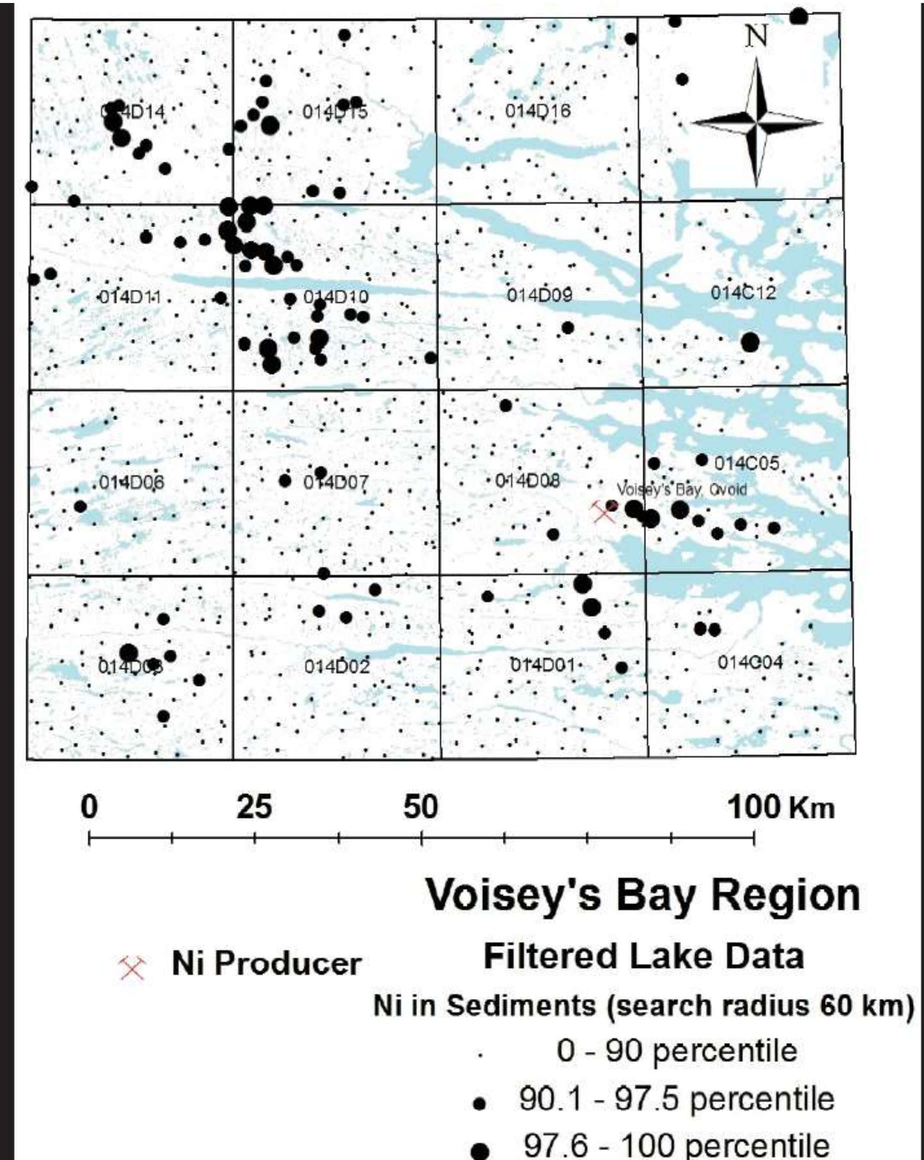
Only 1% of the Labrador samples (out of 18,000) have Cu > 85% AND Ni > 95%




Combining Cu AND Ni anomalies enhances the anomalous signature of the deposit to a very considerable extent

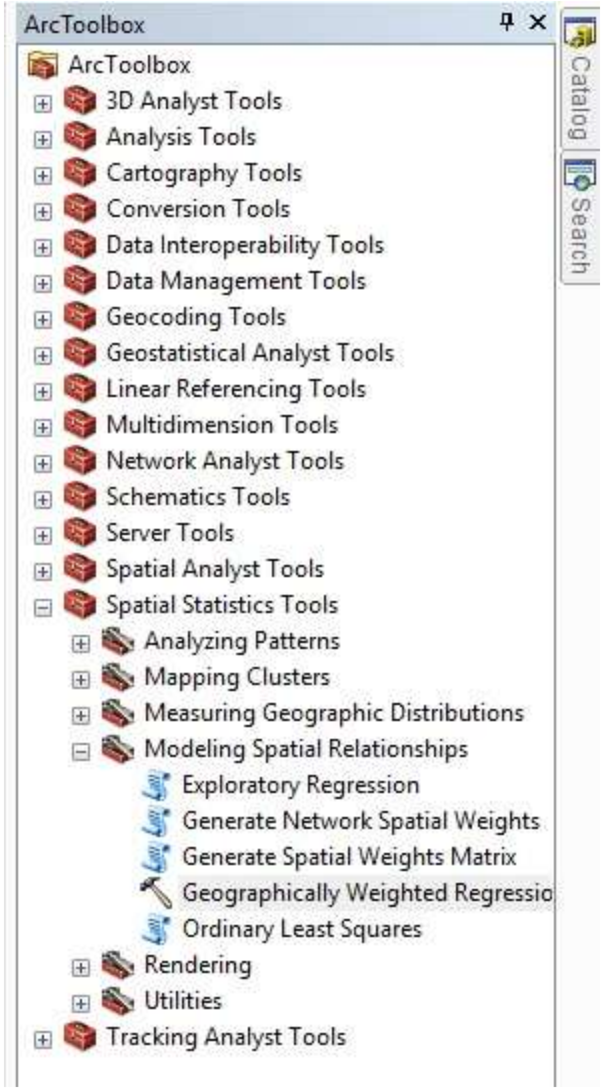
**It's claimed that GWR enabled the identification of an anomaly in lake sediments associated with Voisey's Bay, although the input parameters are not known …**



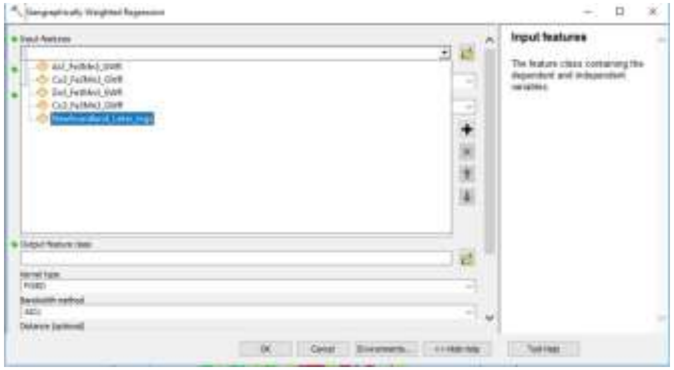




**… other methods, such as "filtering" the data, can also enhance the lake-sediment signal at Voisey's Bay; however they also identify other strong anomalies that are apparently (despite extensive exploration) not associated with mineralization.**

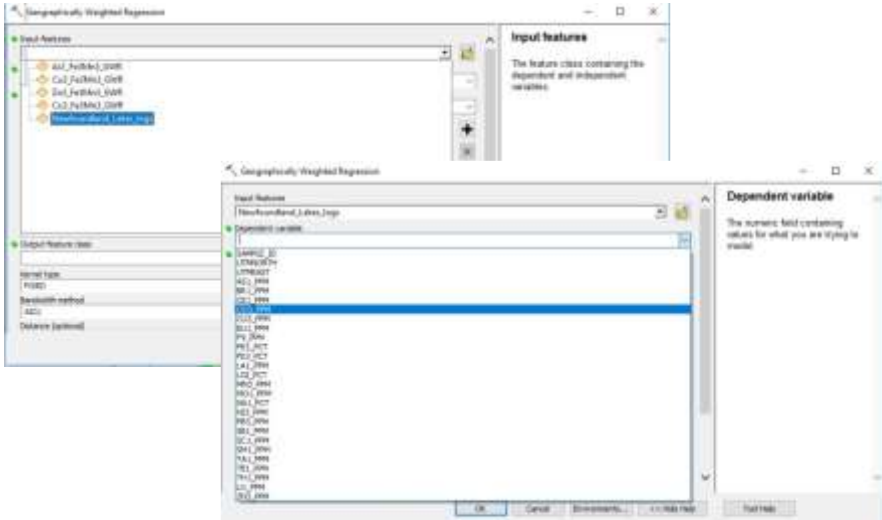
# How it’s done



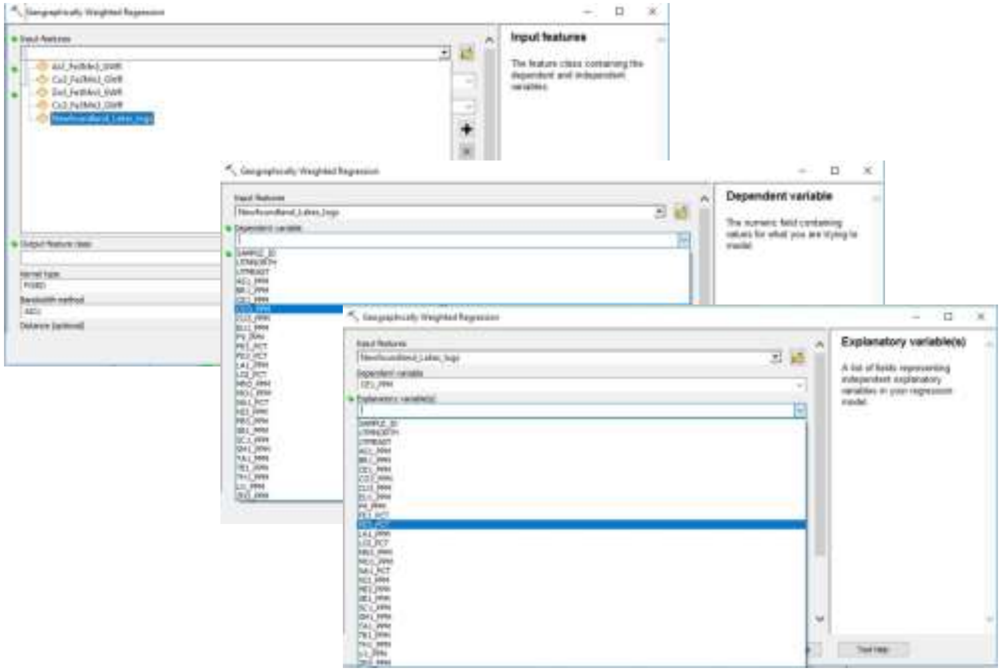

**GWR is an option within the "Modeling Spatial Relationships" module of the ArcGIS toolbox.**



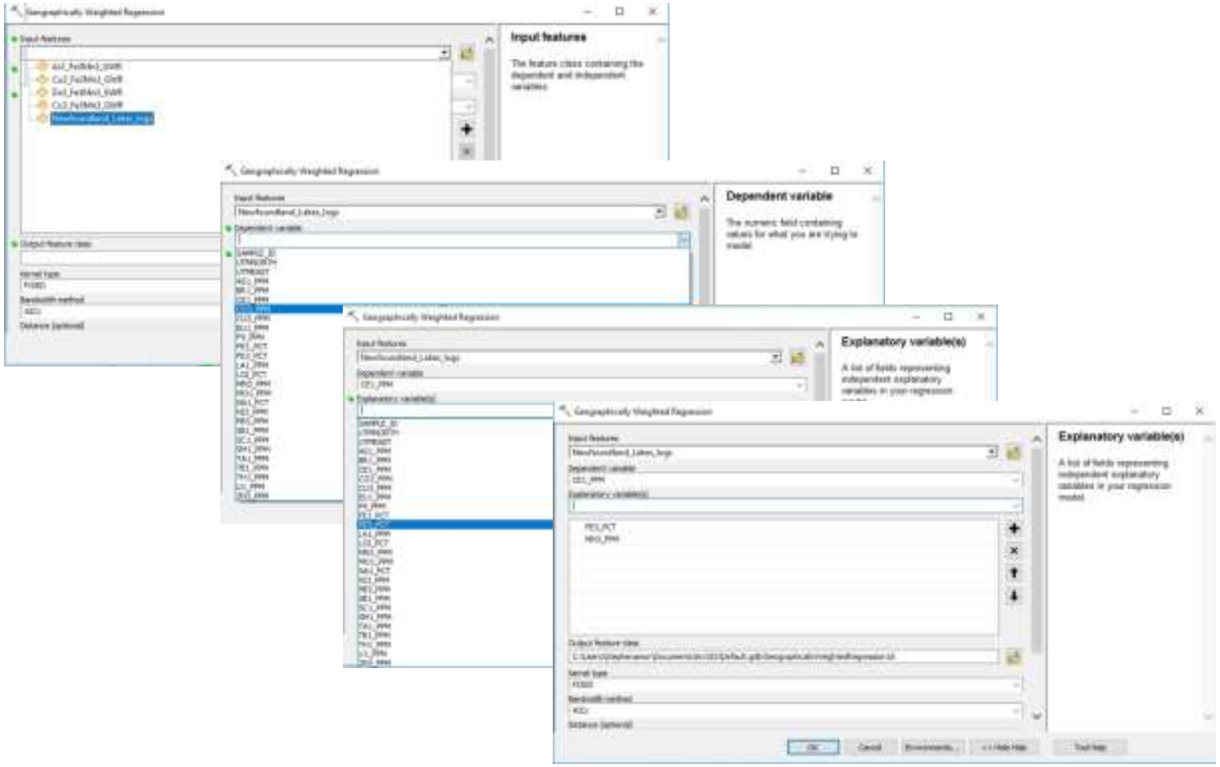

"Input features" refers to the data from which the regressions will be calculated



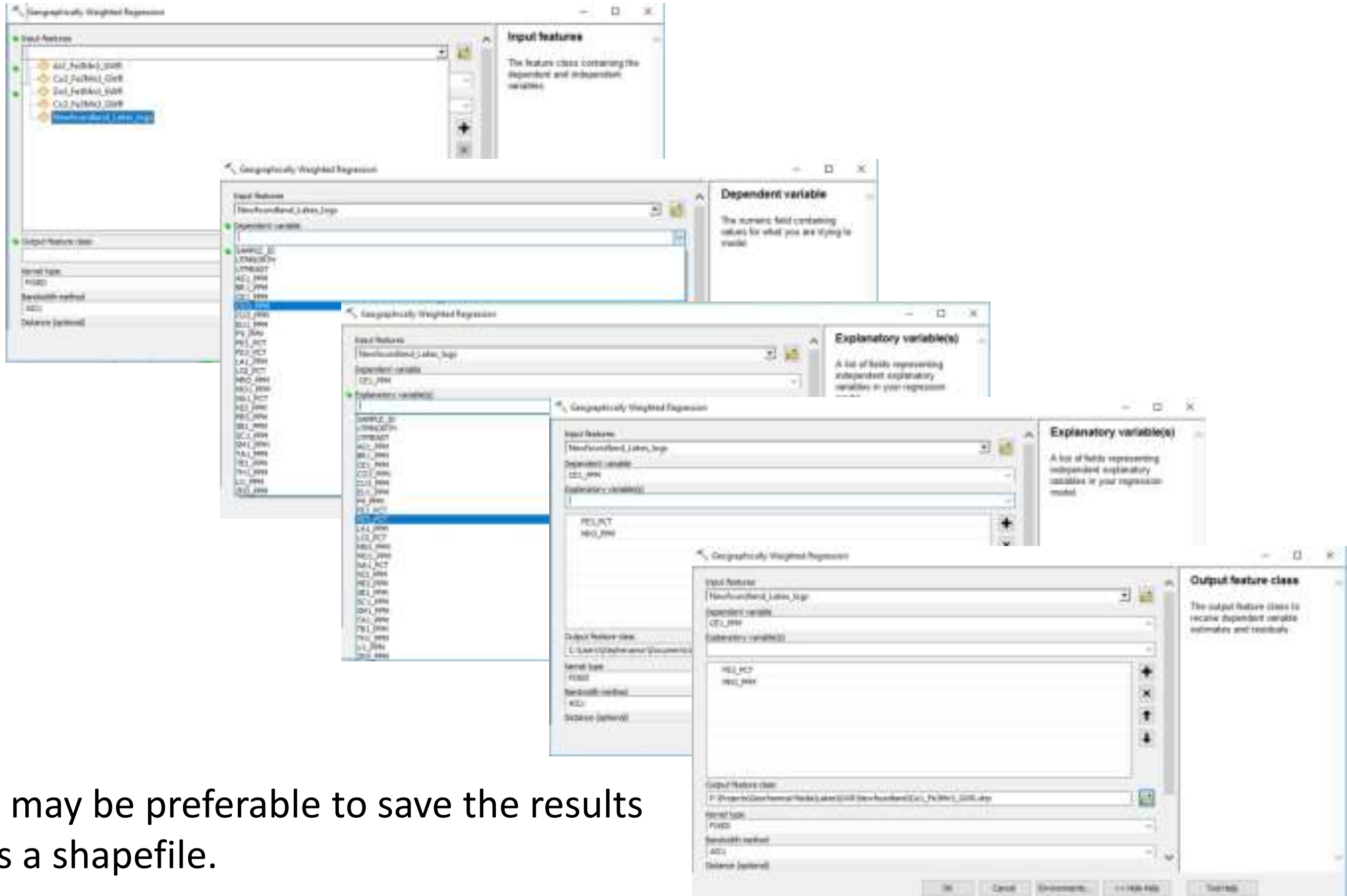

The “dependent variable” is the variable to be predicted (on the left in the regression equation)



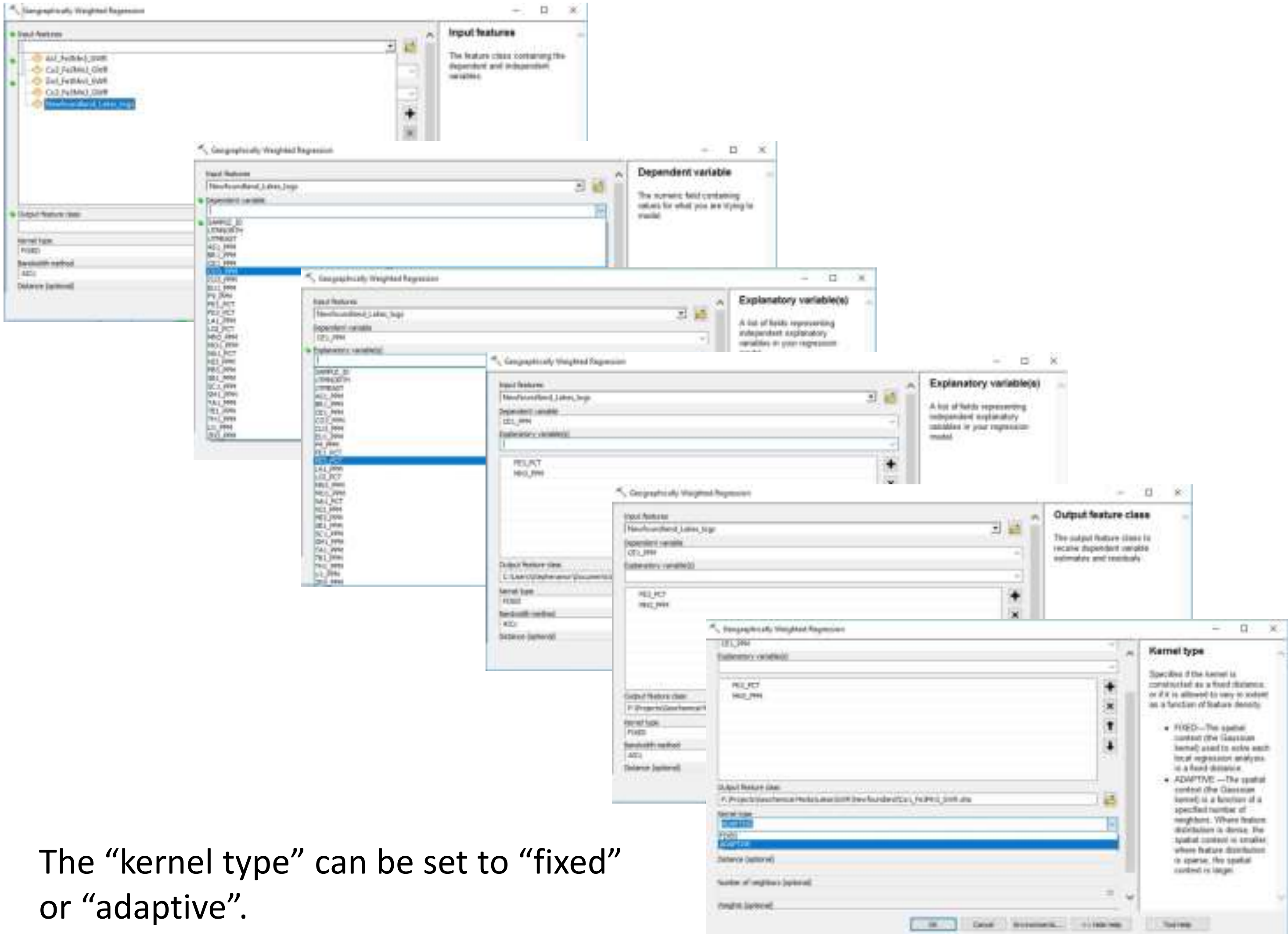

The “explanatory variables” are used to predict values of the dependent variable. Also referred to as “independent” or “predictor variables”



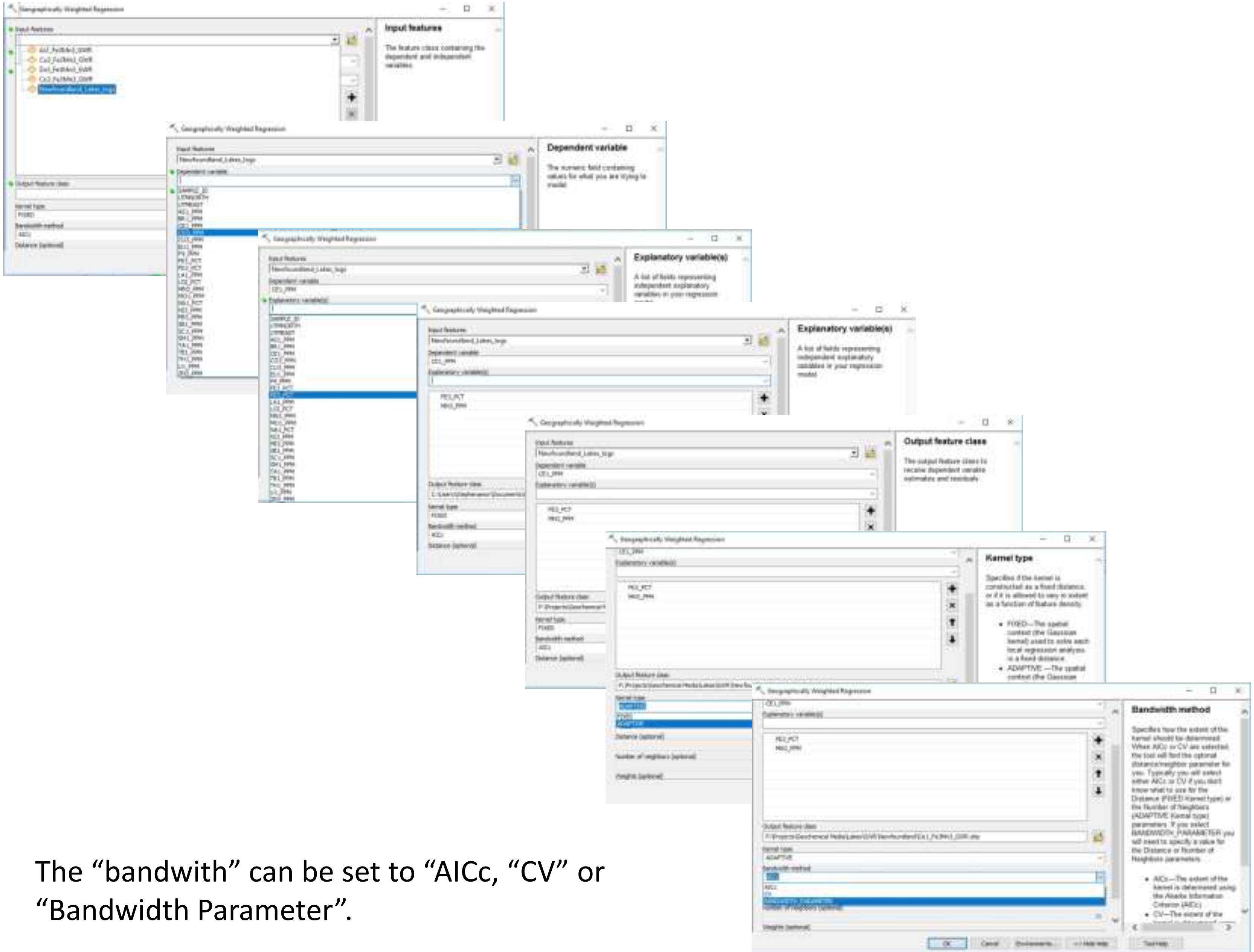

The “output feature class” holds the results of the GWR. The default is to add it to the ArcGIS GDB database.



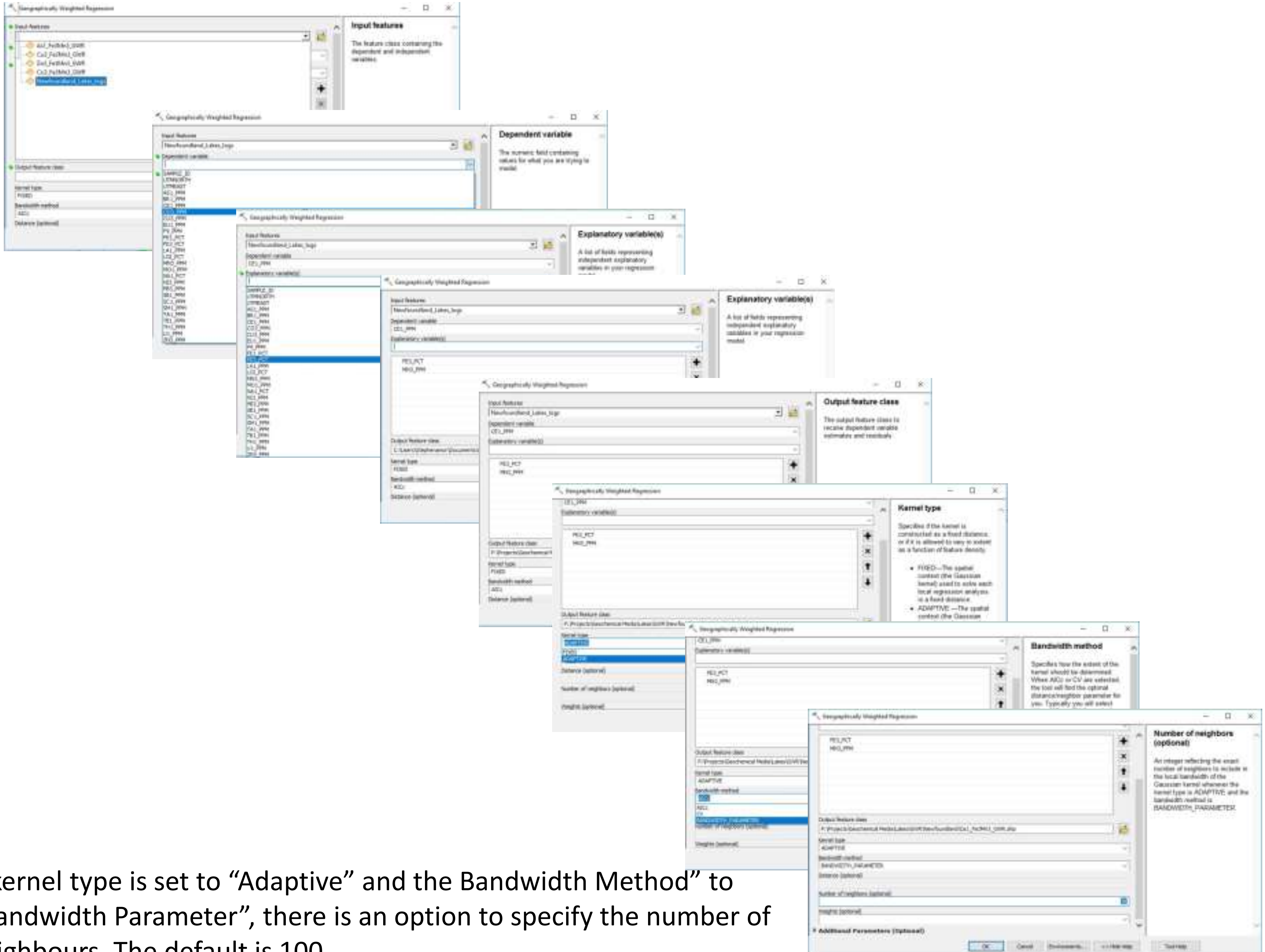

It may be preferable to save the results as a shapefile.



The “kernel type” can be set to “fixed” or “adaptive”.



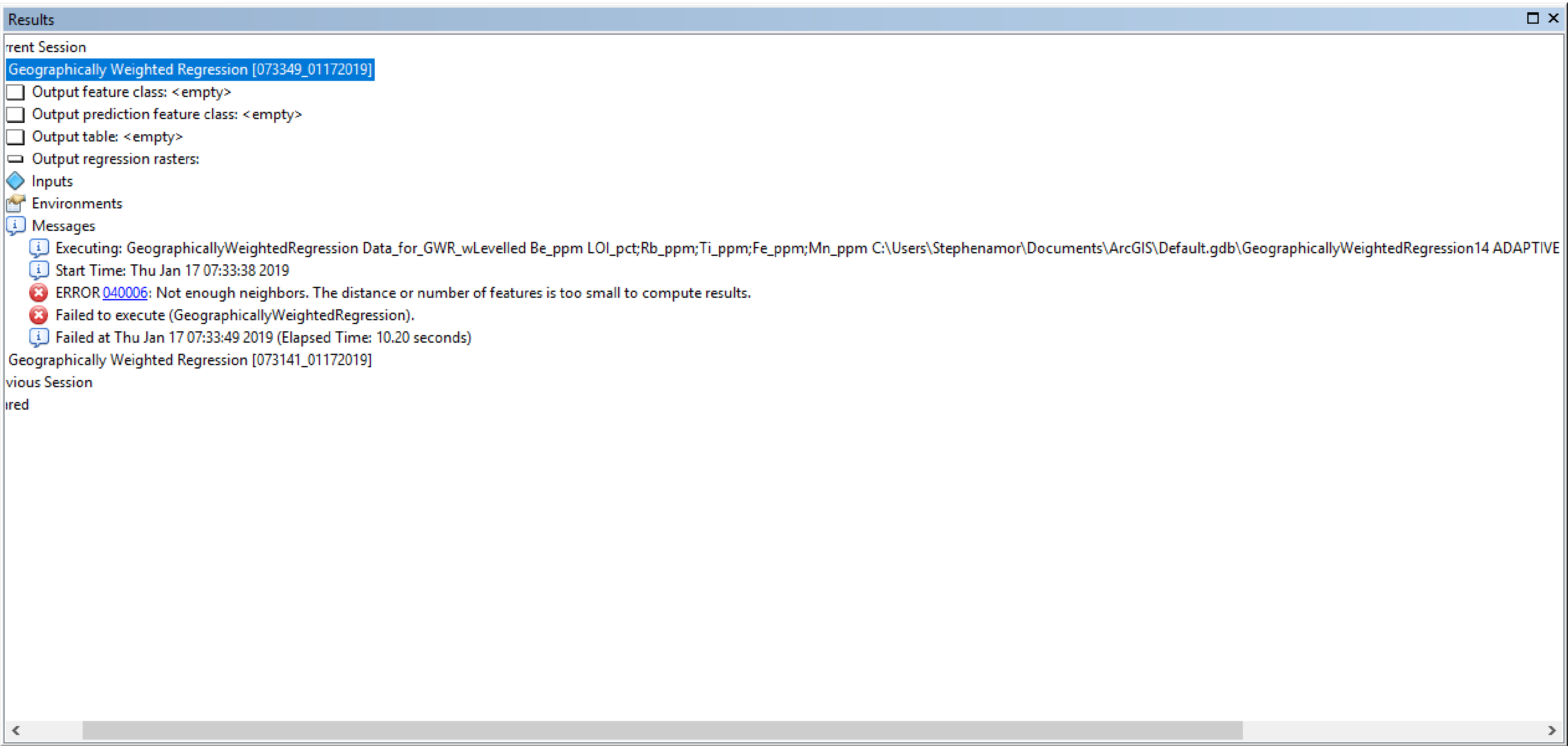

The “bandwith” can be set to “AICc, “CV” or “Bandwidth Parameter”.



If kernel type is set to “Adaptive” and the Bandwidth Method” to “Bandwidth Parameter”, there is an option to specify the number of neighbours. The default is 100

# A common error message

**The message that “the distance or number of features is too small to compute results” is misleading as it is normally given when there are more than 3 or 4 explanatory variables**



# GWS Output

**$R^2$ (strength of relationship)**

**Original dependent variable (log-transformed)**

**Standardized Residual**
**(this is the default variable that ArcGIS plots, but any of these can be plotted)**

LOI, Ti, Rb

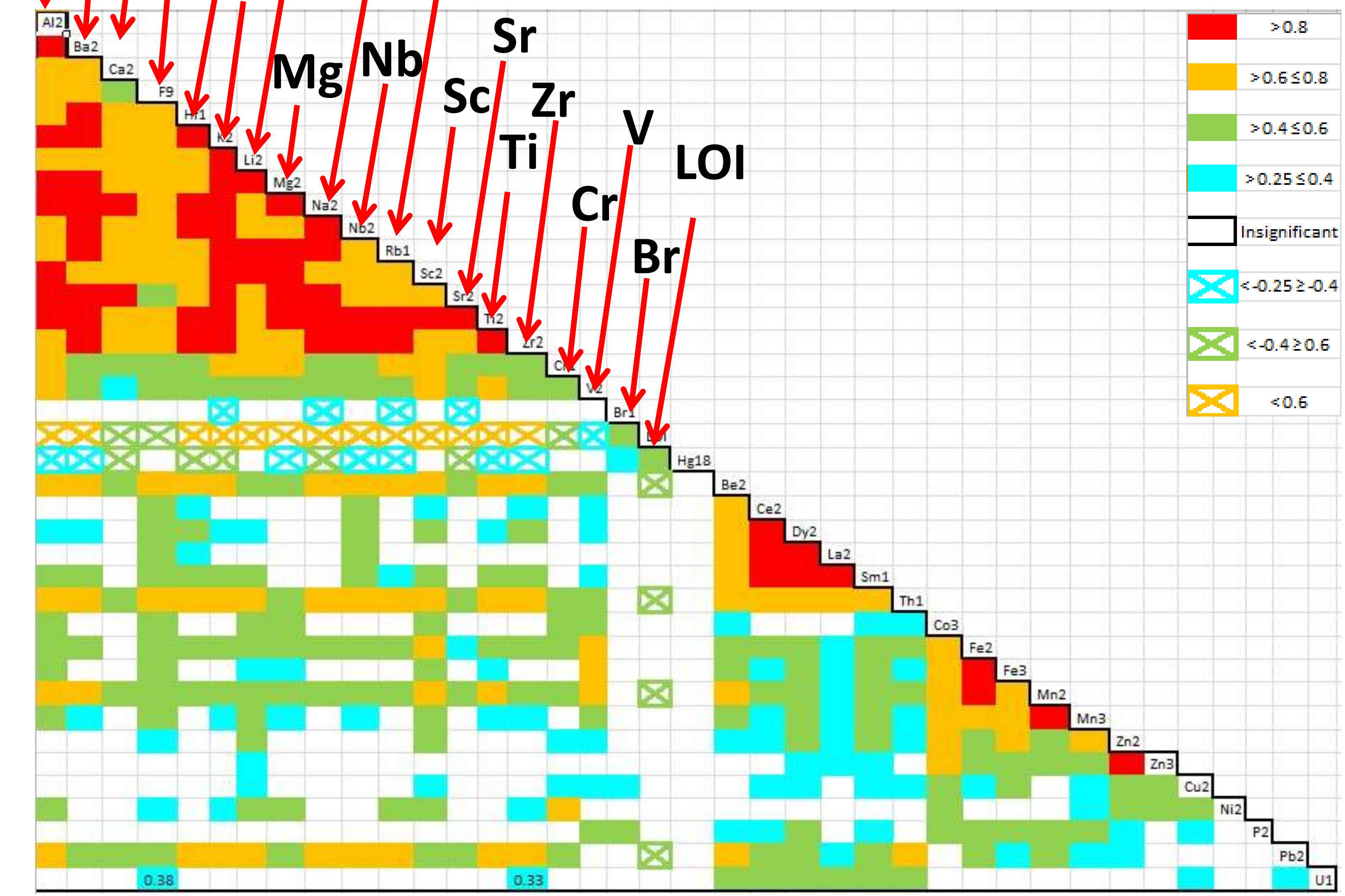
| | | | | | | Intercept and Coefficients | | | | Standard Error | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | LOI | Rb | Ti | | Total | Coefficient | LOI | Rb | Ti | | |
| ID | Shape * | Observed | Cond | LocalR2 | Predicted | Intercept | C1_LOI_pct | C2_Rb_ppm | C3_Ti_ppm | Residual | StdError | StdErr_Int | StdErrC1_L | StdErrC2_R | StdErrC3_T | StdResid | Source_ID |
| 0 | Point | -0.49485 | 51.812899 | 0.488778 | -0.607837 | -0.585004 | 0.423136 | 0.639292 | -0.314148 | 0.112987 | 0.265985 | 0.742864 | 0.181115 | 0.258928 | 0.283427 | 0.424786 | 0 |
| 1 | Point | -0.455932 | 52.766073 | 0.504439 | -0.541344 | -0.569612 | 0.420539 | 0.659394 | -0.32078 | 0.085412 | 0.262701 | 0.739648 | 0.179573 | 0.25417 | 0.282047 | 0.325131 | 1 |
| 2 | Point | -0.537602 | 52.778522 | 0.502588 | -0.572779 | -0.572188 | 0.423841 | 0.654692 | -0.321584 | 0.035177 | 0.264708 | 0.738701 | 0.180078 | 0.252067 | 0.281041 | 0.132891 | 2 |
| 3 | Point | -0.537602 | 53.05045 | 0.505692 | -0.612148 | -0.565363 | 0.423986 | 0.65926 | -0.325182 | 0.074546 | 0.263183 | 0.741747 | 0.180872 | 0.252002 | 0.281851 | 0.283246 | 3 |
| 4 | Point | -0.468521 | 56.099266 | 0.535516 | -0.518703 | -0.561457 | 0.409772 | 0.698894 | -0.320278 | 0.050182 | 0.26348 | 0.755161 | 0.178796 | 0.256081 | 0.289456 | 0.190457 | 4 |
| 5 | Point | -0.585027 | 56.682575 | 0.530496 | -0.543186 | -0.563565 | 0.410191 | 0.692387 | -0.319434 | -0.041841 | 0.258637 | 0.76374 | 0.180615 | 0.255718 | 0.292131 | -0.161775 | 5 |
| 6 | Point | -0.585027 | 57.811299 | 0.541423 | -0.639398 | -0.594302 | 0.410724 | 0.700876 | -0.306291 | 0.054371 | 0.253129 | 0.76588 | 0.178845 | 0.258042 | 0.294131 | 0.214797 | 6 |

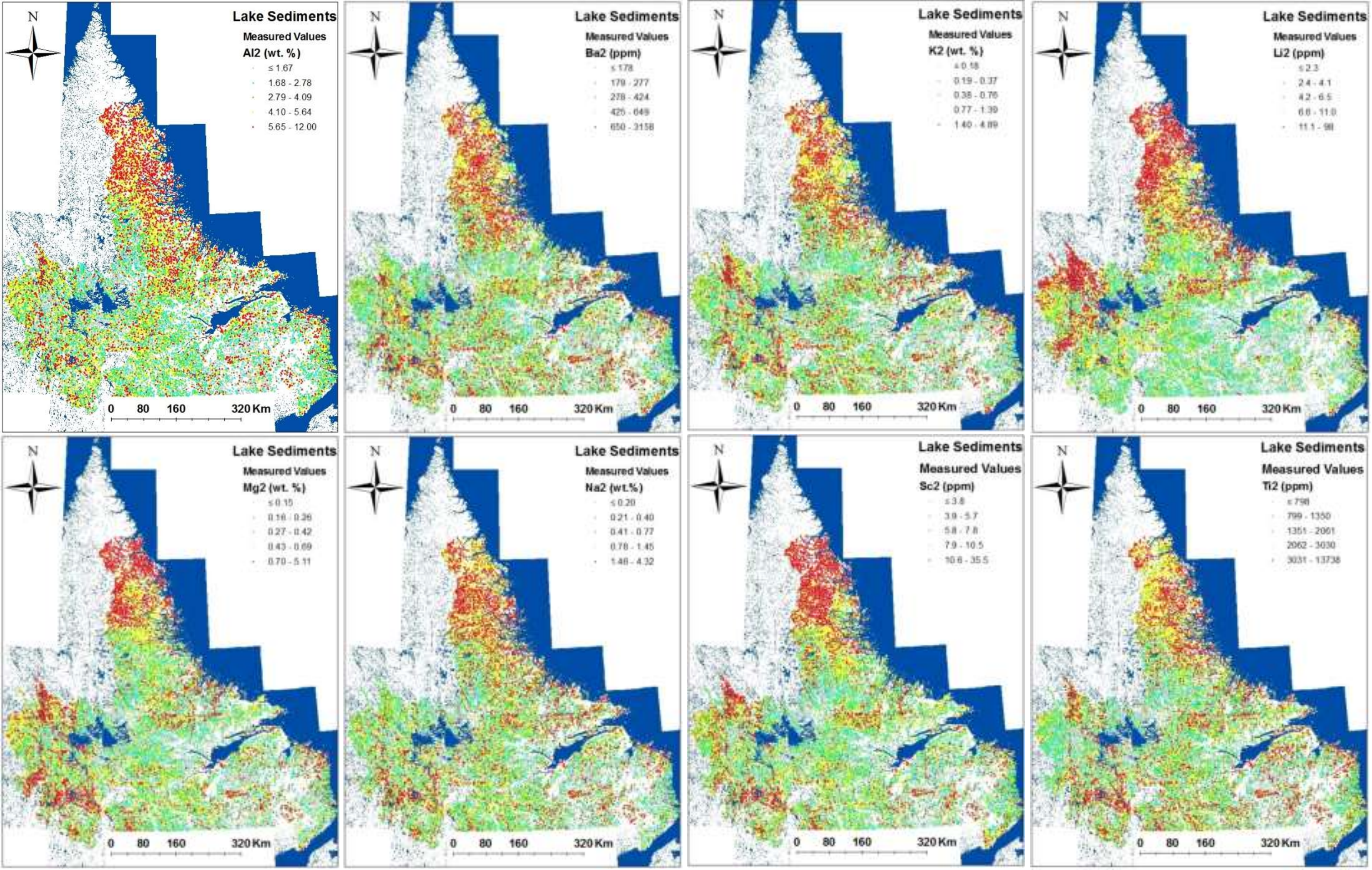
**Dependent variable predicted from regression equation(log-transformed)**

# Earlier (pre-GWR) studies

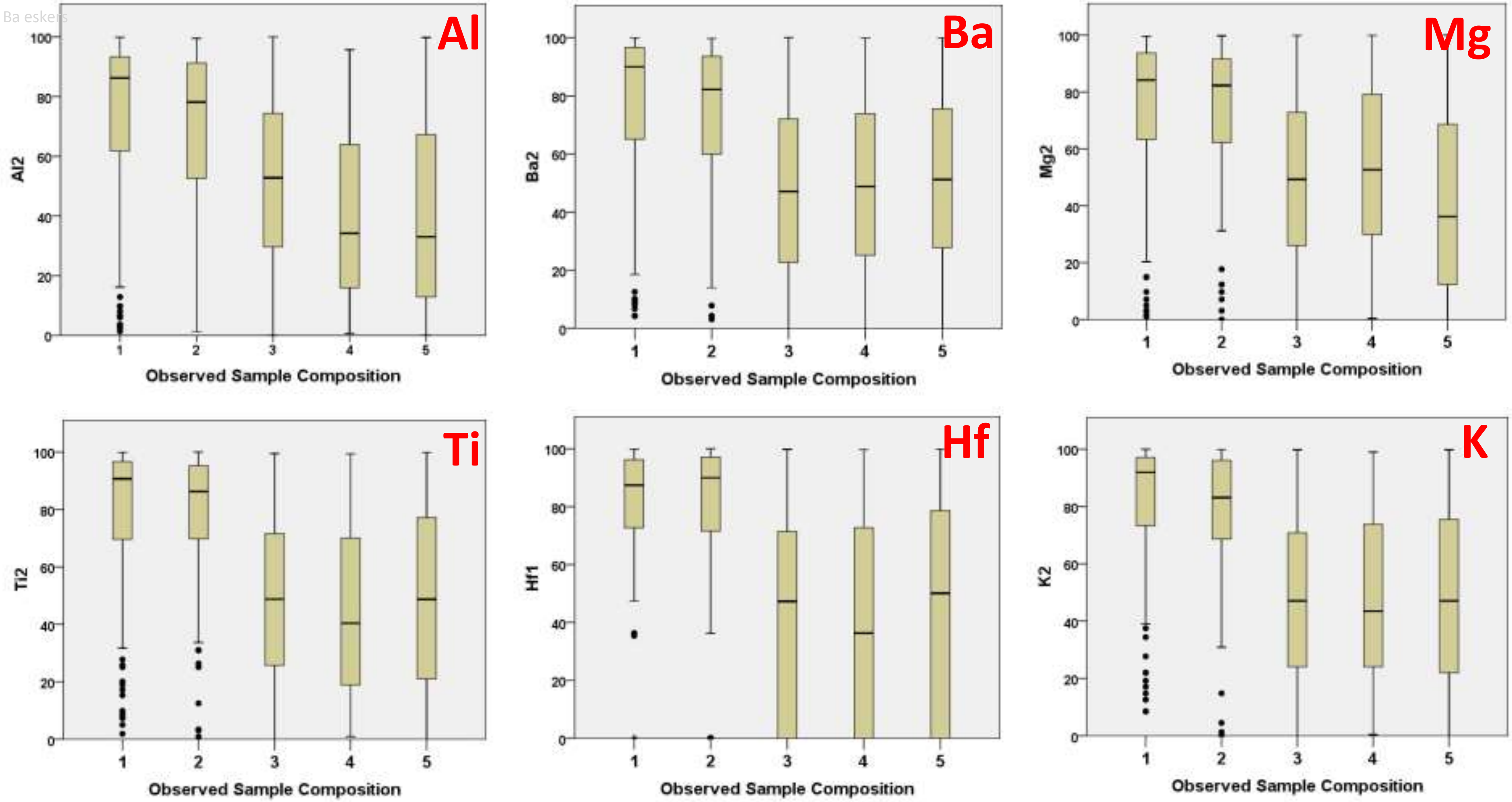
# Correlations in Labrador lake sediments

(strongly correlated elements are not normally co-associated in rocks; co-association must be due to another process)





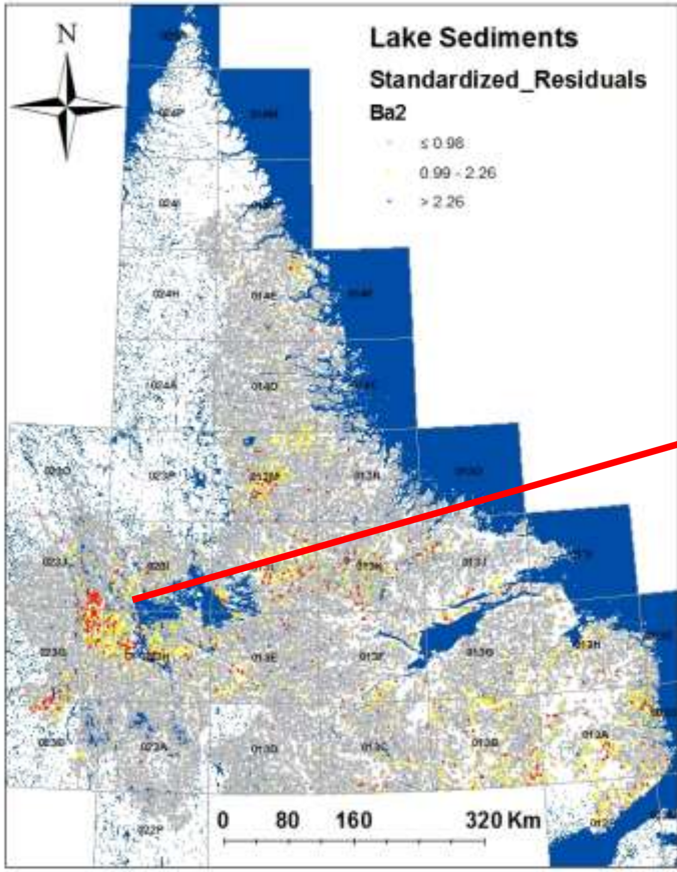

**Areal distributions of the co-associated elements look very similar, at least at regional scale.**

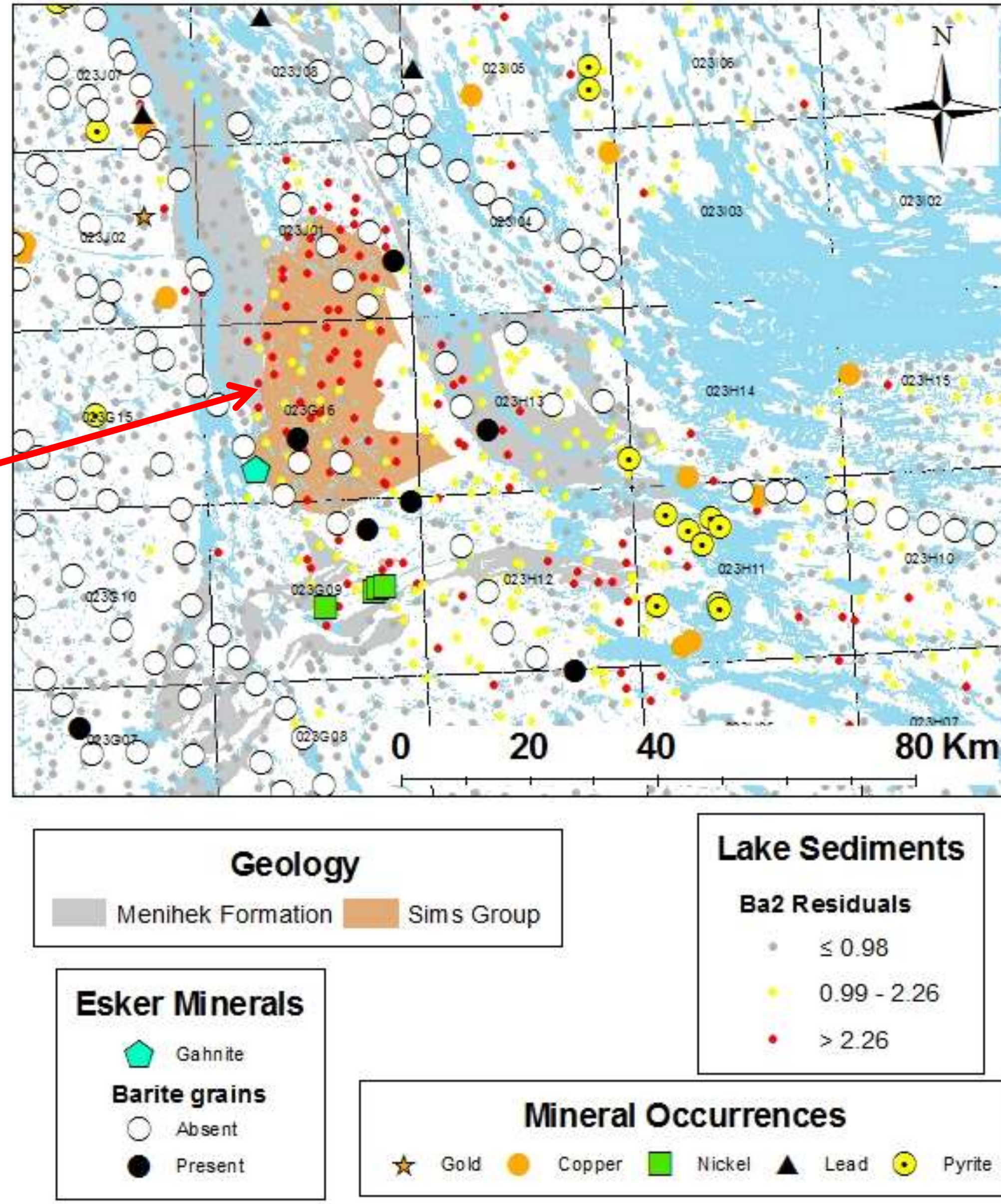



1: Clastic, fine-grained 2: Clastic, coarse-grained 3: Organic ooze 4: Organic, granular 5: Organic, peaty

**Elements in the association tend to be enriched in samples rich in inorganics**

4 equations




**Modelling of clastic component enabled extraction of non-clastic residual component. High Ba residuals coincide with barite in esker samples**

## However, the associations vary slightly, depending on whereabouts in Labrador the lake-sediment samples are selected from.

**Table 2:** Elements in Factor 1 association in four areas of Labrador. Elements common to all three areas are printed in red. Elements in parentheses display negative factor loadings .

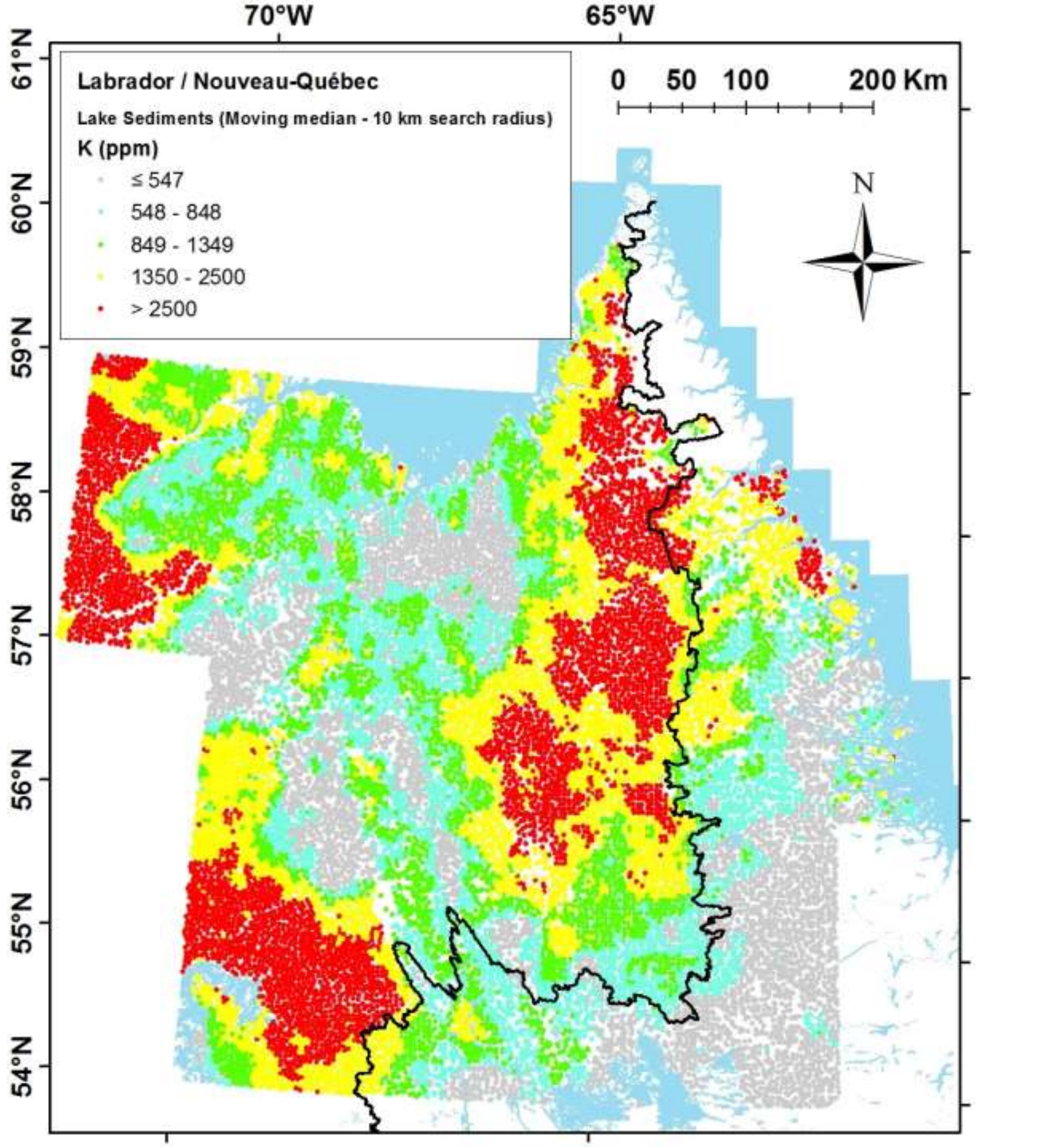
| Area | Elements in Factor 1 with Varimax-rotated loadings (Davis, 1973) greater than \|0.5\|, in descending order of strength | Percentage of total variance accounted by this factor |
|---|---|---|
| SE Labrador | K2, Na2, Sr2, Zr2, Mg2, Hf1, Rb1, Ti2, Ba2, Ca2, Nb2, Al2, Li2, Sc2, F9, Pb2, Cr2, Be2 | 37.1 |
| S-C Labrador | Na2, Sr2, K2, Ti2, Ba2, Mg2, Hf1, Zr2, Ca2, Rb1, Nb2, Al2, Sc2, Li2, Pb2, F9, Cr2, Th1 (LOI) | 38.1 |
| SW Labrador | K2, Ti2, Rb1, Al2, Mg2, Nb2, Zr2, Na2, Hf1, Ba2, Sc2, F9, Be2, V2, Li2, Cr2, Th1, Sr2, Pb2, Fe2, Mn2 (LOI) | 42.6 |
| N Labrador | Na2, Sr2, K2, Ba2, Ti2, Al2, Ca2, Hf1, Rb1, Nb2, Zr2, Mg2, Sc2, Li2 (LOI, Hg18, Br1) | 33.2 |

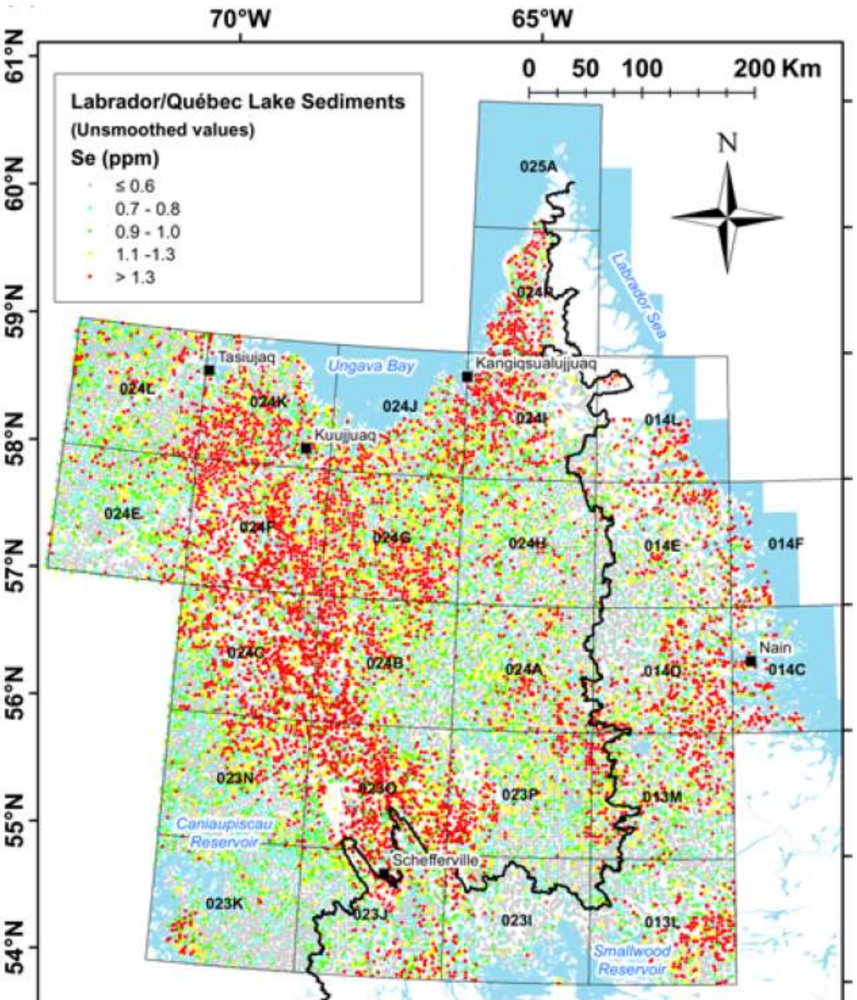
(Table from Amor, S.D., “Geochemical Quantification of the clastic component of Labrador lake sediments and applications to exploration”, Government of Newfoundland and Labrador, Department of Natural Resources, Geological Survey, Open File LAB/1625, 2014.)

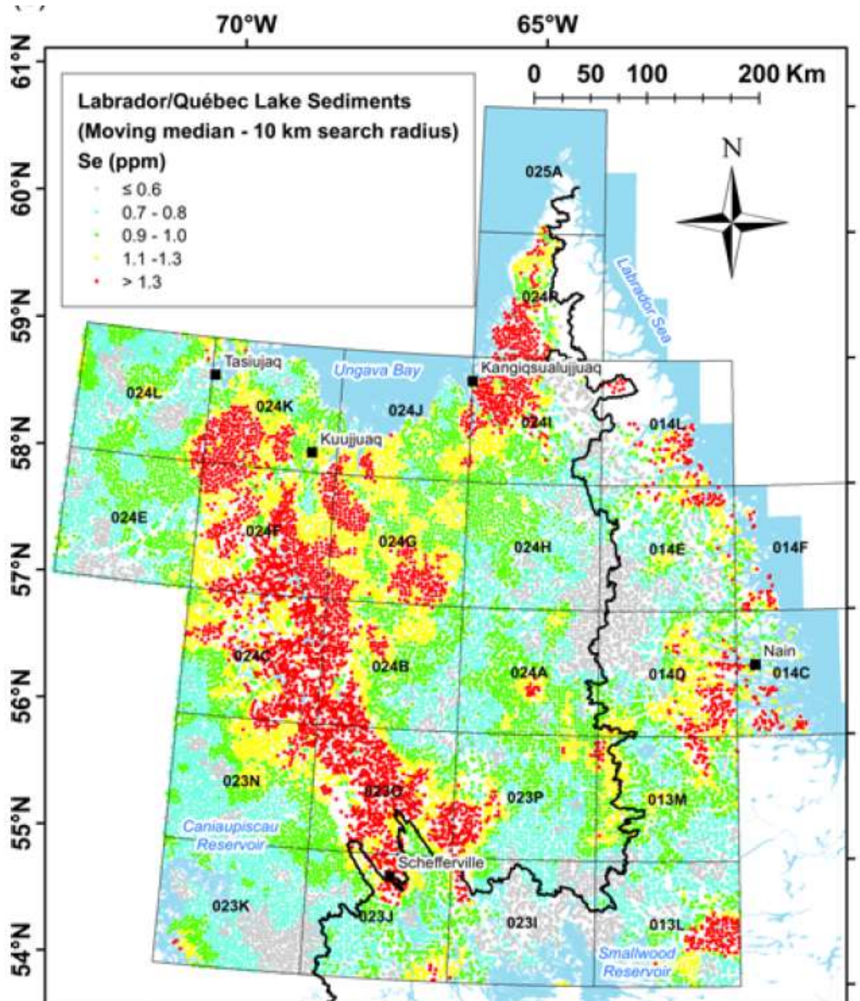
# The Labrador-Québec dataset

**New analyses of old Labrador samples, using same digestion applied to Québec samples, so that data “match” on either side of border**

**K distribution does not reflect bedrock geology**





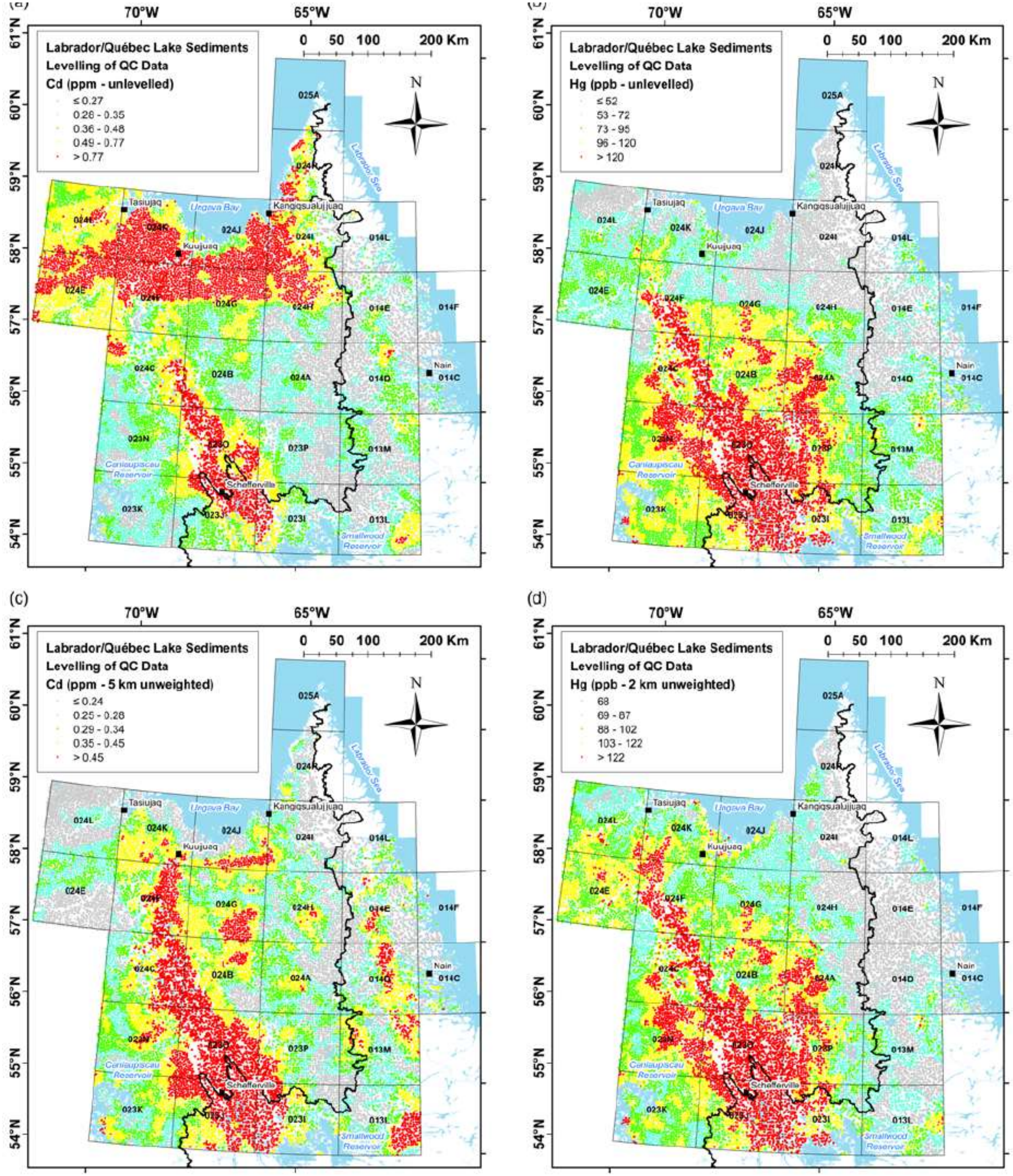




## Data have been smoothed to extract “signal” from “noise” and clarify regional picture

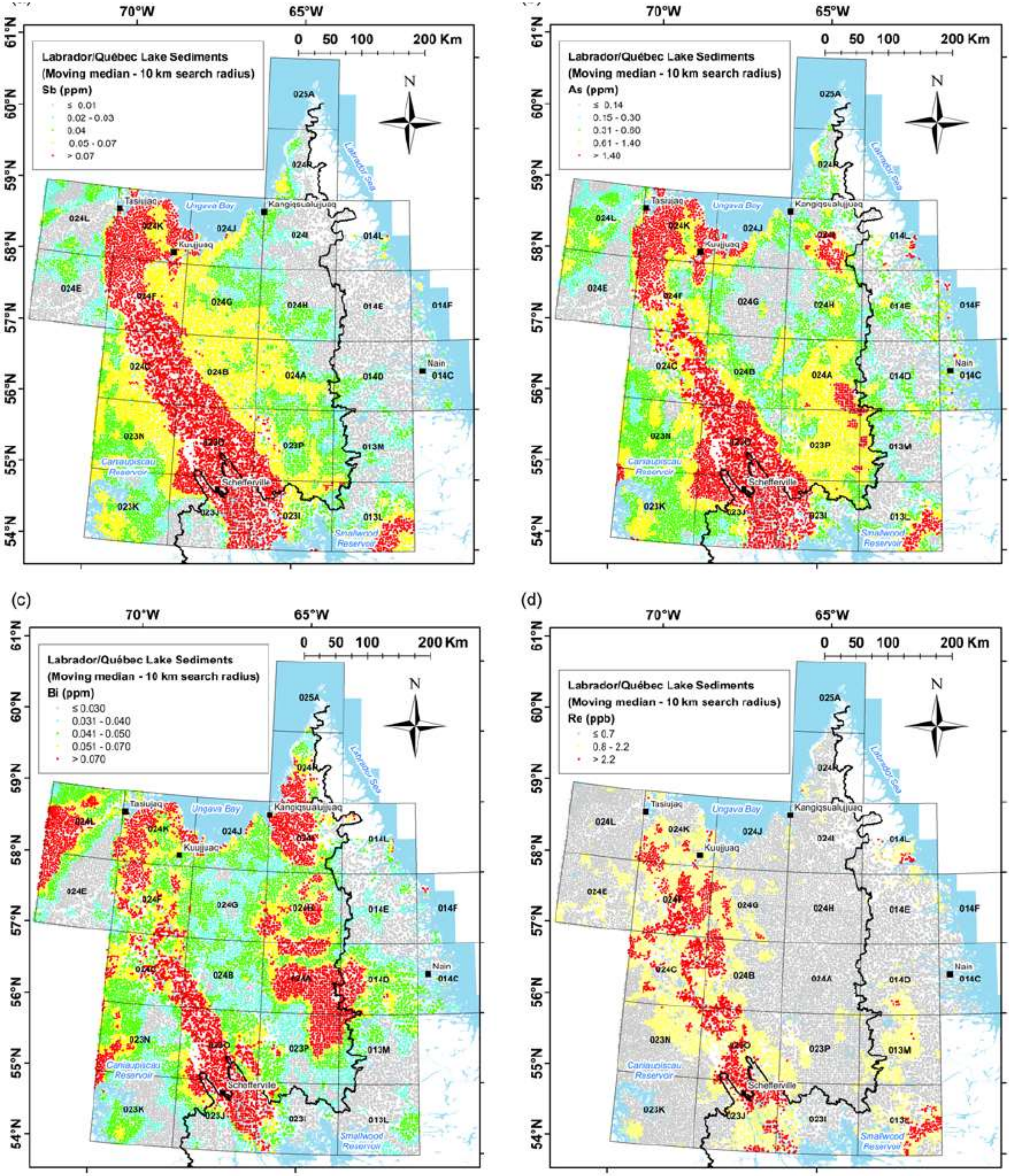
(from Amor, McCurdy and Garrett, “Creation of an atlas of lake-sediment geochemistry of Western Labrador and Northeastern Québec”, Geochemistry: Exploration, Environment, Analysis, Volume 19, Number 1, 2019.)

The trough the trough

**Some levelling was necessary, although not between Labrador and Québec data**

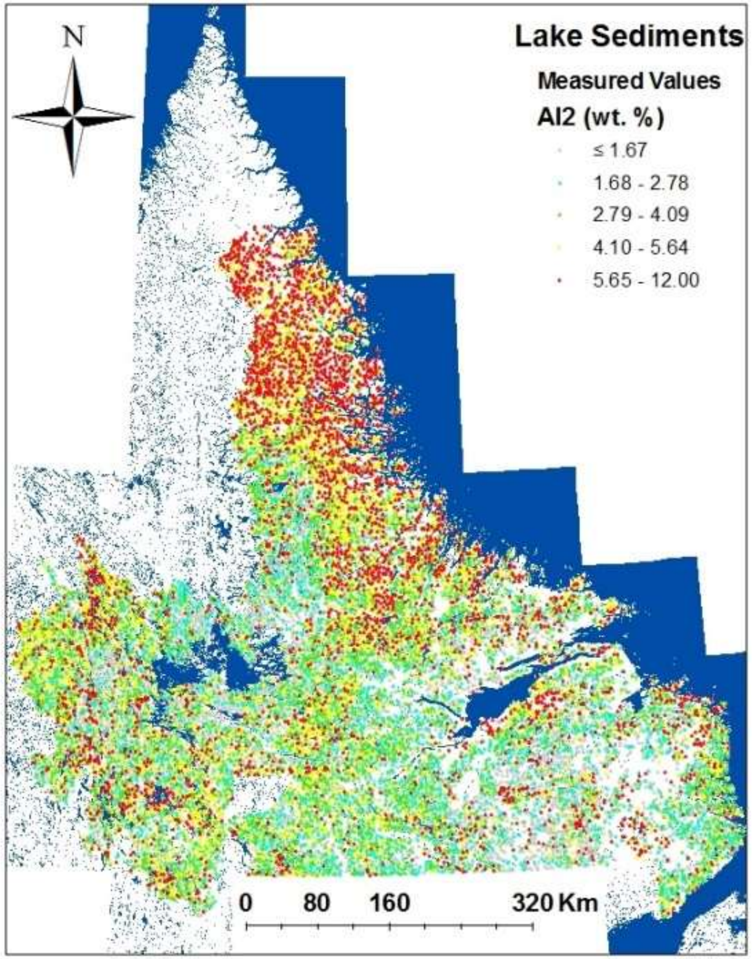


**Some examples of the distinctive geochemistry of the Labrador Trough**



# The Labrador lake dataset

**National Geochemical Reconnaissance samples collected and analyzed by GSC; subsequently analyzed at GSNL for additional elements**

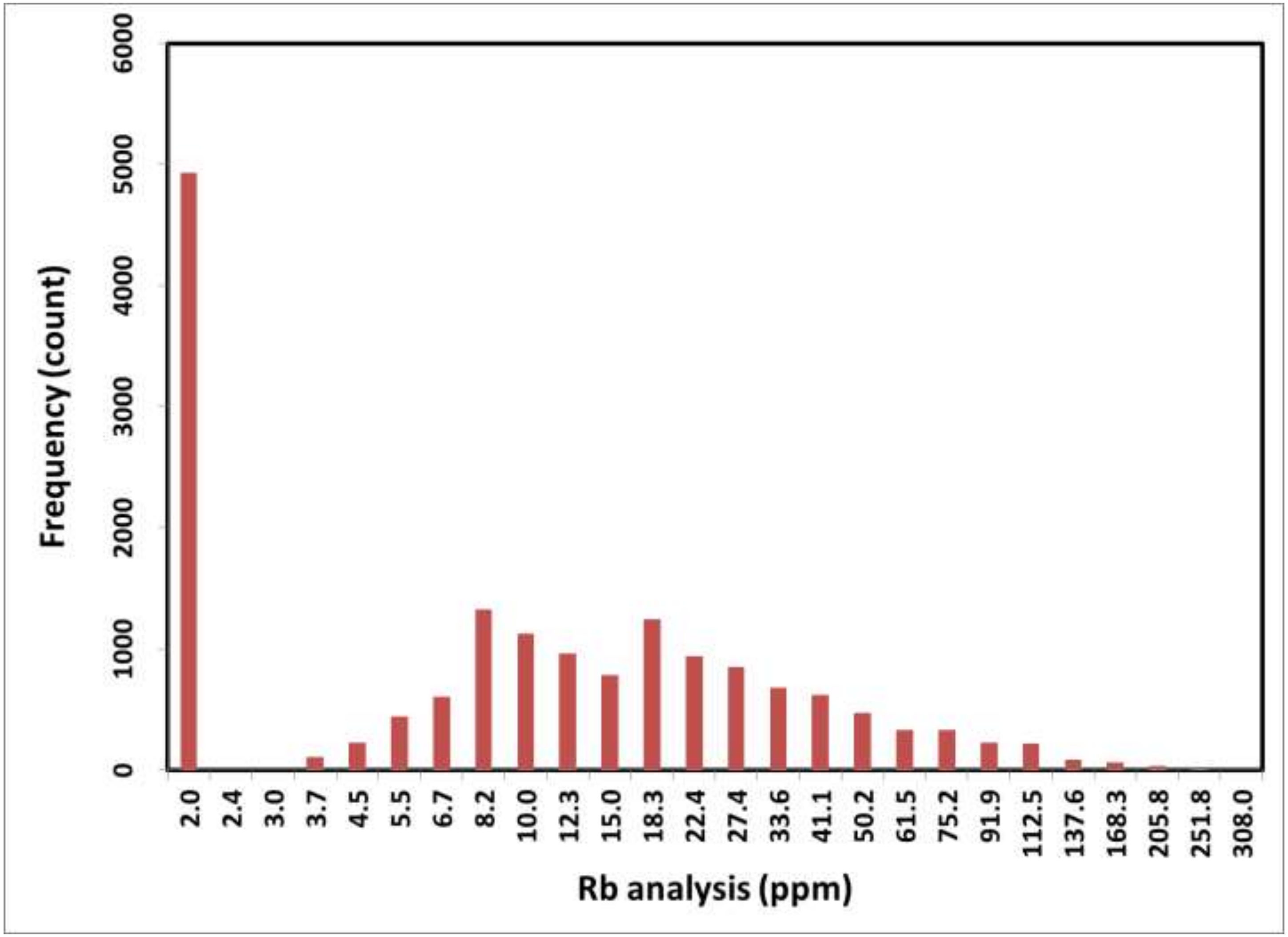



# Distribution of Labrador lake-sediment samples

**(northern Torngats not sampled as stream drainage is better developed)**

# The Newfoundland lake dataset

**Analyzed by GSNL for same elements and by same methodology as NGR (AA, INAA).**

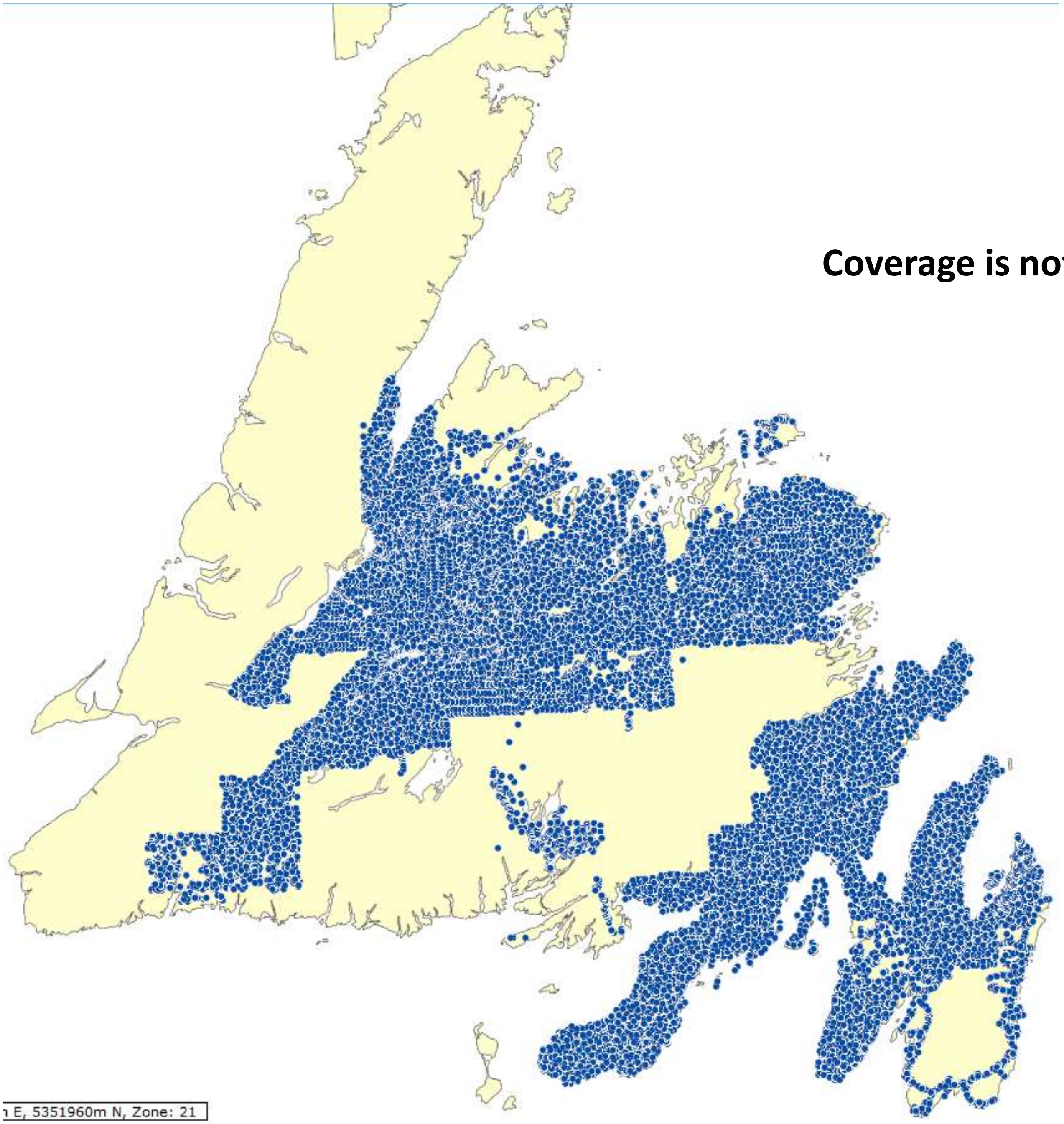



**The frequency distributions of several key elements in the Newfoundland lake-sediment dataset are too severely truncated or censored for use in GWR (or many other statistical procedures)**

# The Newfoundland till dataset

## Analyzed by GSNL by INAA and multiacid/ICP-OES etc.

(considering fluoride only in this study)

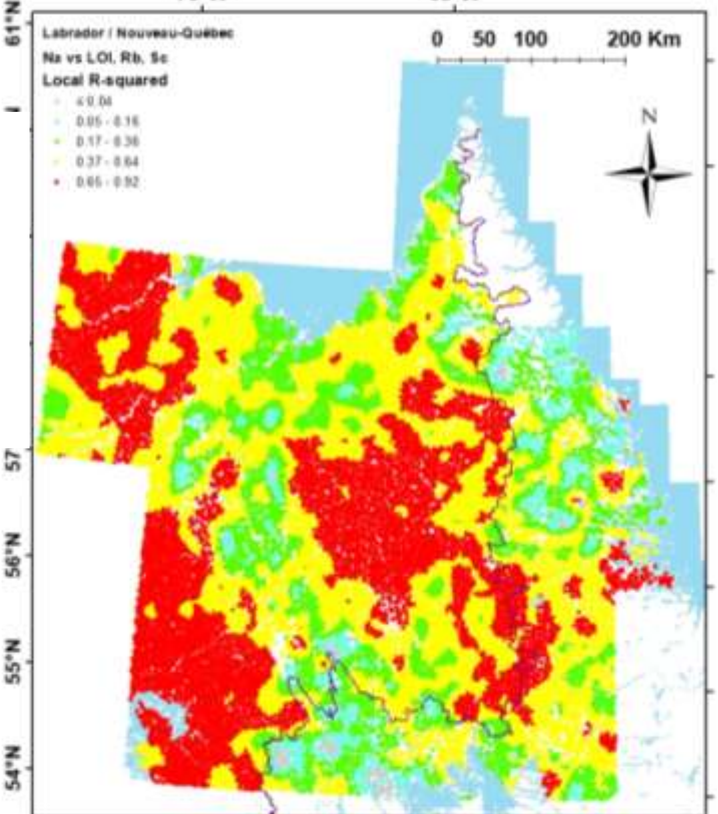


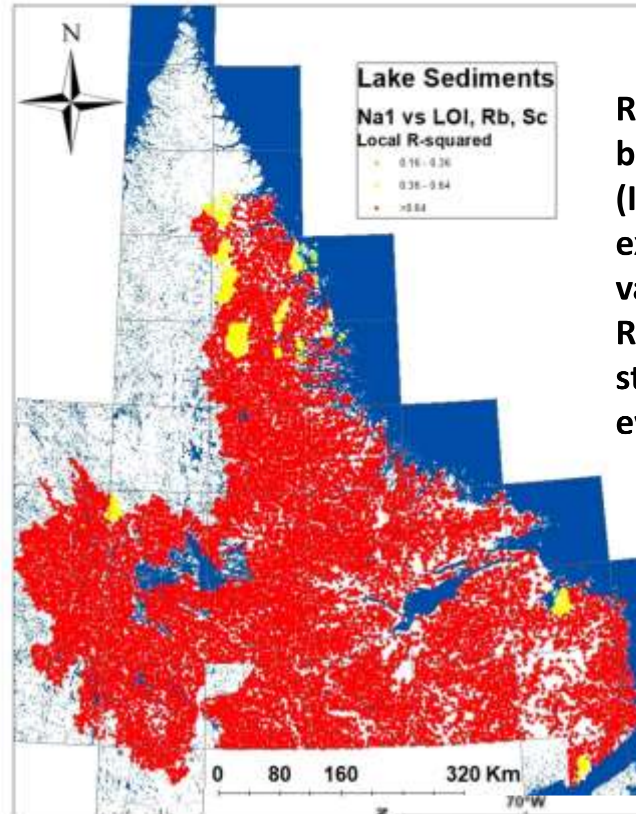

**Coverage is not complete**

# Results

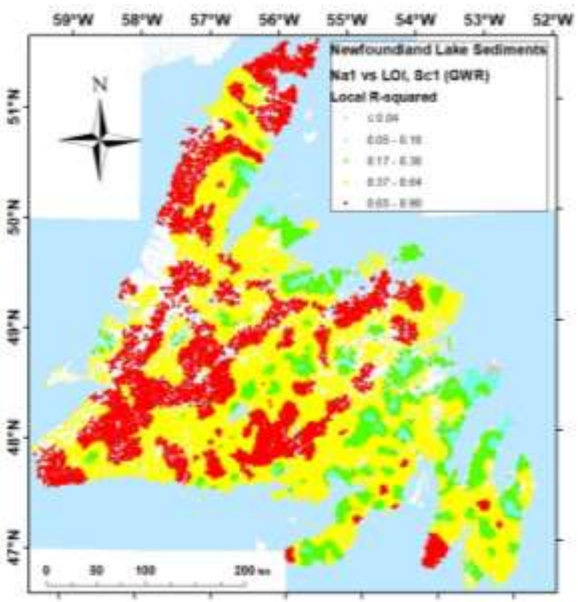
## Local $R^2$ (strength of relationship)

Relationship between *partial* (aqua regia) Na and explanatory variables LOI, Rb, Sc is strong, but variable.



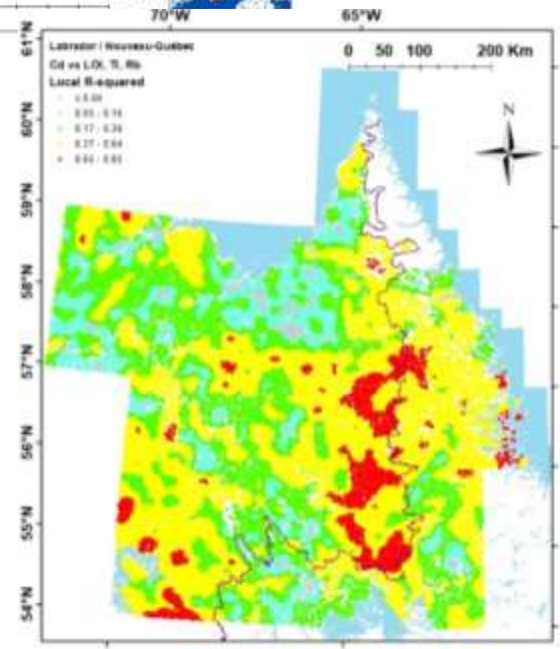




Relationship between *total* (INAA) Na and explanatory variables LOI, Rb, Sc is very strong almost everywhere




Relationship between *total* (INAA) Na and explanatory variables LOI and Sc is variable (no other explanatory variables available)

Relationship between Cd and explanatory variables LOI, Rb, Ti changes sharply at same discontinuity defined by original analyses; demonstrates that "calibration shift" was accompanied by a deterioration of the precision.



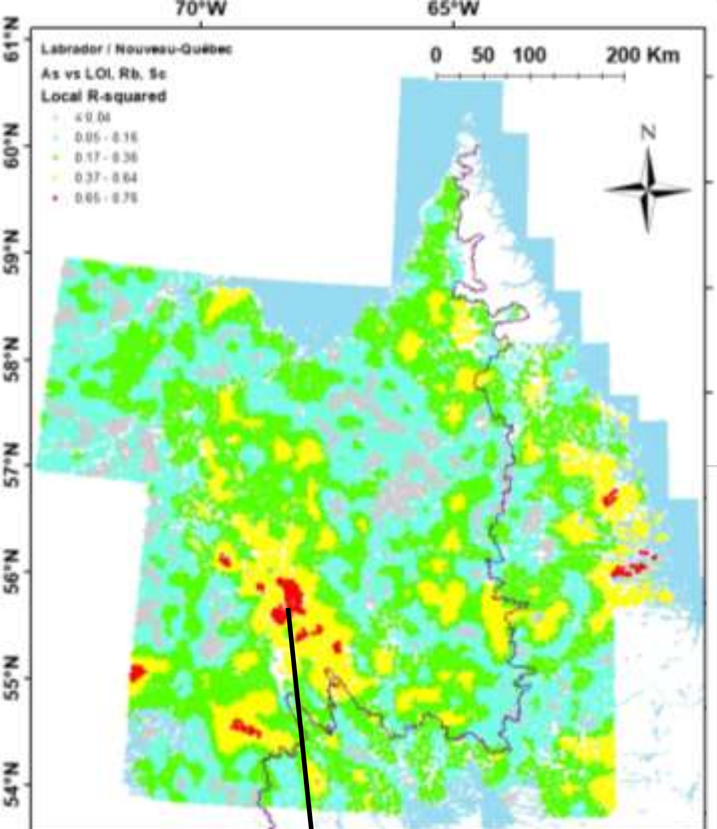
Na,. Na, Na, Cd

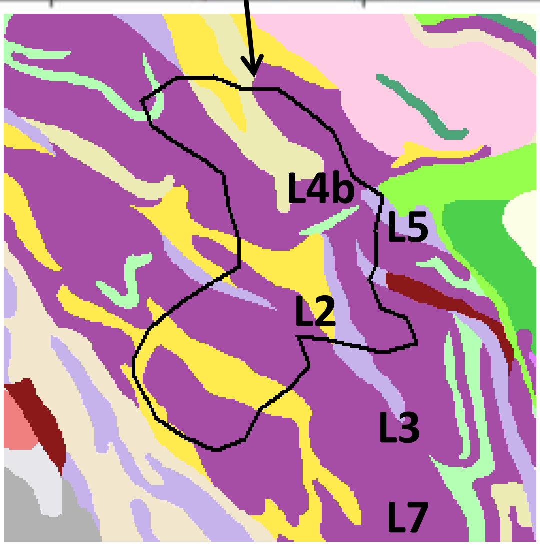
# An example of locally strong association

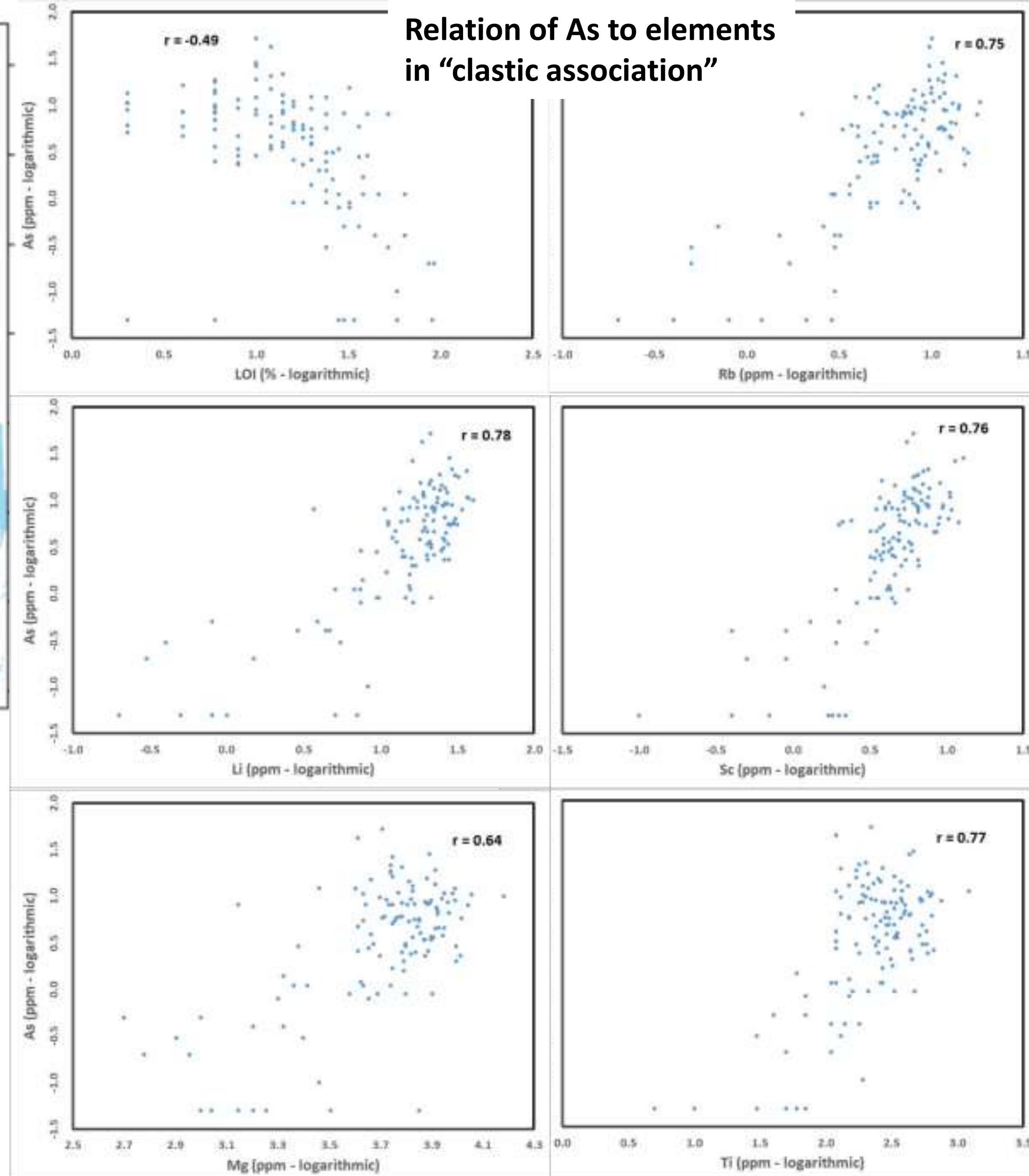
## Arsenic modelled by Rb and Sc

Does the strong correlation indicate that arsenic is locally being dispersed clastically to an unusual extent?

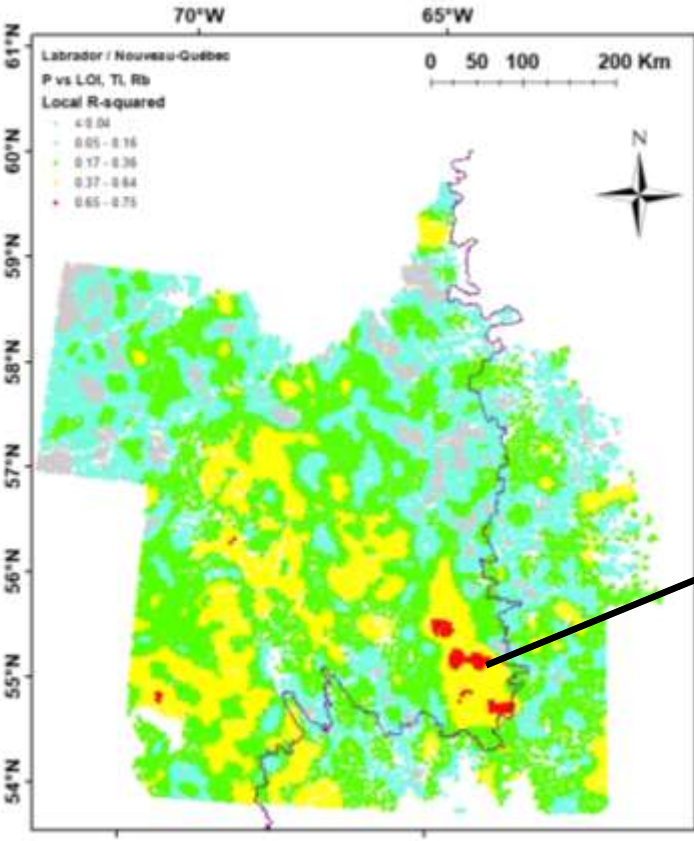
## Relation of As to elements in “clastic association”

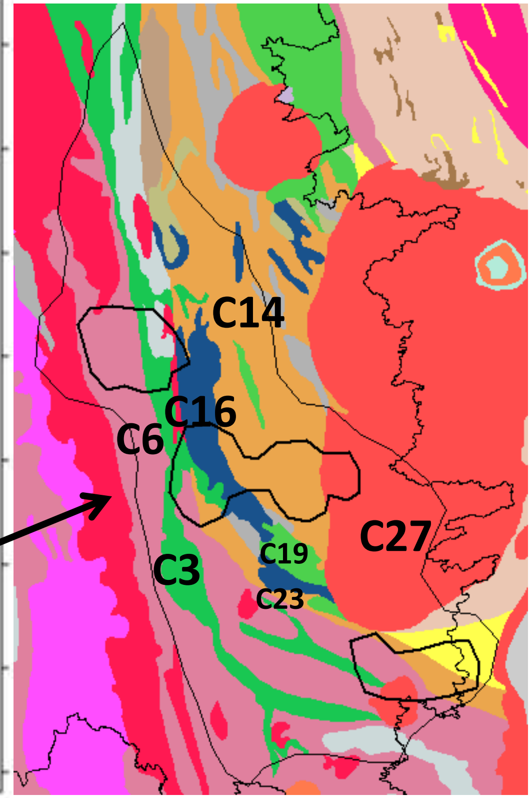


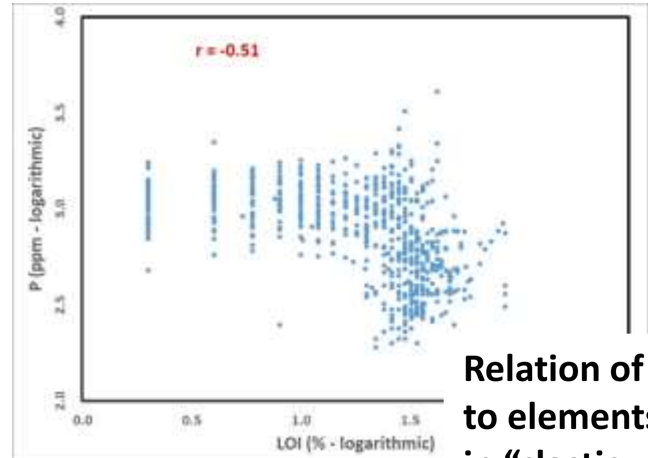



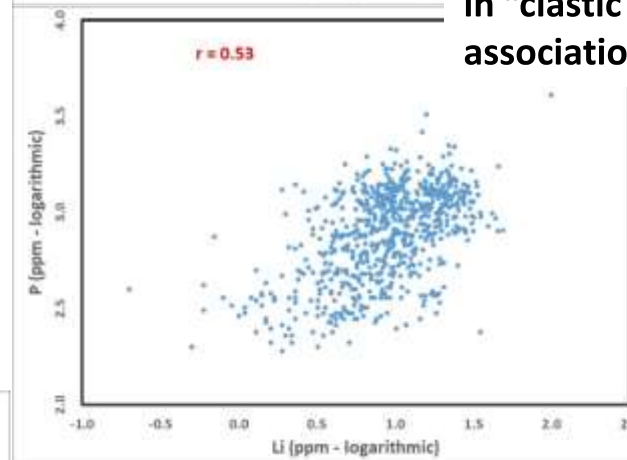

L2: Arenaceous sediments; L3: Denault dolomite; L4b: Baby argillaceous sediments; L5: Bacchus mafic volcanics; L7: Montagnais mafic intrusives

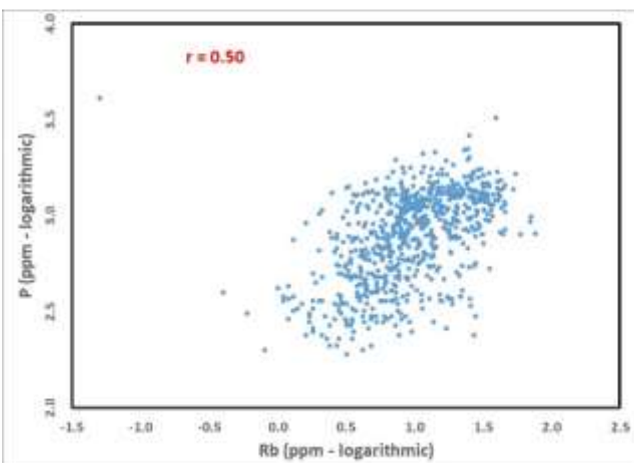


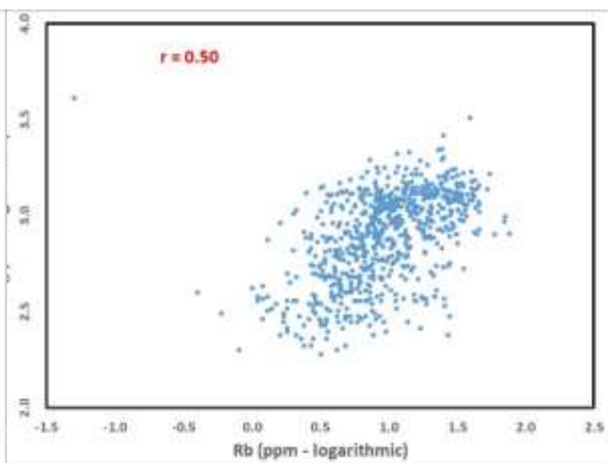



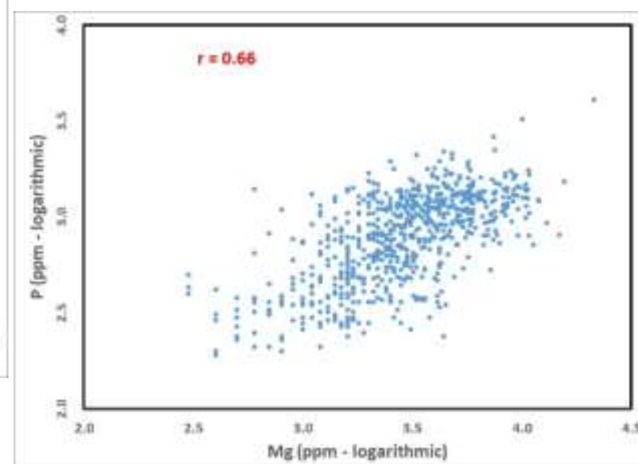




**Relation of P to elements in "clastic association"**



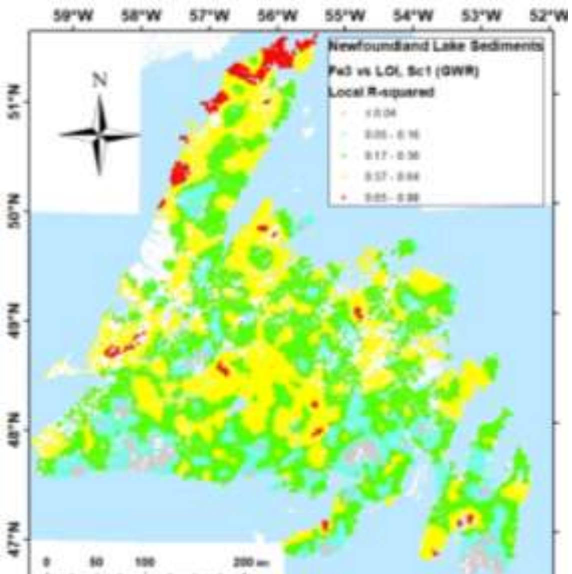



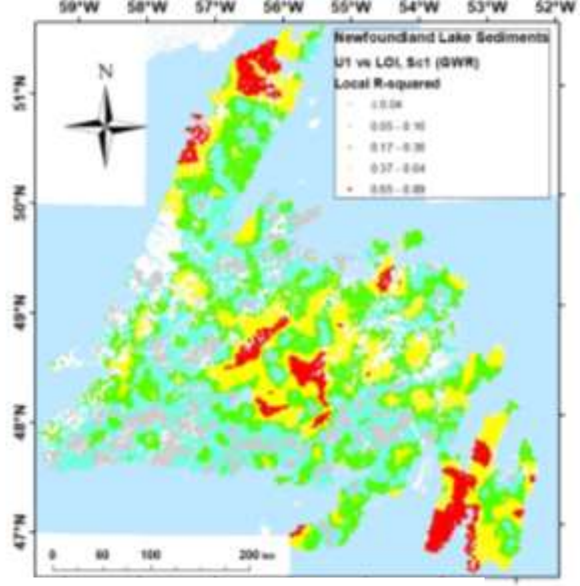




C3: Tunulik mafic volcanics; C6: gneiss and metaplutonic rocks; C14: Mistinibi metasediments; C16: Pallatin suite granite; C19: Pelland Complex granite, syenite etc.; C23: De Pas Batholith; C27: Granite, quartz monzonite, syenite

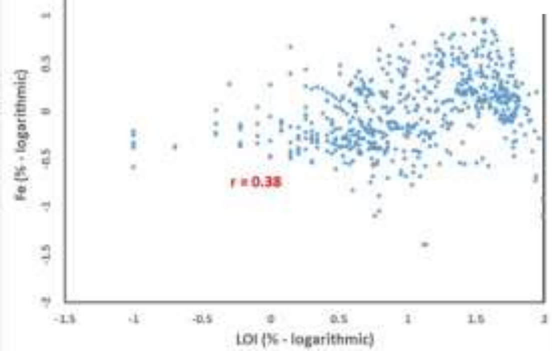
LOI Sc no Rb
LOI positive correlation

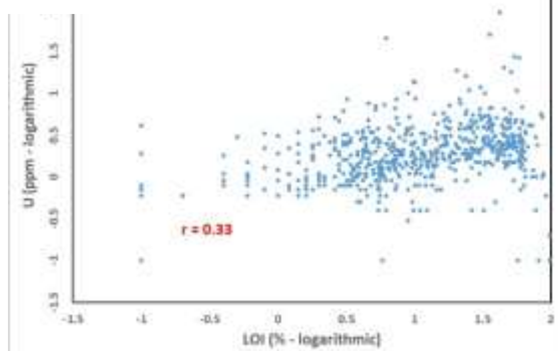
Local $R^2$
(Newfoundland Lakes)

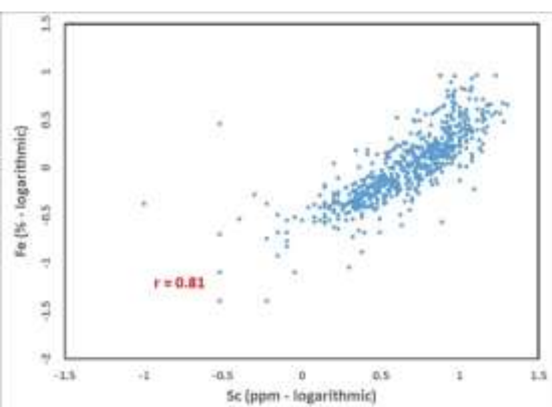


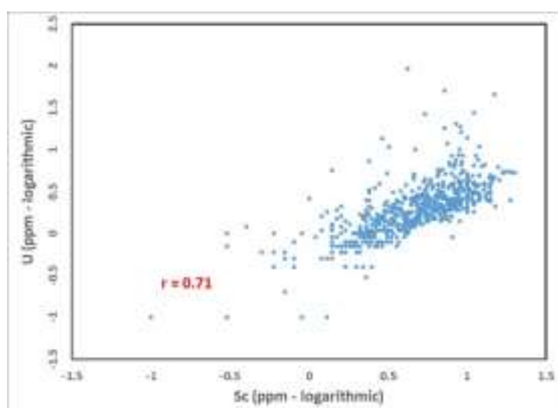



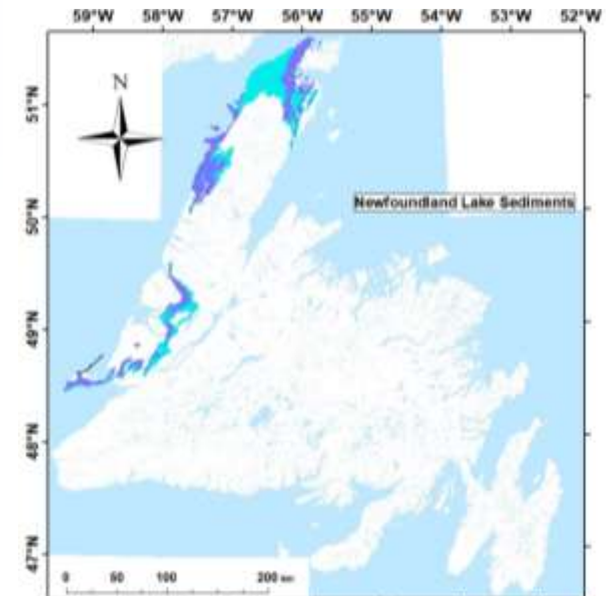

**Relation of Fe and U to elements in “clastic association”**



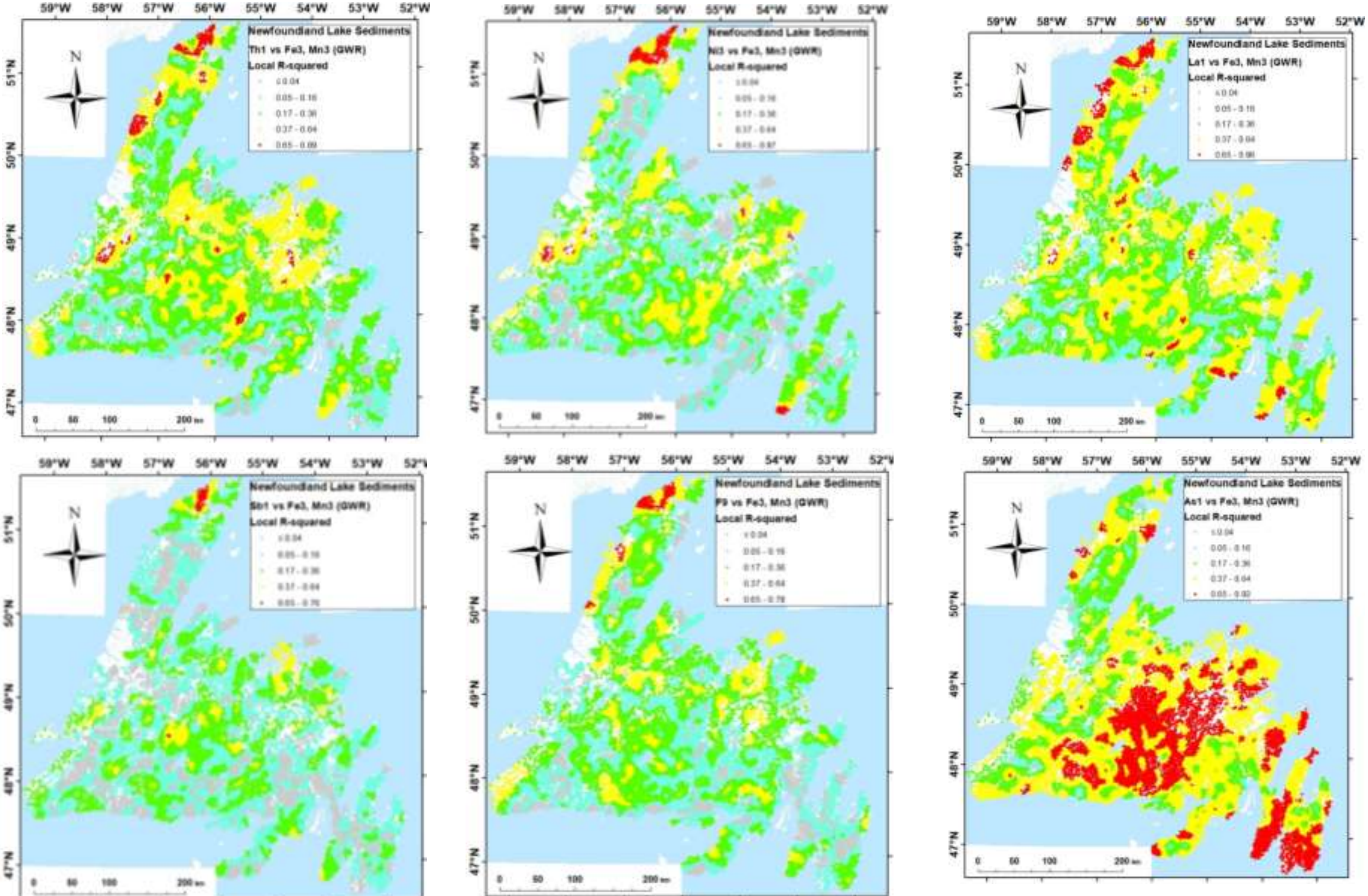



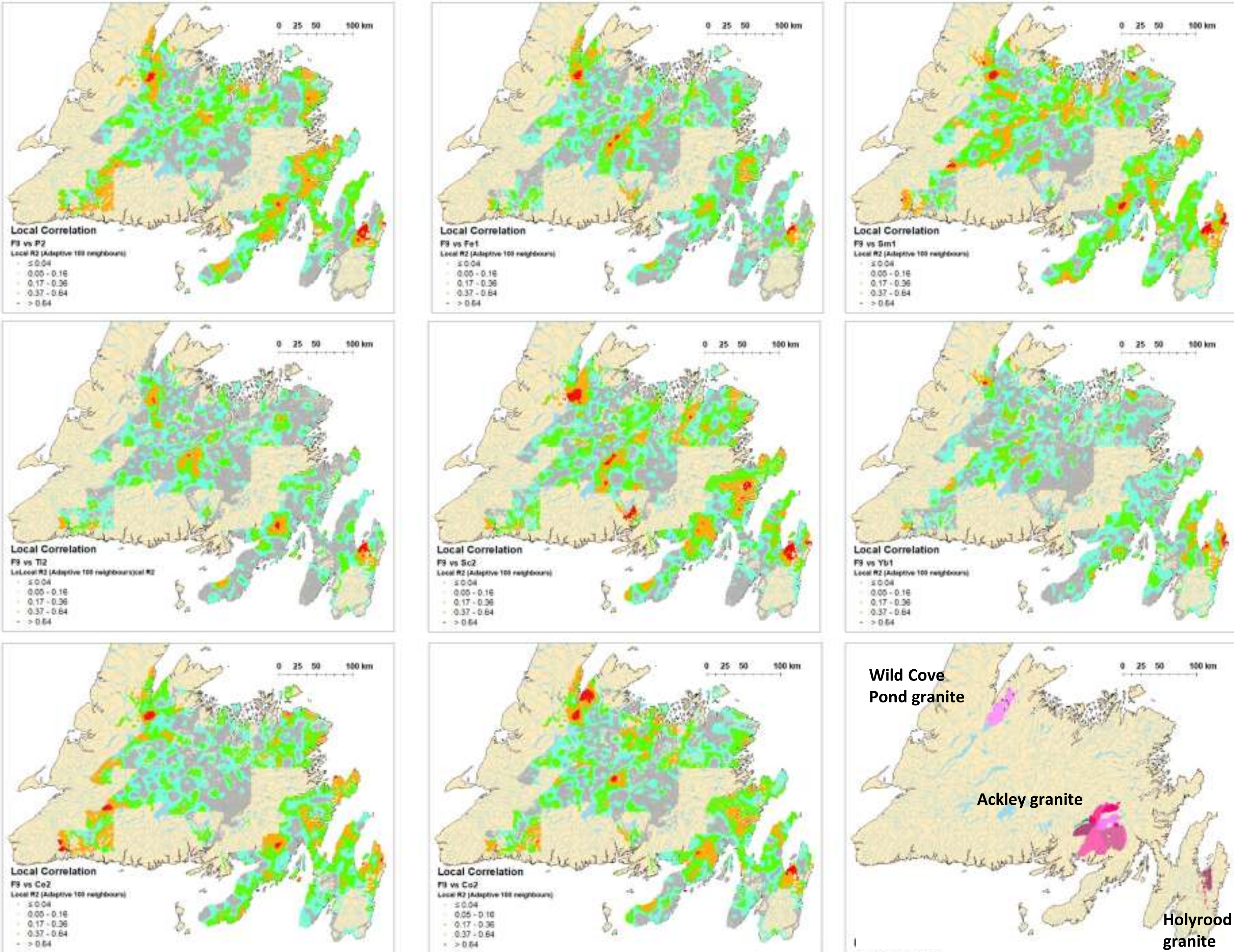




**Relationship between “clastic component” (modelled here by just LOI and Sc) and Fe, and U, is strong in areas underlain by carbonates (note positive correlation between Fe and LOI)**

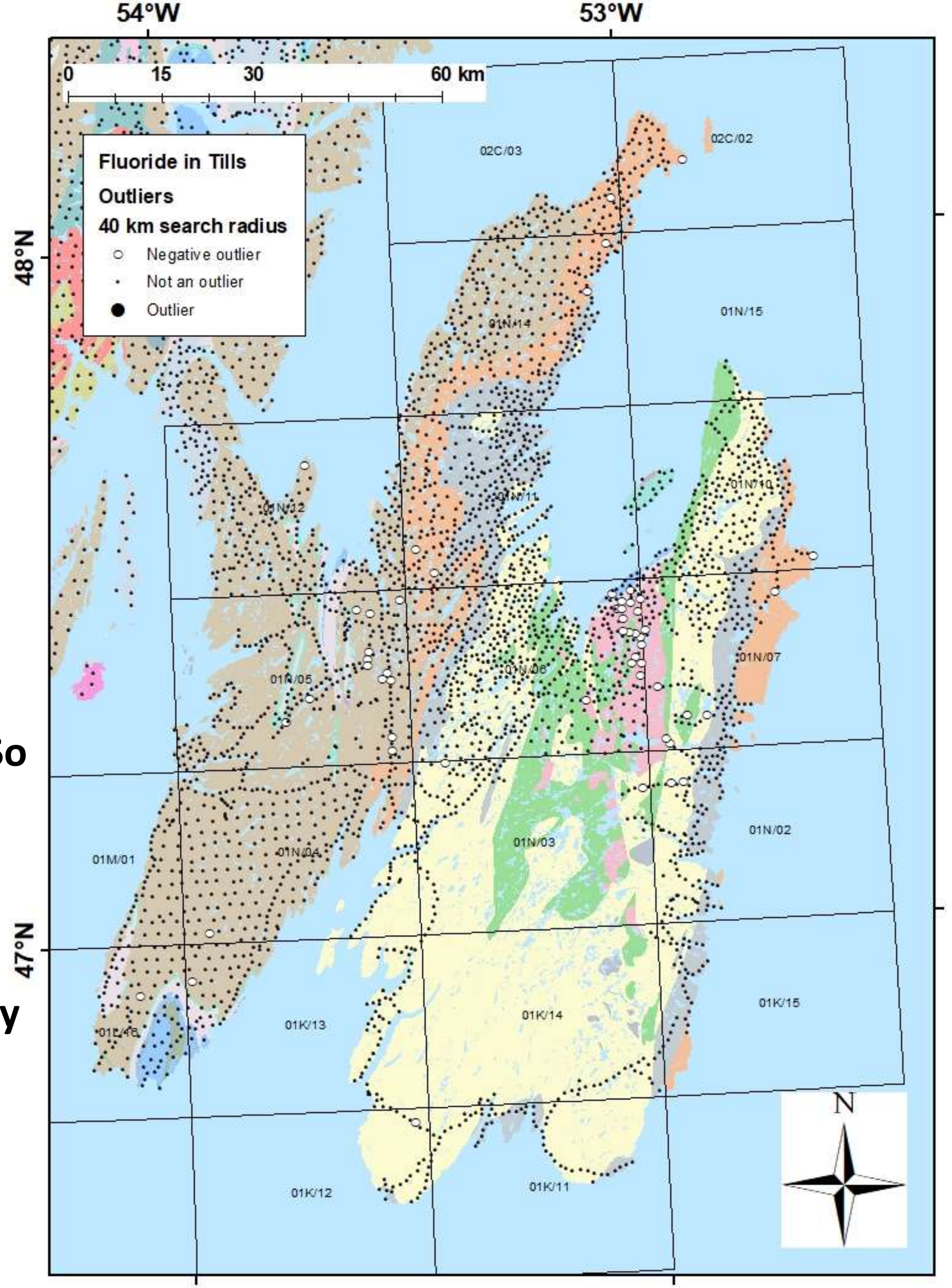



**$R^2$ (relationship strength) of Fe and Mn to various elements (note strong correlation with As in central Newfoundland and southern Avalon)**

# Local $R^2$
# (Newfoundland Tills - Fluoride)


0 25 50 100 km
Local Correlation
F9 vs P2
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Fe1
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Sm1
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Ti2
LoLocal R2 (Adaptive 100 neighbours)cal R2
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Sc2
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Yb1
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Ce2
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Local Correlation
F9 vs Co2
Local R2 (Adaptive 100 neighbours)
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
> 0.64
0 25 50 100 km
Wild Cove
Pond granite
Ackley granite
Holyrood
granite

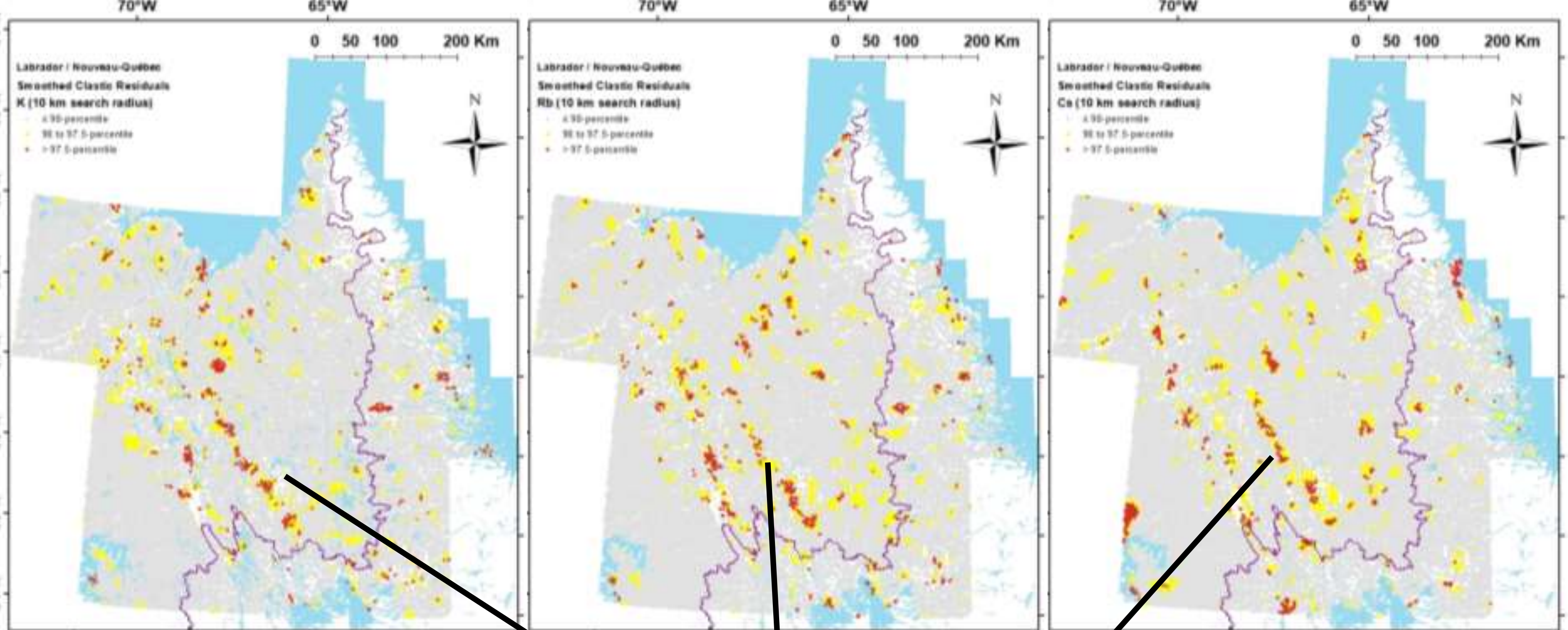
**Although fluoride shows strong correlation with elements like phosphorus (suggesting the presence of fluorapatite), fluoride values are very low in tills collected over the Holyrood granite. So the minerals in which these elements are associated with fluoride are unlikely to be very abundant.**



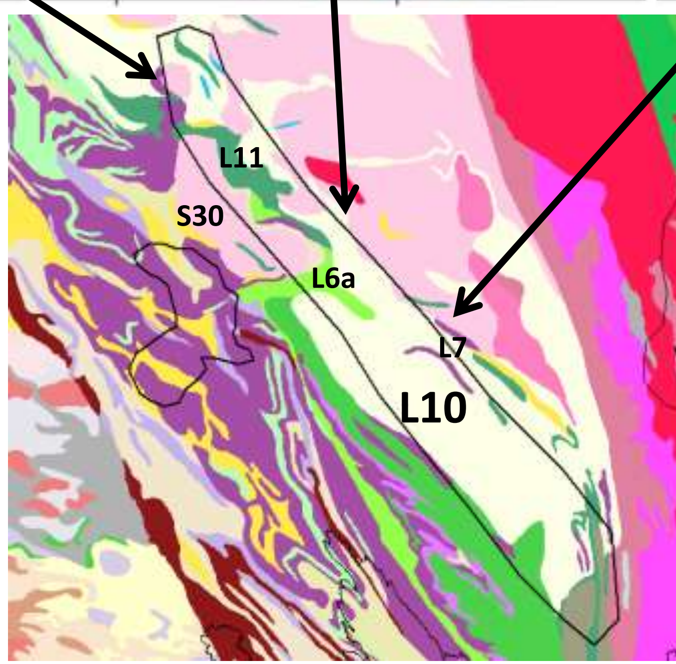
End of Multiple $R^2$. Beginning of

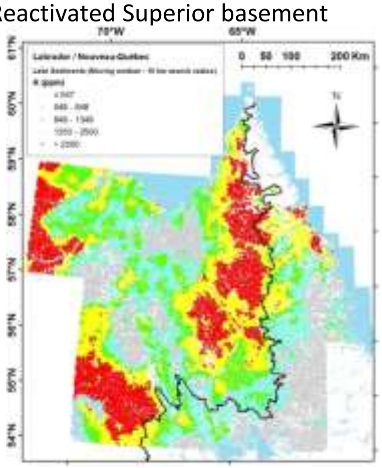
# Residuals

**(Labrador / Québec)**




L6a: Murdoch mafic volcanics; L7: Montagnais mafic intrusives; L10: Rachel-Laporte psammitic schist; L11: amphibolite and mafic gneiss; S30: Reactivated Superior basement

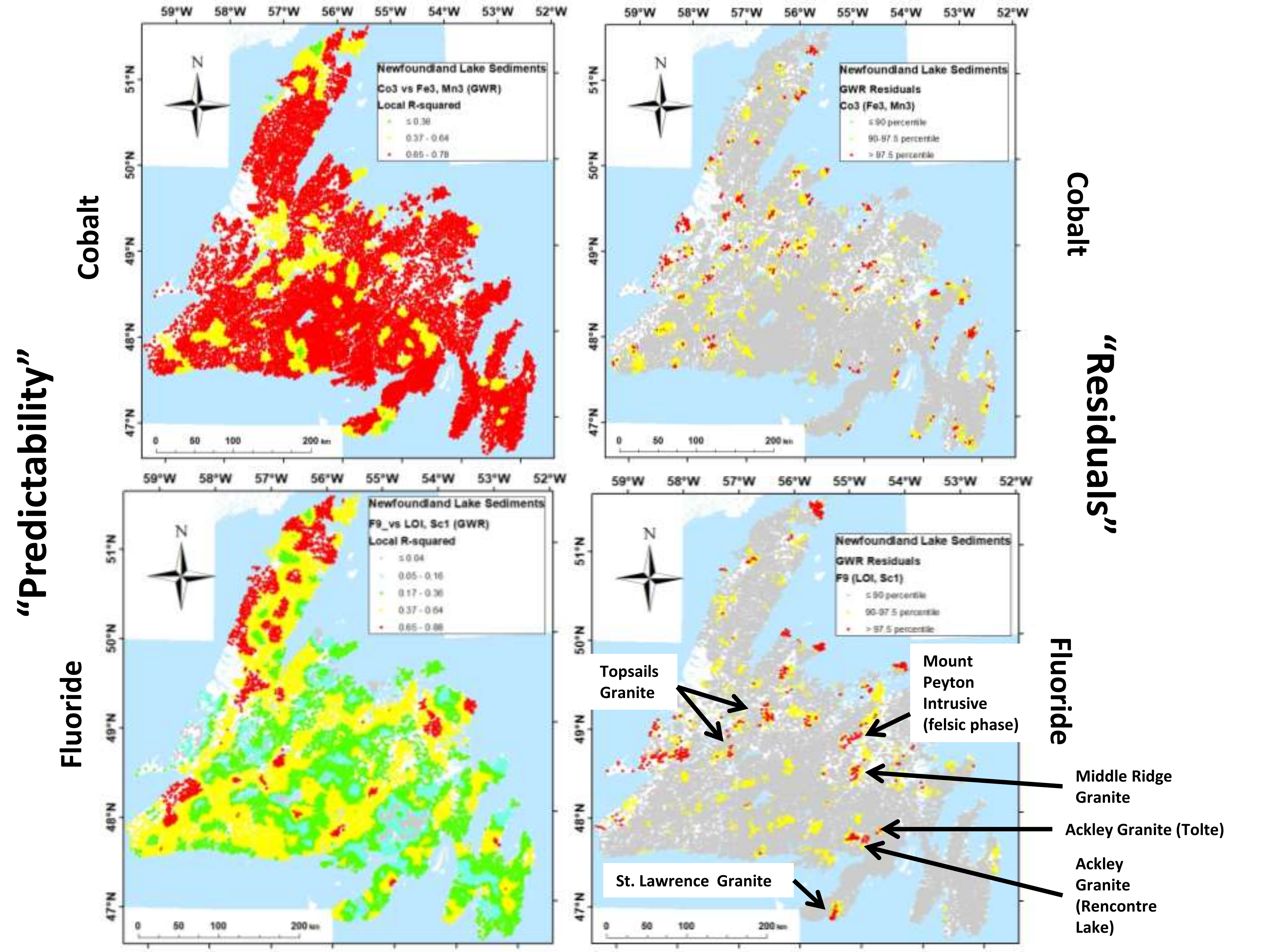



**Raw K values (reminder)**

**GWR residuals of K, Rb and Cs delineate a zone parallel to strike of Labrador Trough.**

# Residuals
**(Newfoundland Lake Sediments)**



Cobalt
"Predictability"
Fluoride
Cobalt
"Residuals"
Fluoride
Newfoundland Lake Sediments
Co3 vs Fe3, Mn3 (GWR)
Local R-squared
≤ 0.36
0.37 - 0.64
0.65 - 0.78
Newfoundland Lake Sediments
GWR Residuals
Co3 (Fe3, Mn3)
≤ 90 percentile
90-97.5 percentile
> 97.5 percentile
Newfoundland Lake Sediments
F9_vs LOI, Sc1 (GWR)
Local R-squared
≤ 0.04
0.05 - 0.16
0.17 - 0.36
0.37 - 0.64
0.65 - 0.88
Newfoundland Lake Sediments
GWR Residuals
F9 (LOI, Sc1)
≤ 90 percentile
90-97.5 percentile
> 97.5 percentile
Topsails Granite
Mount Peyton Intrusive (felsic phase)
Middle Ridge Granite
Ackley Granite (Tolte)
Ackley Granite (Rencontre Lake)
St. Lawrence Granite

**Issues to be addressed:**
**1. Multiple Truncation**

If it is accepted that populations in which more than 10% of the analyses are less than the detection limit are unacceptable, then <u>every</u> such population, of the many thousands (potentially) in the GWR should be excluded. ArcCIS offers no facility to do this

Caption

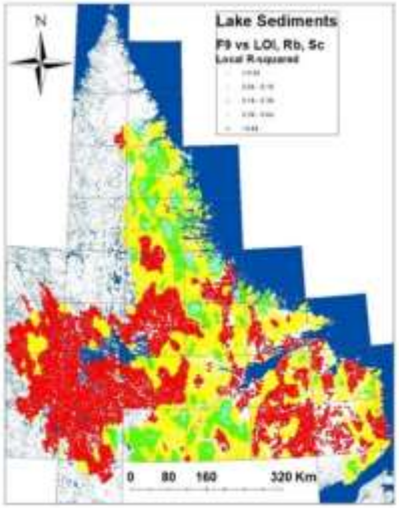
**Issues to be addressed:**

2. **"Excessive predictability"**

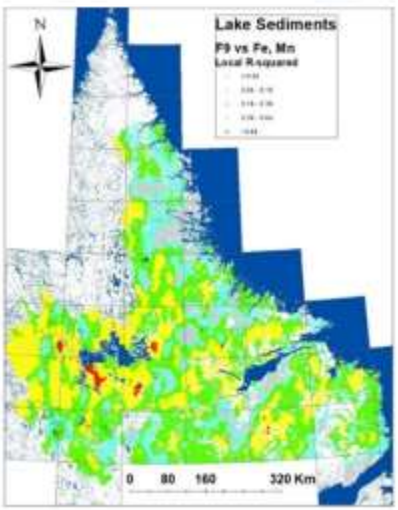
Even the poor predictors (both theoretically and empirically) seem to be capable of recreating the observed values of certain elements; at least, on a regional scale and after the data are smoothed.

LOI, Rb and Sc are much better predictors …

LOI-Rb-Sc



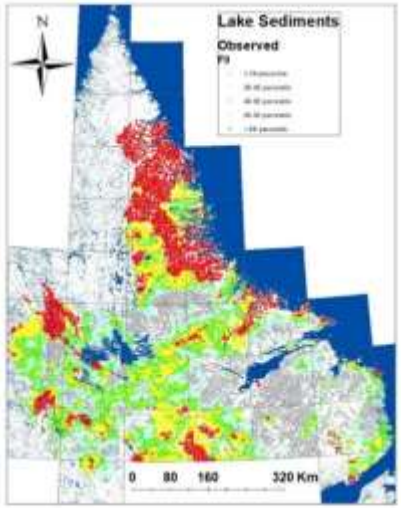

Fe-Mn




… than Fe and Mn. Which is to be expected.

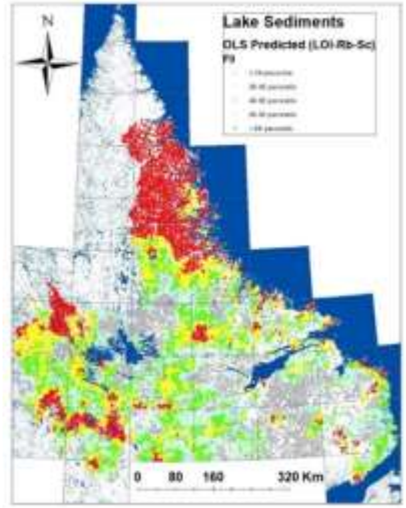
Correlations

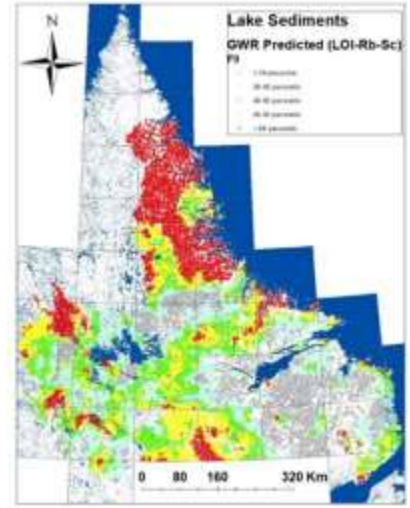
# F9



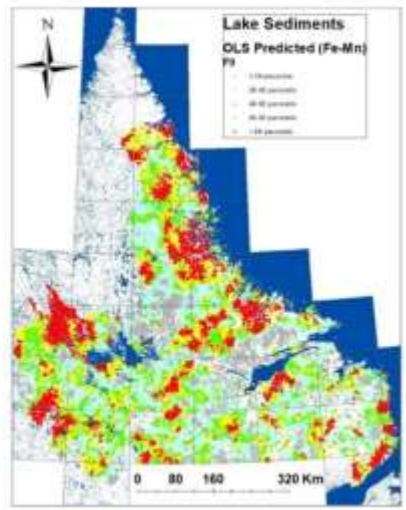

**Observed**

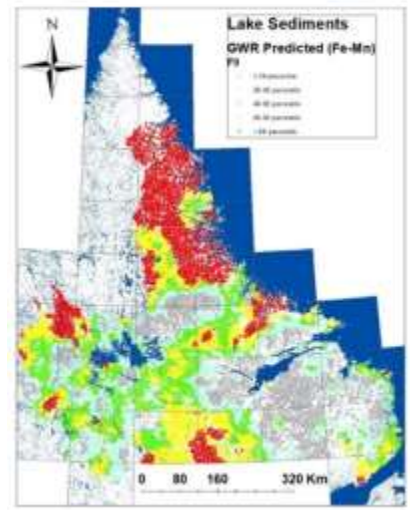
LOI, Rb and Sc are good predictors of the fluoride content (which is to be expected …)




LOI-Rb-Sc




Fe-Mn

**"Global"**

(regression equation derived for entire Labrador dataset)

**GWR**

… but so are Fe and Mn!


Thank
You
Obrigado
Merci
Multumesc
Asante
Grazie
Maake
Spasibo
Kiitos
Welalin
Mamana
Dank Je
Arigato
Matondo
Juspaxar
Vinaka
Kia Ora
Mochchakkeram
Nirringrazzjak
Cam on ban
Chokrane
Kilbos