

Exploring the Relationship between Surface and Subsurface Soil Concentrations of Heavy Metals using Geographically Weighted Regression

S. Deschenes¹, E. Setton², P. A. Demers³ and P. C. Keller⁴

¹ Department of Geography, University of Victoria, Victoria, Canada, steeve@carexcanada.ca

² Department of Geography, University of Victoria, Victoria, Canada, eleanor@carexcanada.ca

³ Cancer Care Ontario, Toronto, Canada, paul.demers@cancercare.on.ca

⁴ Department of Geography, University of Victoria, Victoria, Canada, pkeller@uvic.ca

Abstract. Geographically Weighted Regression (GWR) is used to analyze the spatial variability of the relationship between the surface and the subsurface (b horizon) soil metal concentration. We used publicly-available soil samples from provincial government websites in Canada. The correlation between the log of concentration levels of the two layers are 0.51 for As, 0.40 for Cd, 0.33 for Cr, 0.52 for Co, 0.38 for Ni, and 0.23 for Pb. Although the correlation results show that the two layers seem to be related, the GWR analysis suggests that other factors might play important role in predicting the surface soil concentration of these metals. For example, only arsenic ($R^2=0.34$) shows no spatial autocorrelation in the residuals. This study proposes that factors (natural and anthropogenic) other than the subsurface concentration itself are controlling the concentration surface levels for all the studied metals in this dataset.

Key Words: Heavy metals, spatial variability, GWR, soil, horizons

Introduction

Toxic metals (As, Cd, Co, Cr, Ni, and Pb) are natural elements present at various concentrations in the environment. In soil, their concentration levels are highly related to parent materials and pedogenesis processes, but their distribution patterns can be greatly altered by human activities (Murray et al., 2004). Anthropogenic emissions, both diffuse and point source, affect the levels of metal concentration in soil and makes the determination of a natural background level difficult (Matschullat et al., 2000). Elevated metal concentration in soil has well-known adverse effects on plants and animals, including humans (Adriano, 2001). For this reason, the capacity to predict potential elevated levels of metals in surface soils can help target efforts to reduce exposure risk by remediation and prevention.

Although the above metals have some similarities, they can behave differently under the same conditions. Thus, the patterns of the concentration levels among the metals are not expected to be the same. In fact, soil properties are known to vary, both vertically and horizontally, which influence the metal mobility. Investigating concentration variability between surface

and subsurface soils can help to determine the degree of relation between the two layers. For soil exposure, the lack of relationship could be an indicator of anthropogenic influence, which in turn may help in finding potential contaminated locations.

The objective of this research is to explore the contribution of subsurface levels in predicting surface levels of selected metals as a precursor to developing more complex surface soil concentration level models at regional or even at national scales.

Methods

Data

The data were obtained in 2010 from the Ministry of Environment of British Columbia (Background Soil Quality Database - 1996) and the Ministry of Northern Development, Mines, and Forestry of Ontario (MRD015 1992-94, MRD021 - 1996, and MRD136 - 2003). The data, which cover several regions of Canada, were available online (<http://www.gov.bc.ca/env/> and http://www.ontario.ca/en/your_government/009883). In is important to note here that these datasets have different

spatial distributions. The Ontario dataset is from geological surveys where large clusters are distributed throughout the province, while the British Columbia survey is more evenly spread across the province. First, the files were sorted by sampling depth (surface and B-horizon) and then by type of digestion and analytical methods. To account for random sampling assumption in linear regression analysis, all transects and clusters (50 to 100 samples within a 1 km² area) were removed from the data set. Blanks, zeros, or values below the detection limit were excluded. Multiple samples taken at the same location for the same depth were averaged. The *Aqua Regia* digestion method was retained for Cd (n = 81), Co (n = 839), Cr (n = 899), Ni (n = 860), and Pb (n = 860) as it was the most widely used method for dissolving the soil samples among the dataset. Following that, Inductively Coupled Plasma (ICP) methods (ICP-MS and ICP-OES) were combined together as they provide comparable results for this type of study, both in accuracy and precision (Baffi et al., 2002). For As (n = 877), the Instrumental Neutron Activation Analysis (INAA) method was retained because not enough samples analyzed by ICP methods were available after removing blanks and the values below the detection limit for both the surface and the subsurface samples. As a consequence, the As dataset only includes the geological surveys from the province of Ontario. Finally, surface and subsurface samples located at the same site were paired together to analyze their spatial relationships. Finally, before the analysis, the concentration data were log- and square-transformed to normalize the distribution.

Analysis

As a first step, correlation analysis was applied to assess the global relationship between the surface and the subsurface layers. Geographically Weighted Regression (GWR) was then used to analyze the spatial non-stationarity in the relationships. GWR is a local form of linear regression that applies a kernel to subset the dataset and calculates a regression equation for each point location using a decay function (Brunsdon et al., 1996). The number of points included per equation is determined by the bandwidth of the kernel, which itself is calculated by Least Squares Cross Validation or Akaike Information Criterion corrected (AICc). The method allows the parameters to vary in space, while showing spatial patterns and providing a better understanding about local phenomena acting on the dependent variable. For this GWR analysis, we determined the number of points included in each equation by an adaptive kernel using AICc it maximizes the number of sample points to include in the regression as described by Brunsdon et al. (1996). Finally, the results were analyzed to assess the validity of the regression residuals for both their normality and spatial autocorrelation.

Results

The correlation between the subsurface and surface concentration levels were 0.51 for As, 0.40 for Cd, 0.33 for Cr, 0.52 for Co, 0.38 for Ni, and 0.23 for Pb (log). The correlations suggest that the subsurface levels could be important predictors of surface levels. However, GWR results showed non-normal residuals for all metals except As, which indicates linear regression is not effective for predicting surface concentrations using only subsurface concentrations. Therefore, only the results of the As model are presented (Table 1). The R-squared for each sample location, based on subsample size of 142 locations (neighbors), varies from 0.00 to 0.63. The overall adjusted R-squared is 0.34. The residuals are normally distributed and not spatially autocorrelated (Moran's I p value is > 0.05).

Table 1. GWR Analysis for Arsenic (log)

Arsenic (n = 877)	Results
Neighbors	142
R ² adjusted	0.34
Intercept range	.0544 to 1.611
Coefficient range	-0.249 – 0.7555
R ² range	0.00 – 0.63
Moran's I p-value	0.929112

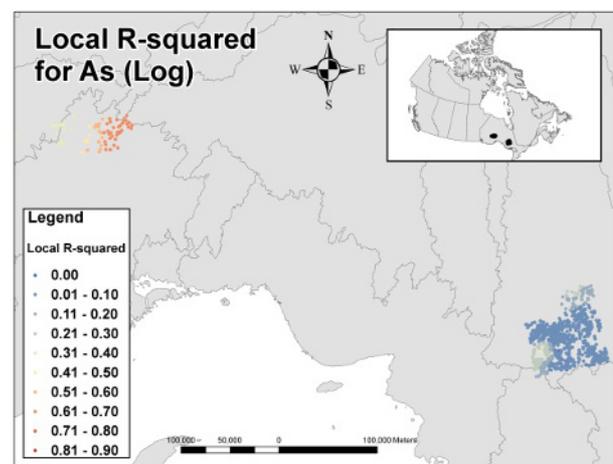


Fig. 1. Spatial Distribution of the Local R² for As

Figure 2 shows the spatial distribution of the regression coefficients (range from -0.249 to 0.7555) for the study area. The spatial variability in the relationship between the surface and the subsurface concentration levels is clearly demonstrated by the patterns on the map. We can observe distinct cluster of negative values with some sample location with coefficients equal to zero and the majority of the points having a positive coefficient. The pattern suggests that for most of the sample sites, the surface level is higher than the subsurface.

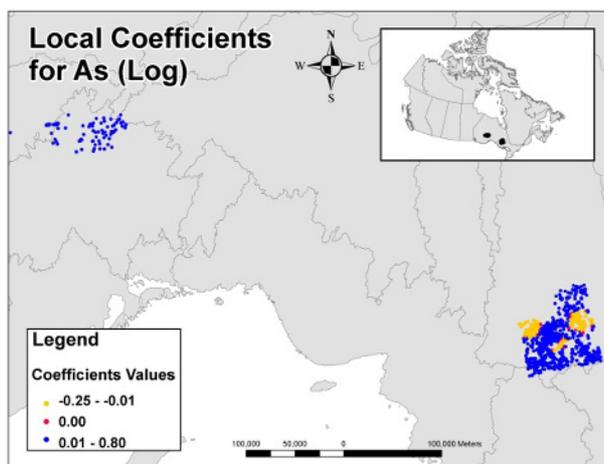


Fig. 2. Spatial Distribution of the regression coefficients
Figure 1 shows the spatial distribution of the R-squared. Note that the As data are in two distinct geographic areas in Ontario, as the original purpose of the sampling was for geological exploration. Most of the R-squared values in the eastern region are below 0.1 suggesting that subsurface concentrations do not predict surface concentrations well here; however in the western region, the R-squared are typically above 0.5.

Discussion

This research explores the relationship between the surface and the subsurface concentration levels of As, Cd, Co, Cr, Ni, and Pb using publicly available data, as a precursor to developing more sophisticated spatial models of surface concentrations at a regional level.

In general, soil properties and conditions vary both horizontally and vertically in soil (Murray et al., 2004), often over very small areas, and may influence the relationship between subsurface and surface levels locally for all studied metals. Our analysis showed that the concentration levels between the two layers are correlated for all the metals, but this varies (0.23 to 0.52). The variation of the correlations is not entirely unexpected. In fact, the specific chemical properties of each metal are different under same soil conditions, which also vary in time and space (Adriano, 2001; Svendsen et al., 2007). Therefore, the relationships between the surface and the subsurface levels should also be different for each metal. For example, Pb and Cd depositions, mainly from anthropogenic sources are most likely to remain bound with the organic matter at the surface, leading to an accumulation of the metals over long periods of time (Steinnes and Friedland, 2006), and so the relationship to subsurface levels may vary widely geographically (Adriano, 2001). For Co, Cr, and Ni, the literature suggests very complicated vertical processes that are not yet, fully understood (Adriano, 2001). For example, elevated concentrations are reported to be generally located in the B- and C-horizons, but higher levels were

also recorded in undisturbed surface layers (Adriano, 2001).

In the case of As, the mobility of the metalloid is better understood and the relationship between the two soil layers is expected to be more consistent. For example, As tends to be more uniformly distributed throughout the soil column (Adriano, 2001). Its surface level is noted to be controlled by the concentration of underlying rocks as well as the atmospheric deposition from near emission sources (Murray et al., 2004). Generally, As concentration would be lower in the upper layer above the B-horizon due to volatilization and its downward leaching movement (Wang and Mulligan, 2006). Moreover, the total content (measured here by INAA analysis) may be a better indicator of concentration relationship between the two layers as the speciation of As influences its mobility in soil (Adriano, 2001).

The GWR analysis of As confirms that the relationship between the surface and the subsurface concentration levels can vary widely geographically. Factors that may be involved in the distribution of As in soil (pH, organic content, and soil texture) were not available for this study, and may be difficult to obtain for large geographic regions.

Moreover, human activities are well-known to affect the metal distribution in soil. In fact, anthropogenic emissions of metal and other human disturbances increase the concentration levels and may change the chemical and physical conditions of the soil. Important sources of local contamination include proximity of mines, industrial emissions, and urbanization (Adriano, 2001). Therefore, spatial data on the location and emission intensity and also from natural sources (geology) may help to improve the prediction of soil surface concentration of As and the other metals.

Finally, the comparison between the surface and the B-horizon might not be adequate. For example, Steinnes and Friedland (2006) proposed to use the C-horizon concentration levels with the surface soil to analyze the spatial variability of metal concentrations to determine background concentration and potential contaminated locations. Unfortunately for this study, the C-horizon values were not widely available. There is also some debate on whether subsurface levels should be related to surface levels in general. Reimann and Garrett (2005) argue that the concentration of metals at the surface cannot be explained by the concentration of the subsurface. The authors explain that the chemical compositions of the different layers are not chemically similar and therefore, they cannot be compared.

Conclusion

The study used correlation analysis and GWR to assess the relationship between the surface and the subsurface of paired-sample concentration levels for As, Cd, Co, Cr, Ni, and Pb. This study shows that the B-horizon is, at

various degrees, related to the surface concentration levels and may contribute to explaining the surface concentration levels. Nonetheless, only As showed a statistically robust relationship between the two soil layers suggesting that other factors contribute to the spatial distribution patterns of the surface concentration levels of the studied metals.

Further research using anthropogenic and natural sources may help providing more information about the spatial distribution of heavy metals in surface soil across Canada.

Acknowledgements

This study is part of CAREX Canada, which is funded by the Canadian Partnership Against Cancer (CPAC).

References

- Adriano DC. Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability, and Risks of Metals. Springer, 2001.
- Baffi C, Bettinelli M, Beone GM, Spezia S. Comparison of different analytical procedures in the determination of trace elements in lichens. *Chemosphere* 2002; 48: 299-306.
- Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 1996; 28: 281-298.
- Matschullat J, Ottenstein R, Reimann C. Geochemical background - can we calculate it? *Environmental Geology* 2000; 39: 990-1000.
- Murray KS, Rogers DT, Kaufman MM. Heavy metals in an urban watershed in southeastern Michigan. *J Environ Qual* 2004a; 33: 163-172.
- Reimann C, Garrett RG. Geochemical background-concept and reality. *Science of The Total Environment* 2005; 350: 12-27.
- Steinnes E, Friedland AJ. Metal contamination of natural surface soils from long-range atmospheric transport: Existing and missing knowledge. *Environmental Reviews* 2006; 14: 169-186.
- Svendsen ML, Steinnes E, Blom HA. Vertical and horizontal distributions of Zn, Cd, Pb, Cu, and Hg in uncultivated soil in the vicinity of a zinc smelter at Odda, Norway. *Soil & Sediment Contamination* 2007; 16: 585-603.
- Wang S, Mulligan CN. Occurrence of arsenic contamination in Canada: Sources, behavior and distribution. *Science of the Total Environment* 2006; 366: 701-721.